Temperature-Aware NBTI Modeling and the Impact of Standby Leakage Reduction Techniques on Circuit Performance Degradation

Article *in* IEEE Transactions on Dependable and Secure Computing · September 2011

DOI: 10.1109/TDSC.2010.41 · Source: DBLP

CITATIONS
21

READS
41

6 authors, including:

Yu Wang
Tsinghua University
157 PUBLICATIONS **1,504** CITATIONS
SEE PROFILE

Hong Luo
Tsinghua University
16 PUBLICATIONS **222** CITATIONS
SEE PROFILE

Ku He
University of Texas at Austin
18 PUBLICATIONS **255** CITATIONS
SEE PROFILE

Huazhong Yang
Tsinghua University
453 PUBLICATIONS **2,605** CITATIONS
SEE PROFILE

Temperature-Aware NBTI Modeling and the Impact of Standby Leakage Reduction Techniques on Circuit Performance Degradation

Yu Wang, *Member, IEEE,* Hong Luo, *Student Member, IEEE,* Ku He, Rong Luo, *Member, IEEE,* Huazhong Yang, *Senior Member, IEEE,* Yuan Xie, *Member, IEEE*

***Abstract*— As technology scales, Negative Bias Temperature Instability (NBTI), which causes temporal performance degradation in digital circuits by affecting PMOS threshold voltage, is emerging as one of the major circuit reliability concerns. In this paper, we first investigate the impact of NBTI on PMOS devices and propose a temporal performance degradation model that considers the temperature variation between active and standby mode. We then discuss the resemblance between NBTI and leakage mechanisms, and find out that the impact of input vector and internal node on leakage and NBTI is different; hence, leakage and NBTI should be optimized simultaneously. Based on this, we study the impact of standby leakage reduction techniques (including input vector control, sleep transistor insertion) on circuit performance degradation considering active and standby temperature differences. We demonstrate the potential mitigation of the circuit performance degradation by these techniques.**

***Index Terms*— Negative Bias Temperature Instability (NBTI), leakage reduction, temperature-aware NBTI modeling, circuit performance degradation.**

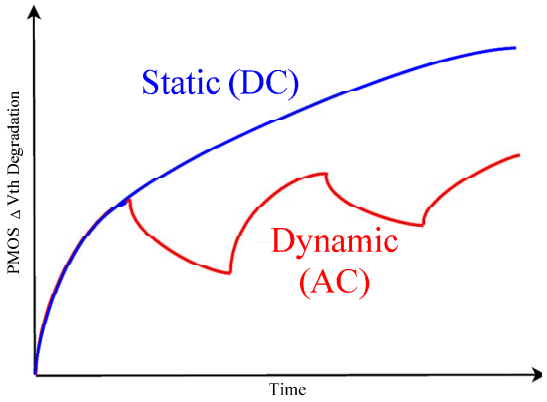

Fig. 1. The conceptual illustration of the PMOS ΔV_{th} Degradation under DC and AC stress conditions, which shows the difference between static NBTI and dynamic NBTI.

I. Introduction

CIRCUIT reliability is one of the major concerns in VLSI circuits and systems designs. Negative Bias Temperature Instability (NBTI), which has deleterious effect on the PMOS transistor threshold voltage and the drive current of semiconductor devices, is emerging as a major reliability degradation mechanism [1]. **IBM has pointed out that NBTI is the most severe degradation mechanism** in the circuit life time domain as CMOS technologies scale beyond a 90-nm minimum lithographic linewidth, and the impact of NBTI on circuit delay is about 15% on 65nm technology node [2].

NBTI occurs when PMOS transistors in circuits are stressed under negative gate voltage (i.e., $V_{gs} = -V_{dd}$) at elevated temperature, causing a shift in threshold voltages [3], and resulting in the degradation of device performance [1]. Bias temperature stress under constant voltage (i.e., under *DC stress condition*, and is called *static NBTI*) leads to rapid device performance degradation. However, under actual *AC stress condition*(i.e., *dynamic NBTI*) [4] [5], when stress is periodically removed (i.e., the PMOS transistor is under *stress* or *relaxation* condition alternatively), the degradation of device parameters is partially recovered, which leads to less severe parameter shifts over a long time compared to that under DC stress condition. Fig. 1 is the conceptual illustration of the difference between DC stress NBTI and AC stress NBTI.

The PMOS threshold voltage degradation due to dynamic NBTI depends on the sequence of stress and relaxation applied to the gate. Furthermore, in a digital circuit system, the degradation of different timing paths not only depends on the signal probabilities and activity factors of the gates, but also on the temperature variations, which are caused by the circuit mode transitions (between active and standby). Currently, the timing specifications of logic blocks are designed by leaving a certain safety timing margin that accounts for NBTI-induced performance degradation. Kumar [6] takes the signal probabilities and activity factors of gates into account during the estimation of temporal degradation of various paths in digital circuits, under the assumption of worst case temperature. However, during circuit operations, the temperature varies when the circuit mode changes between active and standby modes. It has been shown that the NBTI-induced degradation is faster and the recovery is slower, under higher temperature [5]. Hence, the estimation of circuit performance degradation due to NBTI with a worst-case temperature assumption could

Manuscript received .

This work was supported by National Key Technological Program of China under contracts, No. 2008ZX01035-001, National 863 project of China (No. 2009AA01Z130), National Natural Science Foundation of China (No.60870001), and Tsinghua National Laboratory for Information Science and Technology (TNList) Cross-discipline Foundation. Yuan Xie's work was supported in part by NSF 0643902 and 0916887.

Yu Wang, Hong Luo, Ku He, Rong Luo and Huazhong Yang are with the Department of Electronic Engineering, TNList, Tsinghua University, Beijing, 100084, China. (e-mail:yu-wang@mail.tsinghua.edu.cn)

Yuan Xie is with CSE Department, Pennsylvania State University, University Park, PA, USA. (e-mail: yuanxie@cse.psu.edu)

Digital Object Indentifier 10.1109/TDSC.2010.41

be too pessimistic.

On the other hand, the leakage current in the circuits also depends on the gate overdrive. For example, the leakage current of a logic gate also depends on the internal states that are determined by the signal probabilities and the activity factors of the gate. Further, the leakage current and NBTI are both affected by the threshold voltage V_{th}. Higher V_{th} leads to smaller leakage current and smaller V_{th} degradation due to NBTI (the higher the initial V_{th} is, the smaller NBTI-induced circuit delay degradation will be.) [7] [8]. The standby leakage current reduction techniques, such as input vector control and sleep transistor insertion, can change the gate overdrive which has large influence on the NBTI effect. Therefore, the standby leakage reduction techniques could be potentially used to mitigate the circuit performance degradation due to NBTI.

Our paper distinguishes itself in the following aspects:

(1) We study the impact of NBTI on the temporal performance degradation of combinational circuits considering temperature difference between active and standby mode, and propose an analytical model that can analyze NBTI-induced degradation taking the time ratio and temperature changes between active and standby mode into account.

(2) We propose that NBTI and leakage mechanisms both depend on gate overdrive, and this fact can be used to mitigate the NBTI effect when the leakage current has been controlled during the circuit standby time. An NBTI and leakage co-optimization platform is proposed to simultaneously analyze and optimize the leakage and NBTI-induced degradation for digital circuits.

(3) We evaluate two standby leakage reduction techniques: input vector control and sleep transistor insertion. From our experimental results, the sleep transistor insertion technique is very efficient in NBTI mitigation, while the input vector control (IVC) technique is somehow less effective. The internal node control method, however, could be a potential way to mitigate the NBTI effect.

- We first show that the dependencies of NBTI and leakage on input vector are different for different gates. Our experimental result shows that the impact of IVC approach on NBTI-induced performance degradation is not that effective if the standby temperature is low (330K). Moreover, the internal node control [9], [10] can be used as a more potential way to mitigate the NBTI effect.
- The impact of NBTI on the PMOS sleep transistor (ST) is studied, and a NBTI-aware PMOS ST sizing method is proposed to ensure the circuit reliability. Our investigation shows that the ST insertion technique is efficient to mitigate NBTI effect on circuit performance degradation, because the V_{gs} of PMOS transistor is nearly zero during circuit standby mode, so that the voltage of each node is nearly V_{dd} and no PMOS transistor is negatively biased.

The rest of the paper is organized as follows. Section II reviews previous NBTI models and the standby leakage reduction techniques (mainly the input vector control and sleep transistor insertion techniques). In Section III we describe the NBTI model considering the temperature variation between the active mode and the standby mode. In Section IV, the resemblance between NBTI and leakage mechanism is first discussed in order to show that the NBTI-induced performance degradation may be mitigated through the standby leakage reduction techniques. Two standby leakage reduction techniques (input vector control and sleep transistor insertion) are investigated, and their potential impact on circuit performance degradation is analyzed according to the simulation results on ISCAS85 circuits. Section V concludes the paper and shows some discussion.

II. Related Works

The related work can be classified into two categories: (1) NBTI modeling and mitigation techniques; (2) Leakage reduction techniques.

A. NBTI modeling and mitigation techniques

NBTI has been known since the very early days of MOS device development. Goetzberger *et al.* [11], [12] were one of the first groups to show detailed characteristics of negative bias and temperature stress. In late 90's, the NBTI effect is aggravated due to the usage of nitride oxides [13]. Further, the t_{ox}^{-1} dependence of interface-trap generation implies that NBTI becomes more severe for ultra-thin oxides due to the technology scaling. The bias temperature instabilities exist in both PMOS and NMOS devices, whenever a negative bias or a positive bias is applied. Nevertheless, applying negative bias stress (i.e NBTI condition) to a PMOS device brings the most deleterious impact on the threshold voltage. So people mainly focus on the mechanisms of NBTI on PMOS devices.

The previous work about PMOS NBTI mainly focused on the analysis of the threshold voltage degradation and the impact on the drive current of semiconductor devices [4], [14]. Some researchers devoted themselves to study how process affects the NBTI [13], [15]. Several works [16]–[18] have addressed the issue of describing physical models for NBTI, and providing analytical expressions for the first stress and relaxation phases.

Built on top of device level NBTI modeling, researchers studied the circuit-level performance degradation models for NBTI. Paul *et al.* [19] proposed an estimation method of circuit degradation due to NBTI in digital circuits. A DC stress NBTI model was used, and therefore only an upper bound of total delay variation can be obtained. Some analytical models that evaluate NBTI effect with multi-cycle AC stress were then proposed to help designers estimate the circuit performance degradation due to NBTI [6], [20]. A recursive process was used to evaluate the NBTI effect [6] considering the signal probabilities and activity factors of gates, while Vattikonda *et al.*'s predictive NBTI model [8], [20] described the effect of various process and design parameters.

Based on these analytical circuit degradation models, a few researchers have investigated design techniques to mitigate the NBTI-induced performance degradation. Kumar *et al.* [21] studied the impact of NBTI on the read stability of SRAM cells and proposed a simple bit flipping technique to recover the static noise margin of SRAM cells. Paul *et al.* [22] presented an NBTI-aware sizing algorithm to ensure the reliability of

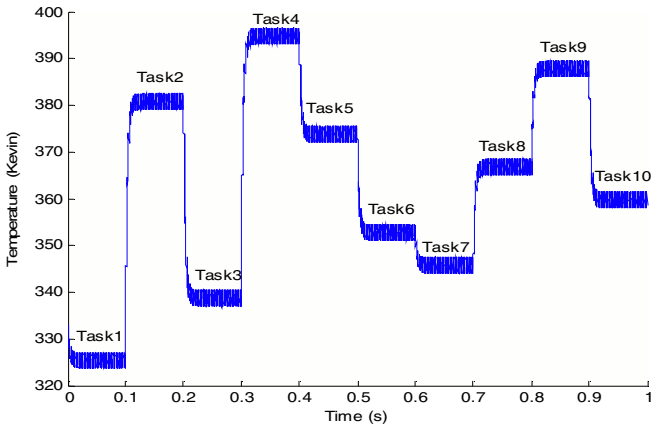

Fig. 2. Thermal profiles of running a task set on a typical processor.

nano-scale circuits. Kumar's work dedicated to memory design issues; while Paul's work was based on a DC stress NBTI model without considering the circuit operation modes. Jaume Abella *et al* [23] proposed an IVC technique, using special input vectors alternatively during idle periods. Any given input would always degrade the same transistors, so they preferred to alternate several inputs that degrade different PMOS transistors, thus the maximum degradation of any PMOS is reduced with practically no cost. Recently, some adaptive techniques were proposed to mitigate the NBTI effect during the circuit lifetime, such as [24], [25].

NBTI can be easily handled by simple guard-banding at a very low cost in the current technology [26], such as 90nm. However, for long term stress model or permanent degradation that can not be recovered for high-k, NBTI-induced degradation is quite serious and cannot be ignored, so it is necessary to mitigate the NBTI-induced degradation for the logic circuits.

Previous analytical NBTI models for the circuit performance degradation are based on the worst-case temperature assumption [6], [8], [19], [20]. However, during circuit operations, the circuit temperature varies between high temperature(for example 400K) and low temperature(for example 330K) when the circuit switches between active and standby modes. As we can see from [27], the power profile of different tasks for Intel's Montecito processor has a very large power variation, from 68W to 126W. Assuming a typical air cooling condition [28], Fig.2 gives out the temperature curve of the processor executing a task set, which contains different tasks with random power profile ranges from 10 to 130W. It is shown that the processors temperature varies in a range from 60 to 110 Centigrade. Therefore, in this paper, we propose an NBTI model considering the temperature variation due to the change of the circuit operation mode: active and standby mode. We also assumes that the temperature switches fast from active time temperature to standby time temperature under a typical air cooling condition.

B. Leakage reduction techniques

Leakage current increases drastically with technology scaling and has become a major contributor to the total IC power [29], especially in high performance systems with predefined power budget. Since not every application requires a fast circuit to operate at the highest performance level all the time, modules in which the computation is burst are often idle. Thus, there is an opportunity to reduce the leakage power consumed by such circuits during the standby mode.

Circuit techniques to reduce leakage currents can be broadly categorized as *design time techniques* and *runtime techniques* [29]. Dual V_{th} assignment [30] is a typical design time technique. Runtime techniques include: 1) *standby leakage reduction techniques*, which put the module into a low leakage mode when the computation is not required; and 2) *active leakage reduction techniques*, which slow down the system to reduce the leakage when maximum performance is not needed. In this paper, we mainly focus on the runtime standby leakage reduction techniques.

The IVC technique [9], [10], [31]–[33] is based on the well-known transistor stacking effect: a CMOS gate's subthreshold and gate oxide leakage current varies dramatically with the input vector applied to the gate [34], [35]. Basically in an IVC technique, the minimum leakage vector (MLV) is used with the help of standby signals to reduce both subthreshold and gate oxide leakage current, when the circuit enters the standby mode. When the MLV is manipulated during the circuit standby mode, the internal state of each node in the circuit is set to be either 0 or 1, such that the circuit standby leakage is minimized.

Sleep transistor insertion [36]–[42] is to place a sleep transistor (ST) between the gates and the power/ground (P/G) net in a circuit in order to put it into sleep mode when the circuit is standby. One example is the block based ST insertion (BBSTI) technique [36]–[39] that put a large block of gates into sleep mode using a single large sleep transistor. The existing literatures on BBSTI techniques present some details in clustering gates into blocks in order to optimize the leakage current and ST size. Another example is the Fine-grain ST insertion (FGSTI) technique [40]–[42], in which sleep transistors are assigned to each standard cell, also showed some advantages over the BBSTI technique, such as guaranteed circuit functionality and improved circuit noise margins.

III. NBTI MODELING CONSIDERING TEMPERATURE VARIATION

In this section, we first review the previous NBTI model, and then propose our model that considers temperature variation, and finally analyze the device and circuit delay degradation based on our improved model.

Note that the simulation results in the following sections are based on a standard cell library constructed using the PTM $90nm$ bulk CMOS model [43]. $V_{dd} = 1.0V$, $|V_{th}| = 220mV$ are set for all the transistors in the circuits. The operation time is set to be $3 \times 10^8 s$ (about 10 years). T_{active} and $T_{standby}$ are first set to 400K and 330K for both leakage current estimation and NBTI-induced degradation.

All ISCAS85 benchmark circuit netlists are synthesized using a commercial synthesis tool and mapped to the 90nm

standard cell library. A static timing analysis (STA) tool [44] is used to calculate the circuit performance degradation, with the usage of our NBTI model.

A. NBTI-induced V_{th} degradation

A shift in the PMOS transistor threshold voltage ΔV_{th} is proportional to the generation of interface traps due to NBTI, which can be expressed as [22]:

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

where m represents equivalent V_{th} shifts due to mobility degradation for a given technology, and $N_{it}(t)$ is the interface trap density due to NBTI. In this paper, we choose the reaction-diffusion model [3] from various NBTI device models to describe $N_{it}(t)$:

$$\frac{dN_{it}}{dt} = k_f(N_0 - N_{it}) - k_r N_{it} C_H(0,t) \tag{2}$$

$$\frac{dN_{it}}{dt} = -D_H \frac{\partial C_H}{\partial x}\bigg|_{x=0} \tag{3}$$

where the mobile diffusing species are assumed to be neutral H atoms, and N_0 is the concentration of initial interface defects. The parameters k_f and k_r are constant dissociation rate and self-annealing rate, respectively. When the device is in recovery phase, k_f becomes zero, but k_r is unchanged. The parameter C_H is the concentration of H atoms, and D_H is the corresponding diffusion coefficient.

The diffusion of H (D_H), which has great impact on N_{it}, follows the equation [3]:

$$\frac{\partial C_H}{\partial t} = D_H \frac{\partial^2 C_H}{\partial x^2} \tag{4}$$

With assumption of quasi-equilibrium and an infinite thick oxide (i.e. t_{ox} is more than diffusion length $\sqrt{4D_H t}$), a solution of Eq. (2)-(4) is presented by [3]:

$$N_{it}(t) = 1.16\left(\frac{k_f N_0}{k_r}\right)^{1/2}(D_H t)^{1/4} = At^{1/4} \tag{5}$$

Eq.(5) describes the NBTI impact under DC stress condition. When the stress is removed after a stress time of t_{stress}, an analytical form for recovery process is proposed by [6]:

$$N_{it}(t) = N_{it}^0\left(1+\sqrt{t/t_{stress}}\right)^{-1} \tag{6}$$

where t is the recovery time and N_{it}^0 is the interface density at the beginning of recovery.

B. Temperature-aware NBTI modeling for V_{th} degradation

The above models have described both the stress and recovery phases of the NBTI degradation; however, in order to estimate the performance degradation of a circuit, the NBTI model should handle multiple cycles of stress and recovery phases. We use the analytical NBTI model proposed in [6] to handle multi-cycle AC stress condition; and the creation of interface traps after n cycles of AC stress can be evaluated by a recursive formula. The generation of interface traps assuming DC stress over the first period (i.e. a stress of time τ) is denoted as $N_{it}^0 = A\tau^{1/4}$; so after n cycles of AC stress, the interface traps can be expressed as [6]:

$$N_{it}[(n+1)\tau] = \frac{\beta}{1+\beta} + \frac{N_{it}^0}{1+\beta}\left[c + \left(\frac{N_{it}(n\tau)}{N_{it}^0}\right)^4\right]^{1/4} \tag{7}$$

and

$$N_{it}(\tau) = \frac{c^{1/4}}{1+\beta}N_{it}^0 \tag{8}$$

where τ is the period time, c is duty cycle of stress phase, and $\beta = \sqrt{\frac{1-c}{2}}$.

In order to get a fast NBTI model, we first simplify the recursion based on Eq.(7) and (8) considering a large enough cycle number n:

$$S_1 = \frac{c^{1/4}}{1+\beta} \tag{9}$$

$$S_{n+1} = S_n + \frac{c}{4(1+\beta)S_n^3} \tag{10}$$

and

$$N_{it}(n\tau) = S_n N_{it}^0 = S_n \cdot A\tau^{1/4} \tag{11}$$

From Eq. (1), the shifts of threshold voltage can be expressed as:

$$\begin{aligned}\Delta V_{th}(n\tau) &= (1+m)\frac{q}{C_{ox}} \cdot S_n \cdot A\tau^{1/4} \\ &= K_V S_n \tau^{1/4}\end{aligned} \tag{12}$$

where K_V is a constant related with E_{ox} and temperature, S_n is controlled by duty cycle.

Previous NBTI models only consider the situation that the temperature is a constant (which is around 400K). However, in practice, the circuit operation condition can change between active and standby states, by using various low power techniques, such as power supply gating or clock gating, to reduce power dissipation. The circuit temperature is related to the power density and its distribution (assuming that the physical layout and the heat dissipation capability–such as chip package and heat sink–are fixed) [45], [46]. When a circuit switches between active and standby mode, power dissipation changes (and therefore power density changes), causing the temperature of the chip to change accordingly, and the temperature usually converges to steady-state very fast (in the order of milliseconds). In this paper, we assume the steady-state temperature during the active mode and standby mode are T_{active} and $T_{standby}$, respectively (Generally speaking, $T_{standby}$ is lower than T_{active}), and define RAS to be the ratio of active and standby time (RAS), which is an indication of how long the circuit will be running at T_{active} and $T_{standby}$. When the circuit operation mode switches between active and standby mode, the circuit operation temperature changes between T_{active} and $T_{standby}$ frequently, hence the impact on V_{th} due to NBTI should be different from the case under a constant high operation temperature condition.

The temperature variation may have impact on various terms in NBTI modeling, such as the mobility factor, activation energy terms, and diffusion coefficient, etc.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

First, the impact of temperature on mobility of interface traps is small; since with the technology scaling, the electrical field is big and results in the saturation of the mobility due to all kinds of scattering. However, the scattering caused by interface traps can be ignored.

Second, we investigate the activation energy terms. The interface trap generation N_{it} is related to dissociation rate k_f, self-annealing rate k_r, and diffusion coefficient D_H, and all these three parameters depend on the temperature:

$$D_H = D_{H0}\exp(-E_D/k_BT) \quad (13)$$

$$k_f = k_{f0}\exp(-E_f/k_BT) \quad (14)$$

$$k_r = k_{r0}\exp(-E_r/k_BT) \quad (15)$$

From Eq. (5), the overall activation energy can be represented as:

$$E_A = \frac{1}{4}E_D + \frac{1}{2}(E_f - E_r) \quad (16)$$

Because $E_f - E_r \approx 0$, $E_A \approx \frac{1}{4}E_D$ [47], we can assume that the temperature dependency of interface trap generation is only in terms of the diffusion coefficient of H atom (D_H) in Eq.(5).

If a triangle diffusion profile [3] is used to model D_H, the effect of H atom diffusion under $T_{standby}$ lasting for a stress time of $t_{standby}$, equals to the effect under T_{active} for a stress time of $t'_{standby}$, where $t'_{standby} = D_{standby} \times t / D_{active}$ ($D_{standby}$ and D_{active} denote diffusion coefficients under standby and active mode, respectively). Therefore, the equivalent stress time t^{eq}_{stress} for each cycle can be expressed as:

$$t^{eq}_{stress} = ct_{active} + t_{standby}\frac{D_{standby}}{D_{active}} \quad (17)$$

where c is the input signal duty cycle of active time. The equivalent recovery time $t^{eq}_{recovery}$ can be considered similarly.

The equivalent duty cycle c^{eq} and period time τ^{eq} can be derived as:

$$c^{eq} = \frac{t^{eq}_{stress}}{t^{eq}_{stress} + t^{eq}_{recovery}} \quad (18)$$

$$\tau^{eq} = t^{eq}_{stress} + t^{eq}_{recovery} \quad (19)$$

With the equivalent duty cycle c^{eq} and period time τ^{eq}, the ΔV_{th} considering the time and temperature changes between active and standby mode can be evaluated using Eq. (9)-(12).

Fig. 3 shows the impact on ΔV_{th} with different active and standby time ratios (RAS). The temperature of the highest line is $T_{standby}$=T_{active}=400K; the temperature of others is $T_{standby}$=330K. Fig. 4 shows the impact on ΔV_{th} with $T_{standby}$. The active and standby time ratio is set to be 1:5. These two figures show that RAS and the standby mode temperature ($T_{standby}$) have great impact on the V_{th} degradation due to NBTI. The trend of ΔV_{th} in Fig. 4 fits well with the previous data of NBTI under temperature variation [48].

The V_{th} degradation with different active and standby time ratios is shown in Table I. The total time is set to $3 \times 10^8 s$ and the signal probability during the active mode is set to 0.5. When $T_{standby}$=T_{active}=400K, the ΔV_{th} is increasing with a decreasing active and standby time ratio, since the total time under stress condition is increased. However, the ΔV_{th} is decreasing when $T_{standby}$=330K with a decreasing active and standby time ratio, since the total time under lower temperature is increased. The largest gap between the ΔV_{th} is about 9.4mv when the active and standby time ratio is 1:9. Meanwhile, when $T_{standby}$=370K, the ΔV_{th} is nearly the same, which means there exists an interval of $T_{standby}$, whose corresponding ΔV_{th} is not sensitive with RAS variation. That is to say, when $T_{standby}$ is near 370K, the impact of RAS on ΔV_{th} is very small so that when RAS changes, the V_{th} degradation is approximatively the same.

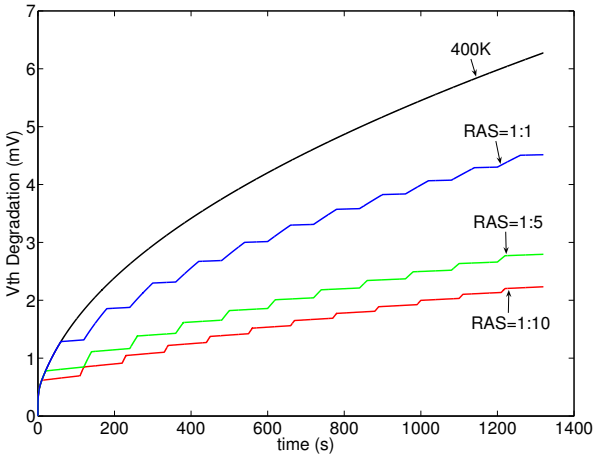

Fig. 3. V_{th} degradation with different active and standby time ratio (RAS). T_{active} = 400K, the input signal probability is set to 0.5 in the active mode; in the standby mode, the input of PMOS is set to 0, which refers to the worst case V_{th} degradation.

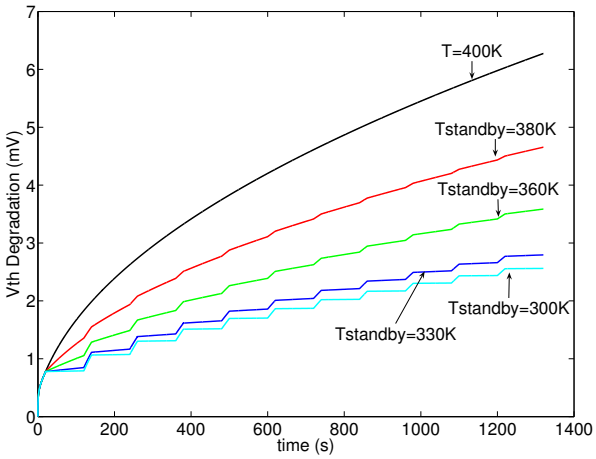

Fig. 4. V_{th} degradation with different $T_{standby}$. The active and standby time ratio RAS is set to be 1:5 and input signal probability is set to 0.5 in the active mode; in the standby mode, the input of PMOS is set to 0, which refers to the worst case V_{th} degradation.

Our temperature-aware NBTI modeling is represented by the active/standby temperature and RAS, when we try to model the performance degradation of a circuit. The temperature of a circuit will converge into steady-state in several milliseconds or less; meanwhile the steady-state temperature and convergence time depend on the runtime applications and the heat dissipation mechanism. In this paper, T_{active} and

TABLE I

ΔV_{th}(MV) UNDER DIFFERENT ACTIVE AND STANDBY MODE RATIO. T_{active} = 400K, INPUT SIGNAL PROBABILITY IS SET TO 0.5 IN THE ACTIVE MODE; IN THE STANDBY MODE, THE INPUT OF PMOS IS SET TO 0, WHICH REFERS TO THE WORST CASE V_{th} DEGRADATION.

RAS	9 : 1	7 : 1	5 : 1	3 : 1	1 : 1	1 : 3	1 : 5	1 : 7	1 : 9
$T_{standby}$=400K	22.09	22.20	22.38	22.74	23.77	24.85	25.26	25.48	25.63
$T_{standby}$=370K	21.63	21.62	21.53	21.49	21.29	21.24	21.22	21.20	21.19
$T_{standby}$=330K	21.26	21.16	20.99	20.63	19.36	17.65	16.88	16.45	16.16

$T_{standby}$ are assumed to be steady-state temperatures during the circuit active and standby time. Assuming a typical air cooling condition [28], we further assume that the temperatures (T_{active} and $T_{stadnby}$) are not affected RAS as we discussed in Section II.A.

C. Gate and circuit performance degradation

In circuit timing analysis, a combinational circuit can be modeled as a directed acyclic graph (DAG) $G = (V, E)$. A vertex $v \in V$ represents a CMOS gate from the given library, while an edge $(i, j) \in E$, $i, j \in V$ represents a connection from vertex i to vertex j.

The delay of a gate v can be approximately expressed as [22]:

$$d(v) = \frac{C_L V_{dd}}{I_d} = \frac{K}{(V_g - V_{th})^\alpha}$$
$$K = \frac{C_L V_{dd}}{\mu C_{ox} W_{eff}/L_{eff}} \quad (20)$$

where C_L is the load capacitance, V_{dd} is the supply voltage; V_g and V_{th} are the gate voltage and the threshold voltage of a transistor, respectively; α is the velocity saturation index, whose value ranges from 1 to 2; μ is the mobility, C_{ox} is the oxide capacitance, and L_{eff} and W_{eff} are the channel length and the transistor width, respectively.

Hence, the delay degradation $\Delta d(v)$ for gate v can be derived as:

$$\Delta d(v) = \frac{K}{(V_g - V_{th} - \Delta V_{th})^\alpha} - \frac{K}{(V_g - V_{th})^\alpha}$$
$$= \left((1 - \frac{\Delta V_{th}}{V_g - V_{th}})^{-\alpha} - 1 \right) d(v) \quad (21)$$

We use Taylor series expansion on the right side of Eq.(21), neglect the higher order terms, such that,

$$\Delta d(v) = \frac{\alpha \Delta V_{th}}{V_g - V_{th0}} \times d(v) \quad (22)$$

where V_{th0} is the original transistor threshold voltage and $d(v)$ is the original delay of gate v. There might be several ΔV_{th} of different PMOS's in one gate. In such cases, we just select the largest one to calculate the gate delay degradation, which is the worst case delay degradation. The shift in the transistor threshold voltage, ΔV_{th}, can be derived using Eq. (12). The signal probability for each edge in the circuit is derived statistically by simulating a large number of input vectors.

Given a time interval, we can have the corresponding gate delay degradation from Eq.(22). A static timing analysis tool [44] is used to compute the max delay of the circuit with all the gates' temporal degradation information derived by our temperature-aware NBTI model.

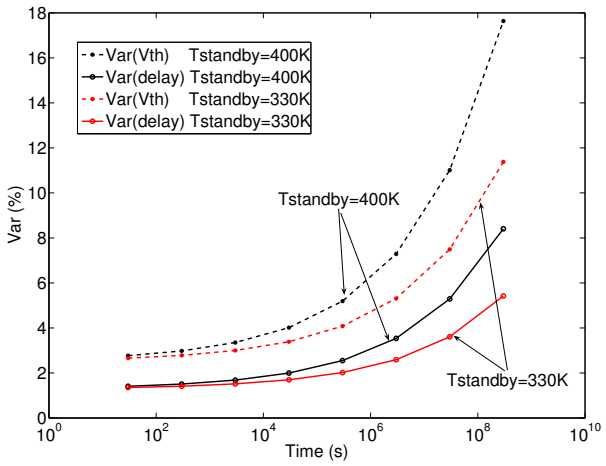

Fig. 5. Comparison between PMOS V_{th} degradation and C432 performance degradation.

In the circuit degradation analysis, we set all the internal nodes' states to '0' during the standby mode, to investigate the worst case circuit degradation. Of course, in a realistic design, there exists no such input vector that makes the internal nodes all 0's, this assumption is only used to calculate the maximum possible degradation and the potential of NBTI mitigation techniques. Fig. 5 shows the performance degradation of ISCAS85 C432 benchmark with time under different standby mode temperature. The circuit degradation is much less than the V_{th} degradation under a same standby mode temperature. And the standby mode temperature difference leads to considerable circuit delay difference.

IV. IMPACT OF STANDBY LEAKAGE REDUCTION TECHNIQUES ON CIRCUIT PERFORMANCE DEGRADATION

From the above section, we show that both RAS and temperatures ($T_{standby}$/T_{active}) are important during NBTI-induced performance degradation analysis in the circuits that have standby modes. Meanwhile in the circuit standby mode, another critical issue, which people have been fighting against for years, is the leakage current. Both NBTI-induced circuit performance degradation and circuit leakage current are affected by temperature and some common circuit parameters related to gate overdrive [7], [20], such as gate input voltage and threshold voltage. Such observation motivates us to investigate the impact of existing standby leakage reduction techniques on

the circuit performance degradation, and study the potential of using the existing leakage reduction techniques to mitigate NBTI-induced performance degradation considering the temperature variation due to different RAS's.

A. The resemblance between NBTI and leakage mechanism

Both NBTI and leakage mechanism have common dependency on technology and design parameters related to gate drive. In this subsection, we focus on the the internal node dependency and the V_{th} dependency analysis:

- **Internal node dependency.** The NBTI effect on a PMOS depends on V_{gs} and the stress time (duty cycle) which are all related to the input state of a gate. Table I also shows that different duty cycle and temperature leads to different circuit performance degradation. On the other hand, the subthreshold leakage and gate leakage all depend on the input state of a gate due to the stacking effect [34], [35]. For example, Table II lists the overall leakage current in NOR2, NOR3, and INV gates under different input combinations, and the temperature is $400K$. We can see that leakage current varies between different input vectors. Furthermore, Table II lists the $\Delta Delay$ due to NBTI effect of these gates under different standby time input vectors. Here RAS is 1:9, T_{active}=400K and $T_{standby}$=330K. The NBTI-induced $\Delta Delay$ also varies between different input vectors. Notice that in this table, the impact of input vector on leakage and delay degradation is same, however, this conclusion is not true for all gate types. We simulate the typical cells (NAND/AND, NOR/OR, INV, BUF) in the library, and find out that for NAND/AND/INV gates, the input vector for least leakage will lead to worst NBTI-induced delay degradation; for NOR/OR gates, the input vector for least leakage will also lead to best case NBTI-induced delay degradation [49].
- V_{th} **dependency**. A higher V_{th} will lead to a smaller performance degradation due to NBTI [20]:

$$N_{it}(t) \propto T_{ox}\sqrt{C_{ox}(V_{gs}-V_{th})}\cdot exp(\frac{E_{ox}}{E_0})\cdot exp(\frac{-E_a}{k_bT}) \quad (23)$$

Meanwhile, the subthreshold leakage decreases exponentially with a higher V_{th}; the gate-tunneling leakage also decreases with a lower E_{ox} caused by a higher V_{th}. Thus, leakage reduction techniques that adjust V_{th} in the design phase or the runtime phase (such as dual V_{th} assignment and dynamic V_{th} scaling) may mitigate the circuit performance degradation due to NBTI.

Based on the above observations, in the following subsections, we show our NBTI/leakage analysis and optimization flow; and then investigate the impact of two existing standby leakage reduction techniques on NBTI-induced performance degradation. First, an input vector control technique which takes NBTI impact into account is proposed and investigated; and the potential impact of internal node control technique is further discussed. Second, the impact of NBTI on the PMOS sleep transistor (ST) is analyzed, then a PMOS ST sizing method considering NBTI effect is proposed. The impact of sleep transistor insertion on NBTI induced circuit degradation shows that this technique is efficient to mitigate NBTI effect on circuit performance degradation.

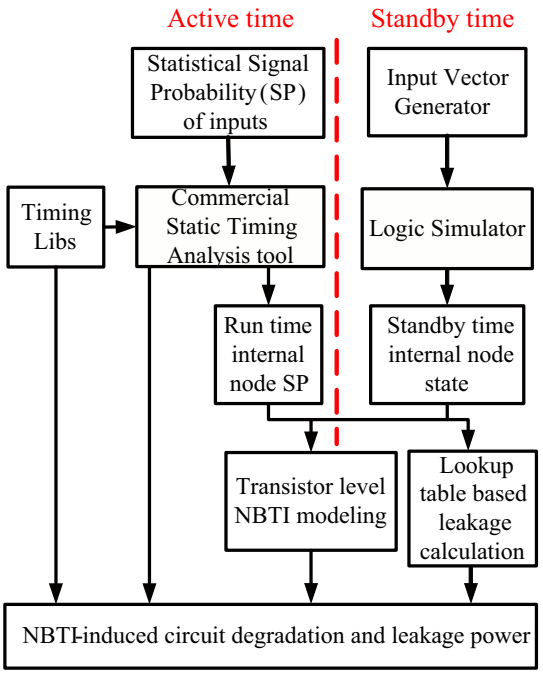

Fig. 6. Our NBTI/leakage analysis and optimization flow.

B. The NBTI/leakage analysis and optimization flow

Fig.6 shows our NBTI/leakage analysis and optimization flow which is capable to analyze and optimize NBTI effect and leakage power in the circuit simultaneously. When the circuit is in an active mode, the statistical information for input Signal Probability (SP) is used to generate the internal node SP. When the circuit is in a standby mode, logic simulator is used to generate the voltage level of each internal node. The active time internal node SP and the standby time internal node states are used to estimate the NBTI-induced V_{th} degradation through temperature-aware transistor level NBTI model. The leakage power is estimated based on the input vector aware leakage lookup tables. Our flow helps evaluate the NBTI and leakage mitigation techniques, such as input vector control and sleep transistor insertion.

Because the inputs of our flow include: circuit netlists, technology libs and NBTI modelings, this flow can deal with different circuits under different technology libs and NBTI models. Moreover, we can modify the timing calculation and input vector generation module to integrate more advanced algorithms for timing analysis and input vector generation methods.

C. Impact of IVC technique

Since NBTI and leakage current both depend on internal states of the circuits (i.e., the input states of gates), we first try to mitigate NBTI-induced performance degradation using existing IVC technique which minimize the standby leakage current(B.1) and find out the impact of IVC technique is

TABLE II

LEAKAGE CURRENT AND NBTI-INDUCED DELAY DEGRADATION COMPARISON UNDER DIFFERENT INPUT VECTORS(90NM): A) NOR2 B) NOR3 C) INV. THE TEMPERATURE OF LEAKAGE CURRENT ESTIMATION IS SET TO 400K; DURING NBTI ANALYSIS, RAS IS SET TO 1:9, T_{active}=400K AND $T_{standby}$=330K.

A) NOR2

Input	Leakage(nA)	$\Delta Delay(\%)$
00	2095.1	5.3
01	466.8	4.1
10	290.9	3.9
11	45.7	2.6

B) NOR3

Input	Leakage(nA)	$\Delta Delay(\%)$
000	3218.4	7.6
001	525.2	6.6
010	332.4	6.9
011	53.3	5.9
100	311.7	6.4
101	53.5	5.4
110	50.8	5.8
111	27.2	4.7

C) INV

Input	Leakage (nA)	$\Delta Delay(\%)$
0	1122.8	0.9
1	368.6	0.5

```
Probability based MLV selection algorithm
  0 Generate N random input vectors (IV)
  1 Select an MLV set S within a leakage range
  2 Compute the probability of each primary input
  3 Use the probability to generate new IVs
  4 Calculate circuit leakage using new IVs
  5 If the circuit leakage current is converged
       Output the MLV set
  6 Else
       Jump back to step 1
```

Fig. 7. A probability-based algorithm to select an MLV set.

not very impressive(B.2). Hence, the internal node control techniques [9], [10] is then discussed and shown to be a potential way to mitigate NBTI effect during circuit standby mode(B.3).

1) IVC technique implementation: A leakage lookup table is created by simulating all the gates in the standard cell library under all possible input patterns. Thus the leakage current $I_{leakage}(v)$ can be expressed as:

$$I_{leakage}(v) = \sum_{IN} I_l(v,IN) \times Prob(v,IN) \tag{24}$$

where $I_l(v,IN)$ and $Prob(v,IN)$ are the leakage current and the probability of gate v under input pattern IN.

Finding MLV is proved to be NP-complete; both exact and heuristic approaches have been proposed to search for the MLV [31]–[33]. In this paper, we first find a set of MLV using a simple probability based method; then investigate the impacts of different MLV on the performance degradation due to NBTI; and finally MLV that simultaneously achieve the minimum circuit performance degradation and the maximum leakage reduction rate are selected.

The pseudo-code for our probability-based algorithm to select an MLV set is shown in Fig.7. The probability based algorithm begins by generating N random vectors (line 0); and the leakage current of each vector in the MLV set is within a given range of the minimum leakage current in the set (line 1). Next, for each primary input, the probability is calculated by the number of 1s out of the total number of vectors (line 2). New vectors are generated using the calculated probabilities (line 3). The leakage current of each new input vectors is calculated and the MLV set is updated (line 4). The probabilities for all primary inputs will converge to either 0 or 1, and it means that there is no probability of generating other vectors. So this is the convergence point of circuit leakage current and the algorithm is halted (line 5 and line 6). The probability-based algorithm is just showing a simple example for leakage and NBTI co-optimization. There are a lot of techniques on how to choose the best input vector for leakage reduction. These techniques can be easily modified to target at NBTI mitigation or leakage and NBTI co-optimization.

The algorithm is integrated in the input vector generation module in Fig. 6. Using a circuit logic simulator, the internal state of each edge can be derived for each MLV. The $\Delta d(v)$ for a given period of time of each gate v is evaluated referring to Eq.(22) in Section 2.3. Based on these information, the static timing analysis tool is used to get the overall circuit delay degradation for each MLV in the MLV set. We choose the MLV with the minimum circuit delay degradation to be the one used in the circuit standby mode.

2) Impact of IVC technique: We use probability based MLV selection method to select a set of MLV, in which the leakage current differences of any MLVs are within 4% of the original circuit leakage current. The impact of these MLVs are investigated to find an MLV with minimum circuit performance degradation.

In Table III, we show that the minimized $\Delta delay$ using our IVC technique is about on average 4.3% of the circuit delay; while the performance impact difference of different MLV is about 0.14% of the original circuit delay and 3% of the $\Delta delay$ (In this experiment, the active and standby time ratio RAS= 1:5; the $T_{standby}$=330K), which reveals the impact of MLVs generated by IVC method on the circuit delay.

In Table III, the numbers in "MLV diff" column are very small, because the standby mode temperature is much lower than the active mode temperature. The small value in this column indicates that the input vector control has insignificant effect on mitigating the NBTI effect, which is one of the conclusions we drawn from this paper. However it will be larger with a higher standby mode temperature or if we use a static NBTI model or consider the permanent degradation in the standby circuit degradation model.

TABLE III
IMPACT OF IVC TECHNIQUE ON CIRCUIT PERFORMANCE DEGRADATION. THE LEAKAGE DIFFERENCE OF THE MLV SET IS WITHIN 4% OF THE TOTAL CIRCUIT LEAKAGE.

Circuit	Gate number	Nominal delay(ns)	$\Delta delay$ (%)	MLV's delay diff(%)
c432	169	2.69	4.09	0.22
c499	204	1.99	4.49	0.16
c880	383	2.29	4.17	0.20
c1355	548	2.39	4.28	0.15
c1908	911	2.46	4.19	0.10
c2670	1279	2.99	4.60	0.17
c3540	1699	3.48	4.35	0.11
c5315	2329	2.94	4.24	0.20
c6288	2447	8.54	4.41	0.09
c7552	3566	2.30	4.19	0.09

According to the relationship of RAS vs ΔV_{th} under different $T_{standby}$ shown in Table I, if $T_{standby}$=330K and RAS is set to be 1:1, the delta delay number will be larger, because the total time under high temperature increase. Meanwhile, MLVs not only reduce the leakage of the circuit, but also show lower temporal degradation compared to the worst case when all the internal nodes are set to be 0.

3) Potential of internal node control: The timing and area overhead of the IVC technique, which is caused by the flip-flop at the primary inputs of the circuits, can be neglected in a large digital circuit design; however, for large circuits, the internal states can not be well controlled by the primary input vectors, thus the leakage variance due to different input combinations is not very large, and the MLV's may not result in a significant leakage reduction. Similarly, for larger circuits, the MLV impact on both leakage current and NBTI induced circuit performance degradation is smaller. Different MLV's may not result in large difference of impact on circuit degradation, which is shown in Table III.

Lin *et al.* [9] pointed out that if the internal node deep in the circuit can be manipulated, greater leakage current reduction can be achieved. If the internal nodes can be controlled to reduce the leakage during the circuit standby mode, they can be also controlled to relieve the NBTI impact. Assuming all the PMOS's in the critical paths and near-critical paths are driven by the supply rail during the circuit standby mode (i.e., all PMOS devices are driven by '1'), the circuit performance degradation will be minimized.

We further evaluate the difference of the maximized performance degradation (all the PMOS devices are driven by '0') and the minimized performance degradation (all the internal nodes are driven by '1'). This circuit delay difference compared with the worst case performance degradation is defined as the potential of the internal node control technique.

In Table IV, we show the delay degradation of ISCAS85 benchmarks and the potential of internal node control under different standby mode temperature. The active and standby ratio is set to 1:9. The best case $\Delta delay$, which is derived from setting all the internal nodes to '1', is around 3.32% of the original circuit delay under all the standby temperatures, because the temperature has negligible effect on NBTI relaxation phase. As we can see from the table, when $T_{standby}$ varies from $330K$ to $400K$, the worst case $\Delta delay$, which is derived by setting all the internal nodes to '0', increases from 4.05% to 7.35%. Therefore, the potential of the internal node control increases from 18.1% to 54.9%. Thus the potential of internal node control is larger when the standby temperature becomes higher. Furthermore, the potential will be smaller with a larger active and standby time ratio, because the total time that spends in the standby mode will be decreased.

Not all the internal nodes on the critical or near critical paths can be practically set to '1' or '0' by internal node control techniques; so this potential can be a reference of the largest performance saving by applying internal node control techniques.

D. Impact of sleep transistor insertion technique

In this subsection, we first argue that using PMOS transistors in a sleep transistor insertion design needs to consider the NBTI effect. A method to decide the size of a PMOS ST considering NBTI effect is proposed. We then discuss the circuit performance degradation variation due to the usage of sleep transistor insertion.

1) NBTI-aware PMOS sleep transistor sizing: In sleep transistor insertion technique, sleep transistors are used for each group of gates. To find the optimum size of the ST, it is necessary to find the vector that causes the worst case current in that group of gates. This requires simulating the circuit under all possible input values, which is impossible for large circuits. All the previous literatures focused on how to reduce the ST area penalty along with a remarkable leakage saving: Kao [37] described a method to decrease the size of sleep transistors based on the mutual exclusion principle; Anis [38] presented several fast heuristic techniques for efficient gate clustering; Long [39] proposed a distributed sleep transistor network (DSTN) approach which assumes that all the sleep devices are connected to further reduce the area penalty. The size of a PMOS ST is decided as following steps [39], where only one PMOS ST is used.

The load dependent delay $D_{w/o}$ of a gate v without ST is given by:

$$D_{w/o}(v) = \frac{KC_L V_{dd}}{(V_{dd} - V_{thlow})^{\alpha}} \tag{25}$$

where K is a proportional constant, C_L is the load capacitance, V_{thlow} is the threshold voltage in the low V_{th} module, and α is the velocity saturation index for modeling short channel effects [50]. When the sleep transistor is present and the source drain voltage drop is V_{ST}, the gate propagation delay increases to

$$D_w(v) = \frac{KC_L V_{dd}}{(V_{dd} - V_{ST} - V_{thlow})^{\alpha}} \tag{26}$$

Following Eq.(21)(22), the increase in propagation delay can be derived as,

$$\Delta D(v) = \frac{\alpha V_{ST}}{V_{dd} - V_{thlow}} \times D_{w/o}(v) \tag{27}$$

TABLE IV
DELAY DEGRADATION OF ISCAS85 BENCHMARKS UNDER NBTI AND POTENTIAL OF INTERNAL NODE CONTROL (RAS = 1/9).

ISCAS85 Benchmark Circuits	Internal node = 1	Internal node = 0					
	$\Delta delay$ (%)	$T_{standby} = 330K$ $\Delta delay$(%)	Potential (%)	$T_{standby} = 360K$ $\Delta delay$(%)	Potential (%)	$T_{standby} = 400K$ $\Delta delay$(%)	Potential (%)
c432	3.11	3.87	19.8	5.04	38.4	7.31	57.5
c499	3.45	4.19	17.4	5.23	33.9	7.39	53.2
c880	3.21	3.93	18.3	5.08	36.7	7.32	56.1
c1355	3.28	4.01	18.2	5.13	36.0	7.34	55.3
c1908	3.17	3.95	19.8	5.09	37.7	7.33	56.8
c2670	3.51	4.26	17.8	5.28	33.6	7.41	52.7
c3540	3.29	4.07	17.7	5.16	35.1	7.36	54.5
c5315	3.57	3.99	17.4	5.11	35.5	7.34	55.1
c6288	3.24	4.28	16.6	5.29	32.6	7.41	51.9
c7552	3.32	3.94	17.7	5.08	3.62	7.33	55.7
Average	3.32	4.05	18.1	5.15	35.6	7.35	54.9

Notice that when the circuit is active, the sleep transistor's input is "0", hence it is influenced by NBTI, so the gate delay needs some margin that the circuit delay will satisfy the performance requirement. So if the performance degradation has a upper bound, that is $\Delta D(v)/D_{w/o}(v) < \eta$ (in this paper, we use η = 0.05), V_{ST} is rewritten as,

$$\alpha V_{ST} < \eta(V_{dd} - V_{thlow}) \tag{28}$$

$I_{ON}(v)$ is the current flowing through ST in gate v during the active mode, which can be expressed as given by [41]:

$$I_{ON}(v) = \mu_p C_{ox}(W/L)_{ST}(V_{dd} - V_{thST})V_{ST} \tag{29}$$

where V_{thST} is the threshold voltage of the PMOS ST. Thus the $(W/L)_{ST}$ can be expressed as,

$$(W/L)_{ST} > \frac{\alpha I_{ON}(v)}{\eta \mu_p C_{ox}(V_{dd} - V_{thST})(V_{dd} - V_{thlow})} \tag{30}$$

Note that $(W/L)_{ST}$ is regarded as the area of the ST in this paper because transistors are implemented with minimum length in conventional designs. Moreover, for a BBSTI technique, η for each block is assumed to be the same; for an FGSTI technique, η can be different according to different slack attribute of each gate. However, previous work to decide the ST size did not take the NBTI effect into account: V_{thST} increases with time so that the current flowing through the PMOS ST decreases, which will slowdown the circuit performance. Referring to Eq.(30), the area of ST considering NBTI effect $(W/L)_{ST/NBTI}$ can be expressed as:

$$\begin{aligned}(W/L)_{ST/NBTI} &= (\frac{\Delta V_{th}}{V_{dd} - V_{thIN} - \Delta V_{th}} + 1) \times (W/L)_{ST} \\ &= (\Delta(W/L) + 1) \times (W/L)_{ST}\end{aligned} \tag{31}$$

Here $(W/L)_{ST}$ is decided by the Eq.(30), which is a constant if every process parameter and performance requirement is decided. Therefore, the size of a safety PMOS ST depends on the threshold voltage degradation ΔV_{th} (from Eq. (12)) and the initial V_{th} of the PMOS ST.

The worst case is that the circuit is active all the time, thus the PMOS ST is always negative biased under a very high temperature. We show the PMOS ST threshold degradation ΔV_{th} with different RAS and initial V_{th} in Fig.8. Because the input of PMOS ST during standby time is 1 in order to gate the circuit, the threshold degradation ΔV_{th} is not influenced by the standby temperature variations. Thus, we set $T_{active} = 400K$, $T_{standby} = 330K$. The PMOS ST ΔV_{th} increases with a larger RAS and a smaller initial V_{th}, thus the largest ΔV_{th} is $30.3mV$, when the initial $V_{th} = 0.20V$, $RAS = 9/1$; the smallest PMOS ST ΔV_{th} is $6.7mV$, when the initial $V_{th} = 0.40V$, $RAS = 1/9$.

Meanwhile, the change of PMOS ST size $\Delta(W/L)$ is shown in Fig.9 following Eq.(31). A larger ΔV_{th} leads to a larger $\Delta(W/L)$. Therefore, the largest $\Delta(W/L)$ is 3.94%, when the initial $V_{th} = 0.20V$, $RAS = 9/1$; the smallest $\Delta(W/L)$ is 1.13% when the initial $V_{th} = 0.40V$, $RAS = 1/9$.

We have to notice that, with technology scaling down, smaller initial ST V_{th} will be used to enlarge the headroom for the gated logic, thus the impact of NBTI on the PMOS ST will be larger.

2) Circuit performance degradation analysis using sleep transistor insertion: There are three kinds of sleep transistors: footer (only NMOS ST), header (only PMOS ST), and both footer and header, Fig.10 shows the example of the three conditions. In the standby time, the ST's are turned off, and all the internal nodes will be charged or discharged depends on the ST types. If all the footers are turned off, all the internal nodes are high voltage, so that there is no PMOS transistor that is negatively biased, thus the NBTI-induced degradation is mitigated. We will analyze the internal states during the standby time in detail to investigate the impact of these three conditions on circuit performance degradation.

- Footer: The footer itself has no NBTI effect. When the footer is turned off, the inner nodes of the slept circuit will be charged up to a voltage near V_{dd}. This process will

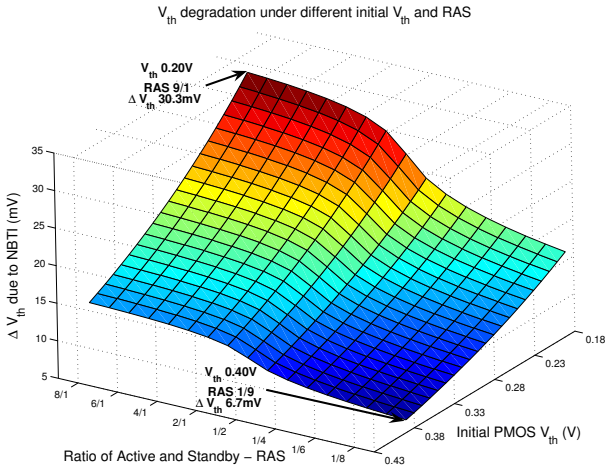

Fig. 8. V_{th} degradation comparison under different initial V_{th} and RAS.

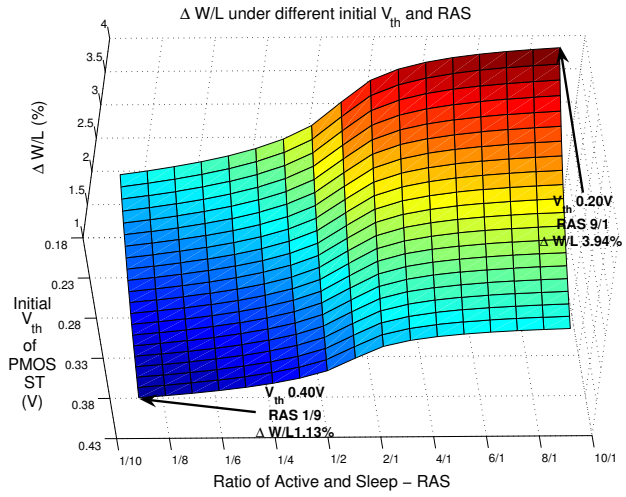

Fig. 9. $\Delta(W/L)$ comparison under different initial V_{th} and RAS.

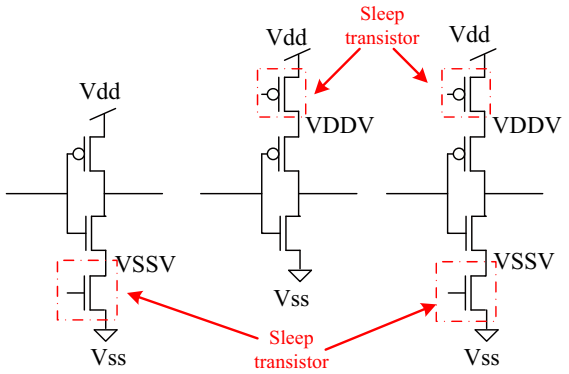

Fig. 10. Example of sleep transistor insertion on an inverter: footer, header, both footer and header.

lower the gate-to-source voltage differences of the PMOS transistors in the circuit, which alleviates the NBTI effect of these PMOS transistors in the standby mode. At the same time, the temperature goes down, along with the gate-to-source voltage differences' going down, the footer can provide more NBTI recovery for the inner circuit. Therefore, using footer as the sleep transistor not only decreases the leakage power, but also slows down the NBTI aging effect.

- Header: Using PMOS ST as a header is not a very good choice in the sleep transistor schemes. As a victim of NBTI, the header will be slowed down by aging. As a result, the V_{th} increase of the header slows down the signal propagation through the circuit, which may cause timing problems. However, since the inner voltage of the circuit during the sleep mode is discharged to a very low voltage near the ground. The header can alleviate the NBTI effect of the inner PMOS transistors in standby mode.
- Footer and header: Using the footer and header at the same time can save more energy, however, as discussed above, the NBTI has no impact on footer but does affect the header a lot. The performance of the circuit will be decreased because of the header's NBTI effect. Since the V_{dd} and ground are both cut off, the PMOS transistors will not be negative biased.

Therefore, the sleep transistor insertion leads to $V_{gs} \approx 0$ for all the internal PMOS transistors. The circuit performance degradation is almost the same as the best case of the internal node control, which is analyzed in the input vector control subsection. Fig.11 shows the potential analysis of C432 with/without ST insertion technique. The worst case $\Delta Delay$ decreases from 7.31% to 3.87% as the $T_{standby}$ decreases from $400K$ to $330K$ when we do not use sleep transistor insertion. However, we also show the circuit degradation numbers with ST inserted under different timing constraints ($\eta = 5\%, 3\%, and 1\%$). It is true that sleep transistor insertion will affect the circuit delay at time0 (additional delay of 5%, 3%, and 1%), however our results show that there exist conditions that we will have a faster circuit at time=10year even if we inserted STs. Hence, lower η will leads to lower degradation at time 10years but high leakage current at the standby time, since the sleep transistor size will be larger. Consequently, we should consider the leakage saving and NBTI degradation simultaneously when we insert sleep transistors.

V. Conclusions and Discussions

In this paper, we propose an improved temporal NBTI-induced performance degradation model for digital circuits. The standby mode temperature and the active and standby time ratio, which have significant impact on circuit performance degradation due to NBTI, are considered in our model. For the first time, we discuss the resemblance between NBTI and leakage mechanisms and then evaluate the impact of different standby leakage reduction techniques on NBTI-induce performance degradation. The impact of the input vector control on performance degradation due to NBTI is limited by the inefficiency of controlling the internal nodes and the low standby time temperature. However, the potential impact of

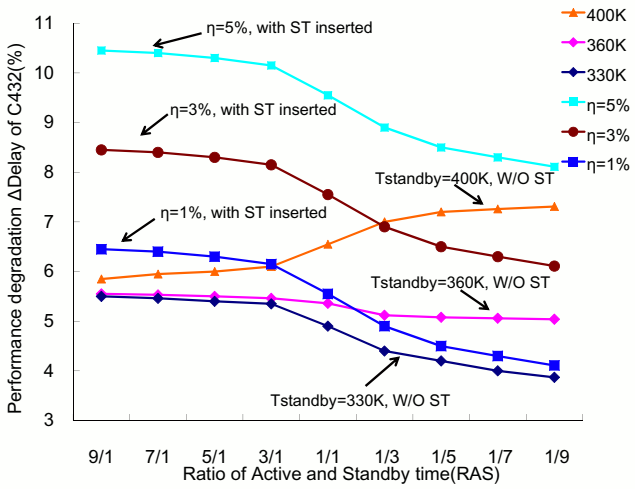

Fig. 11. Circuit degradation of C432 with ST insertion technique (three cases are the circuit without ST insertion under different $T_{standby}$; the other three cases are the circuit with ST insertion under different Time0 timing constraint).

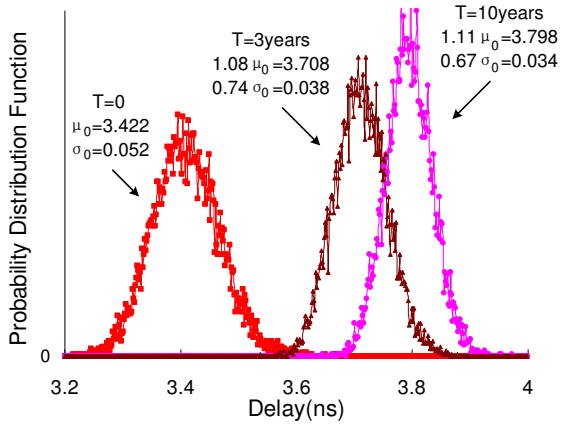

Fig. 12. Circuit performance distribution considering both variation and NBTI induced degradation (C880 in ISCAS85 benchmarks).

internal node control technique may be as large as 54.9%. The impact of NBTI on the PMOS sleep transistor (ST) is analyzed, and a PMOS ST sizing method considering NBTI effect is proposed. Further the impact of sleep transistor insertion on NBTI shows that this technique is efficient to mitigate NBTI effect on circuit performance.

Notice that we only show the potential of the standby leakage reduction techniques on NBTI mitigation. The NBTI model is not limited to R-D based AC NBTI model, since our analysis bases on these facts: 1) NBTI effect depends on the temperature, 2) NBTI and leakage mechanisms both depend on the gate overdrive, 3) the circuit temperature varies between active and standby modes.

Variation is a very important side effect along with the technology scaling, if variations(for example, V_{th} variations and process variations) are considered in the NBTI model, the circuit delay becomes a distribution but not a deterministic value. Along with the circuit life time, this distribution changes monotonously as shown in fig 12. The lower bound (μ-3σ) of delay after 3 years(about 3.599ns) is even larger than the upper bound (μ+3σ) of delay at time 0(about 3.579ns), so NBTI degradation is quite serious. In [51], the authors analyzed the impact of process variations on NBTI-induced degradation. Their results showed that the mean value of gate delay increases with stress time, while the variance decreases, since a lower V_{th} leads to a faster degradation rate, and thus larger V_{th} increase. This phenomenon compensates static process variations and reduces the variance during the stress period. Based on this conclusion, if we may further integrate our temperature-aware NBTI model into a statistical analysis platform, the model will be still effective; meanwhile, the leakage and NBTI co-optimization techniques proposed in this paper will also be effective on a statistical platform.

In this paper, our technique is implemented on high performance devices, actually the main purpose of this work is to study the impact of leakage control methods on NBTI effect. As a matter of fact, the "LP" library usually uses a typical higher "VTH" for standard cells. According to [20], for such "LP" standard cells with higher VTH, the impact of NBTI is much smaller than the case of high performance application. In addition, the temperature for high performance applications are usually higher than those batter powered systems, and consequently, the NBTI degradation is much more serious. That is why we focus on high performance application in this paper. However, the trend of the effect of these techniques should be similar.

VI. Acknowledgment

The authors would like to thank Sanjay Kumar from MNU, Wenping Wang and Prof. Yu Cao from Arizona State University for great discussions on the NBTI modeling and mitigation techniques. Especially thank should be given to Xiaoming Chen who helps me with the paper finalization.

References

[1] V. Huard, M. Denais, and C. Parthasarathy, "NBTI degradation: From physical mechanisms to modelling," *Microelectronics Reliability*, vol. 46, no. 1, pp. 1–23, 2006.

[2] S. Nassif, K. Bernstein, D. Frank, A. Gattiker, W. Haensch, B. Ji, E. Nowak, D. Pearson, and N. Rohrer, "High performance cmos variability in the 65nm regime and beyond," *IEDM 2007.*, pp. 569–571, Dec. 2007.

[3] J. Stathis and S. Zafar, "The negative bias temperature instability in MOS devices: A review," *Microelectronics Reliability*, vol. 46, no. 2-4, pp. 270–286, 2006.

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

[5] S. Mahapatra, P. Bharath Kumar, T. Dalei, D. Sana, and M. Alam, "Mechanism of negative bias temperature instability in CMOS devices: degradation, recovery and impact of nitrogen," in *Tech. Dig. Intl. Elec. Dev. Meeting*, 2004, pp. 105–108.

[6] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "An Analytical Model for Negative Bias Temperature Instability," in *IEEE/ACM Intl. Conf. on Computer-Aided Design*, 2006.

[7] K. Roy, S. Mukhopadhay, and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer cmos circuits," *Proceedings of the IEEE*, vol. 91, no.2, pp. 305–327, Feb. 2003.

[8] S. Bhardwaj, W. Wang, R. Vattikonda, Y. Cao, and S. Vrudhula, "Predictive modeling of the nbti effect for reliable design," in *Proc. CICC*, 2006.

[9] L. Yuan and G. Qu, "A combined gate replacement and input vector control approach for leakage current reduction," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, vol. 14, no. 2, pp. 173–182, 2006.

[10] H. Rahman and C. Chakrabarti, "An Efficient Control Point Insertion Technique for Leakage Reduction of Scaled CMOS Circuits," *IEEE Trans. on Cir. & Syst. II: Express Briefs*, vol. 52, no. 8, pp. 496–500, 2005.

[11] A. Goetzberger and H. Nigh, "Surface charge after annealing of alcsio2csi structures under bias," *Proceedings of the IEEE*, vol. 54, p. 1454, 1966.

[12] A. Goetzberger, A. Lopez, and R. Strain, "On the formation of surface states during stress aging of thermal sicsio2 interfaces," *Journal Electrochem Soc*, vol. 120, pp. 90–96, 1973.

[13] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *Journal of Applied Physics*, vol. 94, no. 1, pp. 1–18, 2003.

[14] S. Mahapatra and M. Alam, "A predictive reliability model for PMOS bias temperature degradation," in *Tech. Dig. Intl. Elec. Dev. Meeting*, 2002, pp. 505–508.

[15] B. Zhu, J. Suehle, J. Bernstein, and Y. Chen, "Mechanism of dynamic nbti of pmosfets," in *Integrated Reliability Workshop Final Report, 2004 IEEE International*, 2004, pp. 113–117.

[16] V. Huard and M. Denais, "Hole trapping effect on methodology for dc and ac negative bias temperature instability measurements in pmos transistors," in *Annual Proceedings - Reliability Physics (Symposium)*. Phoenix, AZ., United States: Institute of Electrical and Electronics Engineers Inc., Piscataway, United States, 2004, pp. 40–45.

[17] M. Alam, "A critical examination of the mechanics of dynamic nbti for pmosfets," in *Tech. Dig. Intl. Elec. Dev. Meeting*, 2003, p. 14.4.1C14.4.4.

[18] S. Mahapatra, P. Kumar, and M. Alam, "Investigation and modeling of interface and bulk trap generation during negative bias temperature instability of p-MOSFETs," *IEEE Trans. on Elec. Dev.*, vol. 51, no. 9, pp. 1371–1379, 2004.

[19] B. Paul, K. Kang, H. Kufluoglu, M. Alam, and K. Roy, "Impact of NBTI on the temporal performance degradation of digital circuits," *Electron Device Letters, IEEE*, vol. 26, no. 8, pp. 560–562, 2005.

[20] R. Vattikonda, W. Wang, and Y. Cao, "Modeling and Minimization of PMOS NBTI Effect for Robust Nanometer Design," in *Proc. of Design Automation Conference*, 2006, pp. 1047–1052.

[21] S. Kumar, C. Kim, and S. Sapatnekar, "Impact of NBTI on SRAM Read Stability and Design for Reliability," in *Intl. Symp. on Quality Electronic Design*, 2006, pp. 210–218.

[22] B. Paul, K. Kang, H. Kufluoglu, M. Alam, and K. Roy, "Temporal Performance Degradation under NBTI: Estimation and Design for Improved Reliability of Nanoscale Circuits," in *Proc. of Design, Automation and Test in Europe*, vol. 1, 2006, pp. 1–6.

[23] J. Abella, X. Vera, and A. Gonzalez, "Penelope: The nbti-aware processor," *MICRO 2007*, pp. 85–96, Dec. 2007.

[24] L. Zhang, R. P.Dick, and L. Shang, "Scheduled voltage scaling for increasing lifetime in the presence of nbti," *ASP-DAC 2009.*, pp. 492–497, 2009.

[25] S. Kumar, C. Kim, and S. Sapatnekar, "Adaptive techniques for overcoming performance degradation due to aging in digital circuits," *ASP-DAC 2009.*, pp. 284–289, 2009.

[26] K. Kang, S. Gangwal, S. Park, and K. Roy, "Nbti induced performance degradation in logic and memory circuits: how effectively can we approach a reliability solution?" *ASPDAC 2008*, pp. 726–731, March 2008.

[27] S. Naffziger, B. Stackhouse, T. Grutkowski, D. Josephson, J. Desai, E. Alon, and M. Horowitz, "The implementation of a 2-core, multi-threaded itanium family processor," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 1, pp. 197–209, Jan. 2006.

[28] K. Skadron, M. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan, "Temperature-aware microarchitecture," *ISCA 2003.*, pp. 2–13, 2003.

[29] A. Agarwal, S. Mukhopadhyay, R. A., K. Roy, and C. Kim, "Leakage power analysis and reduction for nanoscale circuits," *IEEE MICRO*, pp. 68–80, 2006.

[30] Q. Wang and S. Vrudhula, "Algorithms for minimizing standby power in deep submicrometer, dual-vt cmos circuits," *IEEE Trans on CAD*, vol. 21, Issue. 3, pp. 306–318, March 2002.

[31] A. Abdollahi, F. Fallah, and M. Pedram, "Leakage current reduction in CMOS VLSI circuits by input vector control," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, vol. 12, no. 2, pp. 140–154, 2004.

[32] F. Gao and J. Hayes, "Exact and heuristic approaches to input vector control for leakage power reduction," in *IEEE/ACM Intl. Conf. on Computer Aided Design*, 2004, pp. 527–532.

[33] R. Rao, F. Liu, J. Burns, and R. Brown, "A heuristic to determine low leakage sleep state vectors for CMOS combinational circuits," in *IEEE/ACM Intl. Conf. on Computer Aided Design*, 2003, pp. 689–692.

[34] M. Johnson, D. Somasekhar, and K. Roy, "Models and algorithms for bounds on leakage in cmos circuits," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 18, no. 6, pp. 714–725, 1999.

[35] D. Lee, W. Kwong, D. Blaauw, and D. Sylvester, "Analysis and minimization techniques for total leakage considering gate oxide leakage," in *Proc. of Design Automation Conference*, 2003, pp. 175–180.

[36] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-v power supply high-speed digital circuit technology with multithreshold-voltage cmos," *IEEE J. Solid-State Circuits*, vol. 30, no. 8, pp. 847–854, Aug 1995.

[37] J. Kao, S. Narendra, and A. Chandrakasan, "Mtcmos hierarchical sizing based on mutual exclusive discharge patterns," in *Proc. Design Automation Conf.*, 1998, pp. 495–500.

[38] M. Anis, S. Areibi, and M. Elmasry, "Dynamic and leakage power reduction in mtcmos circuits using an automated efficient gate clustering technique," in *Proc. Design Automation Conf.*, 2002, pp. 480–485.

[39] C. Long and L. He, "Distributed sleep transistor network for power reduction," *IEEE Trans on VLSI*, vol. 12, Issue. 9, pp. 937–946, Sept. 2004.

[40] B. Calhoun, F. Honor, and A. Chandrakasan, "A leakage reduction methodology for distributed mtcmos," *IEEE J. Solid-State Circuits*, vol. 39, no. 5, pp. 818–826, May 2004.

[41] V. Khandelwal and A. Srivastava, "Leakage control through fine-grained placement and sizing of sleep transistors," in *Proc. Int. Conf. on Computer-Aided Design*, 2004, pp. 533–536.

[42] Y. Wang, H. Lin, R. Luo, H. Yang, and H. Wang, "Simultaneous fine-grain sleep transistor placement and sizing for leakage optimization," in *Proc. Int. Symp. on Quality of Electronic Designs*, 2006, pp. 723–728.

[43] Nanoscale Integration and Modeling (NIMO) Group, ASU. Predictive Technology Model (PTM). [Online]. Available: http://www.eas.asu.edu/~ptm/

[44] Y. Wang, H. Yang, and H. Wang, "Signal-path level dual-vt assignment for leakage power reduction," *Journal of Circuits, System and Computers*, vol. 15, no. 2, pp. 179–216, 2006.

[45] F. Kreith, Ed., *CRC Handbook of Thermal Engineering*. CRC Press, 2000.

[46] S. Kevin, S. Mircea R., H. Wei, V. Sivakumar, S. Karthik, and T. David, "Temperature-aware computer systems: Opportunities and challenges," in *IEEE Micro*, vol. 23, no. 6, 2003, pp. 52–61.

[47] A. T. Krishnan, S. Chakravarthi, P. Nicollian, V. Reddy, and S. Krishnan, "Negative bias temperature instability mechanism: The role of molecular hydrogen," *Applied Physics Letters*, vol. 88, no. 15, pp. 153 518–, 2006.

[48] C. Liu, M. Lee, C.-Y. Lin, J. Chen, K. Schruefer, J. Brighten, N. Rovedo, T. Hook, M. Khare, S.-F. Huang, C. Wann, T.-C. Chen, and T. Ning, "Mechanism and process dependence of negative bias temperature instability (nbti) for pmosfets with ultrathin gate dielectrics," in *Electron Devices Meeting, 2001. IEDM Technical Digest. International*, 2001, pp. 39.2.1–39.2.4.

[49] Y. Wang, X. Chen, W. Wang, V. Balakrishnan, Y. Cao, Y. Xie, and H. Yang, "On the efficacy of input vector control to mitigate nbti effects and leakage power," *Quality of Electronic Design, 2009. ISQED 2009. Quality of Electronic Design*, pp. 19–26, March 2009.

[50] T. Sakurai and A. Newton, "Alpha-power law mosfet model and its applications to cmos inverter delay and other formulas," *IEEE J. Solid-State Circuits*, vol. 25, pp. 584–594, Apr. 1990.

[51] W. Wang, V. Reddy, B. Yang, V. Balakrishnan, S. Krishnan, and Y. Cao, "Statistical prediction of circuit aging under process variations," *Custom Integrated Circuits Conference, 2008. CICC 2008. IEEE*, pp. 13–16, Sept. 2008.

Yu Wang (S'05-M'07) Yu Wang received his B.S. degree in 2002, and then Ph.D. degree with honor in NICS Group, Electronics Engineering Department, Tsinghua University in 2007, supervised by Prof. Huazhong Yang (Tsinghua University) and Prof. Yuan Xie (Penn. State University). He is now an Assistant Prof. in E.E. Dept., Tsinghua University. Dr. Wang's research mainly focuses on fast circuit analysis, low power circuit design methodology, reliability-aware circuit design methodology, application specific FPGA design, and on-chip communication strategies for MPSOC. Dr. Wang has authored and co-authored over 50 papers in refereed journals and conferences. Dr. Wang is also TPC member for several conferences, such as ICCAD, ISQED, ISVLSI, ISLPED and etc.

Hong Luo Hong Luo received his B.S. degree in 2003 and PhD. degree in 2009 in Circuits and Systems Division, E.E. Dept, Tsinghua University. His research interests includes leakage current modeling and optimization, reliability aware modeling.

Ku He Ku He received the B.S. degree in electronic engineering in 2004, and the M.E. degree in electronic engineering in 2007, both from Tsinghua University. He is currently working for the Ph.D degree in University of Texas - Austin, on the track of computer engineering. His interests include statistical and robust circuit optimization.

Rong Luo Rong Luo received her double B.S. in Engineering Physical and Electronic Engineering from Tsinghua University in 1992 and her Ph.D. from Tsinghua University in 1997. Currently, she is an associate Professor in the Department of Electronic Engineering, Tsinghua University, Beijing. Now, her research work is mainly on SoC Design Technology, VLSI Design and Embedded System Design Technology.

Huazhong Yang (M'97-SM'00) Huazhong Yang was born in Ziyang, Sichuan Province, P.R.China, on Aug. 18, 1967. He received B.S. degree in microelectronics in 1989, M.S. and Ph.D. degree in electronic engineering in 1993 and 1998, respectively, all from Tsinghua University, Beijing. In 1993, he joined the Department of Electronic Engineering, Tsinghua University, Beijing, where he has been a Full Professor since 1998. Dr. Yang was recognized as '2000 National Palmary Young Researcher by NSFC. His research interests include chip design for communication and multimedia applications, synthesis of analog integrated circuits (IC), power estimation and synthesis of digital ICs, noise and delay estimation of deep submicron ICs, yield enhancement, optimization and modeling. Dr. Yang has been in charge of several projects, including projects sponsored by the 863 program, NSFC, the 9th five-year national program and several international cooperation projects. Dr. Yang has authored and co-authored over 100 technical papers and 6 books.

Yuan Xie Yuan Xie is Associate Professor in Computer Science and Engineering department at The Pennsylvania State University, University Park, PA. He received the B.S. degree in electronic engineering from Tsinghua University in Beijing, the M.S. and Ph.D. degrees in electrical engineering from Princeton University. He was a recipient of the SRC Inventor Recognition Award in 2002, NSF CAREER award in 2006, and IBM Faculty Award in 2009. He also received Best Paper Award in ASP-DAC 2008. He has published more than 100 technical papers in the areas of computer architecture, design automation, VLSI design, and embedded system.