On the Efficacy of NBTI Mitigation Techniques

Tuck-Boon Chan*, John Sartori†, Puneet Gupta* and Rakesh Kumar†
† ECE Dept., University of Illinois at Urbana-Champaign. {sartori2,rakeshk}@illinois.edu
* EE Dept., University of California, Los Angeles. {tuckie,puneet}@ee.ucla.edu

***Abstract*—Negative Bias Temperature Instability (NBTI) has become an important reliability issue in modern semiconductor processes. Recent work has attempted to address NBTI-induced degradation at the architecture level. However, such work has relied on device-level analytical models that, we argue, are limited in their flexibility to model the impact of architecture-level techniques on NBTI degradation.**

In this paper, we propose a flexible numerical model for NBTI degradation that can be adapted to better estimate the impact of architecture-level techniques on NBTI degradation. Our model is a numerical solution to the reaction-diffusion equations describing NBTI degradation that has been parameterized to model the impact of dynamic voltage scaling, averaging effects across logic paths, power gating, and activity management. We use this model to understand the effectiveness of different classes of architecture-level techniques that have been proposed to mitigate the effects of NBTI. We show that the potential benefits from these techniques are, for the most part, smaller than what has been previously suggested, and that guardbanding may still be an efficient way to deal with aging.

I. Introduction

Device degradation due to NBTI has become a major concern [2], [6], [10]. NBTI manifests itself as an increase in the magnitude of V_{th} whenever a PMOS transistor is negatively biased. This causes delay to increase, and if not properly provisioned for, can result in timing violations. Recently, studies have proposed techniques at various design abstraction levels, from the circuit level [8], [12], [14]–[16], [27], [28] to the architecture level [1], [7], [13], [18]–[20], [22], to alleviate the impact of NBTI-induced degradation.

At the architecture level, techniques have been proposed to bias input vectors to mitigate aging [1], enhance throughput at the expense of aging in a multi-core environment [13], monitor and adapt to estimated processor lifetime [19], [20], perform aging-aware scheduling [18], and apply voltage scaling [22] or power gating [7] to mitigate the effects of aging.

The techniques proposed by previous architecture-level works, as well as their evaluations, are primarily based on device-level analytical models [20], [22], [24], [25]. While these analytical models do well at estimating the impact of NBTI degradation on the speed of a *device*, we argue in this paper that they are not general enough to model the wide range of adaptations and operating scenarios employed by architecture-level NBTI-mitigation techniques. Unfortunately, these models have been applied, as is, in the previous evaluations. Thus, the accuracy of these evaluations may be limited. This is especially true considering that, as we will show, conclusions related to NBTI are strongly dependent on the nature of NBTI degradation.

We make the following contributions.

- We develop a flexible, adaptable numerical simulation engine for NBTI-induced aging, based on the reaction-diffusion model, that allows us to emulate NBTI degradation and the impact of aging mitigation techniques under various operating conditions, including different voltage scaling, power gating, and activity management scenarios.
- We revisit techniques aimed at mitigating the effects of NBTI-induced aging, evaluate their effectiveness in our adaptable simulation framework, and identify any potential limitations in previously accepted conclusions about the techniques.
- We contribute a better, more confident understanding of how architecture-level techniques impact NBTI degradation and demonstrate that the potential benefits from NBTI mitigation at the architecture-level are, in most cases, smaller than what has previously been suggested.

The rest of the paper is organized as follows. Section II discusses the basics of NBTI degradation and modeling as well as modeling-related limitations of the previous works on NBTI mitigation. Section III describes our flexible, numerical modeling framework that can be used to better estimate the impact of architecture-level techniques on NBTI degradation. Section IV describes the methodology used in this paper to re-evaluate the effectiveness of previous architecture-level techniques using the proposed modeling framework. Section V presents results. Section VI summarizes the paper.

II. Background and Related Work

A. NBTI Overview

NBTI manifests itself as an increase in $|V_{th}|$, and consequently, an increase in logic delay, whenever a PMOS transistor is under stress ($|V_{gs}| > |V_{th}|$). Relaxation of the stress ($V_{gs} = 0$) can recover only part of the V_{th} degradation [3], causing an overall increase in delay over time (*NBTI degradation*). If not appropriately provisioned for, increased delay can result in timing failures on critical logic paths. NBTI degradation is frequency independent [3], [24] but increases with supply voltage (V_{dd}) and temperature. Also, due to the underlying physical phenomena that cause NBTI, the degradation is "front-loaded" by nature. As illustrated in Figure 5, this means that the rate of degradation is rapid in the early lifetime and slows down considerably under continued stress. Front-loaded degradation is a general characteristic of NBTI, independent of the process. For example, Figure 6 shows the front-loaded nature of NBTI degradation for three different processes.

Traditionally, guardbanding has been used to protect against NBTI. I.e., operating frequency is reduced or supply voltage is increased to account for degradation over the lifetime of a design, such that there are no timing violations due to aging during the lifetime. Unfortunately, guardbanding incurs a throughput or power cost over the entire lifetime of a design, even though NBTI degradation does not fully accumulate until the end of the lifetime. As such, several dynamic, architecture-level approaches (discussed in Section II-B) have been proposed to mitigate NBTI degradation. Evaluation of architecture-level approaches to mitigate NBTI degradation is typically based on analytical degradation models, like Equation 1 [22]:

$$\Delta V_{th} = A_{NBTI} \cdot \tau_{ox} \cdot \sqrt{C_{ox}(V_{dd} - V_{th})} \cdot e^{\frac{V_{dd}-V_{th}}{\tau_{ox}E_0} - \frac{E_a}{kT}} \cdot t_{stress}^{0.25} \quad (1)$$

where t_{stress} is stress time, τ_{ox} is oxide thickness, C_{ox} is gate capacitance per unit area, E_0, E_a, and k are fitting constants, and A_{NBTI} is a constant that depends on the aging rate.

The above model describes NBTI degradation over time at the device level. Using a device-level model to evaluate architecture-level techniques may limit the accuracy of evaluations, since device-level models do not account for scenarios like dynamic voltage scaling, averaging effects across logic paths, and different activity and power management schemes used in architecture-level techniques. In the next section, we discuss specific classes of architecture-level NBTI

978-3-9810801-7-9/DATE11/©2011 EDAA

mitigation techniques and the limitations of device-level models in characterizing their impact.

B. Architecture-level Techniques for Mitigating NBTI

1) Dynamic Voltage Scaling: Dynamic voltage scaling (DVS) has been proposed as a technique to mitigate aging in modern processors. Previous works [8], [16] have proposed that rather than using a fixed guardband over the entire lifetime of a processor, aging can be reduced by using a lower supply voltage early in a processor's lifetime and increasing the voltage as necessary to counteract the effects of aging. Facelift [22] is a specific application of DVS in which the supply voltage is only adapted once during the processor's lifetime to switch the processor from a slow aging mode to a high speed mode. Bubblewrap [13] uses techniques based on Facelift to enhance performance in a multi-core processor.

A limitation that may lead to inaccuracies in these works is that they manipulate the NBTI degradation relationship (e.g., Equation 1 [22]) by changing V_{dd} without modifying the time-dependent aging rate ($k \times t^{0.25}$). When V_{dd} is changed, the time t must be redefined to an equivalent value on the new aging curve defined by the new voltage. We describe this with a specific example in Section III.

Another issue with DVS-related works is that they demonstrate that lifetime can be significantly extended by using DVS. Intuitively, this makes sense, because the rate of degradation decreases with voltage. Due to the front-loaded nature of NBTI, however, power or aging benefits of using a lower voltage are possible in the early lifetime, but degradation soon converges to that found in the guardbanded case. We will show in Section V, that DVS cannot significantly extend processor lifetime for any case we studied.

2) Lifetime Awareness: Other works [19], [20] make a case for processors that monitor and adapt to the estimated processor lifetime, based on operating conditions, in order to ensure that a processor reaches a desired lifetime target before failing. These papers model aging such that failures are averaged over the entire lifetime, which assumes that degradation happens steadily over processor lifetime, rather than in a front-loaded nature. This may lead to inaccuracies, especially since we find that the benefits of NBTI mitigation techniques strongly depend on the nature of NBTI degradation. In fact, we show in Section V that benefits of lifetime-aware adaptation may not be significant if a realistic degradation model is used.

3) Dynamic Instruction Scheduling: Some works have suggested policies for scheduling instructions to control or limit aging by controlling the activity factor or utilization of functional units [18]. We find that benefits are highly sensitive to the processor configuration and amount of available hardware redundancy, which determine how much functional units will be stressed during the processor's lifetime. Since these works have not considered the sensitivity of benefits to such parameters, the generality of conclusions may be limited. In fact, we show in Section V that due to the front-loaded nature of NBTI, degradation on functional units converges after the early lifetime, and in order to achieve a significant (15%) reduction in degradation, a functional unit must be inactive for the majority (99%) of its lifetime.

4) Power Gating: Power gating [7] has been proposed as a technique to mitigate aging, since PMOS stress is removed during periods of power gating. The benefits of power gating are highly sensitive to the fraction of time that a circuit spends in sleep mode. In fact, we observe that the front-loaded nature of NBTI causes degradation to converge quickly unless the majority of the lifetime is spent in sleep mode. Typically, substantial performance degradation must be accepted to achieve such high power gating factors.

III. Proposed NBTI Model

A reaction-diffusion (R-D) model is often used to explain the NBTI phenomenon [4]. The R-D model states that the V_{th} shift in a negatively biased PMOS is driven by inversion layer holes interacting with hydrogen-passivated Si atoms. The energized holes can break $Si-H$ bonds at the Si/SiO_2 interface, creating an interface trap and a H atom. The formation of interface traps and the H atom diffusion mechanism are described by the following differential equations [4]:

$$\begin{aligned}
&\text{Reaction at surface}\\
\frac{\partial N_{it}(t)}{\partial t} &= k_f[N_0 - N_{it}(t)] - k_r N_{it}(t) C_H(x=0,t),\\
\frac{\partial N_{it}(t)}{\partial t} &= -D\frac{\partial C_H(x,t)}{\partial x}|_{x=0} + \frac{\delta}{2}\frac{\partial C_H(x,t)}{\partial t},\\
&\text{Diffusion in silicon oxide or poly}\\
D\frac{\partial^2 C_H(x,t)}{\partial x^2} &= \frac{\partial C_H(x,t)}{\partial t}
\end{aligned} \tag{2}$$

where $N_{it}(t)$ is the number of interface traps per unit area at time t, k_f is the dissociation rate of $Si-H$ bonds, k_r is the annealing rate of $Si-H$ bonds, D is the diffusion coefficient of the diffusing hydrogen species (H or H_2), N_0 is number of initial $Si-H$ bonds at the interface when $t=0$, $C_H(x,t)$ is the number of hydrogen (atoms or molecules) per unit area at location x and time t, and δ denotes the interface thickness. The generation of N_{it} causes a V_{th} shift, given by [4]:

$$\Delta V_{th} = \frac{qN_{it}}{C_{ox}}, \tag{3}$$

where q is elementary charge and C_{ox} is PMOS gate capacitance.

Before describing the details of our model, we present two quantitative examples of factors that cause the results from previously used modeling approaches to deviate from the results provided by our numerical model (several such factors were discussed in Section II-B). The first factor is inadequate assessment of the impact of dynamic voltage scaling on NBTI degradation. Figure 1 compares the degradation profiles for a V_{dd} level switch without redefining the equivalent degradation time corresponding to the new V_{dd} level (analytical method) and our numerical simulator. The degradation rate of the analytical method clearly deviates from that obtained from numerical simulation. The deviation arises because the analytical equations do not model the physical degradation phenomenon as our numerical model does. Changing the voltage without changing the time is like instantaneously changing the internal state of the device to reflect a time in the future (voltage increase) or the past (voltage decrease). Such deviation may lead to inaccurate evaluations during degradation analysis.

Another factor that causes the results to deviate is that the previous analytical models approximate signals in circuits as AC signals [4], [14], [24], [25], [27], [29]. However, these AC signals do not resemble typical digital signals in CMOS circuits, like the one illustrated in Figure 3. Note that due to the inverting nature of CMOS logic, a logical one (relaxation state) at a node implies a logical zero (stress state) at the next node. Specifically, if PMOS at one node are relaxed, PMOS at the subsequent node are under stress, or vice versa. For circuit and architecture-level analysis, it is necessary to model the inverting stress and relaxation states in CMOS logic, because there is an averaging effect in the degradation along a path (timing analysis) or across an entire design (power analysis).

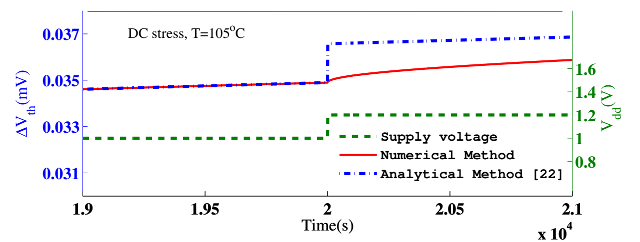

Fig. 1: Applying the analytical degradation model out of context can lead to significant deviation from the numerical solution.

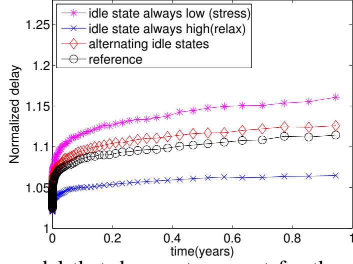

Fig. 2: A signal model that does not account for the averaging effect of CMOS logic (idle always high or low) can cause $\pm5\%$ delay estimation error. The signal pattern with alternating idle states reduces estimation error to less than 1%.

To study the impact of the averaging effect, we simulate NBTI degradation of an eleven stage inverter chain with different modeling approaches. The inverter chain is driven by a periodic waveform of 0.05s AC signal followed by 0.95s DC signal in every cycle. We obtain the exact delay degradation (reference) by calculating V_{th} degradation for each PMOS according to its bias condition. Although this method is accurate, obtaining the exact signal of every node in a modern VLSI is not practical. For architectural NBTI analysis, it is common to estimate V_{th} degradation of a PMOS with an approximate waveform and assume all PMOS in a circuit degrade by the same amount [13], [18], [22]. To model the averaging effect, we use the waveform in Figure 4 with alternating states in every idle period during device-level NBTI simulation. The idle and active periods in Figure 4 correspond to the DC and AC signals of the inverter chain. For comparison, we also simulate test cases with waveforms that have always high or low signals during the idle period. The results in Figure 2 show that ignoring the averaging and inverting signal pattern can cause a $\pm5\%$ difference in delay estimation. The error is reduced to less than 1% when we use the waveform in Figure 4. To ensure that we estimate NBTI effects accurately under the various operating conditions, such as those used in architecture-level NBTI mitigation techniques, we solve Equation 2 numerically in our experiments.

A. A Flexible, Numerical Implementation of the R-D Model

Since the trap generation rate is usually small compared to the dissociation and annealing rates [3], [25], i.e.,

$$\frac{\partial N_{it}(t)}{\partial t} \approx 0, \text{ and}$$
$$N_{it(t)} << N_0,$$

Equation (2) reduces to

$$C_H(x=0,t)N_{it}(t) \approx \frac{k_f}{k_r}N_0,$$
$$D\frac{\partial^2 C_H(x,t)}{\partial x^2} = \frac{\partial C_H(x,t)}{\partial t}.$$

To solve these differential equations numerically, we discretize the equation based on the finite difference method with spatial Δx and temporal Δt increments to obtain the following equations:

$$\alpha = \frac{D\Delta t}{\Delta x^2}$$

$$C_H(x_0,t_i) = \begin{cases} [\frac{k_f N_0}{k_r N_{it}(t_i)}]^S & \text{if device under stress} \\ 0 & \text{if device under relaxation} \end{cases} \quad (4a)$$

$$\mathbf{C}_H(t_{i+1}) = \mathbf{W}\mathbf{C}_H(t_i), \quad (4b)$$

$$N_{it}(t_{i+1}) = S[1,1,\ldots,1]\mathbf{C}_H(t_i+1), \quad (4c)$$

$$\mathbf{C}_H(t) = \begin{bmatrix} C_H(x_0,t) \\ \vdots \\ C_H(x_n,t) \end{bmatrix},$$

$$\mathbf{W} = \begin{bmatrix} 1-\alpha & 1 & 0 & 0 & 0 & \ldots \\ 1 & 1-2\alpha & 1 & 0 & 0 & \ldots \\ 0 & 1 & 1-2\alpha & 1 & 0 & \ldots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix},$$

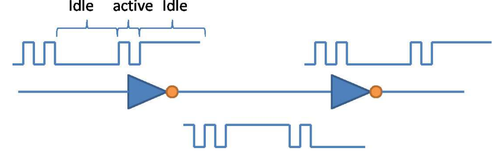

Fig. 3: Signal in a typical digital system.

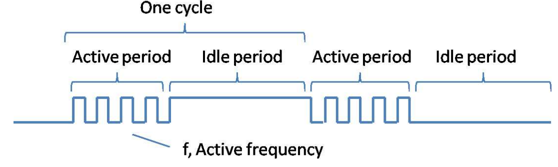

Fig. 4: Periodic signal model for NBTI degradation estimation.

where n is the total number of grid points (locations in oxide or polysilicon normal to the channel surface), x_j denotes the jth location along the one-dimensional space of the Si/SiO_2/polysilicon stack, and t_i denotes the ith time step, $\mathbf{C}_H$ is a column vector that represents the hydrogen profile in PMOS. Parameter S is included to account for different diffusion species ($S=1$ for H and $S=2$ for H_2) [25].

At each time step, we calculate the value of N_{it} from the previous diffusion profile using Equations 4b-4c. Then, we update the hydrogen density value at the interface, $C_H(x_0,t_i)$, depending on the signal at the time, using Equation 4a. We notice that $C_H(x_0,t_i)$ changes very slowly when the time step is small. To reduce simulation time, we approximate $C_H(x_0,t_i)$ as a fixed value for k time steps. I.e.,

$$C_H(x_0,t_i) = \begin{cases} [\frac{k_f N_0}{k_r N_{it}(t_i)}]^S & \text{if device under stress} \\ 0 & \text{if device under relaxation} \end{cases} \quad (5a)$$

$$\mathbf{C}_H(t_{i+k}) = \mathbf{W}^k\mathbf{C}_H(t_i), \quad (5b)$$

$$N_{it}(t_{i+k}) = S[1,1,\ldots,1]\mathbf{C}_H(t_i+k), \quad (5c)$$

This method is different than applying a larger time step, as it implicitly calculates the changes of the hydrogen diffusion profile over k time steps. As a result, it reduces the computation time by a factor of k (with a little overhead to pre-compute $\mathbf{W}^k$) with no loss in accuracy.

The dependency between NBTI degradation and the field applied to a device is given by [31]

$$N_0 = AC_{ox}(V_{gs}-V_{th})exp(E_{ox}/E_0),$$
$$E_{ox} = V_{gs}/\tau_{ox},$$

where τ_{ox} is oxide thickness, and A and E_0 are fitting parameters. Note that V_{gs} only affects N_0. This allows us to model a dynamic change in supply voltage by applying the corresponding V_{gs} value when we evaluate Equation 5a. In this paper, we calibrated the parameters in our model to a 65nm commercial process, $k_r = 10^3 nm^3 s^{-1}$, $k_f = 0.01 s^{-1}$, $A = 5.93\times10^9 nm^{-2}$, $E_0 = 1.5V$, and $D = 29.288 nm^2 s^{-1}$ at $T = 105^o$C. Figure 5 shows that our NBTI estimation using either Equation 5 or Equation 4 is consistent with measurement data in [26]. In Figure 6, we compare Id_{sat} degradation for our model to the silicon measurements in [5] and [30]. The figure shows that our model has higher degradation compared to the other two processes throughout a 10 year lifetime. Therefore, our experiment setup is more likely to amplify the significance of NBTI mitigation techniques, since the potential benefit of these techniques increases with the magnitude of NBTI degradation.

We have developed an open source simulation framework for our model and made it available for download.[1] We hope that the

[1] The setup is publicly available for download at http://nanocad.ee.ucla.edu/Main/DownloadForm

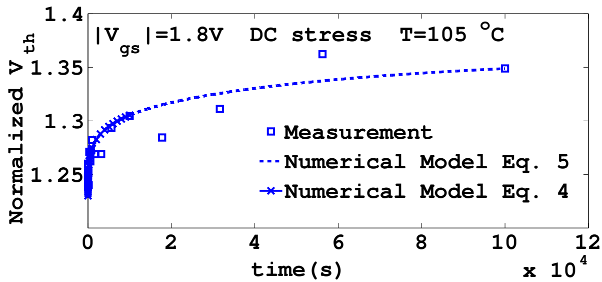

Fig. 5: Our NBTI model vs. measurement data in [26].

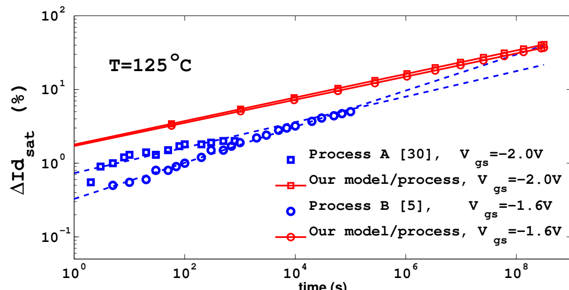

Fig. 6: Id_{sat} degradation of different processes under DC stress.

simulator can be leveraged to enhance research toward other NBTI mitigation techniques.

IV. Experimental Methodology

To evaluate the impact of architecture-level techniques on NBTI degradation, we performed a study based on a commercial 65nm technology with 1V nominal supply voltage. To estimate device level NBTI degradation, we apply Equation 5 with the waveform illustrated in Figure 4. We use signal period = 1s, active frequency = 10kHz, and activity factor = 0.5 in all experiments, unless otherwise specified. When signal frequency is greater than 100Hz, NBTI degradation is frequency independent [3], [24]. Therefore, we use a low frequency (10kHz) during active periods in our experiments to reduce simulation time. To estimate NBTI under worst case operating conditions, all simulations use T = 105^oC. At room temperature, NBTI degradation, likewise the potential benefit of mitigation techniques, reduces by 30%.

To model architecture-level performance, we assume that all PMOS devices experience the same degradation obtained from device-level simulation. Then, we estimate system level performance degradation by measuring the maximum delay among the top ten critical paths of our benchmark circuit (the OpenSPARC T1 processor [21]).

A. Dynamic Supply Voltage Tuning

To emulate DVS in our simulation framework, we have made the simulator able to dynamically switch from one voltage to another. Whenever there is a change in supply voltage, we use the corresponding V_{dd} value when evaluating Equation 5a.

To model dynamic voltage adaptation for our test processor (the OpenSPARC T1 [21]), we first synthesize, place, and route the design and perform STA to extract the critical paths. To find the guardbanded voltage for the processors, we run simulations for different supply voltages until we find the voltage that accounts for NBTI degradation over the lifetime of the processor (10 years). To find the DVS voltage profile over the lifetime of the processor, we begin a new simulation starting from the nominal voltage determined during STA. During the simulation, we use SPICE to check the delay of the critical paths every five minutes and increase V_{dd} by 5 mV any time the critical path delay reaches the clock period minus a safety margin.

B. Power Gating

During power gating, all PMOS devices are in relaxation state, which is equivalent to a high signal at all circuit nodes. Therefore, whenever power gating is applied, we always set the idle period signal to high instead of alternating the signal after each cycle. Since power gating is not applicable in the active period, the active period signal is a regular AC signal.

To compare the guardbanded frequency[2] for different power gating factors, we extract and characterize the critical paths of the processor, simulate degradation for a 10 year lifetime with different power gating factors, and perform SPICE simulations to find delay (frequency).

C. Activity Management

To emulate different activity profiles, we adjust the active period of the input signal (Figure 4) in our NBTI degradation simulator. Since the waveform is periodic, we only need to setup waveform parameters once during the initialization step of our simulator. The activity factor is defined by active time/(active time + idle time). We also evaluate activity management through adapting the processor configuration. Adapting an architecture to reduce NBTI degradation can change the activity factors of on-chip structures. If the activity factors of critical paths are reduced, the NBTI guardband can be reduced. Degradation mitigation can be enhanced if activity management is used to enable more power gating. Note that reduced activity typically corresponds to reduced throughput, so throughput may be traded for guardband reduction. To evaluate processor adaptation strategies, we performed a design space exploration (varying the same parameters as [19], [20]) in which we varied the number of integer and floating point arithmetic units (1,2,4,8), and the size of the instruction window and commit width (to match the total number of arithmetic units), and measured the throughput and processor activity for each case.

We use the following procedure to model the effects of activity management and architectural adaptations on NBTI degradation.

1) Using an architecture-level simulator (SMTSIM [23]), characterize the throughput of the processor and activity of critical on-chip structures for different architectural configurations and different (SPEC) benchmarks (mcf, twolf, art, parser, ammp, swim, equake, wupwise).
2) Using SP&R results for the processor, extract the critical paths and perform degradation simulations to measure processor degradation and lifetime for different activity factors, including those found in step (1). Perform the simulations for power gating and no power gating to evaluate both scenarios.
3) Compare degradation for different activity factors and degradation vs. throughput for processor configurations with different activity factors for the critical structures.

V. Results

A. Dynamic Supply Voltage Tuning

The rate of NBTI degradation is very fast during the early lifetime and slows down exponentially as time increases. The rate of degradation can be represented by a power law function, i.e.,

$$\Delta V_{th} \propto \text{ scalar } \times t^{\text{time_exponent}}. \quad (6)$$

Equation 6 clearly shows that the NBTI degradation rate is a strong function of the time exponent, which usually has a value between 0.16 and 0.25 [4], [5], [11], [25]. To show the front-loaded nature of NBTI degradation, we solve Equation 6 for two common time exponent values. Table I shows that 50% of the total V_{th} degradation occurs within the the first few months of a device's 10 year lifetime. This implies that any DVS scheme must perform voltage adjustments mainly in the early lifetime (first few days or weeks). As a result,

[2]The guardbanded frequency, is the frequency that accounts for NBTI delay degradation over the lifetime of the processor (10 years).

TABLE I: % NBTI degradation vs. time exponent (lifetime = 10 years).

Time exponent	0.16	0.25
Time to 50% degradation	1.6 months	7.5 months
Time to 90% degradation	62.1 months	78.7 months

the supply voltage increases quickly during the early lifetime, rapidly closing the gap between the starting voltage level and that of a simple guardbanding approach. DVS, which incurs substantial implementation overheads in terms of hardware and control mechanisms, has little impact on aging, power, or energy after the early lifetime. Note that the degradation rate is slower when the time exponent increases. Therefore, we expect the benefit of DVS to increase for a process with a higher time exponent.

Although the overall lifetime energy reduction achievable with DVS is limited, DVS can reduce the peak power dissipation of a chip, which is useful to relax chip packaging constraints. This happens because devices have lower V_{th} in early lifetime, which requires a lower V_{dd} to meet timing. Applying a lower V_{dd} in early lifetime reduces power compared to applying a higher supply voltage, required in a simple guardbanding method.

Figure 7 shows how the supply voltage increases over time for DVS. During early lifetime, NBTI degradation occurs rapidly, and V_{dd} increases quickly to compensate for increasing V_{th}. However, after the early lifetime, the difference between the adaptive voltage and the supply voltage in the guardband case is small. Degradation has slowed down, and voltage switches are few and far between. Thus, we observe that DVS has benefits during early lifetime, but benefits swiftly degrade afterward. Observe also from Figure 7 that using DVS does not allow any significant extension of processor lifetime. This is because degradation for both the DVS and guardbanding cases converges, so the DVS supply voltage also converges to the guardband voltage.

Power savings for DVS follow the same trend. Figure 7 also shows the power reduction of DVS compared to guardbanding. Savings are significant during early lifetime, but limited afterward. Using the DVS strategy, we observed a total (10 year) lifetime energy savings of 7% with respect to guardbanding. Note that this is an optimistic upper bound on energy savings, since we do not add any implementation overhead for DVS hardware and control. Although we must pay the area and potentially the control overhead for the entire lifetime of the processor, we only receive significant power benefits during the early lifetime. These results show significantly less benefits than several previous works suggested. Discrepancies in results are due to the previously discussed limitations in previous modeling approaches.

Note also that we are assuming that DVS is able to control the voltage at a very fine granularity (5 mV). Due to the overhead and difficulty of multi-constraint signoff for a large number of operating voltages and the cost of implementing fine-grained voltage control, voltage scaling at such a fine granularity may be infeasible. In modern DVS designs, only a few voltage levels are available, and scaling at such a coarse granularity significantly degrades DVS power and energy benefits, especially since voltage must be significantly higher during the early lifetime when benefits are greatest.

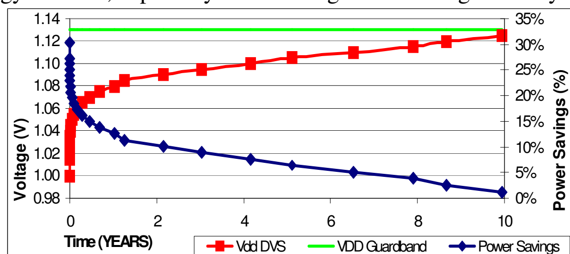

Fig. 7: In the DVS case (5 mV voltage scaling granularity), supply voltage approaches the guardband voltage rapidly in the early lifetime. Since the difference in supply voltage is small, power savings for DVS can be significant initially, but are limited after the first few months of operation.

B. Power Gating

To understand the benefits of power gating, we characterize degradation on the critical paths of the OpenSPARC T1 processor and observe how much degradation can be reduced for different power gating factors. Figure 8 shows the results. Generalizing for the processor, critical structures must be power gated to reduce the guardband (increase frequency, reduce voltage, or reduce area). High power gating factors may be feasible in designs where activity is naturally low.[3] However, such power gating factors are typically accompanied by significant throughput reduction, limiting the feasibility of power gating as an NBTI mitigation technique.

Note that these results are optimistic, since we have not added any overheads for power gating. Typically, power gating requires idle periods (on the order of tens of cycles [9]) to utilize sleep mode without increasing energy, and incurs area and power overheads for power gating circuitry and significant performance and energy overheads for saving and restoring state when entering and exiting sleep mode. We also assume perfect power gating in the sense that every cycle not spent computing can be spent in sleep mode, even though, e.g., an activity factor of 0.5 could mean one cycle of activity followed by one cycle of rest, in which case any benefits from power gating would be impossible.

C. Activity Management

It is well known that NBTI increases the V_{th} of PMOS during each stressing phase and that part of the V_{th} shift is recovered during each relaxation phase. There are many published works that manipulate signals on circuit nodes, assign specific input vectors [1], [14], [15], [18], [27], [28], and optimize circuits during synthesis [15] to reduce NBTI degradation. Many published studies neglect the important fact that a CMOS circuit is always inverting. I.e., a relaxation phase at a node implies that there is a complementary NMOS driving its fanout node to a stressing phase (if the circuit is not power gated). This means that putting a circuit block in idle mode does not help to reduce NBTI degradation, but actually exaggerates the degradation with a sustained period of stress.

Figure 9 shows that NBTI degradation only varies slightly for different activity factors. This means that NBTI mitigation techniques based on signal manipulation are limited in effectiveness. While we observed virtually no benefit from managing the activity of a circuit alone, recall that the benefit of power gating increases when activity factor is reduced and more time can be spent in power gating

[3] High power gating factors needed to significantly mitigate aging may be possible in low-power embedded regimes. However, such designs are already naturally less susceptible to aging, due to lower V_{dd} and higher V_{th}.

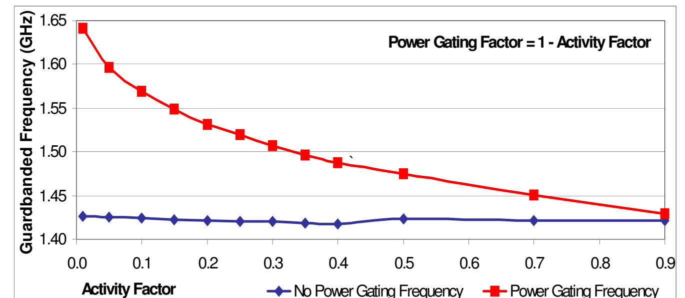

Fig. 8: The guardbanded frequency (that accounts for delay degradation over a 10 year lifetime) can be increased by up to 15% when power gating is used to mitigate NBTI degradation. However, this corresponds to guardbanding the critical regions of the processor for 99% of the processor's lifetime. In order to achieve more than 5% improvement in the frequency (reduction in delay degradation) the power gating factor must be over 60% (6 years) of the 10 year lifetime.

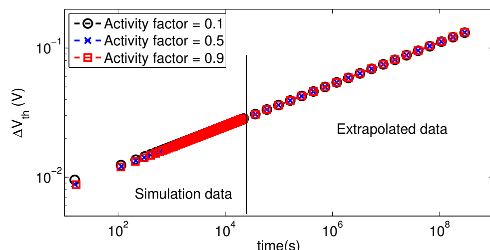

Fig. 9: Because of the complementary nature of CMOS logic, NBTI degradation is insensitive to circuit activity, and thus, there is little to no benefit available from managing activity to reduce NBTI degradation.

mode. Figure 10 shows how the 10 year guardband frequency can be increased when the processor configuration is adapted to reduce activity and allow more power gating. In our design space exploration of processor configurations, we observed that adapting the processor configuration can reduce activity by up to 61%. However, while this reduction in activity incurs a significant performance cost (up to 60% reduction in throughput) the additional frequency benefit (reduced delay degradation) with respect to the baseline processor is only up to 4%. As we observed in Section V-B, significant reduction of NBTI degradation cannot be achieved unless the power gating factor is very high, due to the significant recovery period required to overcome the front-loaded NBTI degradation. While activity management can significantly reduce processor activity and allow more power gating, the performance overhead may be substantial, and the additional power gating is not enough to significantly reduce NBTI degradation. Thus, we observe that activity management may be limited in effectiveness as an NBTI mitigation technique.

Note that our analysis of processor adaptation is based on average processor activity, such that we do not adapt the processor configuration within the phases of a benchmark. Such fine-grained adaptation could potentially produce a better tradeoff in terms of throughput, but considering the previous results and conclusions, we do not expect additional benefits to be significant.

VI. Conclusions

Recent works have proposed architecture-level techniques to mitigate the growing problem of NBTI degradation in next-generation digital circuits. Analysis of these techniques has been based on analytical device-level models that were not designed to model the impact of dynamic architecture-level techniques. To address this limitation, we provide a flexible numerical model of NBTI degradation based on reaction-diffusion that can be adapted to model mechanisms like voltage scaling, power gating, and activity management that are employed by architecture-level techniques. We use our model to evaluate NBTI mitigation techniques and analyze their potential benefits and limitations. Our study of previously proposed NBTI mitigation techniques has demonstrated that achievable benefits from architecture-level mitigation techniques may be significantly less than previously reported, and that guardbanding may still be the most efficient way to deal with aging. Moreover, there is significant random variation in the NBTI degradation [12], which is not accounted for in the mitigation techniques. Such statistical variation of the NBTI process results in an additional random V_{th} degradation on top of the average degradation and may further reduce the reported benefits. Although this paper discusses NBTI, similar conclusions are expected for PBTI-affected processes (e.g., high-k), as PBTI is also typically described by a R-D model [17].

We realize that this work is a brief study of the subject. Our ongoing work includes (1) extending the analyses in greater depth, including the dependence of results on process/technology, power-gating and DVS spatial granularities, as well as overheads introduced by the mitigation techniques and (2) parallelizing the numerical simulator to reduce NBTI degradation simulation runtime.

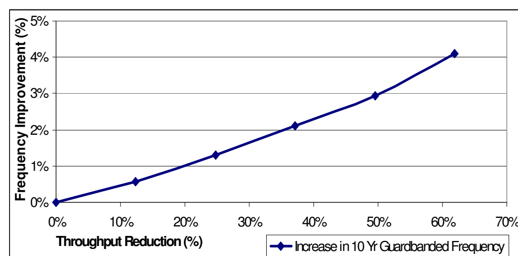

Fig. 10: We can achieve additional aging reduction by adapting the processor configuration to reduce activity and allow more power gating. However, to reduce degradation and increase the guardbanded frequency (that accounts for delay degradation over a 10 year lifetime) by a small amount, we may have to tolerate significant performance degradation.

VII. Acknowledgments

This work is sponsored by GSRC, SRC, and NSF. Also, we would like to thank Dr. Yu Cao for his valuable inputs.

References

[1] J. Abella, X. Vera, and A. Gonzalez. Penelope: The nbti-aware processor. In *MICRO 40*, pages 85–96, 2007.
[2] M. Agarwal, B. Paul, Z. Ming, and S. Mitra. Circuit failure prediction enables robust system design resilient to aging and wearout. In *IOLTS*, page 123, 2007.
[3] M. Alam. A critical examination of the mechanics of dynamic nbti for pmosfets. pages 14.4.1 – 14.4.4, 2003.
[4] M. Alam and S. Mahapatra. A comprehensive model of pmos nbti degradation. *Microelectronics Reliability*, 45(1):71–81, 2005.
[5] H. Aono, E. Murakami, K. Okuyama, et. al.. Modeling of NBTI degradation and its impact on electric field dependence of the lifetime. In *Reliability Physics Symposium Proceedings, 2004.*, pages 23–27, 2004.
[6] S. Borkar. Electronics beyond nano-scale cmos. In *DAC*, pages 807–808, 2006.
[7] A Calimera, E. Macii, and M. Poncino. Nbti-aware power gating for concurrent leakage and aging optimization. In *ISLPED*, pages 127–132, 2009.
[8] X. Chen, Y. Wang, Y. Cao, Y. Ma, and H. Yang. Variation-aware supply voltage assignment for minimizing circuit degradation and leakage. In *ISLPED*, 2009.
[9] Z. Hu, A. Buyuktosunoglu, V. Srinivasan, V. Zyuban et. al.. Microarchitectural techniques for power gating of execution units. In *ISLPED*, pages 32–37, 2004.
[10] V. Huard, M. Dennis, and C. Parthasarathy. Nbti degradation: From physical mechanisms to modelling. *Microelectronics Reliability*, 46(1):1–23, 2006.
[11] K. Jeppson and C. Svensson. Negative bias stress of mos devices at high electric fields and degradation of mnos devices. *Journal of Applied Physics*, 48.
[12] K. Kang, S. Park, et. al. Estimation of statistical variation in temporal nbti degradation and its impact on lifetime circuit performance. In *ICCAD*, 2007.
[13] U. Karpuzcu, B. Greskamp, and J. Torrellas. The bubblewrap many-core: popping cores for sequential acceleration. In *MICRO 42*, pages 447–458, 2009.
[14] S. Kumar, C. Kim, and S. Sapatnekar. Impact of nbti on sram read stability and design for reliability. In *ISQED*, pages 210–218, 2006.
[15] S. Kumar, C. Kim, and S. Sapatnekar. Nbti-aware synthesis of digital circuits. In *DAC*, pages 370–375, 2007.
[16] S. Kumar, C. Kim, and S. Sapatnekar. Adaptive techniques for overcoming performance degradation due to aging in digital circuits. In *ASPDAC*, 2009.
[17] N. Sa, J. Kang, H. Yang, X. Liu, et. al.. Mechanism of positive-bias temperature instability in sub-1-nm tan/hfn/hfo2 gate stack with low preexisting traps. *Electron Device Letters*, 26(9):610 – 612, 2005.
[18] T. Siddiqua and S. Gurumurthi. A multi-level approach to reduce the impact of nbti on processor functional units. In *GLSVLSI*, pages 67–72, 2010.
[19] J. Srinivasan, S. Adve, P. Bose, and J. Rivers. The case for lifetime reliability-aware microprocessors. In *ISCA*, page 276, 2004.
[20] J. Srinivasan, S. Adve, P. Bose, and J. Rivers. Lifetime reliability: Toward an architectural solution. *IEEE MICRO*, 25(3):70–80, 2005.
[21] Sun. *Sun OpenSPARC Project*.
[22] A. Tiwari and J. Torrellas. Facelift: Hiding and slowing down aging in multicores. In *MICRO 41*, pages 129–140, 2008.
[23] D. Tullsen. Simulation and modeling of a simultaneous multithreading processor. In *22nd Annual Computer Measurement Group Conference*, 1996.
[24] R. Vattikonda, W. Wang, and Y. Cao. Modeling and minimization of pmos nbti effect for robust nanometer design. pages 1047–1052, 2006.
[25] W. Wang, V. Reddy, et. al. Compact modeling and simulation of circuit reliability for 65-nm cmos technology. *Device and Materials Reliability*, 2007.
[26] W. Wang, V. Reddy, B. Yang, V. Balakrishnan, S. Krishnan, and Y. Cao. Statistical prediction of circuit aging under process variations. pages 13 –16, 2008.
[27] W. Wang, S. Yang, S. Bhardwaj et. al.. The impact of nbti on the performance of combinational and sequential circuits. In *DAC*, 2007.
[28] Y. Wang, X. Chen, W. Wang, et. al.. On the efficacy of input vector control to mitigate nbti effects and leakage power. In *ISQED*, pages 19–26, 2009.
[29] Y. Wang, H. Luo, K. He, et. al.. Temperature-aware nbti modeling and the impact of input vector control on performance degradation. DATE, 2007.
[30] L. Yang, M. Cui, J. Ma et. al.. Advanced spice modeling for 65nm CMOS technology. In *Solid-State and Integrated-Circuit Technology*, 2008.
[31] Cao Yu. Department of Electrical Engineering, Arizon State University. personal communication, 2010.