# Modeling and Analyzing NBTI in the Presence of Process Variation

Taniya Siddiqua, Sudhanva Gurumurthi, Mircea R. Stan†
Dept. of Computer Science, †Dept. of Electrical and Computer Engg., University of Virginia
{taniya,gurumurthi,mircea}@virginia.edu

## Abstract

*With continuous scaling of transistors in each technology generation, NBTI and Process Variation (PV) have become very important silicon reliability problems for the microprocessor industry. In this paper, we develop an analytical model to capture the impact of NBTI in the presence of PV for use in architecture simulations. We capture the following aspects in the model: i) variation in NBTI related to stress and recovery due to workloads, ii) temporal variation in NBTI due to Random Charge Fluctuation (RCF) and iii) Random Dopant Fluctuation (RDF) due to process variation. We use this model to analyze the combined impact of NBTI and PV on a memory structure (register file) and a logic structure (Kogge-Stone adder). We show that the impact of the threshold voltage variations due to NBTI and PV over the nominal degradation can hurt the yield of the structures. Due to the combined effect of NBTI and PV across different benchmarks, 26 to 117 bits fail in a 8Kb size register file and the execution delay increases by 18% to 28% in a Kogge-Stone adder. We then discuss the implications of these results for architecture-level reliability techniques.*

## 1 Introduction

We are in the era of multicore processors and it is expected that the number of the processing cores on a chip will steadily increase over the next decade, driven by Moore's law. While technology scaling paves the way for high performance multicore processors, the scaling has a dark side too: silicon reliability. Processors are becoming increasingly susceptible to a variety of silicon reliability problems, from soft errors and process variation to several hard error phenomena, which can cause permanent damage to the processor. One important hard error phenomenon is Negative Bias Temperature Instability (NBTI), which affects the lifetime of PMOS transistors. NBTI occurs when a negative bias is applied at the gate of a PMOS transistor and causes an increase in the threshold voltage of the device. In terms of its impact on microprocessor circuits, this increase in the threshold voltage degrades the speed of the transistors and therefore degrades the speed of the circuit in which they are used, eventually causing the circuit to violate timing constraints [16, 11]. Such a timing violation will cause the circuit to behave incorrectly and cause the processor itself to fail. Moreover, the impact of NBTI is exacerbated by Process Variation (PV). PV is the variation in the transistor attributes (length, width, oxide thickness) caused during the fabrication of the integrated circuits and manifests itself as threshold voltage variations which results in variability in circuit performance and power. Processors have to be designed to provide adequate protection against both these problems.

Both NBTI and PV have received attention in the architecture community in recent years and several mitigation techniques have been proposed for each [1, 21, 18, 19, 20, 7]. Since both NBTI and PV affect the threshold voltage of devices, these two problems should not be addressed in isolation. To come up with the appropriate mitigation techniques, it is important to accurately gauge the impact of both NBTI and PV and factor-in the impact of the workloads that run on the processor as well. For this purpose, an analytical model is required which captures the impact of both NBTI and PV in a coherent way and which is suitable for use in architecture level analyses.

There have been several efforts in developing analytical models for NBTI and PV at the circuit-level. However, these models are suitable only for analyzing NBTI and PV effects over a very short time span and are not readily usable for architecture simulations. Architects, on the other hand, study microprocessor reliability by executing different program benchmarks and extrapolate the collected statistics over a much longer timescale (typically, 7-10 years). Throughout the benchmark execution, utilizations of the microarchitectural structures vary. Also, the interactions among the structures, the inputs to each structure, and bits stored within them change over the course of execution of a benchmark. The analytical model for NBTI and PV should be able to factor-in all these "variations" to be usable in architecture simulations to gain correct and holistic insight into these inter-related silicon reliability phenomena. In this paper, we leverage the prior research on NBTI and PV modeling from the circuits community to develop a model that captures the interactions between these two reliability phenomena and which is usable at the architecture-level.

There are different sources of variation inherent in NBTI and PV that affect the PMOS threshold voltage. One source of variation in the threshold voltage due to NBTI is *workload variation* which is caused by executing different workloads on the processor. This variation is due to changing patterns of utilization of the microarchitectural structures and changes in the bit patterns within the structures. Another factor lies in the silicon process, known as the Random Charge Fluctuation (RCF), which causes a *temporal variation* in threshold voltage on top of the workload variation. Along with the variations due to NBTI, each device also has Random Dopant Fluctuations (RDF) due to *process variation* (details of the sources of these variations are discussed in the next section). The analytical model we have developed accounts for all these variations.

We make the following contributions in this paper:

- We develop an analytical model to capture both NBTI and PV for use in architecture simulations.
- We use this model to show the combined impact of NBTI and PV on a memory structure (register file) and a logic structure (Kogge-Stone adder).
- We show that the impact of the threshold voltage variations due to NBTI and PV over the nominal degradation can hurt the yield of the structures. Due to the combined effect of NBTI and PV across different benchmarks, 26 to 117 bits fail in a 8Kb size register file and the execution delay increases by 18% to 28% in a kogge-stone adder. We then discuss the implications of these results for architecture-level reliability techniques.

The outline of the rest of this paper is as follows. The next section gives a brief overview of the different sources of threshold voltage variation due to NBTI and PV. Section 3 discusses the related work. The analytical model for NBTI and PV is described in Section 4. The experimental methodology is described in Section 5. The results are presented in Section 6 and Section 7 concludes this paper.

## 2 Overview of NBTI and PV

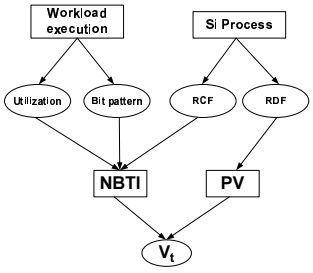


**Figure 1. Different sources of $V_t$ variation in PMOS devices.**

Figure 1 shows the overall picture of the different sources of variation in PMOS threshold voltage degradation due to NBTI and PV. We now describe how NBTI gets affected by workloads that run on the processor and the silicon process.

NBTI affects the lifetime of PMOS devices and occurs when a logic input of '0' (i.e., a negative bias) is applied at the gate of the transistor. This negative bias leads to the generation of interface traps at the $Si/SiO_2$ interface which originate from the $Si-H$ bonds. The $Si-H$ bonds break during the stress condition of negative bias and form interface traps or trapped charges, which cause the threshold voltage of the transistor to increase. However, some of the interface traps can be eliminated by applying a logic input of '1' to the device, putting the device into a "recovery mode".

As shown in Figure 1, the impact of NBTI is affected by several factors. In a real processor, different microarchitctural structures exhibit different utilization patterns based on the characteristics of the workloads that exercise them. On top of the overall utilization of the structures, all the PMOS devices within each processor structure are stressed in different ways throughout the workload execution due to the varying data bit patterns (gate inputs of the devices) within them. Therefore, workload execution leads to a variation in the threshold voltage degradation, which we call *workload variation*. The third factor lies in the silicon process, known as Random Charge Fluctuation (RCF), which causes a *temporal variation* on top of the workload variation. Recent observations on PMOS devices with small gate areas show that the threshold voltage degradation is a subject to random fluctuations [9, 2]. These fluctuations increase as a function of stress time. The source of this behavior is the formation of a random number of trapped charges, which can occur at random locations across the gate. Such random fluctuations of trapped charges result in a variation in the threshold voltage degradation and needs to be considered when studying NBTI. We call the impact of NBTI which considers only the structure utilization and does not capture the effect of the workload variation and temporal variation as *static NBTI*.

Furthermore, the degradation in processor lifetime due to NBTI is exacerbated by Process Variation (PV). Process variations can be broadly categorized into two groups: inter-die and intra-die variations [10]. Due to inter-die variations, the same device on a die can have different characteristics across various dies, whereas, due to intra-die variations, transistors can have different characteristics within a single die. There are two more subcategories of intra-die variation: systematic and random variations. Due to systematic variations, transistors close to each other are expected to have relatively similar parameters (channel length and oxide thickness) when compared to those farther away on the die. On the other hand, random variation is mostly caused by RDF. Due to RDF, transistors can have mutually independent $V_t$ variation with respect to each other, regardless of their spatial location. We consider only the effect of RDF

in this work, for two reasons. First, RDF is expected to be the major contributor to transistor threshold voltage variations in the sub-65nm technology [10]. Second, we look at individual processor microarchitectural structures where the devices within them are spatially proximate. The analytical model we develop accounts for the combined effect of workload and temporal variation due to NBTI in the presence of RDF.

## 3 Related Work

Several recent studies have proposed techniques for mitigating NBTI to improve processor lifetime. [1] proposes and evaluates an NBTI-aware processor design with strategies for writing special values to both combinational and storage blocks whenever they are idle to reduce NBTI stresses. [21] uses temperature-based job-scheduling on a multicore processor along with $V_{dd}$ and $V_t$ control to hide the effects of aging due to NBTI. [3] proposes a low-cost NBTI-Aware DVFS framework to reduce energy consumption and increase the lifetime of the processor. [18] presents a multi-level optimization approach, combining techniques at the circuit and microarchitecture levels, for reducing the impact of NBTI on the functional units of a high-performance processor core. [19] proposes a technique that allows both PMOS devices in the memory cell to be put into the recovery mode by slightly modifying the design of conventional SRAM cells. All these works study the impact of static NBTI. While all of them focus on NBTI mitigation, there are several other studies which aim for PV tolerant system design. [20] presents an architectural framework that applies cycle-time stealing to the pipeline to tolerate PV. [7] proposes a scheme of adjusting the clock speed of a processor based on the instruction-level parallelism of the program phases to achieve overall performance improvement to address PV.

All the aforementioned studies concentrate on either NBTI or PV without considering their interaction. There have been several studies on the combined effect of NBTI and PV. [9] proposes a compact circuit-level $V_t$ model that captures the impact of temporal NBTI variations in the presence of PV and shows how temporal $V_t$ variations can affect the lifetime and performance of different circuit topologies. [4] presents a methodology to develop standard cells which can be used in timing critical sections of circuit, and [12] designs a comprehensive IC reliability analysis framework. Both [4] and [12] consider the combined impact of static NBTI and PV. Finally, [8] proposes NBTI (static and workload variation) and PV tolerant micrarchitecture design techniques to improve processor lifetime.

While all these prior works study some combinations of NBTI (static, or temporal variation, or workload variation) with or without PV, to our knowledge no prior work has holistically analyzed the combined effect of temporal and workload variations on top of static NBTI with process variation.

## 4 An analytical model for NBTI and PV

There have been several efforts in developing an analytical model for NBTI based on the reaction-diffusion model [22, 9]. These models have been extended to address dynamic temperature and voltage variations in [23, 3] and are suitable for use in circuit-level simulations. However, these models cannot be directly used for architecture-level simulations. This is because these models assume continuous stress on the PMOS devices in a circuit and do not capture scenarios where there are multiple sequences of varying stress/recovery times, which is the case when real workloads run on the processor. In this work, we present a compact analytical model that is suitable for both circuit and architecture simulations and also takes into account the effect of PV. Before we present our model, we first discuss why the existing models cannot be directly used in architecture simulations.

### 4.1 Challenges Posed by Existing NBTI models for Architecture Simulation

To explain the problems with the existing NBTI models, we choose the model presented in [22] which is widely used in the circuit literature. However, the problems we discuss apply to the use of other circuit-level NBTI models too.

The period of time when the PMOS transistor is negatively biased is known as the *stress phase*. The increase in $V_t$ due to stress is given by the following equation:

$$\Delta V_{ts}=(\frac{qt_{ox}}{e_{ox}})^{\frac{3}{2}}.K_1.\sqrt{C_{ox}(V_{gs}-V_t)}.e^{\frac{-E_a}{4kT}+\frac{2(V_{gs}-V_t)}{t_{ox}E_{01}}}.T_0^{-0.25}.t_{stress}^{0.25}$$

where $t_{stress}$ is the time under stress, $V_t$ is the threshold voltage at that time, $V_{gs}$ and $T$ are the supply voltage and temperature respectively, $t_{ox}$ is the oxide thickness, $C_{ox}$ is the gate capacitance per unit area and $K_1$, $E_a$, $T_0$, $E_{01}$ and $k$ are constants.

When a logic input of '1' is applied at the gate ($V_{gs}$ = 0), the transistor turns off eliminating some of the interface traps. This is known as the *recovery phase*. The final increase in $V_t$ after considering both the stress and recovery phases is given by:

$$\Delta V_t = \Delta V_{ts}.(1 - \frac{2\xi_1 t_{ox} + \sqrt{\xi_2 e^{\frac{-E_a}{kT}} T_0 t_{rec}}}{(1+\delta)t_{ox} + \sqrt{e^{\frac{-E_a}{kT}}(t_{stress}+t_{rec})}})$$

where $t_{rec}$ is the recovery time and $\xi_2$, $\xi_1$ and $\delta$ are constants.

This model has two limitations when used in an architecture simulation, which we will now explain:

**i) The model is not additive:**

Let us consider a hypothetical scenario where a PMOS device is stressed for $t_1$, $t_2$ and $t_3$ units of time and the degradations in threshold voltage due to these stress events

are $V_t(t_1)$, $V_t(t_2)$ and $V_t(t_3)$ respectively. If the device is stressed for $t_1$ time units followed by $t_2$, and $t_3$ is equal to ($t_1$ + $t_2$), then we expect $V_t(t_3) = V_t(t_1) + V_t(t_2)$. However, the model does not possess the additive property. Figure 2 describes this behavior. The x-axis is the time and y-axis is the threshold voltage. In the figure, once stress time of $t_1$ is applied initially, the model computes the threshold voltage to reach A. At this point, if stress time of $t_2$ is applied, instead of reaching a value of B, the model computes it to be C, whereas the value should be B after ($t_1$ + $t_2$) units of time. The reason for this problem is because $\Delta V_{ts}$ has an exponential relationship with the stress time and we know that $(t_1 + t_2)^x \neq ({t_1}^x + {t_2}^x)$. Therefore, using the model, we do not achieve the expected value of $V_t(t_3)$ which should be equal to [$V_t(t_1)$ + $V_t(t_2)$].

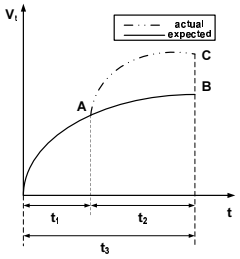


**Figure 2. The existing circuit-level NBTI model lacks the additive property.**

**ii) The model is not usable with multiple stress/recovery events:**
To understand this problem, let us consider another hypothetical example where a PMOS device gets stressed followed by a single sequence of stress and recovery events. Figure 3 illustrates this situation. From the figure, with the first set of stress and recovery events, the $V_t$ reaches a value of B and A respectively. At this point of time, with a subsequent stress event, the threshold voltage degradation pattern should follow the pattern of the first stress event starting from point A, since both instants have a $V_t$ value of A. Therefore, after applying the second stress for $t_1$ time units, the final $V_t$ value should be, B whereas the model provides a different value of C. Thus, the model is not able to capture multiple sequences of stress/recovery events properly. The reason for this problem is because the model uses the instantaneous $V_t$ value as the history of degradation. From the stress phase equation, we can see that the value of $\Delta V_{ts}$ depends on the value of $V_t$. For a fixed stress time, the model would produce different $\Delta V_{ts}$ values for different $V_t$ values. In this hypothetical scenario, the first stress event uses the nominal $V_t$ value and the second stress event uses the degraded $V_t$ value. Therefore, the two stress events of $t_1$ time unit starting from point A produce two different values.

Both these properties need to be modeled correctly for an architecture level analysis of NBTI degradation. Since the architecture simulations update the $V_t$ values of the devices due to NBTI at different points of time throughout the execution of a workload, the lack of the additive property in the model results in incorrect estimation of the $V_t$ degradation. Also, real workloads show varying patterns of stress/recovery for different structures within the processor. Hence, the assumption of continuous stress on the PMOS devices does not capture the realistic scenario. The next section discusses how to modify the model to address these two limitations.

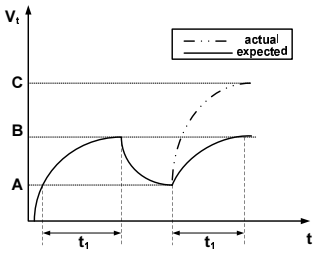


**Figure 3. The existing circuit-model NBTI model does not capture multiple sequences of stress and recovery events.**

### 4.2 Adapting the NBTI model for Circuit and Architecture Simulation

From the original model, we know that $\Delta V_{ts}$ is a function of voltage, temperature, instantaneous $V_t$, and $t_{stress}$, whereas, $\Delta V_t$ after recovery is a function of $\Delta V_{ts}$, $t_{stress}$ and $t_{rec}$. We can rewrite the original model as:

$$\Delta V_{ts} = f_{stress}(V, T, V_t).t_{stress}^{0.25} \quad (1)$$

$$\Delta V_t = \Delta V_{ts}.f_{rec}(t_{stress}, t_{rec}) \quad (2)$$

In this model, the $V_t$ value represents the history of stress and recovery events (the total degradation). The main idea behind our proposed model is to represent the degradation history in terms of the equivalent stress time experienced by the PMOS device. Since the existing model is applicable for a single stress/recovery event, we transform the previous multiple stress/recovery events into a single stress event and use that equivalent stress time with the new stress/recovery event. Note that in this case, we always use the nominal $V_t$ value. After the aforementioned modification, we get the following model:

$$\Delta V_{ts} = f_{stress}(V, T).(t_{equi-stress} + t_{stress})^{0.25} \quad (3)$$

$$\Delta V_t = \Delta V_{ts}.f_{rec}[(t_{equi-stress} + t_{stress}), t_{rec}] \quad (4)$$

where $t_{equi-stress}$ is the equivalent stress time resulting from previous stress and recovery events and $t_{equi-stress}$ =

0 at t = 0. Now we discuss how to calculate the value of $t_{equi-stress}$.

From equation 1, we get $\Delta V_{ts}$, which is the increase in threshold voltage due to the stress time $t_{stress}$. If we reorganize equation 1, we find the following:

$$t_{stress} = [\frac{\Delta V_{ts}}{f_{stress}(V,T,V_t)}]^4 \quad (5)$$

This equation expresses the stress time experienced due to the previous stress/recovery events when the increase in threshold voltage is known. Using equation 5 with the nominal $V_t$ and the given $\Delta V_t$, which is a result of previous stress and recovery events, we can calculate the $t_{equi-stress}$:

$$t_{equi-stress} = [\frac{\Delta V_t}{f_{stress}(V,T)}]^4 \quad (6)$$

After combining equations 3, 4 and 6, we get the following final model:

$$\Delta V_{ts} = f_{stress}(V,T).\{[\frac{\Delta V_t}{f_{stress}(V,T)}]^4 + t_{stress}\}^{0.25} \quad (7)$$

$$\Delta V_{tf} = \Delta V_{ts}.f_{rec}(\{[\frac{\Delta V_t}{f_{stress}(V,T)}]^4 + t_{stress}\}, t_{rec}) \quad (8)$$

where $\Delta V_{tf}$ is the final threshold voltage degradation and $\Delta V_t$ is the threshold voltage degradation due to previous stress and recovery events and $\Delta V_t$ = 0 at t = 0.

Note that, this model captures the effect of voltage and temperature variation as well. Since the equivalent stress time is also a function of voltage and temperature, whenever there is a variation in either voltage or temperature or both, the equivalent stress time gets calculated under the new stress condition.

### 4.3 Capturing the impact of Workload Variation, Temporal Variation, and PV

The NBTI $V_t$ model presented in the previous section assumes the nominal or static degradation for each device without considering the workload variation or temporal variation. As described in Section 1, in a realistic scenario, the nominal NBTI for each structure is impacted by the workload execution due to the variation in the utilization of the structure and its bit patterns. While executing a workload, for a given structure, we track the stress/recovery patterns for each device within that structure. Using the model presented in the previous section, we get a $V_t$ distribution (Standard Deviation = $\sigma_{ARCH}$). This results in multiple groups of devices where all the devices within each group experience similar stress/recovery patterns and have similar final $V_t$ values.

Moreover, as mentioned in Section 1, the temporal variation in the underlying degradation process due to RCF causes additional variation on top of the workload variation. From [13], if a group of devices are stressed in a similar way, the variation caused by RCF is:

$$\sigma_{RCF} = \sqrt{\frac{K.t_{ox}.\Delta V_{tf}}{A_g}}$$

where, $\sigma_{RCF}$ is the standard deviation of the $V_t$ distribution, $A_g$ is the gate area of the device, $t_{ox}$ is the oxide thickness, $\Delta V_t$ is the nominal degradation due to NBTI and $K$ is a constant. Since workload variation results in multiple groups of devices experiencing similar kinds of stress patterns, temporal variation within each group of devices results in several $V_t$ distributions. After combining all the distributions, we get a final $V_t$ distribution which captures the effect of both workload and temporal variation (Standard Deviation = $\sigma_{(ARCH+RCF)}$).

Furthermore, to combine the effect of PV, we know from [13]:

$$\sigma_{RDF} = \frac{\alpha}{\sqrt{A_g}}$$

where, $\sigma_{RDF}$ is the standard deviation of the $V_t$ distribution due to RDF, $A_g$ is the gate area of the device, and $\alpha$ is a constant.

Finally, combining the effect of NBTI (static, workload and temporal variation) and PV, we get the following standard deviation:

$$\sigma_{(PV+NBTI)} = \sqrt{{\sigma_{(ARCH+RCF)}}^2 + {\sigma_{RDF}}^2} \quad (9)$$

This completes the model. From the equations 7 and 8, we get the mean $V_t$ degradation and equation 9 gives the $V_t$ standard deviation.

## 5 Experimental Setup

To carry out the architecture simulations, we use the M5 simulator [5]. We simulate a 4-wide issue core, which runs at a 3 GHz clock frequency and is representative of cores that is used in multicore processors today. We use the 32nm process with a supply voltage of 0.9V. We assume the initial threshold voltage of the PMOS devices to be 0.2 V and the service life of the processor to be 7 years [21]. Our workloads consist of benchmarks from the SPEC CPU2000 benchmark suite [24]. We present simulation results for 8 representative benchmarks - 4 integer and 4 floating-point. The benchmarks are compiled for the Alpha ISA and use the reference input set. We perform detailed simulation of the first 100-million instruction SimPoint for each benchmark

[17]. Our circuit-level simulations are performed using the Cadence Virtuoso Spectre circuit simulator [6] taking the technology parameters of 32nm process from the Predictive Technology Model [14]. In this paper, we focus on the impact of NBTI and PV on one memory structure - the register file (RF) and one logic structure - the Kogge-Stone Adder (KSA). The RF is a 128x64 size SRAM array made up of 6T bitcells and the KSA is implemented for 64-bit inputs.

*RF Reliability Metric:* NBTI and PV affect the read and write delays and the read Static Noise Margin (SNM) of the SRAM cells. Previous work [11] has shown that the SNM is the one that is most heavily affected by NBTI. Therefore we use SNM as the reliability metric for the RF.

*KSA Reliability Metric:* Since NBTI affects the threshold voltages of PMOS devices in the KSA, the delay of the KSA increases, which could potentially cause a timing violation. Therefore we use delay as the reliability metric for the KSA.

Before exercising the RF and the KSA with workloads, the SNM and the delay of the RF and KSA respectively are already degraded because of PV. We calculate this degraded SNM distribution and delay by using the Spectre circuit simulator. The SNM and delay degrades further after the structures get exercised by the workloads, due to NBTI. We capture this impact by tracking the stress and recovery cycles on all the PMOS devices in the RF and the KSA over the course of the architecture simulation and extrapolate the statistics to calculate the final degradation in $V_t$ after the 7-year service life. We calculate the mean $V_t$ and the different standard deviation values due to temporal, workload, and the combined variations for both the RF and the KSA. We then feed these values into the Spectre circuit simulator, and calculate the degraded SNM distributions of the RF and delays of the KSA.

## 6 Results

We now quantitatively analyze the effect of NBTI in the presence of PV in RF and KSA. We evaluate four different conditions: i) *RDF*: considering only the impact of RDF without the effect of NBTI, ii) *RCF+RDF*: considering the impact of NBTI only with the temporal variation on top of the RDF effect, iii) *ARCH+RDF*: considering the impact of NBTI only with the workload variation on top of the RDF effect, and finally, iv) *ARCH+RCF+RDF*: considering the impact of NBTI with both the temporal and workload variation on top of the RDF effect.

### 6.1 RF Results

We now explain the impact of NBTI and PV on the RF by means of an example. We first show the $V_t$ distributions under different conditions. From the simulations, we calculate the following standard deviations: ($\sigma_{RDF}$, $\sigma_{(RCF+RDF)}$, $\sigma_{(ARCH+RDF)}$ and $\sigma_{(PV+NBTI)}$). Figure 4 shows the $V_t$ distributions of the RF for one of the benchmarks we evaluate - $mcf$. Initially, before the workload is executed, the $V_t$ distribution is due to RDF (the leftmost distribution in the figure). But once the workload is executed and the stress/recovery statistics on the RF are extrapolated to 7 yrs, the $V_t$ distribution shifts to the right due to NBTI. As the figure indicates, the effect of temporal variation in the presence of RDF merely causes a shift in the mean of the distribution, but once the workload variation is factored in, the distribution widens. In order to understand why the width increases, we need to understand how the $V_t$ of the PMOS devices get affected by workload and temporal variation. As mentioned in Section 4.3, workload variation results in multiple groups of devices which experience similar stress patterns, leading to similar $V_t$ values. However, because of the temporal variation, each group of devices ends up in a $V_t$ distribution. Therefore, when we take into account all the $V_t$ values in the structure, we get a wider distribution. It is important to note that without considering the effect of RDF, the distributions due to NBTI with temporal, workload, and the combined variations would be much narrower. Hence it is important to consider the effect of NBTI in the presence of PV along with temporal and workload variation to avoid any significant error in the lifetime estimation of the structure. Now we show how the $V_t$ distributions affect the yield of the RF, using the RDF as the baseline.

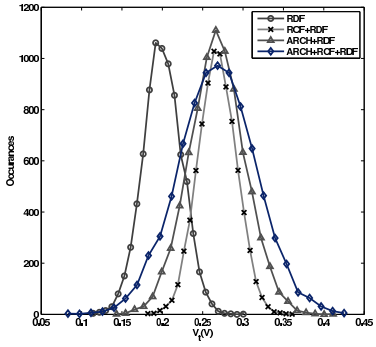


**Figure 4. $V_t$ distributions of the RF due to RDF, temporal, workload and combined variation for the $mcf$ benchmark.**

The required design coverage ($N_\sigma$) of a memory is a function of the target yield and the memory density and is expressed by the following equation [15]:

$$N_\sigma = \phi^{-1}(Y_{mem}{}^{\frac{1}{N_{bits}}})$$

where $\phi^{-1}$ is the inverse standard normal cumulative distribution, $Y_{mem}$ is the yield of the memory, and $N_{bits}$ is the total number of bitcells in the memory. Once the design coverage is calculated, from the expected SNM distribution (baseline: $\mu_{SNM-RDF}$, $\sigma_{SNM-RDF}$), the minimum allowed SNM can be calculated as:

$$SNM_{min} = \mu_{SNM-RDF} - N_\sigma * \sigma_{SNM-RDF}$$

Under each NBTI and PV condition, we count the number of bitcells whose SNM values are less than $SNM_{min}$. We denote this number as $\#bit_{fail}$.

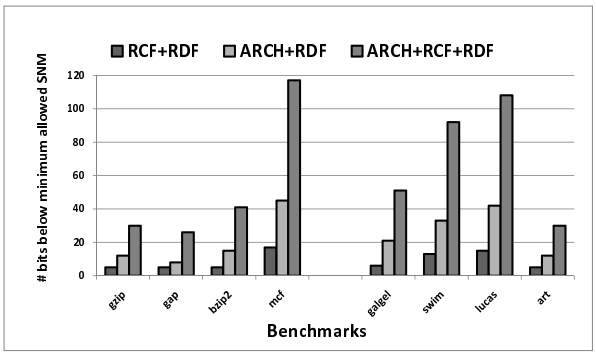


**Figure 5. Number of bits experiencing SNM below the minimum allowed value in a RF due to temporal, workload and the combined variation for the different benchmarks.**

Figure 5 shows the $\#bit_{fail}$ in the RF under three different conditions (*RCF+RDF*, *ARCH+RDF*, and *ARCH+RCF+RDF*) for different benchmarks. $\#bit_{fail}$ ranges from 5 to 17 for the *RCF+RDF* condition where only the temporal variation is considered in the presence of PV. It ranges from 8 to 45 for the *ARCH+RDF* condition, whereas it ranges from 26 to 117 for the *ARCH+RCF+RDF* condition. As expected from the $V_t$ distributions, this result shows that the impact of the temporal variation alone is less than the impact of the workload variation, whereas the combined effect is much greater than the sum of the individual effects. This is due to the widening of the $V_t$ distribution, as explained before. It is also important to note that the effects of the variations vary significantly across the benchmarks. $mcf$, $lucas$ and $swim$ benchmarks have large $\#bit_{fail}$ values (117, 108 and 92 respectively) under the *ARCH+RCF+RDF* condition. The reason behind this is due to workload variations. $mcf$, $lucas$ and $swim$ experience much higher $\sigma_{ARCH}$ as compared to the other benchmarks because of the bit patterns and the long residence times of the bits in each register. Generally, we find that most of the registers tend to have more 0's in the higher order bits and a random mix of 0's and 1's in the lower order bits, which contribute to the variability of the stress/recovery patterns of the register bits. Also, these benchmarks experience high L2 cache miss rate which causes stalls in the processor pipeline. Therefore, the contents of the register files do not get updated often. As a result, some bits tend to experience more stress whereas others experience less stress. Because of this, the bits in the RF experience high workload variation. The impact of workload variations, combined with temporal and process variations leads to a higher failure rate.

## 6.2 KSA Results

To explain the impact of NBTI and PV on the KSA, we again begin with the $V_t$ distributions under different conditions. Figure 6 shows the $V_t$ distributions of the KSA for $mcf$. The leftmost $V_t$ distribution in the Figure is due to the RDF and this distribution gets shifted to the right because of NBTI. Similar to the RF, the effect of temporal variation and the workload variation in the presence of RDF is less than their combined impact. However, unlike the RF, the curves for the temporal variation and the workload variation are close to each other. The reason why the workload variation does not contribute to $V_t$ changes significantly beyond the temporal variation is because of the circuit design of the KSA. Based on the inputs to the KSA, bits propagate through the internal nodes of the circuit. The inherent design of the circuit generates internal node values of 0's and 1's within the structure in a balanced manner, which produces a comparatively smaller workload variation. Overall, the combined effect of NBTI and RDF is significant, similar to the RF. We now show the implication of the $V_t$ distributions on the delay of the KSA.

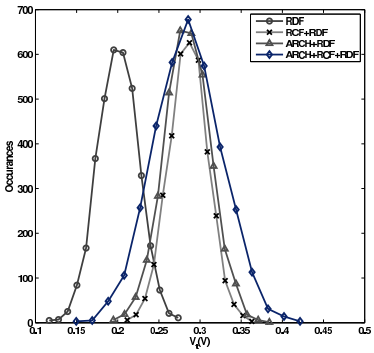


**Figure 6. $V_t$ distributions of the KSA due to RDF, temporal, workload and combined variation for the $mcf$ benchmark.**

As before, we use the *RDF* condition as our baseline. We calculate the percentage increase in delay with respect to the baseline for the other three conditions to analyze the impact of different variations due to NBTI.

Figure 7 shows the percentage increase in delay in the KSA with respect to the baseline due to NBTI in the presence of PV for three different conditions for different benchmarks. The increase in delay ranges from 9% to 15% for the *RCF+RDF* condition, 11% to 20% for the *ARCH+RDF* condition, whereas it ranges from 18% to 28% for the *ARCH+RCF+RDF* condition. Just like the RF behavior, this result also shows that the impact of the temporal variation is less than the impact of the workload variation. Unlike RF, in this case the combined effect is not higher than the sum of the individual effects. This is because of the cancelling effect of the variations in the same timing paths of the logic structure. Each timing path of the structure con-

sists of many PMOS devices which have different threshold voltages and the effect of the slower devices gets offset to some extent by the faster devices. Although, the combined effect of the workload and temporal variation causes an increase in the delay for each benchmark, this impact does not vary significantly across the benchmarks. Again, the reason behind this relates to the circuit design of the KSA which balances the values of 0's and 1's within the structure and reduces the impact of the variability in the utilization and bit patterns on the KSA across the different benchmarks.

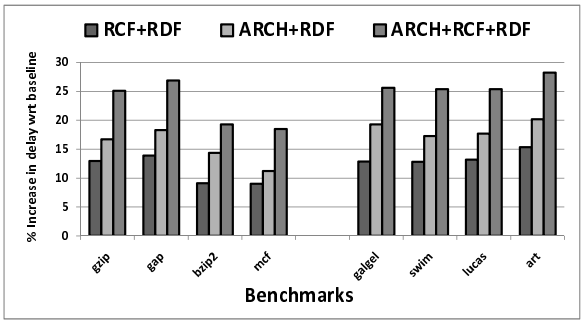


**Figure 7. Percentage increase in delay in a KSA due to temporal, workload and the combined variation for different benchmarks.**

### 6.3 Implications of the Results

- As the results indicate, both PV (RDF) and NBTI have a significant impact on $V_t$. More importantly, as Figures 4 and 6 show, if we consider only the impact of RDF or only Static NBTI (as is the case in a large number of architecture studies [1, 21, 18, 19, 20, 7]), then one does not get an accurate picture of the impact of these related reliability phenomena on the $V_t$ distributions. For example, if only RDF is considered, then the shift in the mean of the $V_t$ distribution due to NBTI is not captured. Even within NBTI, unless both temporal and workload variations are accounted for, the widening of the $V_t$ distribution will not be captured. It is important to capture these behaviors accurately in order to select appropriate guardbands and also develop effective mitigation techniques.
- While RDF and RCF depend on the underlying process, we can observe that the combined impact of RCF (temporal variation) and workload variation on lifetime reliability is significant. Since both temporal variation and workload variation strongly depend on the stress and recovery patterns on microarchitectural structures and also the bits that flow through them, there is large scope for NBTI mitigation at the architecture-level. However, it is important to develop and evaluate such mitigation techniques in way that is cognizant of the interaction between PV, temporal variation, and workload variation. The model that we have presented in Section 4 can be used to carry out such studies.

## 7 Conclusion

NBTI and PV are very important silicon reliability problems facing processor designers. In this paper, we develop an analytical model that captures both NBTI and PV for use in circuit and architecture simulations. We capture the following aspects in the model: i) variation in NBTI due to workloads, ii) temporal variation in NBTI and iii) process variation. We use this model to analyze the combined impact of NBTI and PV on a memory structure (register file) and a logic structure (Kogge-Stone adder). We show that the impact of the threshold voltage variations due to NBTI and PV both need to be captured in order to get an accurate view of silicon reliability.

## 8 Acknowledgments

We thank Adam Cabe for his valuable feedback. This research was supported in part by gifts from Intel.

## References

[1] J. Abella et al. Penelope: The NBTI-Aware Processor. *MICRO*, 2007.
[2] M. Agostinelli et al. Random Charge Effects for PMOS NBTI in Ultra-Small Gate Area Devices. *IRPS*, 2005.
[3] M. Basoglu et al. NBTI-Aware DVFS: A New Approach to Saving Energy and Increasing Processor Lifetime. *ISLPED*, 2010.
[4] S. Basu and R. Vemuri. Process Variation and NBTI Tolerant Standard Cells to Improve Parametric Yield and Lifetimes of ICs. *ISVLSI*, 2007.
[5] N. Binkert et al. The M5 Simulator: Modeling Networked Systems. *IEEE Micro*, 26(4):52–60, July 2006.
[6] Cadence Virtuoso Spectre Circuit Simulator. http://www.cadence.com/products/cic/spectre_circuit/.
[7] E. Chun et al. Shapeshifter: Dynamically Changing Pipeline Width and Speed to Address Process Variations. *MICRO*, 2008.
[8] X. Fu et al. NBTI Tolerant Microarchitecture Design in the Presence of Process Variation. *MICRO*, November 2008.
[9] K. Kang et al. Estimation of Statistical Variation in Temporal NBTI Degradation and its Impact on Lifetime Circuit Performance. *ICCAD*, 2007.
[10] K. Kang et al. Statistical Timing Analysis Using Levelized Covariance Propagation Considering Systematic and Random Variations of Process Parameters. *TODAES*, 2006.
[11] S. Kumar et al. Impact of NBTI on SRAM Read Stability and Design for Reliability. *ISQED*, 2006.
[12] Y. Lu et al. Statistical Reliability Analysis Under Process Variation and Aging Effects. *DAC*, 2009.
[13] S. Pae et al. Effect of BTI Degradation on Transistor Variability in Advanced Semiconductor Technologies. *TDMR*, 2008.
[14] Predictive Technology Model. http://www.eas.asu.edu/~ptm/.
[15] M. Rahma et al. Reducing SRAM Power Using Fine-Grained Wordline Pulsewidth Control. *TVLSI*, 2009.
[16] D. Schroder and J. Babcock. Negative Bias Temperature Instability: Road to Cross in Deep Submicron Silicon Semiconductor Manufacturing. *Journal of Applied Physics*, 94(1), July 2003.
[17] T. Sherwood et al. Automatically Characterizing Large Scale Program Behavior. *ASPLOS*, October 2002.
[18] T. Siddiqua and S. Gurumurthi. A Multi-Level Approach to reduce the Impact of NBTI on Processor Functional Units. *GLSVLSI*, 2010.
[19] T. Siddiqua and S. Gurumurthi. Recovery Boosting: A Technique to Enhance NBTI Recovery in SRAM Arrays. *ISVLSI*, 2010.
[20] A. Tiwari et al. ReCycle: Pipeline Adaptation to Tolerate Process Variation. *ISCA*, June 2007.
[21] A. Tiwari and J. Torrellas. Facelift: Hiding and Slowing Down Aging in Multicores. *MICRO*, November 2008.
[22] W. Wang et al. Compact modeling and simulation of circuit reliability for 65-nm cmos technology. *TDMR*, 2007.
[23] B. Zhang and M. Orshansky. Modeling of NBTI-Induced PMOS Degradation under Arbitrary Dynamic Temperature Variation. *ISQED*, 2008.
[24] SPEC CPU2000. http://www.spec.org/cpu2000/.