

# Intrinsic NBTI-Variability Aware Statistical Pipeline Performance Assessment And Tuning\*

Balaji Vaidyanathan<sup>†‡</sup>, Anthony S. Oates<sup>‡</sup>, Yuan Xie<sup>†</sup>

<sup>†</sup>Department of Computer Science and Engineering, Pennsylvania State University, PA 16801, USA

<sup>‡</sup>Technology Reliability Physics Dept., R&D, Taiwan Semiconductor Manufacturing Company, Hsinchu, Taiwan

E-mail: {vbalaji@tsmc.com,aoates@tsmc.com,yuanxie@cse.psu.edu}

**Abstract**—Random process variation and variability intrinsic to PMOS Negative Bias Temperature Instability (NBTI-induced statistical variation) are two major reliability concerns as transistor dimensions scales with technology. Previous works have studied these two sources of variation separately at device and circuit level. We study the impact of the interaction between intrinsic PMOS NBTI variability and time0 process variability on circuit delay spread. A statistical pipeline timing error model is proposed including both the variability sources to predict its impact on pipeline stage count. It is shown that a wide difference in statistical timing response to intrinsic NBTI variability exists among different circuits. Traditional design time NBTI-aware delay guard-banding is proved to be statistically insufficient in pipelines and an excess of 2x guard-band needs to be incorporated at the end of 10 years. However, the guard-band is shown to be reduced by 30% when the dynamic cycle time stealing technique is employed.

## I. INTRODUCTION

Sources of variations (process manufacturing conditions, and random dopant fluctuations) adversely affect the timing yield of the product at time0 (fresh after manufacturing) that degrades with time due to aging phenomenon like Negative Bias Temperature Instability (NBTI) in PMOS [1]. The most challenging source of variability to handle at time0 is the intrinsic random dopant fluctuation that needs statistical solution, as they cannot be controlled by any extrinsic means. Intrinsic variability in NBTI induced aging is the other source of random variation that has gained prominence with device dimension scaling beyond 45nm threatening the timing integrity of Static Random Access Memory (SRAM) and analog circuits during its operational lifetime [2].

PMOS devices of similar dimensions on the same die under identical NBTI stress condition experiences different NBTI induced drain current ( $I_d$ ) and threshold voltage ( $V_t$ ) degradation due to the intrinsic NBTI variability [2]. The random nature of the trap creation in gate dielectric and the substrate/gate-dielectric interface are being considered as the

\*This work is partly supported by NSF 0916887, NSF 0643902, and a grant from SRC.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICCAD'09, November 2–5, 2009, San Jose, California, USA.  
Copyright 2009 ACM 978-1-60558-800-1/09/11...\$10.00.

cause for such behavior that becomes more pronounced as gate length and gate oxide dimension scales with technology [2], [3]. Hence, apart from the extrinsic factors that affect a PMOS NBTI aging like its gate voltage, temperature and its gate input signal activity (that can be controlled or can be used as knobs to mitigate NBTI), presence of intrinsic variability introduces randomness in NBTI aging that needs statistical solution. Effect of intrinsic NBTI variability (or NBTI-induced statistical variation) is widely studied for SRAMs read stability issues and mismatch issues in analog circuits [2], [3]. Kang et al. [4] studied the time0 process and NBTI variability impact on SRAMs, simple library, and inverter chain circuits with technology scaling, however, the interaction of NBTI with time0 process variation was not analyzed for complex circuits and micro architectural components. In our previous work [5], we studied the interaction of time0 process variation and NBTI at circuit level however the intrinsic NBTI variability was not considered. Similarly, at the micro architectural level Fu et al. [6] proposed techniques to leverage the positive interaction between NBTI and process variation, however the intrinsic variability in NBTI was not considered.

There is plethora of study on time0 manufacturing and random process variability focusing on modeling, characterization, and simulation both at device and circuit design level [7], [8]. Particularly at circuit design level, the research focuses on statistical CAD algorithms and process variation aware designs [7]. However, the underlying mechanism that causes time0 process variations and the intrinsic NBTI variation during product operation is different [2]. Hence the interaction of such uncorrelated variability mechanisms create an additional layer of unpredictability in circuit delay during its operational lifetime. In this regard the work makes following contributions.

- 1) An analytical framework to explain the effect of intrinsic NBTI variability on statistical gate delay is developed.
- 2) A model is developed to track the effect of intrinsic NBTI variability on statistical pipeline timing during its lifetime and its impact on the pipeline stage count.
- 3) Finally, various statistical pipeline timing error improvement techniques are analyzed and dynamic cycle time stealing technique is proved to be an efficient way to handle intrinsic NBTI variability.

## II. BACKGROUND AND EXPERIMENTAL SETUP

The NBTI induced PMOS  $V_t$  degradation is considered to be a combination of interface traps and fast saturating hole-

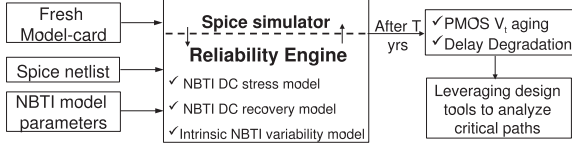


Fig. 1. Circuit NBTI simulation setup

traps in advanced technology (Equation 1) [1]. The interface trapping induced PMOS  $V_t$  degradation ( $\Delta V_{t\_it}$ ) is modeled as shown in Equation 2 in accordance with the reaction diffusion theory, while the hole-trapping ( $\Delta V_{t\_h}$ ) mechanism saturates at nominal voltages within tens of milliseconds [1]. Additionally, circuits operating at and beyond 1MHz of frequency does not encounter hole-trapping due to its large time constant [9]. At long stress periods interface traps dominates aging and hence we will consider Equation 2 in our work to track NBTI aging.

$$\Delta V_t = \Delta V_{t\_it} + \Delta V_{t\_h} \quad (1)$$

$$\Delta V_{t\_it} = \Delta V_{t0} * e^{A * E_{ox}} * e^{-E_a / K_B T} * t^n \quad (2)$$

Where fitting parameters ( $\Delta V_{t0}$ , and  $A$ ), activation energy ( $E_a$ ), Boltzmann constant ( $K_B$ ), electric field across the gate oxide ( $E_{ox}$ ), operational temperature ( $T$ ), stress time ( $t$ ), and time exponent ( $n = 1/6$  in accordance with reaction-diffusion theory) are used in modeling the NBTI behavior due to slow interface trapped charges [1]. One has to also incorporate the recovery model to understand the AC (or dynamic activity) behavior of the NBTI induced  $V_t$  stress in PMOS transistors. Universal recovery model is used in our analysis as proposed by Kaczer et al. [10] that follows from Equation 3.

$$r(\xi) = 1 / (1 + B \xi^\beta) \quad (3)$$

$$\xi = \frac{1}{DF} - 1 \quad (4)$$

$$\Delta V_{t\_AC} = R * r(\xi) + P \quad (5)$$

Where  $DF$  is the duty factor,  $B$  is the scaling parameter and  $\xi$  is the dispersion parameter [10]. The total NBTI induced  $\Delta V_{t\_AC}$  shift is considered to be a summation of permanent ( $P$ , permanent interface traps) and recoverable ( $R$ , recoverable interface traps) component (Equation 5), while the  $r(\xi)$  describes the duty factor and technology dependence of the recovery as shown in Equation 4. NBTI AC/DC factor derived from the above stress/recovery models is fed into the spice simulator for circuit lifetime extraction (Figure 1).

To predict complex digital circuit lifetime we synthesize circuits to the digital libraries and extract critical paths covering top 10% of the max delay. We statically calculate the  $V_t$  degradation at each transistor using 50% static signal probabilities at the circuit primary inputs (note only primary inputs see 50% static probability while the internal node probabilities depends on its connectivity). The NBTI induced  $V_t$  degradation is incorporated by adjusting the DELVTO parameter in HSPICE (using public domain BSIM4 model-card [11]) to obtain aged critical path delay. Intrinsic NBTI lifetime variability (model explained in Section III) is incorporated into the aging extraction framework (using DELVTO parameter) to track the product delay aging distribution with stress time.

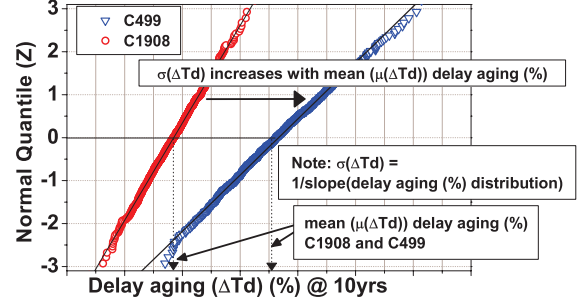


Fig. 2. Delay aging (%) spread in ISCAS85 circuits due to intrinsic NBTI variability

Process variation can be subdivided into global and local variation. Global variation encompasses inter-die, inter-wafer, and inter-lot variation, while the local variation covers the within-die (WID) variations. WID variation has random and correlated component. The sources of variation are modeled through Leff, Weff, Tox, and  $V_t$  (assuming Gaussian distribution) to precisely analyze statistical circuit characteristics. Monte-Carlo simulation (1000 runs) using HSPICE simulator (with public domain BSIM model-card [11]) is performed to obtain 3-sigma variance on transistor and circuit parameters.

### III. MOTIVATION: EFFECT OF INTRINSIC NBTI VARIABILITY ON CIRCUIT MEAN DELAY AGING WITH TECHNOLOGY SCALING

NBTI stress in PMOS is known to affect both the mean and variance of the induced  $\Delta V_t$  shift, which is considered to be a concern in SRAM and analog circuits [3] due to  $V_t$  mismatch problems. Based on the Poisson assumption for total NBTI induced charge distribution [2], we use Equation 6 to track the variation ( $\sigma(\Delta V_t)$ ) in  $\Delta V_t$  shift, given its mean ( $\mu(\Delta V_t)$ ).

$$\sigma(\Delta V_t) = \sqrt{\frac{K * T_{ox} * \mu(\Delta V_t)}{A_{Gox}}} \quad (6)$$

Where,  $T_{ox}$  is effective gate oxide thickness and  $A_{Gox}$  is its area, and  $K$  is an empirical constant equal to 1 [2]. Firstly, we assessed the effect of intrinsic NBTI variability induced  $\Delta V_t$  variation ( $\sigma(\Delta V_t)$ ) on circuit delay aging spread ( $\sigma(\Delta T_d)$ ). Figure 2 shows an increase in spread of the circuit delay aging ( $\sigma(\Delta T_d)$ ) with the mean circuit delay aging ( $\mu(\Delta T_d)$ ). This can be explained using Equation 6 that predicts an increase in the spread of NBTI induced  $V_t$  aging ( $\sigma(\Delta V_t)$ ) with the mean  $V_t$  shift ( $\mu(\Delta V_t)$ ). Additionally, using a simple linear gate delay model shown in Equation 7 relating  $V_t$  and  $T_d$  linearly, one can understand the similarity in gate delay and  $V_t$  behavior with intrinsic NBTI variability.

$$T_d = B * V_t + C \quad (7)$$

Secondly we analyze the sensitivity of circuits to intrinsic NBTI variability with technology scaling. Technology scaling has been incorporated by scaling the  $T_{ox}$  and  $A_{Gox}$  with generation [12] while maintaining  $K=1$ , and  $\mu(\Delta V_t)$  constant (which means that the PMOS devices across technology have the same mean NBTI lifetime [2] which is an optimistic

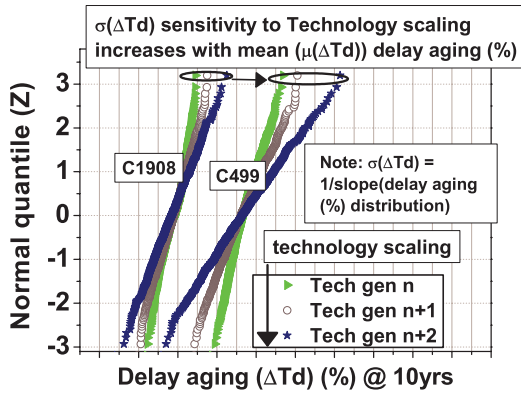


Fig. 3. Sensitivity of the delay aging (%) spread in ISCAS85 circuits to the  $A_{Gox}$  (that scales down exponentially with technology scaling)

assumption and hence does not affect our interpretation). Figure 3 shows that circuit (C499) with higher mean delay aging has a wider tail separation compared to its lesser aging counterpart (C1908).

These two findings have important implications both within and across technology. Firstly, the variation in higher delay aging circuits is larger compared to its lower aging counterpart within a particular technology node (Figure 2). Secondly the difference between the statistical spread in delay aging ( $\sigma(\Delta T_d)$ ) of a lower and higher aging circuits scales exponentially with technology (as evident from Equation 6 and Figure 3) thus exacerbating the discrepancy in delay aging among different circuits on the same die. The rest of the paper is organized into sections that consist of definition of the term "fall-out" to understand the interaction of intrinsic NBTI variability with time0 process variation; modeling the interaction at gate, circuit, and at pipeline level and verification using HSPICE based Monte-Carlo simulations; and finally assessing pipeline performance hit and proposing a dynamic cycle stealing technique and analyzing its benefits comparing with static guard-banding technique.

#### IV. EFFECT OF INTRINSIC NBTI VARIABILITY ON STATISTICAL DEVICE PARAMETERS

One can generically assume that the manufactured products to consist of transistors with their parameters ( $I_{dsat}/V_t$ ) falling within a certain range of spread defined by 3-sigma variation (or 99.9% worst-case/best-case value). However, this statistical spread shifts with device NBTI aging leading to more devices shifting out of the time0 statistical  $3\sigma$  of fresh device parameters (Figure 4). Thus NBTI induced aging could lead to a shift in the mean, sigma, or both in the device parameters as well as the circuit delay. To capture the combined effect of NBTI induced shift in both mean and sigma of statistical PMOS device parameter like drain saturation current ( $I_{dsat}$ ) and also the circuit parameter (delay) device parameters we arbitrarily define a term called fall-out (Figure 4). The fall-out is another way of looking at the device or circuit aging due to NBTI in a statistical domain and hence should not be confused with a failure indicator.

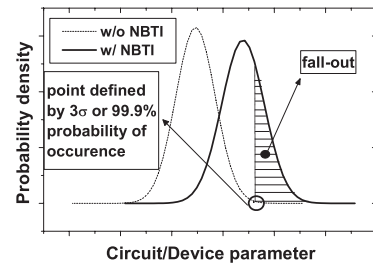


Fig. 4. Circuit delay or device parameters ( $I_{dsat}$  or  $V_t$ ) falling-out due to NBTI (Illustration)

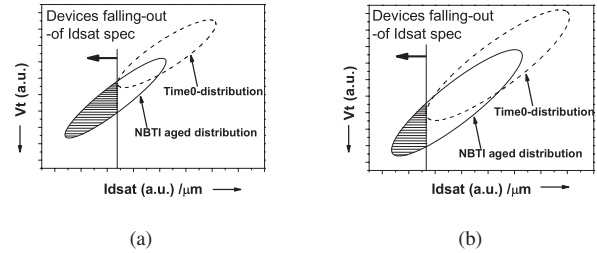


Fig. 5. (Illustration)  $I_{dsat}$  versus  $V_t$  distribution of fresh devices shifting due to identical NBTI stress in (a) larger width PMOS, and (b) smaller width PMOS

Figure 5 briefly illustrates the effect of static NBTI aging (without intrinsic NBTI variability) on the statistical  $I_{dsat}$  and  $V_t$  spread of PMOS device with different gate widths. We call the PMOS devices, with their  $I_{dsat}$  moving out of the  $3\sigma$  of fresh PMOS statistical  $I_{dsat}$  distribution as a result of NBTI, as  $I_{dsat}$ -fall-outs.  $I_{dsat}$  is chosen as fall-out indicator, as it is one of the major decider of transistor delay. Illustration in Figure 5(a) and 5(b) shows that the percentage fall-out for a larger width PMOS device is relatively higher compared to smaller width PMOS. It is well known that random dopant fluctuation induced time0 process variation increases with decreasing transistor dimensions [8]. In other words, smaller dimension devices will have larger time0 random process variation masking NBTI induced shift and vice versa for the larger dimension devices. This illustration explains an instance where one can observe the interaction between static NBTI aging and random process variation. However when statistical nature of NBTI (due to intrinsic NBTI variability) is also considered, the interaction becomes complex and hence the need for modeling.

#### V. EFFECT OF INTRINSIC NBTI VARIABILITY ON STATISTICAL SINGLE STAGE GATE DELAY

PMOS width and NBTI induced  $\Delta V_t$  shift has substantial effect on the fall-outs due to NBTI at the transistor level. At the circuit level NAND, NOR, and INV can be considered as basic building blocks. PMOS NBTI aging differs in the way it affects the delay aging of these three basic gates. It is well known that NBTI aging impacts NOR gate more than the NAND, and INV gate delay. The reason being that the PMOS stacking in NOR is more vulnerable to delay aging due to stack effect [13]. In this regard, the rise-delay fall-out of basic gates is analyzed applying global and local variation

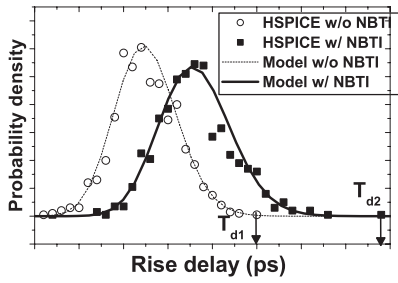


Fig. 6. INV rise delay ( $T_d$ ) PDF distribution (model and HSPICE based Monte-Carlo simulation) before and after NBTI aging in INV PMOS

(we assume that all the PMOS in the Inverter gate are fully correlated and the same for NMOS) to the individual gates. We have derived our delay fall-out model based on the gate delay model (following an alpha-power law based CMOS inverter delay model [14], where alpha is assumed to be equal to 1) shown in Equation 8.

$$T_d = \frac{A}{V_{gs} - V_t - \Delta V_t} \quad (8)$$

where,

$$A = \frac{C_{total} * V_{dd}}{n * \mu * C_{ox}} \quad (9)$$

Where,  $C_{total}$  is the output load capacitance,  $n$  is a fitting parameter,  $\mu$  is PMOS device mobility, and  $C_{ox}$  is the gate oxide capacitance. Assuming a Gaussian distribution for  $V_t$  (Equation 10) and  $\Delta V_t$  (Equation 11), one can derive the Probability Distribution Function (PDF) of  $T_d$  (Equation 12) to analyze the effect of time0 process variation and intrinsic NBTI variability induced  $V_t$  shift on the  $T_d$  distribution and the ensuing rise-delay fallouts.

$$f(V_t) = \frac{1}{\sigma_{V_t} \sqrt{2\pi}} e^{-\frac{(V_t - \mu_{V_t})^2}{2\sigma_{V_t}^2}} \quad (10)$$

$$f(\Delta V_t) = \frac{1}{\sigma_{\Delta V_t} \sqrt{2\pi}} e^{-\frac{(\Delta V_t - \mu_{\Delta V_t})^2}{2\sigma_{\Delta V_t}^2}} \quad (11)$$

$$f(T_d) = \frac{A}{T_d^2 (\sigma_{V_t} + \sigma_{\Delta V_t}) \sqrt{2\pi}} e^{-\frac{(V_{gs} - \frac{A}{T_d} - (\mu_{V_t} + \mu_{\Delta V_t}))^2}{2(\sigma_{V_t}^2 + \sigma_{\Delta V_t}^2)}} \quad (12)$$

$$F(T_d) = \frac{1}{2} \left( 1 + erf \left( \frac{V_{gs} - \frac{A}{T_d} - (\mu_{V_t} + \mu_{\Delta V_t})}{\sqrt{2(\sigma_{V_t} + \sigma_{\Delta V_t})}} \right) \right) \quad (13)$$

It can be observed from Figure 6 that the PDF of  $T_d$  distribution (Equation 12) is non-Gaussian with a right tail. Additionally the NBTI variability induced  $\mu_{\Delta V_t}$  and  $\sigma_{\Delta V_t}$  increase not only shifts the mean of inverter rise delay ( $\mu_{T_d}$ ) but also increases its spread ( $\sigma_{T_d}$ ). Thus explaining the non-usage of Gaussian model (that would have under-estimated the fall-out) for the inverter rise-delay to derive its delay fall-out with aging. Equation 13 shows the Cumulative Distribution Function (CDF) of the aged inverter rise delay, which is used to derive the fall-out in Equation 14.

$$fall - out(T_d) = F(T_{d2}) - F(T_{d1}) \quad (14)$$

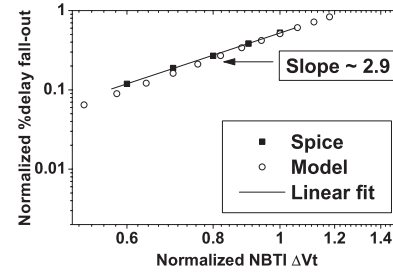


Fig. 7. Inverter rise delay fall-out (model and HSPICE based prediction) due to NBTI aging in INV



Fig. 8. Synthetic pipeline with ISCAS85 circuits

Where  $T_{d1}$  corresponds to the inverter rise-delay with 99.9% probability of occurrence before NBTI aging and  $T_{d2}$  corresponds to the inverter rise-delay with 99.9% probability of occurrence after NBTI aging (Figure 6). Hence with NBTI aging (increasing  $\mu_{\Delta V_t}$  and  $\sigma_{\Delta V_t}$  while keeping  $\sigma_{V_t}$ ,  $A$ , and  $V_{gs}$  as constants), the inverter rise-delay fall-out increases following a trend that is characteristic of the error function  $erf$  shown in Equation 13 (Figure 7).

It is interesting to note that the resultant single-stage inverter rise delay fall-out could be approximated with a power-law function for a considerable range of NBTI induced delay fall-out to conveniently propagate the model up the circuit level (dealt in Section VI). With the current understanding of the statistical delay fall-out model one can extend it to take care of NOR and NAND gates (whose fall-out versus NBTI induced  $\Delta V_t$  shift have power-law slopes of 3.1 and 2.3 respectively) with appropriate  $V_t$ ,  $A$ , and  $B$  parameters to match with the HSPICE Monte-Carlo prediction.

## VI. EFFECT OF INTRINSIC NBTI VARIABILITY ON STATISTICAL CIRCUIT AND PIPELINE DELAY

In this section we analyze the effect of intrinsic NBTI variability and time0 process variability on statistical circuit delay to derive a circuit delay fall-out model during its lifetime. The derived per-stage delay fall-out model is used to assess pipeline delay fall-out. We use a synthetic pipeline consisting of four stages constructed out of ISCAS85 benchmark circuits (Figure 8) with feedback from the last to the first stage.

Based on our statistical reliability framework (Figure 1) we calculate fall-outs for each of these pipeline stages as follows. We synthesize complex logic circuits (ISCAS85 benchmarks) using basic NBTI characterized libraries (INV/NAND/NOR) of varying transistor widths, stacks and fingers. Spice netlist of the circuits were augmented with Leff, Weff, Tox, and  $V_t$  variations that incorporate global (variation across die) and local (random, and correlated component within die) variations. The above-mentioned spice parameters that were used to model the total variation are assigned mean and standard deviations (following a Gaussian distribution) such

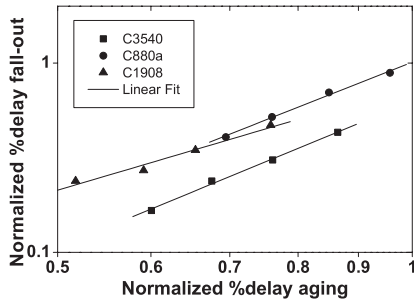


Fig. 9. ISCAS85 benchmark critical path delay aging versus delay fall-out (HSPICE based Monte-Carlo prediction) due to NBTI (Note: Local process variation component is completely random (correlation factor=0) among all the critical path libraries in this simulation; All %delay aging values are normalized to %delay aging value of C499 at 10yrs)

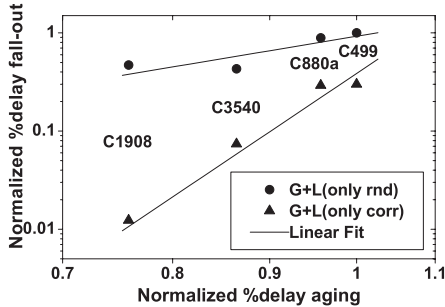


Fig. 10. ISCAS85 benchmark critical path delay aging and fall-out at the end of 10yrs (HSPICE based Monte-Carlo prediction) due to NBTI (Note: G (L) denotes global (local) variation; All delay fall-out values (%delay aging and %delay guard-band values) used in this paper are normalized to the %delay fall-out value (%delay aging value) of C499 (with local variation component that is completely random (correlation factor=0) among all the critical path libraries) at 10yrs)

that larger width transistors would observe a 10%  $I_{dsat}$  shift from the mean at the 3-sigma point.

Further the local variation (mismatch) dependence on  $L_{eff}$  and  $W_{eff}$  is modeled based on the empirical expression proposed by Asenov et al. [8]. In our simulation either we allow the circuits to have total random local variation or total correlated local variation to understand the difference between the NBTI interaction with the two extremes though in reality there will be a mix of both. HSPICE based Monte-Carlo simulation were performed on time0 and time10yrs NBTI aged circuits. NBTI characteristics are set such that a PMOS with 50% input duty cycle in a circuit will have mean  $V_t$  shift ( $\mu_{\Delta V_t}$ ) = 45mV at the end of 10yrs. However, note that the actual stress seen by a PMOS will be different based on its input activity.

Two main observations can be drawn from the NBTI induced fall-out prediction for the ISCAS85 benchmarks shown in Figure 9 and 10. Namely, the power-law dependence of circuit delay fall-outs on NBTI induced delay aging (Figure 9) and delay standard deviation of its critical path (Figure 10). The behavior of circuit fall-outs to NBTI induced delay shift can be understood from a simple inverter rise-delay fall-out behavior without loss of generality. Delay fall-out has a power law dependence on NBTI induced PMOS  $\Delta V_t$  (Figure 7). Additionally, based on a first order linear transistor delay ( $T_d$ ) approximation model shown in Equation 7 (note that the

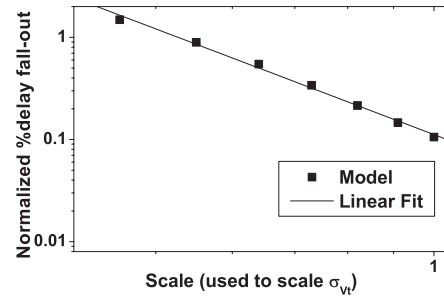


Fig. 11. Inverter rise delay fall-out (model from Equation 14 at the end of 10yrs due to NBTI aging (Note: scale value is used to scale down the time0 sigma of PMOS  $V_t$  distribution ( $\sigma_{V_t}$ ) used in Equation 13 and 14)

statistical mean value of  $T_d$  can be approximated with a linear dependence model for the range of  $\Delta V_t$  shift due to NBTI leading to 10%  $T_d$  shift), one can derive a linear relation between the gate delay response and NBTI induced PMOS  $\Delta V_t$ . Hence delay fall-out is a power-law function of the %mean delay aging that matches with the HSPICE simulation based prediction in Figure 9.

Secondly, NBTI induced delay fall-outs are smaller for circuits with local variation component that is completely correlated (correlation factor=1) among all the critical path libraries compared to the completely random local variation counterpart (Figure 10). This observation be explained based on the fact that a completely correlated local variation leads to higher time0 variation in statistical circuit delay compared to the completely random local variation case [15]. Additionally, Figure 11 shows a power-law dependence of circuit delay fall-out on the time0 circuit delay spread (sigma) based on the single-stage INV rise delay fall-out model (Equation 14). Hence there is a difference between the ISCAS85 delays fall-outs with completely correlated and random local variation case that explains the HSPICE simulation based prediction shown in Figure 10. An important implication of this understanding is that the tightening or reduction of time0 circuit delay variation leads to more NBTI induced fall-outs and hence the necessity to include more circuit delay guard-band during product operational lifetime.

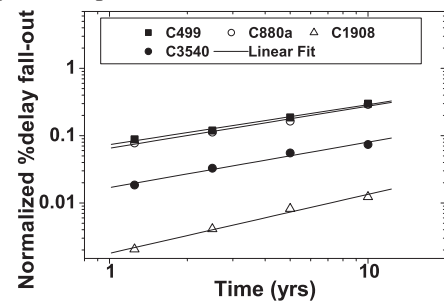


Fig. 12. ISCAS85 benchmark critical path delay fall-out during its operational lifetime (HSPICE based Monte-Carlo prediction with local variation component that is completely correlated (correlation factor=1) among all the critical path libraries)

We qualitatively interpreted the dependence of circuit delay fall-out on its delay aging and time0  $V_t$  variation. However, it is also important to track the dependence of circuit delay fall-out on the stress time to obtain a time dependent circuit delay

fall-out model. This will help in tracking time dependent delay fall-outs for complex systems like pipelines (refer Section VI-A). The delay fall-out of the circuit follows a power law (slope  $\approx 0.5$  to  $0.6$ ) behavior with time (Figure 12). This can be explained as follows. Figure 7 shows the power-law dependence of rise-delay fall-out of single stage INV on the NBTI induced PMOS  $\Delta V_t$  (slope=2.9). Similarly the slope for NOR gate is 3.1 (most of the synthesized ISCAS85 critical path gates are comprised of INV and NOR gates) as explained in Section V. Additionally we know that the NBTI induced  $\Delta V_t$  is related to stress time through power law slope=0.166 (Equation 2). Hence transitively, the dependence of single-stage gate rise-delay fall-out on stress time is a power law (with the power-law slope  $\approx 0.5$  ( $\approx 3.0 * 0.166$ )). Hence the single-stage gate delay fall-out model predicted slope qualitatively explains the fall-out versus stress time simulation behavior shown in Figure 12 for ISCAS85 benchmarks.

#### A. Pipeline Delay Fall-out Model

In this section the per-stage delay fall-out model is used to derive pipeline delay fall-out. The delay fall-out value for each stage ( $f_i$ ) at a given time during its operational lifetime (excluding flip-flop aging) is used to derive the total pipeline delay fall-out ( $F_p$ ) following a simple weakest link based system failure probability representation (Equation 15).

$$F_p = 1 - \prod_{i=1}^n (1 - f_i) \quad (15)$$

Where  $n$  is the total number of stages in the pipeline. Using the derived power-law model (in the previous section) for per-stage delay fall-out with stress time, we can rewrite the pipeline delay fall-out as shown in Equation 16.

$$F_p = 1 - \prod_{i=1}^n (1 - A_i \times t^{0.5}) \quad (16)$$

Where  $A_i$  is the delay fall-out (@ time  $t=1\text{sec}$ ) calculated using the fall-out power-law model for  $i^{\text{th}}$  pipeline stage. Equation 15 and 16 are derived based on two assumptions. Firstly, the stage delay fall-outs ( $f_i$ ) are considered to be totally independent. Secondly, all the stages are assumed to have same time0 delay at 3-sigma point ( $\mu + 3 * \sigma$ ) and the pipeline is designed to accommodate this delay. Such a simple model helps us track the real pipeline delay fall-out dependence on correlation factor and also on the unbalanced nature of per-stage time0 delay distribution.

Figure 13 and 14 shows the HSPICE calculated delay fall-outs for each stage and pipeline, as well as the model prediction of pipeline delay fall-out. The fully random local variation case (Figure 13) shows difference between the model predicted and the HSPICE calculated pipeline delay fall-outs. This can be explained as follows. The pipeline is designed to accommodate 3-sigma delay of stage with maximum spread. Hence there will be slack period ( $slack_i$ ) left for other stages to accommodate NBTI induced delay fall-outs. The resultant per-stage fall-out ( $f_i$ ) contribution to pipeline delay fall-out is re-written with per-stage slack information in Equation 17.

$$F_p = 1 - \prod_{i=1}^n (1 - A_i \times (t - slack_i)^{0.5}) \quad (17)$$

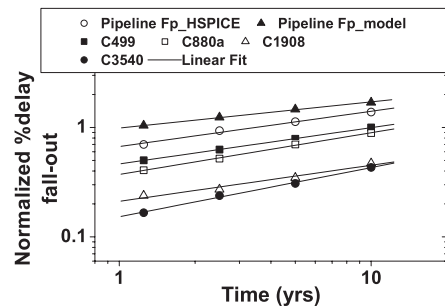


Fig. 13. ISCAS85 benchmark critical path delay fall-out during its operational lifetime (HSPICE based Monte-Carlo prediction with local variation component that is completely random (correlation factor=0) across all the critical path libraries and also across the pipeline stages as well)

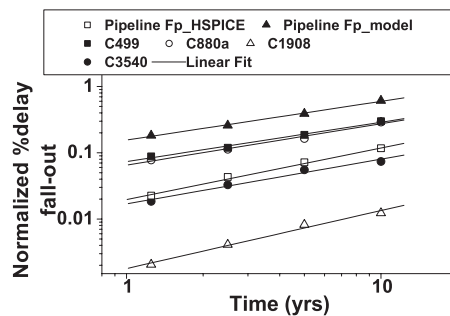


Fig. 14. ISCAS85 benchmark critical path delay fall-out during its operational lifetime (HSPICE based Monte-Carlo prediction with local variation component that is completely correlated (correlation factor=1) among all the critical path libraries and also across pipeline stages as well)

Where  $slack_i$  is the slack time period after which the  $i^{\text{th}}$  pipeline stage starts contributing to pipeline delay fall-outs.

The fully correlated local variation case (Figure 14) shows a much larger difference between the model predicted and the HSPICE calculated pipeline delay fall-outs compared to the fully random local variation case shown in Figure 13. Here, in addition to the excess slack ( $slack_i$ ) in certain stages (explanation is same as for the fully random local variation case elucidated in the previous paragraph), the presence of correlation among the stages makes the delay fall-outs to be mostly correlated as well. As a result, the per-stage delay fall-outs do not contribute separately to the pipeline delay fall-out, instead they occur almost in tandem. High correlation among the pipeline stages invalidates the pipeline delay fall-out model given in Equation 17, and resultant model is given in Equation 18 that lower bounds (due to existence of randomness induced by intrinsic NBTI variability) the HSPICE prediction.

$$F_p = \max(A_i \times (t - slack_i)^{0.5}) \quad (18)$$

Figure 15 shows the new  $F_p$  models (given in Equation 17, and 18) incorporated with correlation and slack information closely matching the HSPICE simulation of pipeline delay fall-out. Under an assumption of uncorrelated local variation in pipeline stage-delays and equal delay fall-out from all stages (i.e.  $A_i$  is equal for all pipeline stages) and  $slack_i = 0$  ( $\forall i = 1, \dots, n$ , where  $n$  is the total number of stages in the pipeline) the Equation 17 can be rewritten as shown in Equation 19

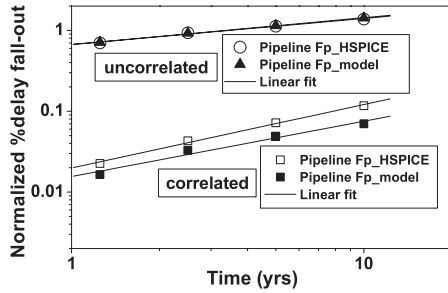


Fig. 15. Delay fall-out for pipeline with completely uncorrelated and correlated local variability components for stage delays (HSPICE versus model prediction) during its operational lifetime.

during the initial periods of pipeline operations (note that the delay fall-out is small during the initial time periods of pipeline operation and hence the approximation is valid).

$$F_p = 1 - \prod_{i=1}^n (1 - A_i \times (t - slack_i)^{0.5}) \approx n \cdot A \cdot t^{0.5} \quad (19)$$

Equation 19 elucidates an important implication of NBTI induced product delay fall-out on the architectural decision of increasing pipeline stages. The total pipeline delay fall-out linearly scales with the number of stages in the pipeline. Additionally, with the proposed delay fall-out model for pipeline one can efficiently track statistical pipeline reliability based on its per-stage reliability model. This helps in making architectural decision towards manufacturing a product with higher statistical lifetime.

## VII. INTRINSIC NBTI VARIABILITY AWARE RUNTIME TECHNIQUES TO MANAGE STATISTICAL PIPELINE PERFORMANCE

A pipeline is designed to accommodate stages of equal delay. However, after manufacturing, time0 process variation might invalidate such an assumption. Though one can adjust the clock cycle delay to meet the timing of the stage with maximum delay, it leads to a pessimistic clock delay assignment. Clock cycle time Stealing (CS) [16] is one of the techniques to deal with time0 process variation in pipelines, by averaging out the imbalance in stage delays thus gaining back the frequency that would otherwise be lost in pessimistic over-design (Figure 16). This averaging is achieved by stealing the slack in one stage into the other. The CS techniques necessitates the presence of tunable delay buffers in the clock network to shift the clock phase, built-in-self-test (BIST) vectors to analyze the critical path timing and to take care of minimum delay path constraints [16].

The afore-mentioned CS technique is used at time0 to average out the manufacturing variation in stage delays. However, one could extend the CS technique (leveraging the already present hardware and algorithm to enable CS technique) to operate at runtime to take care of NBTI by adding per-stage aging-aware guard-band. In this regard, presence of an aging detector is an additional benefit to track the per-stage delay aging with time and add appropriate guard-band

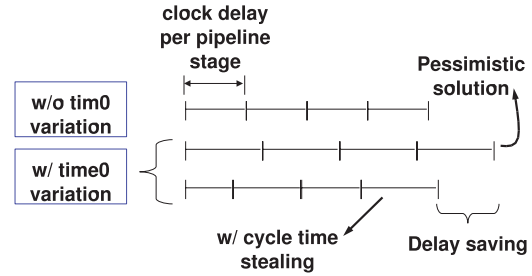


Fig. 16. Post-manufacturing solution to handle time0 process variation in a 4-stage pipeline.

to stage delays pro-actively. Otherwise, timely execution of BIST vectors would be necessary to track stage delay timing failures and add guard-band appropriately. Explanation of aging detector or the BIST vector is beyond the scope of the work and hence we assume their presence. The main aim of the work is to motivate the benefit of dynamic CS technique and assess its advantages through the following experimental setup and study.

Per-stage delay guard-banding (refer Figure 17) HSPICE analysis shows that there is a wide difference in the reduction of delay fall-out among different stages obtained by incorporating delay guard-band. Pipelines can have such large difference among its stage delay fall-outs due to three reasons. Firstly, the difference in time0 delay spread (dependent on manufacturing variation) among the stages affects delay fall-outs differently (refer Figure 10 and 11). Secondly, the difference between the NBTI induced mean delay aging of the stage critical path leads to wide difference in delay fall-outs (refer Figure 9 and 10). Finally, the difference in delay fall-outs among the stages widens with technology scaling (refer Figure 3). Hence there is a need for dynamic CS technique to handle wide differences in statistical delay fall-out among all the stages in a pipeline to obtain optimal pipeline delay fall-out with less performance hit during its operational lifetime.

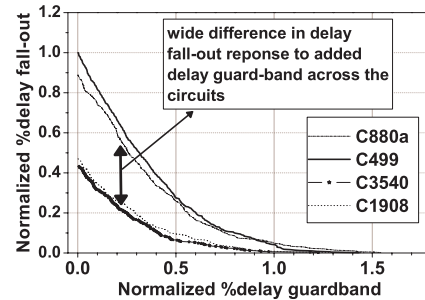


Fig. 17. Decrease in delay fall-out (%) of each stage at the end of 10yrs of the pipeline operation with increase in incorporated delay guard-band (HSPICE prediction for ISCAS85 benchmark circuits) (Note: All %delay guard-banding values used in the paper are normalized to %delay aging value of C499 at 10yrs)

To take care of NBTI traditionally one would add guard-band to pipeline stages at design time that equals maximum statistical mean delay aging of all stages (calculated for 10yrs of operation, shown in x-axis of Figure 10) anticipating the product to meet this delay at 10yrs (Figure 18(a)). However, once the design is manufactured, there will be spread in the stage delays that may exceed the mean delay guard-band

incorporated at time0. This process variation induced spread could be taken care by the CS technique at time0 (Figure 18(b)). Finally, the NBTI aged statistical delay spread at the end of 10yrs (as a result of interaction of intrinsic NBTI variability with time0 process variation) may exceed beyond the traditional mean delay aging guard-band and time0 CS technique. Hence the application of time10yrs dynamic CS technique will be required at the end of 10yrs as shown in Figure 18(c).

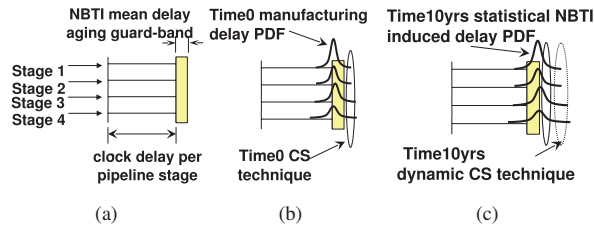


Fig. 18. (a) All 4 stages of the pipeline are designed with equal stage delays and incorporated with a 10yrs NBTI mean delay aging guard-band, (b) time0 manufacturing spread on the stage delays that exceeds the NBTI aging guard-band are taken care by time0 CS technique, and the (c) time10yrs statistical NBTI spread taken care by time10yrs dynamic CS technique.

Figure 19 plots the HSPICE simulation results for the pipeline incorporated with the three techniques mentioned in Figure 17. Firstly, the incorporation of 10yrs NBTI mean-delay aging guard-band that is used by the time0 CS and dynamic CS technique is not entirely sufficient to protect the pipeline from incurring any delay fall-outs. As a result, twice the mean delay aging guard-band (compare the guard-band indicated by the larger downward arrow and the tail of the plot shown in asterisk in Figure 19) has to be incorporated into the pipeline using dynamic CS technique to reduce delay fall-outs to zero at 10yrs. Further, in the absence of dynamic CS technique the pipeline would have required 50% excess guard-band to prevent any delay fall-outs at 10yrs (compare dotted curve with the plot shown in asterisk in Figure 19). In other words, usage of dynamic cycle time stealing technique brings down the guard-band needed by 30% compared with the static guard-banding technique at the end of 10 years. This shows that adding static guard-band at design time is statistically insufficient and justifies the usage of dynamic CS in handling delay fall-outs entirely without any static guard banding.

### VIII. CONCLUSION

In this paper, we established a framework to link the effect of intrinsic NBTI variability to process variation induced statistical delay distribution of library cells, more complex circuits and pipeline systems. Circuit delay fall-out (-of time0 manufacturing  $3\sigma$  statistical spread of product delay) is predicted to have power-law dependence on stress time. The obtained circuit level delay fall-out model is used to derive a stress-time dependent statistical pipeline timing error model, using which the impact of pipeline stage count on pipeline delay fall-out is assessed. Using our statistical model framework, the presence of wide difference among the pipeline stage delay fall-out response to intrinsic variability in NBTI is predicted and shown to worsen with technology scaling. Consequently, we prove that it is statistically insufficient to use traditional NBTI induced mean delay aging value for guard-banding the

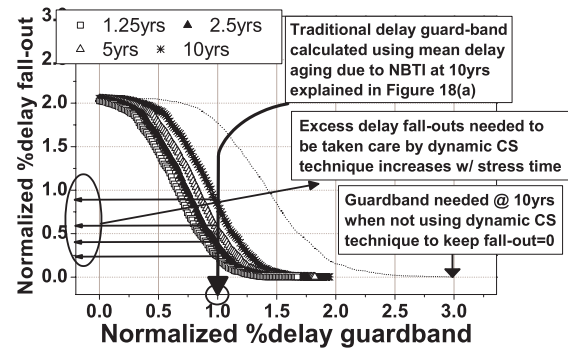


Fig. 19. Decrease in delay fall-out (%) of the pipeline (implemented with dynamic CS technique) with increase in delay guard-band. (Note:- CS means clock Cycle time Stealing)

pipeline and an excess of 2x guard-band would be required to mitigate pipeline time error at the end of 10 years. Benefits of extending cycle time stealing technique to handle pipeline delay fall-outs at runtime is analyzed. It is shown that 30% lesser delay guard-band is needed compared to the pipeline operating without dynamic cycle time stealing technique to prevent statistical pipeline timing errors due to intrinsic NBTI variability at the end of 10 years.

### REFERENCES

- [1] A. E. Islam, H. Kufluoglu, D. Varghese, S. Mahapatra, and M. A. Alam. Recent Issues in Negative-Bias Temperature Instability: Initial Degradation, Field Dependence of Interface Trap Generation, Hole Trapping Effects, and Relaxation. *IEEE TED*, Sept 2007.
- [2] S. Pae, J. Maiz, C. Prasad, and B. Woolery. Effect of BTI Degradation On Transistor Variability in Advanced Semiconductor Technologies. *IEEE TDMR*, Sept 2008.
- [3] S. E. Rauch. Review and Reexamination of Reliability Effects Related to NBTI-Induced Statistical Variations. *IEEE TDMR*, Dec 2007.
- [4] K. Kang, S. P. Park, K. Roy, and M. A. Alam. Estimation of Statistical Variation in Temporal NBTI Degradation and its Impact on Lifetime Circuit Performance. In *ACM/IEEE ICCAD*, 2007.
- [5] B. Vaidyanathan, A. S. Oates, Y. Xie, and Y. Wang. NBTI-Aware Statistical Circuit Delay Assessment. In *IEEE ISQED*, 2009.
- [6] X. Fu, T. Li, and J. Fortes. NBTI Tolerant Micoarchitecture Design in the Presence of Process Variation. In *IEEE MICRO*, 2008.
- [7] J. A. G. Jess, K. Kalafala, S. R. Naidu, R.H.J.M, Otten, and C. Visweswariah. Statistical Timing for Parametric Yield Prediction of Digital Integrated Circuits. *IEEE Tran. On CAD of ICAS*, Nov 2006.
- [8] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva. Simulation of Intrinsic Parameter Fluctuations in Decanometer and Nanometer-Scale MOSFETs. *IEEE TED*, Sept 2003.
- [9] P. Hehenberger, T. Aichinger, T. Grasser, W. Gos, O. Triebel, B. Kaczer, and M. Nelhiebel. Do NBTI-Induced Interface States Show Fast Recovery? A Study Using a Corrected On-the-fly Charge-Pumping Measurement Technique. In *IEEE IRPS*, 2009.
- [10] B. Kaczer, T. Grasser, P.J. Roussel, J. Martin-Martinez, R. O'Connor, B.J. O'Sullivan, and G. Groeseneken. Ubiquitous Relaxation in BTI Stressing-New Evaluation and Insights. In *IEEE IRPS*, 2008.
- [11] <http://www.eas.asu.edu/~ptm>.
- [12] [http://www.itrs.net/Links/2007ITRS/2007\\_Chapters](http://www.itrs.net/Links/2007ITRS/2007_Chapters).
- [13] R. Vattikonda, W. Wang, and Y. Cao. Modeling and Minimization of PMOS NBTI Effect for Robust Nanometer Design. In *ACM/IEEE DAC*, 2006.
- [14] T. Sakurai and A. R. Newton. Alpha-Power Law MOSFET Model and its Application to CMOS Inverter Delay and Other Formulas. *IEEE JSSC*, Apr 1990.
- [15] P. Friedberg, Y. Cao, J. Cain, R. Wang, J. Rabaey, and C. Spanos. Modeling Within-die Spatial Correlation Effects for Process-design Co-optimization. In *IEEE ISQED*, 2005.
- [16] A. Tiwari, S. R. Sarangi, and J. Torrellas. Recycle: Pipeline Adaptation to Tolerate Process Variation. In *ACM ISCA*, 2007.