System-Level Estimation of Threshold Voltage Degradation due to NBTI with I/O Measurements

Soonyoung Cha, Chang-Chih Chen, and Linda S. Milor
Georgia Institute of Technology, 30332, Atlanta, GA, USA, E-mail: scha31@gatech.edu

Abstract—With the scaling of CMOS technology, Negative Bias Temperature Instability (NBFI) and Process Variations (PV) are serious issues for transistors. Normally, degradation due to NBFI is modeled based on test structure data or ring oscillators embedded within product die. In this paper, we present a method to determine the initial average channel length \( L \) and threshold voltage \( V_{th0} \) for individual chips, together with NBFI model parameters through I/O measurements. We determine a relationship between \( \Delta V_{th}, V_{th0}, L \) and ground signal variation and fit models to the simulation results. The voltage of the ground signal is used for the calculation of the delay and amplitude shifts which are used to extract PV and NBFI parameters. Then, we calculate the lifetime for each chip individually using calibrated NBFI models, accounting for process variations. The methodology enables the extraction of NBFI and PV model parameters for individual chips, not just for the manufacturing process, and hence it becomes possible to differentiate chips that have different parameters initially and are more or less vulnerable to NBFI.

I. INTRODUCTION

NBFI is an important wearout mechanism, which affects pMOS devices by shifting the threshold voltage \( V_{th} \). This increase in \( V_{th} \) degrades the operating speed of circuits, and, in extreme cases, causes the circuit to behave incorrectly because of timing constraints \[1\]. Also, difficulties in controlling lithography and doping cause significant PV. Because PV also cause variation in the transistor characteristics (channel length, threshold voltage, oxide thickness), which adds to that caused by NBFI, chip reliability should take into account not only the NBFI-induced \( V_{th} \) shift, but also the initial \( V_{th} \) and channel length in each individual chip \[2\].

During the last decade, the reaction-diffusion (R-D) model has been used to explain the NBFI effect in a pMOS device, which is the breaking and rebonding of hydrogen-silicon bonds at the silicon-gate dielectric interface \[3\]. However, although the R-D model explains the threshold voltage shifts produced by DC bias and temperature stress, it could not explain the speed of recovery and the impact of the dynamic gate bias \[4\]–\[8\]. Therefore, recently, the charge trapping and detrapping (T-D) model is being considered as a more likely explanation of the NBFI mechanism \[4\],\[5\].

Starting with a physical model of NBFI, which characterizes the threshold voltage shift as a function of time, the device-level models can be used to derive gate-level simulation models and system lifetime \[9\]–\[12\]. System-level prediction of lifetime enables the optimization of timing guardbands or circuit adaptation based on a prediction of the increase in delay as a function of time, temperature, and usage. This can be done with look-up tables of frequency and voltage pairs as a function of time under operation \[13\]. This approach assumes that the NBFI model is known, and hence errors result if failure rate parameters or operating history are not estimated correctly. Also, the lifetime of a circuit is jointly affected by NBFI and PV, circuit timing has to consider both effects.

Alternatively, ring oscillators \[14\] or replica critical paths \[15\],\[16\] are used to analyze degradation. However, the ring oscillators or the replica critical paths do not share the same ambient environment because of the difference in their on-die location and the critical paths within a datapath can change as a result of degradation \[17\]. Hence, this approach may not produce an accurate prediction of the aging rate of the true critical paths. An alternative, involving direct monitoring of critical paths \[18\],\[19\], entails some area overhead and the correct selection of the vulnerable critical paths. Another approach involving direct monitoring of aging, involves the use of specially designed latches that detect delay errors, at the expense of significant power overhead per latch \[20\],\[21\].

In this paper, we build a method to find initial average PV parameters using the ground voltage signal. We analyze the relationship between PV parameters and an initial signature signal at time zero to estimate the average values for PV parameters. Also, as mentioned in \[22\], the threshold voltage shift due to NBFI translates into shifts in current, which should be visible by monitoring the signature on the ground terminal during operation. We find a relationship between the \( V_{th} \) shift and the shift in the signature to extract NBFI parameters. In an actual circuit, data are measured at the I/Os using the ground signature. PV parameters are calculated using the initial ground signal, and the NBFI parameters are derived from the threshold voltage shift, extracted from the shift in the ground signal with time. The results are combined to provide a lifetime estimate for each chip. Fig. 1 shows the full process of how we can get the PV and NBFI model for a circuit using measurement data from the ground signature and how we can use two models, one for the NBFI failure rate parameters and one for PV, to estimate the lifetime for each chip, accounting for PV.

We summarize, in Section II, the T-D model of the average and standard deviation of the threshold voltage shift. In Section III, the system level chip simulation results are generated by FastSpice. For the PV effect on a chip, we consider the global variation of \( V_{th0} \) and the length of MOS devices for the whole chip. Then, we extract the signature signal, which is affected by variation. Using the affected signature signal and comparing it with the nominal signature signal, we can extract the delay and amplitude variation. In order to see the degradation effect, we applied the \( \mu(\Delta V_{th}) \) and \( \sigma(\Delta V_{th}) \), extracted in Section II to each stressed pMOS to determine the degradation of the signature signal. The signature signal is analyzed to extract the amplitude and delay shift. We check the relationship between the threshold voltage shifts and amplitude and delay shifts for different circuits and input patterns. Then, we determine the accuracy of our model.
for PV extraction and NBTI degradation. Section IV presents the experimental results for extracting PV and the $V_{th}$ shift for the same chips as a function of stress time. Then, we estimate the lifetime of each chip, based on the extracted model parameters and conclude with a summary in Section V.

II. CHARGE TRAPPING AND DETRAPPING NBTI MODEL

Random Telegraph Noise (RTN), that is, charge trapping and detrapping of oxide defects, has been observed in submicron FETs. Defects, which are located in the oxide layer in pMOSs, capture and emit charges [23]. If a defect captures a charge carrier in the oxide (SiO$_2$), unexpected current noise occurs, and the charged defects affect the mobility and scattering in the device [23],[24].

For a device under constant bias, because charge trapping depends on the Fermi level only [24], the current remains constant on average and does not change with time. In digital IC applications, the gate voltage (bias) and the Fermi level models the impact of the applied voltage. The fluctuation in average threshold voltage is determined by multiplying equation (1) by $\delta$, the shift in threshold voltage due to a single trap [30], to get

$$\Delta V_{th}(t) = \delta n(t)$$

(3)

Therefore, using equation (3), we obtain the theoretical average and standard deviation of the shift in threshold voltage for each pMOS device in our system. Note that the relative variation in the number of defects, $\sigma(n(t))/n(t)$ decreases as a function of time, as does the relative variation in the threshold voltage. This is consistent with experimental observations of delay variations in circuits [31].

Digital circuit operation has stress and recovery periods. The fraction of time under stress is called the duty cycle, $\alpha$. Instead of modeling the shift in threshold voltage for each stress and recovery period separately, it is best to determine an effective Fermi level, as a function of duty cycle [32]:

$$E_{F,\text{eff}}(\alpha) = \alpha E_{F,\text{ph}} + (1 - \alpha)E_{F,\text{eff}}$$

(4)

The effective Fermi level is used to determine the evolution of threshold voltage degradation. Therefore, $\varphi(T, E_{F,\text{eff}})$ can more reasonably adjust the probability of charge capture and emission. For example, it can be seen that as $\alpha$ approaches one, when there is no time for recovery, degradation is much more significant. This is illustrated in Fig. 2, using the U-shaped trap density distribution function for $\varphi(T, \alpha)$ [23],[29].

Fig. 3 illustrates the shift in pMOS threshold voltage for a switching inverter, showing stress and recovery cycles. It also illustrates a ground signature signal, which shows the shift in delay and amplitude as a function of time under stress.

III. SIGNATURE SIGNAL ANALYSIS AND MODELING FOR NBTI AND PROCESS VARIATION

Several case studies involving different chips and test benches have been considered: a Microprocessor, a Floating Point Unit (FPU), and a Finite Impulse Response (FIR) filter.
which is affected by PV. The delay and amplitude variation are threshold voltage shift. Fig. 4 shows an example signature, Fig. 3 shows the ground voltage signature as a function of the components, then it becomes possible to solve for the average variations, \( \Delta \) is used to computed the corresponding principal components, \( \Delta \) as the solution of \( C \)\( \Delta \) = \( \Delta \), \( \Delta \) is a principal component. Then, for an unstressed circuit, the difference between a simulated nominal signature and an experimental signature is computed. The shifts in amplitude and delay are computed, \( \Delta A \) and \( \Delta D \). These values are used to computed the corresponding principal components, \( PC_i \), \( i = 1, \ldots, n \), to form a right hand side vector, \( PC \). If \( C^T \) is the transpose of \( C \), then we solve for a vector of process variations, \( \Delta p \) as the solution of \( \Delta p = (C^T C)^{-1}C^T PC \). Effectively, the process parameters are estimated as a linear combination of shifts in amplitude and delay, i.e.,

The process of extracting PV involves three steps. First, we use a full factorial design to generate a set of 16 experiments at the 16 corners of the process domain. In other words, a maximum and minimum value is set for the nMOS and pMOS threshold voltage domain and the nMOS and pMOS channel length domain. Simulations of the ground signatures are performed for each combination of the corners for \( V_{thp0} \), \( V_{thn0} \), \( L_{p0} \), and \( L_{n0} \). This consists of a 16 simulations.

Second, the signature on the ground terminal is computed for an input pattern. Each maximum and minimum point in the signature is determined and associated with an amplitude and time point, shown in Fig. 5. Hence, several amplitude and time point pairs are computed for each signature signal. Third, we calculate the difference between the time point and amplitude pairs for the nominal signal and the corner signal to form a matrix. The matrix has 16 rows corresponding to 16 corner simulations and one column for each peak point in the signature.

Each signature has many peaks. Hence, the shifts in the peaks produce a large data set of amplitude and delay shifts \( (A_{PV} \text{ and } D_{PV}) \), which is reduced to a few principal components [34]. The principal components are linear combinations of the shifts in delay and amplitude, i.e. \( PC = \sum_j a_{ij} \Delta A_j + b_{ij} \Delta D_j \). They provide dimensional reduction and are computed using the covariance matrix among the shifts in amplitude and delay at the set of peaks. The values of the principal components are computed for each of the 16 corner experiments. From this set of experiments, each principal component is modeled as a function of PV, i.e. the shift from an assumed nominal value for channel length and threshold voltage, \( \Delta L \) and \( \Delta V_{th} \) for nMOS and pMOS devices, i.e. \( PC_i = c_{1i} \Delta L_{rn} + c_{2i} \Delta L_{lp} + c_{3i} \Delta V_{thn} + c_{4i} \Delta V_{thp} \). If there are at least four principal components, then it becomes possible to solve for the average shift in \( \Delta L_{rn} \), \( \Delta L_{lp} \), \( \Delta V_{thn} \), and \( \Delta V_{thp} \). For \( m \) principal components, the matrix \( C \) is \( m \times 4 \), where each row corresponds to a principal component. Then, for an unstressed circuit, the difference between a simulated nominal signature and an experimental signature is computed. The shifts in amplitude and delay are computed, \( \Delta A \) and \( \Delta D \). These values are used to computed the corresponding principal components, \( PC_i \), \( i = 1, \ldots, n \), to form a right hand side vector, \( PC \). If \( C^T \) is the transpose of \( C \), then we solve for a vector of process variations, \( \Delta p \) as the solution of \( \Delta p = (C^T C)^{-1}C^T PC \). Effectively, the process parameters are estimated as a linear combination of shifts in amplitude and delay, i.e.,
Average error rate, δ, of each source(%) & Applied standard deviation to $L$ and $V_{th}$

<table>
<thead>
<tr>
<th>$\delta(\Delta L_i)$</th>
<th>5.96</th>
<th>5.97</th>
<th>5.98</th>
<th>5.99</th>
<th>5.99</th>
<th>6.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta(\Delta L_t)$</td>
<td>5.94</td>
<td>5.97</td>
<td>5.98</td>
<td>6.00</td>
<td>5.99</td>
<td>5.98</td>
</tr>
<tr>
<td>$\delta(\Delta V_{th})$</td>
<td>3.91</td>
<td>3.94</td>
<td>3.94</td>
<td>3.94</td>
<td>3.94</td>
<td>3.94</td>
</tr>
<tr>
<td>$\delta(\Delta V_{thp})$</td>
<td>3.22</td>
<td>3.22</td>
<td>3.22</td>
<td>3.22</td>
<td>3.23</td>
<td>3.22</td>
</tr>
</tbody>
</table>

$$\Delta L = \sum_{i=1}^{n} (a_i \Delta A_i + b_i \Delta D_i), \quad \Delta V_{th} = \sum_{i=1}^{n} (c_i \Delta A'_i + d_i \Delta D'_i). \quad (5)$$

**B. Validation of the Methodology for Extracting PV**

We need to check the validity of our methodology. To mimic a real process, each device is assigned a random initial threshold voltage and channel length. For each sample, an average shift in threshold voltage and channel length for nMOS and pMOS devices is applied. The signature is computed by simulation for each sample and shifts in amplitude and delay are computed for the peak points. From these shifts, the model computed with equation (5) is used to estimate the mean shift in threshold voltage and channel length. The estimation of threshold voltage and channel length is compared with the actual shifted value. Table I shows the average error rate based on several simulation runs.

Within-die random variation can reduce the accuracy in estimating the mean values of threshold voltage and channel length. Hence, we have included the impact of random variation (0 ~ 10%) in Table I. All errors are under 6%, indicating that process parameter values can be computed accurately.

**C. Data Extraction and Modeling for NBTI**

Given a ground voltage signature, for the analysis of NBTI, our method extracts the shift in delay and amplitude as a function of time. Instead of collecting a set of signatures from a full factorial experiment, signatures are generated with an initial random set of process parameters, to mimic a true process, at different time points, to mimic the impact of stress. The delay ($\Delta L_{NBTI}$) and amplitude ($\Delta A_{NBTI}$) sets are computed, at each peak, as illustrated in Fig. 5. The average delay and amplitude as a function of time for NBTI is computed to form the vectors $D_{NBTI}$ and $A_{NBTI}$. Fig. 6(a) shows a signature signal as a function of $V_{th}$ degradation. Fig. 6(b) shows the delay and amplitude degradation due to NBTI for different circuits and test benches at different time points. We find the average $D_{NBTI}$ ($\Delta D_{avg}$) and $A_{NBTI}$ ($\Delta A_{avg}$) shift of peak points in the signature at different stress times, and $\Delta V_{th}$ is computed by regression [34] from the average shifts in the peak points of the signature, i.e.,

$$\mu(\Delta V_{th}) = \mu + \alpha \Delta A_{avg} + \beta \Delta D_{avg} + \gamma \Delta A_{avg} \Delta D_{avg}. \quad (6)$$

where $\alpha$, $\beta$, $\gamma$, and $\varepsilon$ are fitting constants. Fig. 6(c) compares the average shift in threshold voltage with the true shift in threshold voltage, showing that our method is accurate. Given an extracted shift in the threshold voltage as a function of time, computed using the model in equation (6), the model parameters in equation (1), $A$ and $B$ are computed.

**D. Validation of the Methodology for Extracting NBTI Parameters**

We need to check if our methodology can correctly extract NBTI model parameters in equation (1) using signature data. To do this, we assume process-level values are given for NBTI model parameters in equation (1) using signature data. Each device is assigned a random initial threshold voltage. NBTI degradation models are applied to each device, and the signature is simulated. From the signature, the shift in amplitude and delay is computed, and the average shift in threshold voltage is estimated. From the shift in threshold voltage as a function of time, we estimate $A$ and $B$.

Table II shows the average error rate for each coefficient, based on several simulation runs for each of the circuits and test benches using full chip simulations.
TABLE II.
THE ERROR (\(\delta\)) IN COMPUTING THE COEFFICIENTS IN EQUATION (1)

<table>
<thead>
<tr>
<th>Chip Microprocessor FPU FIR</th>
<th>Test Bench</th>
<th>Add</th>
<th>Cache</th>
<th>Add</th>
<th>FTI</th>
<th>Toggle</th>
<th>DC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\delta(A)) (%)</td>
<td>2.742</td>
<td>2.471</td>
<td>2.573</td>
<td>2.147</td>
<td>1.926</td>
<td>1.759</td>
<td></td>
</tr>
<tr>
<td>(\delta(B)) (%)</td>
<td>1.967</td>
<td>1.679</td>
<td>1.846</td>
<td>1.641</td>
<td>1.687</td>
<td>1.576</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 7. Measurement and extraction results from the initial signature signal variation. (a) Signature signal at time zero for several microprocessors and (b) extracted process parameters.

IV. EXPERIMENTAL RESULTS

Our PV and \(V_{th}\) degradation models have been applied to extract PV parameters and NBTI parameters for several microprocessor test chips, which is the same design as one of the simulated circuits. In this case study, the MOSFET gate lengths ranged from 90nm to 500nm and a wide gate width of 5\(\mu\)m was used. The nominal threshold voltage was 0.4V at 25°C. We considered supply voltages of 1.4V, 1.6V, and 1.8V, and temperatures of 40°C, 60°C, 80°C, 100°C, and 120°C. Our purpose is to show that shifts in amplitude and delay are measurable, and hence can be used to extract the shift in threshold voltage. Also, we applied a DC input signal to stress the chips in order to avoid the HCI degradation effect. The dynamic test bench is applied during testing so that ground bounce can be observed.

A. Data Collection and Computation of PV Parameters and Threshold Voltage Degradation

As noted in the previous section, the time zero ground signature signal was recorded with an oscilloscope for extraction of process parameters. Several initial ground signals are shown in Fig. 7(a). We measured the ground voltage signature from nine chips. The average of the nine signatures is assumed to be the nominal case. Using equation (5), the calculated process parameter shifts from nominal are shown in Fig. 7(b).

Next, after intervals of voltage and temperature stress, we monitored the ground signal. The NBTI-induced ground voltage was recorded by an oscilloscope, and \(\Delta\)Amplitude and \(\Delta\)Delay were computed at peaks in the ground signal. In Fig. 8(a), we capture the voltage signature degradation as a function of time for different input voltages and temperatures. The extracted degradation data is shown in Fig. 8(b). Using equation (6), the extracted \(V_{th}\) drift data are shown in Fig. 9. These degradation results are similar to data in the literature [28],[32],[35].

We have used a sampling oscilloscope and have used the “average” function and “bandpass filter” function to filter out the noise and jitter. These functions use 512 samples of the same time point to generate a stable ground signal. The sampling is every 0.2 picosecond, which is sufficient to see the delay degradation. The resulting signal is then processed for amplitude and peak detection.
B. Performance Degradation Analysis

In order to further study the joint impact of NBTI and PV on the microprocessor, we have extracted the critical paths through static timing analysis (STA). A detailed explanation of the methodology is given in [36],[37].

The models we have developed are for test conditions. They are scaled to use conditions by adjusting $g(T, E_p)$. The lifetime is the time until the system suffers from timing violations. NBTI and PV impact these timing margins. Lifetime is a function of frequency because higher operating frequencies have smaller timing margins.

Fig. 10 shows the estimated lifetime of each chip as a function of operating frequency. It can be seen that each chip has unique values for PV and NBTI parameters. PV affects the lifetime intercept (A in Equation (1)). As a result, some chips start out more vulnerable to NBTI because of PV. The extracted NBTI parameters primarily affect the slope. This causes the lifetime of some chips to be more sensitive to frequency.

V. CONCLUSION

This paper presents two models, one is for extraction of initial process parameters, and the other is for extraction of system-level NBTI parameters. We used simulation to develop an equation to extract the difference in the MOS length and threshold voltage from nominal using the initial ground signature. Also, the ground signature is used for extraction of an average threshold voltage shift, from which NBTI failure rate parameters are computed. The method involves measuring shifts in amplitude and delay. The ability to measure shifts in the ground signal has been demonstrated experimentally.

This approach enables the estimation of process variation aware system reliability under BTI degradation on a chip-by-chip basis. It determines an average threshold voltage shift, from which wearout distribution parameters are estimated. This shift depends on actual usage. Hence, if the threshold voltage is monitored periodically over time, lifetime can be estimated while taking into account actual usage. The model can then be used to forecast lifetime based on similar usage. Alternatively, if chips are stressed periodically, the proposed methodology can be used to estimate the remaining lifetime.

Future work will explore extensions that involve estimation of PBTI in nMOS devices in technologies beyond 90nm.

ACKNOWLEDGMENT

The authors thank the Defense Advanced Research Projects Agency (DARPA) under grant HR0011-11-1-0011 for support.

REFERENCES


