# Atomistic approach to variability of bias-temperature instability in circuit simulations

B. Kaczer[1], S. Mahato[1,2], V. Valduga de Almeida Camargo[1,3], M. Toledano-Luque[1,4], Ph. J. Roussel[1], T. Grasser[5], F. Catthoor[1,6], P. Dobrovolny[1], P. Zuber[1], G. Wirth[3], G. Groeseneken[1,6]

[1]imec, Kapeldreef 75, B-3001 Leuven, Belgium
[2]TU Munich, Germany & NTU, Singapore
[3]UFRGS, Brazil
[4]UC Madrid, Spain
[5]TU Wien, Austria
[6]KU Leuven, Belgium

phone: +32 16-281-557, e-mail: kaczer@imec.be

***Abstract*—A blueprint for an atomistic approach to introducing time-dependent variability into a circuit simulator in a realistic manner is demonstrated. The approach is based on previously proven physics of stochastic properties of individual gate oxide defects and their impact on FET operation. The proposed framework is capable of following defects with widely distributed time scales (from fast to quasi-permanent), thus seamlessly integrating random telegraph noise (RTN) effects with bias temperature instability (BTI). The use of industry-standard circuit simulation tools allows for studying realistic workloads and the interplay of degradation of multiple FETs.**

***Bias-temperature instability (BTI), random telegraph noise (RTN), time-dependent variability, single-carrier effects, circuit simulations***

## I. Introduction

It is generally accepted that systematic and statistical variability will have to be considered in the design of future ULSI circuits [1]. This process relies on describing the as-fabricated device parameters in terms of their statistical *distributions*. The challenge is to extend the same concept to circuit operation [2-6]. To that end, the *time dependence of the parameter distributions* during circuit operation, after being thoroughly understood, will need to be inserted into circuit simulators. Reliability assessment of future applications can thus be seen as *time-dependent variability* analysis.

We have previously shown that a large portion of BTI degradation and recovery is due to charging and discharging of gate oxide defects [7]. Here we argue that incorporating *the stochastic* nature of these defects and their *individual impact on FET operation* into existing industry-standard circuit simulation tools should *naturally* guarantee realistic time-dependent variability results under actual circuit workloads. Because only a handful of defects will be present in future nm-sized devices [8], tracking the occupancy of individual defects becomes a feasible approach to circuit reliability assessment. *Aging and degradation is thus viewed as "merely" following the occupation of each defect in time, while considering the impact of the individual charged defects on the respective FET's operation.*

## II. Problem Statement

Large devices employed in the past technologies were assumed to behave identically during BTI stress. Consequently, one device was typically sufficient to project a well-defined lifetime at each stress condition (Fig. 1a) [9]. With the downscaling of device dimensions, the number of defects per device decreases. As a consequence of that and of the stochastic nature of the defects, the projected lifetime spread drastically increases (Figs. 1bc). The example in Fig. 1 is calculated for constant stress only using a custom toolkit. The challenge, addressed here, is to translate this approach to an arbitrary circuit employing arbitrary workloads using industry-established simulation tools.

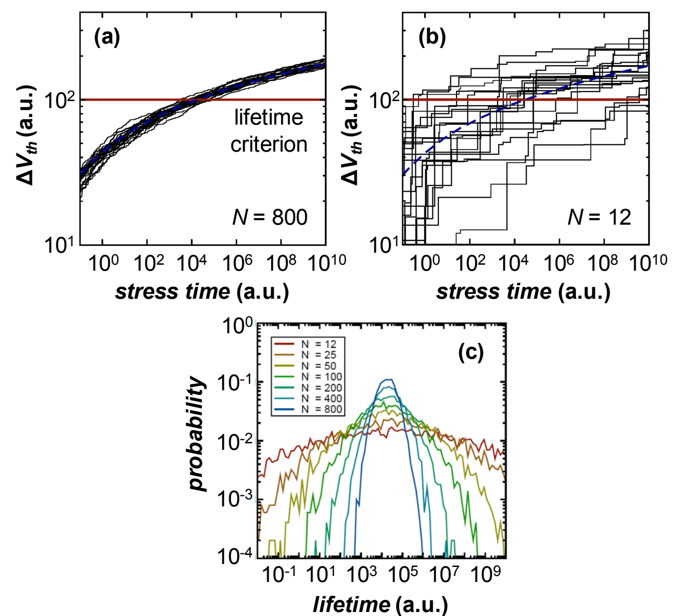


Figure 1. (a) The random properties of many defects $N$ in large devices average out, resulting in a well-defined lifetime while (b) the stochastic nature of a handful of defects in deeply-scaled devices becomes apparent, resulting in large variation in the lifetime, shown in (c). Only capture events are shown for the sake of simplicity.

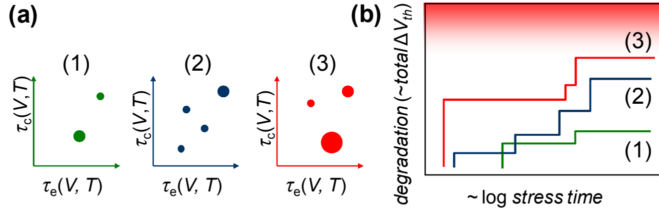


Figure 2. (a) Parameters affecting degradation schematically illustrated for three FET instances (1, 2, and 3). Each device is characterized by the number of defects (circles), their capture and emission times $\tau_c$ and $\tau_e$, and their impact on the device *when charged* (demarked by the size of the circle). (b) Schematic showing the progress of degradation during constant stress in the three devices. As in Fig. 1, only capture events are shown.

for each FET instance with gate length $L_G$ and width $W_G$:
  initial values:
    $V_{th0}$: from Gaussian(<$V_{th0}$>, $\sigma$)
  oxide defects:
    for each defect $i$ = 1 .. $n$ from Poisson($N$), where $N = L_G W_G N_{ot}$:
      initial occupancy state $P_{c0}$ (0 or 1): Eq. 4
      $\Delta V_{th,i}$ when occupied: from Exponential($\eta$), where $\eta \sim 1/L_G W_G$
      $\tau_{c,H}$, $\tau_{e,H}$, $\tau_{c,L}$, $\tau_{e,L}$:
        $\tau_{e,L}$ uniformly distributed on log scale [13]
        capture and emission $\tau$'s weakly correlated [13]
        L and H $\tau$'s connected via voltage dependence [11]

Figure 3. Prescription for including initial (time-zero) and time-dependent variability in the trap-enhanced FET model.

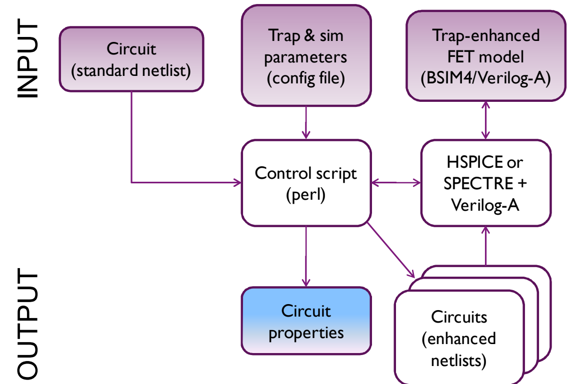


Figure 4. Simulation setup to study time-dependent variability of circuits based on industry-standard tools.

## III. Approach

### A. Assumptions

The general assumptions are schematically illustrated in Fig. 2a [9]. Each device is characterized by the number of defects, their capture and emission times $\tau_c$ and $\tau_e$, and their individual impact on the device when charged. Fig. 2b then schematically illustrates the calculation of Fig. 1b for the 3 devices in Fig. 2a.

These assumptions are more formally given in Fig. 3. Each instance of a FET with gate dimensions $L_G \times W_G$ is initiated with a specific number of defect precursors $n$ taken from a Poisson distribution with mean $N = L_G W_G N_{ot}$. Here, $N_{ot}$ is the gate oxide defect precursor areal density. Each defect can be independently initialized as empty (uncharged) or occupied (charged). In the proof-of-concept version of our simulation framework, only the impact of defects on the FET threshold voltage $V_{th}$ is assumed in the form of a voltage shift $\Delta V_{th}$, taken from an exponential distribution with the mean value $\eta$ [9]. The scaling of $\eta$ with device dimensions is presently under discussion [8,10]—we use the most simple scaling $\eta \propto 1/L_G W_G$.

Furthermore, each defect is characterized by its capture and emission times $\tau_c$ and $\tau_e$. Due to the inelastic nature of the charging mechanism, $\tau_c$ and $\tau_e$ are strongly voltage and temperature dependent [7,11,12]. The voltage dependence of each time constant is simplified in the case of digital circuits to two values, at low (L) and high (H) gate biases. Hence each defect is described by 4 time constants: $\tau_{c,H}$, $\tau_{c,L}$, $\tau_{e,H}$, and $\tau_{e,L}$. The latter is assumed to be uniformly distributed on the log scale between $10^{-9}$ and $10^{9}$ s. $\tau_{c,H}$ and $\tau_{e,H}$ are taken to be weakly correlated with $\tau_{e,L}$, with $<\tau_{c,H}> \sim 0.01 <\tau_{e,L}>$ [13] and $<\tau_{e,H}> \sim 100 <\tau_{e,L}>$ [11]. $\tau_{c,L}$ is assumed to be large [11]. Values of $\tau$'s for intermediate voltages are interpolated from the supplied L and H constants following the behavior given in Fig. 13 of Ref. 11. Impact of temperature is not assumed in this first iteration.

The initial (time-zero) variability is captured by assuming a Gaussian distribution for the initial threshold voltage $V_{th0}$ of each FET. The parameters of this distribution are the mean $<V_{th0}>$ and the variance $\sigma^2$.

### B. Implementation

The simulation framework (Fig. 4) accepts an arbitrary circuit in the form of a standard netlist. Based on the configuration file the control script generates multiple instances of the circuit, *each enhanced with unique random values of defect parameters for each FET* as per Fig. 3 and submits them to the HSPICE or SPECTRE solvers. The crucial component of the framework is the Verilog-A-based BSIM4 FET model *enhanced to simulate the impact of individual defects on the FET's behavior,* specifically, *its ability to follow the occupancy of each defect in every degrading FET*. Its details are discussed in the following section. Finally, the resulting circuit parameters from all instances are statistically analyzed.

### C. Time-dependent defect occupancy

To a good approximation, the capture and emission of charge can be described by first-order kinetics. The probability of a defect capturing charge $P_c$ is then described by the differential equation

$$\frac{\mathrm{d}P_c}{\mathrm{d}t} = \frac{1-P_c}{\tau_c} - \frac{P_c}{\tau_e}. \quad (1)$$

The two terms on the right-hand side of Eq. 1 correspond respectively to capture and emission probabilities. The general solution of this equation is

$$P_c = \frac{\tau}{\tau_c} + \left(P_{c0} - \frac{\tau}{\tau_c}\right) \exp\left(-\frac{t}{\tau}\right), \quad (2)$$

where $\tau^{-1} = \tau_c^{-1} + \tau_e^{-1}$ and $t$ stands for the time elapsed since the trap was occupied with the probability $P_{c0}$.

Our setup determines each defect's occupancy transitions *during the circuit simulation* using

$$P_{p,V} = \frac{\tau_V}{\tau_{p,V}}\left[1 - \exp\left(-\frac{\Delta t}{\tau_V}\right)\right]. \quad (3)$$

Here $p$ stands for process (capture or emission), $V$ the gate voltage (H or L), $\tau_V^{-1}=\tau_{c,V}^{-1}+\tau_{e,V}^{-1}$, and $\Delta t$ the simulation time step (at voltage $V$) [11, 14]. Eq. 3 is implemented in Verilog-A (see Fig. 4) and used to update the occupancy (i.e., *instantiate* it with 0 or 1 based on a randomly generated number) of every trap at every simulation step. The process $p$ is selected depending on the state $P_{c0}$ (0 or 1) of the defect at the beginning of the simulation step. *In this way our setup directly and naturally simulates the impact of individual charge trapping events (such as RTN) on FETs, and hence on the circuit operation.*

*D. Initialization of defect occupancies*

It is, however, computationally unfeasible to simulate the circuit after long operation $t_s$ (e.g. 1 year). To circumvent that, we have derived and experimentally verified [15] an analytical description of trap occupancy after stressing with a square periodic signal with frequency $f$, duty cycle (or duty "factor") $DF$, and duration $t_s$

$$P_c = \frac{B}{A}[1 - \exp(-At_s)]. \quad (4)$$

Here $A$ and $B$ are in general functions of $f$, $DF$, $t_s$, and arbitrary $\tau_{c,H}$, $\tau_{c,L}$, $\tau_{e,H}$, and $\tau_{e,L}$, elaborated in Ref. 15. Eq. 4 is valid for both "slow" (i.e., $\tau$'s >> $1/f$) and "fast" (i.e., $\tau$'s << $1/f$) traps. Fig. 5 graphically shows the behavior of Eq. 4 for a particular set of trap parameters, $t_s$, $DF$, and $f$. Instantiated $P_c$ (i.e., 0 or 1 decided with the probability $P_c$) is input into the circuit simulation as the initial state of each trap.

For the more realistic case of *irregular* workloads, Eq. 4 can be also used to initialize the simulation at time $t_s$ if *workload-equivalent DF* [2,4] is used. Fig. 6a shows that Eq. 4 is still a good approximation for "slow" defects. As will be illustrated below, such defects become dominant at longer operating times. *The framework is thus also capable of integrating the impact of defects with larger time constants into the circuit simulation.* Eq. 4 is obviously not strictly correct for "fast" defects in case of irregular workloads (Figs. 6b and 6c) but may be essentially irrelevant as their occupancy will change rapidly during simulation.

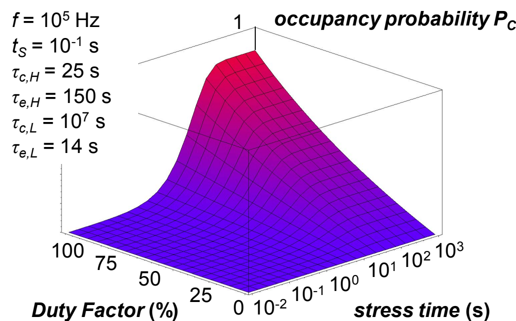


Figure 5. Graphical representation of the analytical description (Eq. 4) of a single trap occupancy probability $P_C$ after periodic AC stress with duty cycle $DF$ lasting $t_S$ for given emission and capture times.

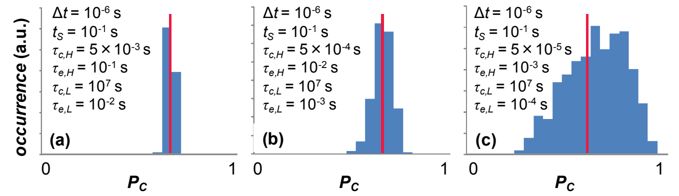


Figure 6. Occupation probabilities $P_C$ (histograms) calculated using Eq. 3 by following 5000 irregular random signals with mean $f$ = 25 kHz and mean $DF$ = 50%. (a) Slow defects generally follow the analytical description for a periodic signal (red vertical lines, Eq. 4), while (b) 10× faster and (c) 100× faster traps do not.

## IV. Demonstration

Our approach is demonstrated on the simple case of a CMOS inverter—the basis of many circuit elements, such as the SRAM cell. Publicly available PTM 16 nm Metal Gate / High-k / Strained-Si technology model card is used [16]. Initial $V_{th0}$ variation is assumed in both FETs of the inverter, while, for demonstration purposes, only the inverter pFET is assumed to be degrading. The studied inverter is sandwiched between two inverters with fixed parameters (Fig. 7) to condition the digital signal.

We first confirm that the threshold voltage shift distribution generated by the framework for a single pFET agrees with the analytical prediction (Fig. 8) [9].

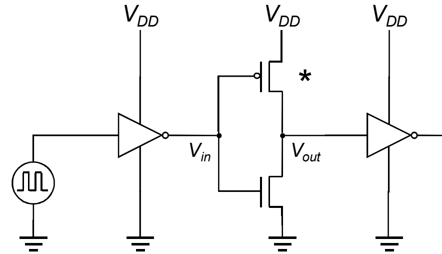


Figure 7. The circuit used in the demonstration consists of a 3-inverter chain. Only the pFET in the middle inverter (denoted with an asterisk) is assumed to be degrading. Parameters: $V_{DD}$ = 0.8 V, $f$ = 1 GHz, $DF$ = 50%, $L_n = L_p$ = 22 nm, $W_n$ = 45 nm, $W_p$ = 90 nm, $V_{th0}$ = ±0.25 V, $\sigma$ = 25 mV, $\eta$ =16 mV, $N_{ot}$ = $5\times10^{11}$ cm$^{-2}$.

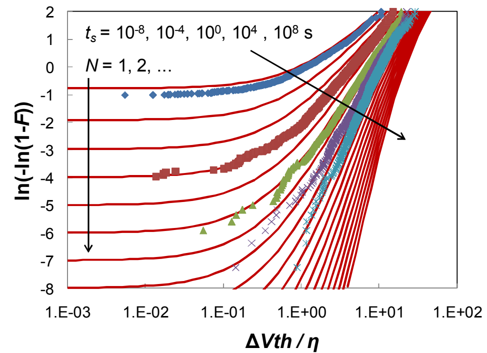


Figure 8. Threshold voltage shift distributions for a single pFET at different stress times $t_S$ generated by the framework (symbols) agree with the analytical description for the average *charged* defect number $N$ (lines) [9].

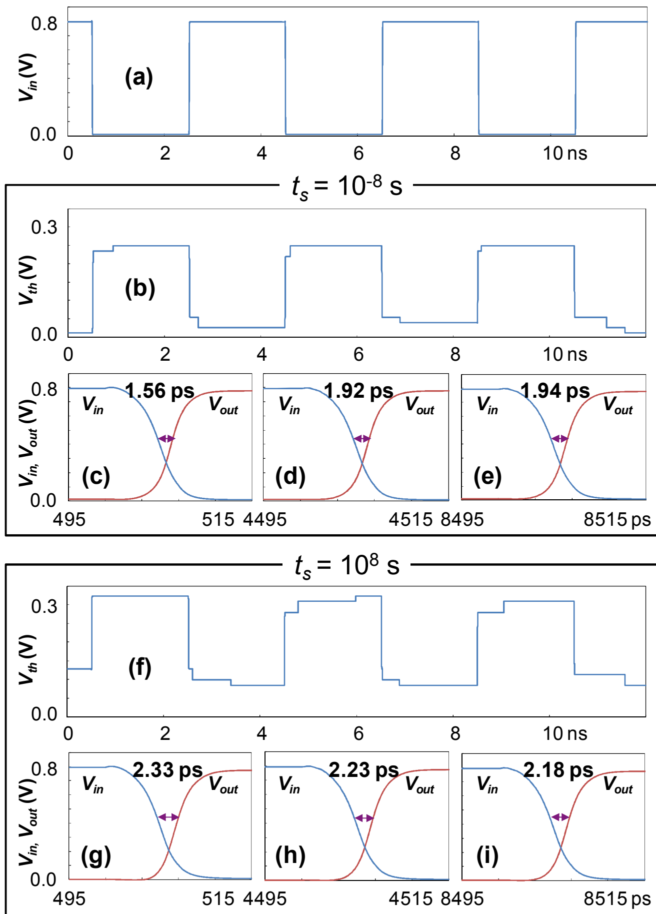


Figure 9. Inverter degradation simulation transient snapshots. (a) Three periods of a $f = 250$ MHz input signal of the studied inverter. Corresponding time-dependent pFET threshold voltage $V_{th}$ (b) after $10^{-8}$ s and (f) $10^{8}$ s. Two fast defects modulate the $V_{th}$ in both cases. Note that due to short channel effects, $V_{th}$ is also modulated by $V_D$, i.e., the output signal of the inverter. (f) $V_{th}$ after $10^{8}$ s is futher degraded with respect to (b) due to the capture of charge in slow defects. Switching delay of the studied inverter during the three periods (c-e) after $10^{-8}$ s and (g-i) $10^{8}$ s.

Fig. 9 then shows snapshots of one instance of the inverter simulation at the beginning of its operating lifetime (Figs. 9b-e) and after $10^{8}$ s (Figs. f-i). Fig. 9b shows the pFET threshold voltage $V_{th}$ behavior at the beginning of the circuit operation. $V_{th}$ is changing as single holes are captured in two fast "RTN" defects when the inverter input is low (i.e., the pFET gate is stressed), and subsequently emitted when the inverter input is high (i.e., pFET $V_G = V_S$). The same two defects are still active $10^{8}$ s into the circuit operation (Fig. 9f); note, however, the pFET $V_{th}$ is further degraded with respect to its initial value (Fig. 9b) due to the charge capture in slow defects. This latter behavior thus naturally emulates the "classical" BTI degradation.

The inverter switching transients are also illustrated in Fig. 9. It is apparent that there is a variation in the inverter switching delay from period to period (see Figs. 9c-e and 9g-i), resulting in the so-called delay jitter. The overall slowdown of the inverter after $10^{8}$ s of operation is also apparent (cf. Figs. 9c-e and 9g-i).

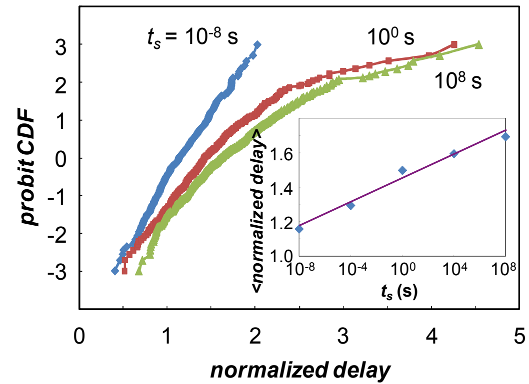


Figure 10. Distributions of normalized inverter delay for three points of the circuit lifetime. Inset: the average values increase as log $t_S$, in agreement with the assumptions (Section III.A).

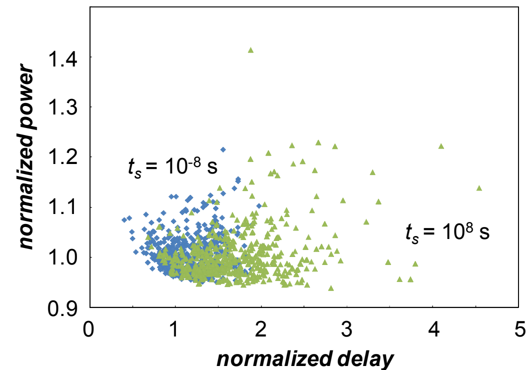


Figure 11. Normalized delay-power plot at the beginning (diamonds) and end (triangles) of circuit life. Shift in the distribution cloud due to increased trap occupation is apparent.

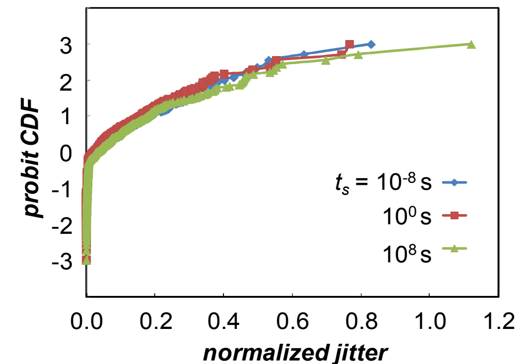


Figure 12. Normalized inverter delay jitter at three points of the circuit lifetime. The delay jitter is defined as the difference between the maximum and the minimum inverter delays simulated over 100 periods. The values are normalized by the nominal delay. Analysis of the impact of fast traps *during* circuit operation (e.g. delay jitter) is possible within the proposed framework.

The ability of the framework to generate statistical distributions of circuit parameters is demonstrated next. 500 instances of the circuit in Fig. 7 are generated at each of several points of the circuit operation $t_s$. Fig. 10 shows the increasing variability of the inverter delay distribution with ongoing circuit operation [3]. Here, the delay is averaged over 100 periods of operation of each circuit instance. All circuit parameters are normalized to the corresponding value of the

inverter without either initial (i.e., $V_{th0}$) or time-dependent variations (i.e., defects). The same simulation results are shown in the energy-delay diagram in Fig. 11. Again, increased variability of the circuit parameters with ongoing operation is apparent.

The ability of our simulation setup to investigate variations in circuit parameters *during operation* at different points of the circuit lifetime is documented in Fig. 12. Unlike the results Figs. 10 and 11, the jitter does not change as the circuit ages. This is because the jitter is caused by fast traps, which are assumed (see Section III.A) to be present in the pFET from the beginning.

Finally, we note that some penalty in terms of simulation speed is incurred by moving to a Verilog-A-based FET description (Table I). However, the additional penalty of following individual defects is minor.

| SPECTRE 7.1.1 + BSIM4.4 | SPECTRE + Verilog-A | SPECTRE + trap-enhanced Verilog-A |
|---|---|---|
| 16.5 s | 69 s | 95 s |
| 24% | 100% | 138% |

Table I: Comparison of circuit simulation speed. Circuit in Fig. 7 is used, with 15 defects in *each* of its 6 FETs (i.e., in total 90 defects) simulated in the "trap-enhanced" case.

.

## V. CONCLUSIONS

We have demonstrated a blueprint for an atomistic approach to introducing time-dependent variability into a circuit simulator, based on the stochastic properties of individual defects and their impact on the FET. The framework is capable of following defects with widely distributed time scales (from fast, i.e., RTS-like, to slow and quasi-permanent, i.e., BTI-like) in a unified manner. The employment of existing industry-standard circuit simulator tools ensures correct combination of the deterministic workload-dependent component with the stochastic modeling aspect while simultaneously incorporating interactions among different devices. The refinement and expansion, be it by other types of defects or by their impact on FET behavior, can be further explored. We expect the framework should prove useful for investigation of the reliability of ULSI circuits.

## VI. ACKNOWLEDGEMENTS

This work was carried out as part of IMEC's Industrial Affiliation Program funded by IMEC's core partners. V.V.A.C. thanks CNPQ Brazil for financial support. M.T.L. thanks the IberCaja Postdoctoral grant program and Spanish MEC (TEC2010-18051) for financial support.

## REFERENCES

[1] M. Miranda Corbalan, B. Dierickx, P. Zuber, P. Dobrovolny, F. Kutscherauer, P. Roussel, and P. Poliakov, "Variability aware modeling of SoCs: from device variations to manufactured system yield", *Int. Symp. Quality Electronic Design (ISQED),* San Jose, CA, March 2009.

[2] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "NBTI-Aware Synthesis of Digital Circuits", *Design and Automation Conference (DAC)*, 2007.

[3] Tong Boon Tang, A. F. Murray, Binjie Cheng, and A. Asenov, "Statistical NBTI-effect prediction for ULSI circuits", *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, p. 2494, 2010.

[4] Wenping Wang, Shengqi Yang, S. Bhardwaj, S. Vrudhula, F. Liu, and Yu Cao, "The Impact of NBTI Effect on Combinational Circuit: Modeling, Simulation, and Analysis", *IEEE T. VLSI Systems* **18**, p. 173, 2010.

[5] V. Huard *et al.*, "Managing SRAM reliability from bitcell to library level", *Proc. Int. Reliab. Phys. Symp. (IRPS)*, p. 655, 2010.

[6] H. Kufluoglu, V. Reddy, A. Marshall, J. Krick, T. Ragheb, C. Cirba, A. Krishnan, and C. Chancellor, "An Extensive and Improved Circuit Simulation Methodology For NBTI Recovery", *Proc. Int. Reliab. Phys. Symp. (IRPS)*, p. 670, 2010.

[7] T. Grasser *et al.*, "Recent Advances in Understanding the Bias Temperature Instability", *Int. Electron Dev. Meeting (IEDM) Tech Dig.*, p. 82, 2010.

[8] A. Asenov, R. Balasubramaniam, A. R. Brown, and J. H. Davies, "RTS Amplitudes in Decananometer MOSFETs: 3-D Simulation Study," *IEEE T. Electron Dev.* **50**, p. 839, 2003.

[9] B. Kaczer, T. Grasser, Ph. J. Roussel, J. Franco, R. Degraeve, L.-A. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, "Origin of NBTI Variability in Deeply Scaled pFETs", *Proc. Int. Reliab. Phys. Symp. (IRPS)*, p. 26, 2010.

[10] A. Ghetti, C. M. Compagnoni, A. S. Spinelli, and A. Visconti, "Comprehensive Analysis of Random Telegraph Noise Instability and Its Scaling in Deca–Nanometer Flash Memories," *IEEE T. Electron Dev.* **56**, p. 1746, 2009.

[11] T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Goes, and B. Kaczer, "The Time Dependent Defect Spectroscopy (TDDS) for the Characterization of the Bias Temperature Instability", *Proc. Int. Reliab. Phys. Symp. (IRPS)*, p. 16, 2010.

[12] T. Grasser, H. Reisinger, P.-J. Wagner, and B. Kaczer, "Time-dependent defect spectroscopy for characterization of border traps in metal-oxide-semiconductor transistors", *Phys. Rev. B.* **82**, p. 245318, 2010.

[13] H. Reisinger, T. Grasser, W. Gustin, and C. Schlünder, "The statistical analysis of individual defects constituting NBTI and its implications for modeling DC- and AC-stress", *Proc. Int. Reliab. Phys. Symp. (IRPS)*, p. 7, 2010.

[14] J. Martin-Martinez, B. Kaczer, M. Toledano-Luque, R. Rodriguez, M. Nafria, X. Aymerich, and G. Groeseneken, "Probabilistic Defect Occupancy Model for BTI", accepted to *Proc. Int. Reliab. Phys. Symp. (IRPS)* 2011.

[15] M. Toledano-Luque, B. Kaczer, Ph.J. Roussel, J. Franco, T. Grasser, C. Vrancken, N. Horiguchi, and G. Groeseneken, "Response of a single trap to AC Negative Bias Temperature Stress", accepted to *Proc. Int. Reliab. Phys. Symp. (IRPS)* 2011.

[16] http://ptm.asu.edu/ .