Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Efficient Estimation of Reliability Metrics for Circuits in Deca-Nanometer Nodes

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μιχαήλ Α. Νόλτσης

Επιβλέπων : Δημήτριος Σούντρης
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2014

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Efficient Estimation of Reliability Metrics for Circuits in Deca-Nanometer Nodes

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μιχαήλ Α. Νόλτσης

Επιβλέπων : Δημήτριος Σούντρης
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 31η Ιουλίου 2014.

Κιαμάλ Πεκμεστζή Καθηγητής Ε.Μ.Π.	Δημήτριος Σούντρης Αναπληρωτής Καθηγητής Ε.Μ.Π.	Ιωάννης Ξανθάκης Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2014

Μιχαήλ Α. Νόλτσης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Abstract

In modern technologies of integrated circuits (IC) and with the downscaling of device dimensions, various degradation modes constitute major reliability concerns. Bias Temperature Instability (BTI) is a representative example, posing as a significant reliability threat in Field-Effect Transistor (FET) technologies and has been known for more than 30 years. At first, the model that tried to explain this phenomenon was based on the Reaction-Diffusion (RD) theory and was developed nearly 30 years ago. Recently, an atomistic model has been proposed, that enables the modeling of BTI in modern technologies.

By observing the amount of software designed to simulate the BTI degradation, tools can be found that are based on the atomistic theory but are computationally prohibitive when it comes to simulating complex circuits consisting of a large number of devices. Tools based on the RD model are unable to accurately capture the BTI-induced degradation, especially in devices with small dimensions. The current thesis is appropriately positioned since it discusses a novel simulation framework that is efficient yet highly accurate. A subset of an embedded Static Random Access Memory (SRAM) is used for verification purposes. The estimation of the functional yield of the circuit over three years of operation will be examined as well as other reliability metrics, such as defects per million (DPM), mean time to failure (MTTF) and failures in time (FIT rate). Finally, the interplay between these metrics is discussed and efficient computation methods are proposed for each one.

Key words

Bias Temperature Instability (BTI), Gate Stack Defects, Reliability, Circuit Simulations, Static Random Access Memory (SRAM), Functional Yield, Mean Time to Failure (MTTF), Defects Per Million (DPM), Failures in Time rate (FIT rate).

Περίληψη

Καθώς η τεχνολογία οδηγεί στην κατασκευή τρανζίστορ ολοένα και μικρότερων διαστάσεων, έχουν εμφανιστεί αρκετά φαινόμενα που επηρεάζουν την αξιοπιστία των ολοκληρωμένων κυκλωμάτων. Ένα από αυτά τα φαινόμενα ονομάζεται “Bias Temperature Instability”, αποτελεί σημαντικό κίνδυνο για την αξιοπιστία των ολοκληρωμένων κυκλωμάτων και έχει παρατηρηθεί εδώ και πάνω από 30 χρόνια . Το πρώτο μοντέλο που προσπάθησε να εξηγήσει αυτό το φαινόμενο εμφανίστηκε πριν από 30 περίπου χρόνια, βασίστηκε στη διάχυση υδρογόνου και ως εκ τούτου ονομάστηκε “Reaction-Diffusion model”. Πριν από μερικά χρόνια δημιουργήθηκε ένα νέο ατομιστικό μοντέλο το οποίο βασίζεται κυρίως στην εμφάνιση ελαττωμάτων στο διηλεκτρικό μεταξύ της πύλης και του καναλιού των FET τρανζιστορ.

Μελετώντας κανείς τη βιβλιογραφία που αφορά στο ατομιστικό αυτό μοντέλο, μπορεί να συναντήσει εργαλεία που προσομοιώνουν με ακρίβεια το μοντέλο αλλά δυστυχώς απαιτούν αρκετό χρόνο για να εκτελεστούν, κάτι το οποίο τα καθιστά απαγορευτικά για εκτενή χρήση. Παράλληλα, υπάρχουν εργαλεία βασισμένα στο μοντέλο της διάχυσης τα οποία βέβαια αδυνατούν να παράξουν σωστά και λεπτομερή αποτελέσματα, κυρίως σε τεχνολογίες μικρών διαστάσεων. Η παρούσα λοιπόν διπλωματική εργασία παρουσιάζει τα αποτελέσματα ενός νέου και καινοτόμου εργαλείου το οποίο βασίζεται στο ατομιστικό μοντέλο, ωστόσο προσομοιώνει αποδοτικά αλλά και με ακρίβεια το φαινόμενο της γήρανσης. Ένα αντιπροσωπευτικό μονοπάτι στατικής μνήμης (SRAM) θα χρησιμοποιηθεί ως παράδειγμα της λειτουργίας του μοντέλου ενώ παράλληλα θα υπολογισθούν, με βάση τα αποτελέσματα των προσομοιώσεων αυτών, μετρικές, σημαντικές για το χαρακτηρισμό της απόδοσης και αξιοπιστίας του κυκλώματος, ενώ παράλληλα θα μελετηθούν λεπτομερώς και οι σχέσεις που τις συνδέουν.

Acknowledgements

First of all, I would like to thank assistant prof. Dimitrios Soudris for supervising this thesis. Having the chance to attend his lectures was of great importance to me, since he gave me the opportunity to attain a strong background in this specific field. Furthermore, his advice and guidance were really helpful for the completion of my undergraduate studies and the start of my professional career.

Moreover, this thesis would not be completed without the mentorship of Dimitrios Rodopoulos. Cooperating with him was one of my best academic experiences since our discussions always inspired and motivated me to work and overcome any difficulties. I am thankful for the time he devoted to my instruction and assistance.

Last but not least, I would like to express my gratitude to my family for their moral and financial support throughout these years.

Michail A. Noltsis,

Athens, July 31, 2014

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

The probability that an electronic circuit shall perform correct service along its operating time is defined as reliability of the respective electronic system [7]. Probability is introduced because predicting the failure mechanisms of any electronic circuit in a fully deterministic way is practically infeasible. Moreover, nowadays, the demand for ultra-large-scale integration in chips from the semiconductor industry led to a dramatic decrease in the size of the devices, amplifying a variety of reliability concerns in digital systems.

One key feature affecting reliability is the voltage and temperature stress. An effect that emerges from this voltage stress and deteriorates from the exposure of the device at elevated temperatures is called Bias Temperature Instability (BTI). The BTI effect is the aging phenomenon that results in an absolute increase of the threshold voltage of a transistor. It is this rise of V_{th} that further causes degradation of the drain current and transconductance of the channel in MOSFETs (Metal-Oxide semiconductor Field-Effect transistor). Therefore, these aging effects engender fluctuations in the delay and power of electronic circuits and may even cause system failures.

In order to capture this aging phenomenon, the community has developed a significant number of tools based on either the Reaction-Diffusion (RD) or the atomistic models. The simulation frameworks that follow the atomistic theory provide highly accurate results, however the computational time needed for long-lasting simulations is inhibitive. The tools based upon the RD theory are incapable of providing precise estimations when simulating small-dimensioned devices. Hence, the usage of tools that effectively model the aging effect but also provide elaborated results in short run-times is crucial and necessary. Thus, the purpose of this thesis is to discuss some preferable features of such tools and present a relevant instantiation.

Specifically, this thesis combines existing tools into an overall framework, which performs estimation of the time-dependent functional yield of a circuit, under BTI-induced degradation. The modules of the framework, which are already in place, are discussed and certain optimizations are proposed for some of them. Also, the mathematical formulation that connects the functional yield with some other important reliability metrics are presented.

In Chapter 2, the established BTI models are introduced. First, the Reaction-Diffusion model is presented along with its disadvantages and drawbacks. Next, follows the elaboration on the transition from the RD model to the atomistic theory. The concept of the defect is explained along with the definitions of capture and emission time constants. Afterwards, the functional yield along with other metrics used in reliability analysis such as defects per million (DPM), mean time to failure (MTTF) and

failures in time rate (FIT rate) are underlined. Lastly, some existing industry tools for BTI modeling are mentioned.

A simulation framework that accurately captures the mechanism of Bias Temperature Instability is discussed in Chapter 3. The components of this framework are demonstrated and the verification of one of them is presented step by step. The improved accuracy of this tool is emphasized in contrast to crude approximations given by other existing frameworks. Also, the synthetic design flow is demonstrated. Finally, the functional yield is converted through mathematical formulation to other reliability metrics.

Moving to Chapter 4, the simulation results of the analyzed framework are presented in detail, with graphs that show the fluctuations of some reliability metrics of the target electronic circuit throughout its lifetime. It must be pointed out that a realistic circuit of a subset of an embedded Static Random Access Memory (SRAM) is used for verification purposes. The functional yield, MTTF, DPM and FIT rate of this circuit are estimated over three years of operation for 2 different cases of V_{dd} configuration. For the first case, V_{dd} is fixed to a constant value while for the second case, a dynamic voltage scaling is used for the simulations.

Finally, in the last Chapter, a number of conclusions that derive from this work are underlined. Also, some topics for future work are proposed that can expand the contributions of this thesis.

CHAPTER 2

Prior Art

2.1 Introduction

This Chapter elaborates on the two existing models that try to explain the BTI phenomenon. Firstly, the Reaction-Diffusion model is reviewed along with its handicap to accurately describe the BTI degradation on downscaled devices. Other insufficiencies of this model are also noted that mostly regard the incapability of the model to explain the relaxation phase. Next, the atomistic theory is presented and an approach towards an atomistic model is reviewed, along with the advantages of this model over the Reaction-Diffusion approach. The main advantages of the model, namely the workload dependency and the capturing of the Random Telegraph Noise (RTN) as well, are noted.

BTI-induced degradation may lead to functional reliability violations, hence it constitutes a valid reliability concern for deca-nanometer circuits. Thus, the next part of the Chapter describes aspects of the reliability analysis and points out some key metrics used in this analysis. For this reason, the terms of functional yield, defects per million, mean time to failure and failures in time are discussed.

Finally, some simulation frameworks and commercial tools that were developed to model BTI degradation are reviewed. These tools are presented briefly while their disadvantages in following the aging effect in small-dimension devices are pointed out.

2.2 RD theory

2.2.1 The Reaction-Diffusion Model

The Reaction-Diffusion model was created 30 years ago and was the first model that tried to explain and predict the BTI effect. It is based on the reaction-diffusion theory, namely on “the breaking and annealing of the Si-H bonds at the Si/oxide substrate and mainly focuses on pFETs” [8]. According to this approach, when the gate of a pFET is under bias stress, then the Si-H bonds are breaking and the H atoms are diffusing towards the gate stack, while holes are taking their place in those bonds. This procedure leads to the creation of interface traps and oxide charge that cause a shift to the pFET’s threshold voltage as shown in Figure 2.1 [1]. A similar approach is expected for the nFETs, but with electrons being the trapped charge in this case. The model also argues that trap distances from the Si/oxide interface determine the portion of this threshold shift with “oxide charge located closer to the oxide substrate leading to a higher threshold voltage shift than charge near the gate-oxide interface” [9].

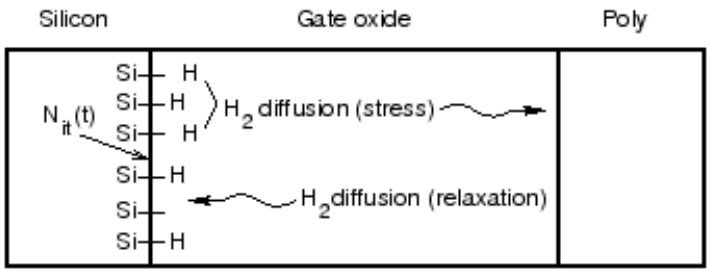

Figure 2.1: Schematic description of the R-D model as shown in [1]

In addition, the conventional RD theory predicts a recovery phase, during which, once the stress is removed from the pFET's gate, a subset of the Si-H bonds is annealed because hydrogen is expected to diffuse back, thus partially restoring the threshold shifts. In fact, according to the model, there are two degradation components : a permanent component, which remains after the bias stress is removed and a reversible component, that recovers with the absence of voltage stress to the gate of the device. In other words, the Reaction-Diffusion model claims partial recovery.

To be more specific, the number of interface traps N_{it} created by detached hydrogen atoms (from the Si-H interface) is defined by the Equation 2.1 [10]:

$$N_{it} = \sqrt{\frac{k_F N_o}{2k_R}}(D_H t)^{\frac{1}{4}} \tag{2.1}$$

where k_F is the dissociation rate constant that creates broken bonds, k_R is the rate of reverse annealing of Si-H bonds, N_o is the total number of Si-H bonds and D_H shows the hydrogen diffusion coefficient.

Also, according to the Reaction Diffusion model, the threshold voltage shift that occurs due to the increase in interface charge is given by [11]:

$$\Delta V_{th}(E_{ox}, t) = (1 + m)\frac{qN_{it}(E_{ox}, t)}{C_{ox}} \tag{2.2}$$

$E_{ox} = \frac{V_{gs}}{C_{ox}}$, q is the elementary charge and C_{ox} presents the gate capacitance and V_{gs} presents the gate voltage.

By combining the Equations 2.1 and 2.2, it can be implied that the change in V_{th} follows an 1/4 exponential dependence throughout the device lifetime. And it is true that in old technologies with large devices, such a shift in the threshold voltage could be experimentally verified [2]. However, in modern technologies and with the downscaling of device dimensions, a huge variation in ΔV_{th} that did not follow the exponential rule mentioned earlier was observed, as shown in Figure 2.2. Moreover, when the research community turned its attention to the partial recovery phase predicted by the RD model and compared it with experimental data, it appeared that the data could not be explained by the RD theory and were opposed to what the theory suggested [12] .

2.2.2 Insufficiencies of the RD Model

One key feature of the RD theory is that a general and universal recovery is anticipated, meaning that "when ΔV_{th} is normalized to its value at the end of the stress, it depends only on the ratio of the

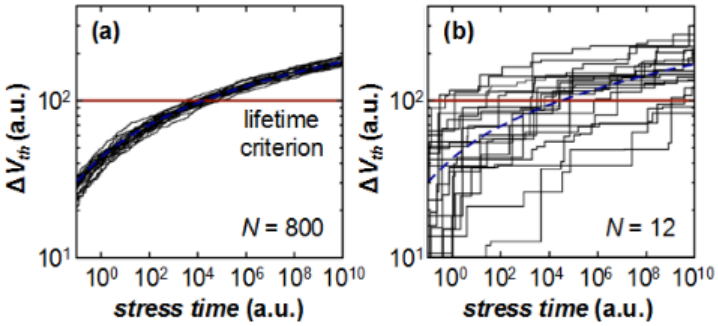

Figure 2.2: Expected threshold voltage shift in devices of (a) older and (b) modern technologies as presented in [2]

stress and relaxation times t_s and t_r" [12]. Neither the forward and backward reaction rates nor the diffusion coefficient D_H have any impact on the recovery. Thus, the normalized RD recovery can be captured approximately by the Equation 2.3 [12]:

$$\frac{\Delta V_{th}(t_s, t_r)}{\Delta V_{th}(t_s, 0)} = \frac{1}{1 + (\frac{t_r}{t_s})^{\frac{1}{2}}} \tag{2.3}$$

Hence, according to the RD theory, a 50% recovery is expected when $\frac{t_s}{t_r} = 1$, namely when the relaxation time equals the stress time. However, it can be seen from experimental data that "recovery in large devices is rather uniform and roughly follows $\log(t_r)$ over the whole experimental time interval" [12]. Thus, this formula proves inconsistent with scientific measurements. Moreover, another problem emerges from the above mathematical formulation: while an insignificant part of the ΔV_{th} is expected to have recovered when $t_r \ll t_s$, all experiments show that "recovery is already in full flight in measurements taken just some few microseconds after the stress is removed" [12], hence the relaxation phase and the ΔV_{th} recovery are not consistent with Equation 2.3.

Summarizing the above claims, it can be concluded that the RD theory may prove rather unsuccessful in predicting BTI, especially in devices of modern technologies. Therefore, in order to precisely predict the BTI degradation another theory was created that captures the bulk of experimental data more accurately.

2.3 Atomistic Theory

2.3.1 The Two-State Atomistic Model

The atomistic model has been recently proposed by the research community in order to capture BTI degradation in a more detailed manner. The defects created from the charge trapping mechanism occupy the central role of the atomistic theory. The essence of this defect-centric model is that, once the defect is created, its charge state can continuously change between two states depending on the temperature and gate stress. This stands in contrast to the RD theory and the term of trapped holes, which are not sensitive to bias changes. Hence, the atomistic theory confronts these defects as switching traps. According to the model, a defect has two states and can be charged or discharged, in other words occupied or unoccupied.

If a defect is unoccupied, it does not add to the ΔV_{th} and has no impact in the aging of the device, but when a defect is occupied, it adds to the threshold shift. It has to be noted that while the RD theory supported the contribution of an oxide trap to the voltage threshold shift to depend on the defect's depth in the oxide, the atomistic model does not follow this assumption. Besides, the thickness of modern oxides cannot support any significant distribution of defects based on distance alone [9]. The atomistic approach supports that every defect is identified with its ΔV_{th} impact on the device -when occupied- that comes from an exponential distribution with a mean value η [2]. All defects are also characterized with their capture and emission time constants τ_c and τ_e [2]. Each FET with gate length L_G and width W_G is populated with a specific number of defect precursors n taken from a Poisson distribution with mean $N = L_G W_G N_{ot}$ [2]. This means that the number of defects of each transistor is strongly depending on its size with larger transistors being dwelled by a bigger number of defects than devices with smaller dimensions.

Whether a defect is captured or not, depends on the capture probability (P_c) of each gate stack defect. "P_c is the probability of a gate stack defect being captured after a specific voltage stress signal (V_{gs}) has been applied to the respective gate" [8]. Because of introducing probability to the scene, the atomistic model has a stochastic nature in dealing with the BTI and does not focus so much on the cite creation and the physical mechanisms responsible for BTI (and the creation of these defects). Nevertheless, this model accurately captures this aging phenomenon in large as well as in small devices were the number of defects decreases dramatically.

As a matter of fact, experiments show that recovery in small-area devices (e.g. $W \times L = 100$ nm $\times$ 100 nm) proceeds in small steps, which are not consistent with a reaction-diffusion model. Rather, they are consistent with the emission of individual holes based on a first-order reaction-rate process. "The properties of these discrete steps are fully consistent with charge trapping observed in the context of Random Telegraph Noise" [12]. RTN is a type of electronic noise that occurs in semiconductors and consists of sudden capture and emission transitions of carriers on traps [13] at random and unpredictable times on a tiny fraction of time. This means that the atomistic model, with its inherent features of stochastic capture and emission of a defect, can accurately predict the impact of the BTI effect and also provide a solid modeling of the RTN.

As mentioned earlier, P_c is an important parameter of the atomistic model. Actually, the capture probability depends on the time constants of each defect and the characteristics of the V_{gs} signal or, to be more specific, its frequency f, its duty factor α and its duration Δt. Time constants are generally voltage and temperature dependent. According to an atomistic approach designed for digital circuits [2], the voltage dependence of each time constant is simplified to the two dominant values: low (L) and high (H) gate bias. Hence, all defects are characterized by four time constants: $\tau_{e,H}, \tau_{e,L}, \tau_{c,H}, \tau_{c,L}$. This approach [2] also assumes the latter to be uniformly distributed on the logarithmic scale between 10^{-9} and 10^{9}s while $\tau_{c,H}$ and $\tau_{e,H}$ are weakly correlated with $\tau_{e,L}$, with $<\tau_{c,H}>$ [1] $\sim 0.01 <\tau_{e,L}>$ and $<\tau_{e,H}> \sim 100 <\tau_{e,L}>$. Lastly, $\tau_{c,L}$ is assumed to be large in comparison with the other 3 constants [2].

The capture and emission of charge is described by first-order kinetics and the probability of a defect capturing charge P_c is assumed to follow the differential Equation 2.4 [2]

$$\frac{dP_c}{dt} = \frac{1 - P_c}{\tau_c} - \frac{P_c}{\tau_e} \tag{2.4}$$

[1] $<\tau>$ presents the mean value of τ.

The two terms of this Equation, τ_c and τ_e represent the capture and emission probabilities respectively. The P_c's dependence on the capture and emission constants can be seen from 2.4. The solution of Equation 2.4 is 2.5 [2]:

$$P_c = \frac{\tau}{\tau_c} + \left(P_{c0} - \frac{\tau}{\tau_c} \right) exp\left(-\frac{t}{\tau} \right) \tag{2.5}$$

where τ^{-1}=$\tau_c^{-1} + \tau_e^{-1}$ and t stands for the time elapsed since the defect was occupied with the probability P_{c0}. The dependency on the initial value of P_{c0} should be expected since the capture probability was described by a first-order differential equation.

For every instance of a simulation at a time t, P_c is evaluated and the occupancy of a defect is resolved based on a comparison between the capture probability and a number r chosen randomly between 0 and 1. The defect is assumed occupied iff $r < P_c$ and its impact to V_{th} is added to the device's V_{th}. A similar concept is followed for the calculation of the emission probability (P_e) for a gate stack defect.

2.3.2 Advantages and boundaries of the atomistic model

While the RD model fails to capture the aging degradation at aggressively downscaled devices, the atomistic model succeeds in providing an approach that can accurately predict the threshold voltage shift due to BTI stress regardless of the size of the device. This atomistic approach introduces time dependent variability based on the stochastic properties of individual defects and their impact on a FET. "This framework is capable of following defects with widely distributed time scales, from fast and RTN-like to slow and quasi-permanent such as BTI, in a unified manner" [2]. Hence, the atomistic approach provides a first step towards RTN modeling as well.

Another important advantage of the atomistic model is the consideration of the workload of the device for the appropriate time inspection [14], an essential feature in reliability analysis of modern ICs. The atomistic model takes into account not only the time of the stress t_s, but also the type of the gate bias applied to the device. This distinct degradation of different input sequences and workload scenarios assists significantly on examining the BTI degradation of systems under realistic workloads, but also provides a strong contribution when considering mitigation techniques.

Simulation frameworks based on this model can also easily incorporate time-zero variability of threshold voltage, a factor that proves of great importance as the size of the dimensions is shrinking. Time-zero V_{th} variability is introduced because from a large set of devices, the assumption that all of them have an identical threshold voltage is unrealistic mainly due to production flaws. After all, even state-of-the-art manufacturing processes cannot produce chips with consistent characteristics, which means that two transistors, fabricated a few dozen nanometers apart on the same piece of silicon, can have different electrical properties [15]. Hence, this initial variability can be captured by the atomistic model by introducing a Gaussian distribution for the initial threshold voltage V_{th0} of each FET.

Nevertheless, the advantages of the atomistic model vanish when it comes to simulating devices over extensive time intervals and the superiority of the model becomes a bottleneck in case of the long-term aging simulations. Simulation frameworks following the atomistic model provide an accurate estimation of the BTI degradation but are computationally prohibitive when it comes to simulating large netlists. Other tools, that require lower memory usage and CPU time, provide only a crude approximation of the results.

To conclude, it seems that the atomistic model can capture more aspects of the BTI effect than the RD model, however the detailed nature of the model becomes a major disadvantage due to increased computational complexity.

2.4 Functional Yield and other Reliability Metrics

It has been made clear by now that BTI-induced degradation constitutes a significant reliability threat in the hardware of digital systems. The aging effect can lead to delays or even functional violations. In the context of this thesis, a failure constitutes a functional violation. Specifically, "a failure is a nonperformance of the operating function" [16], that can be observed by the end-user.

A common metric used in many cases [17], [8] as quality assessment of digital circuits is the functional yield, which refers to the proportion of devices found to perform properly.

Functional Yield. Represents the ratio of the number of the circuits working correctly, to the total circuits examined [18].

$$\text{Functional Yield} = \frac{\text{Number of circuits working properly}}{\text{Total circuits examined}} \tag{2.6}$$

Another metric used in some cases [19] to describe the functional behaviour of a set of ICs is defects per million (DPM). In the context of this thesis, a defect is simply a failed circuit sample.

Defects Per Million. (DPM) is the ratio of the number of "functionally incorrect" circuits, to the total circuits examined, normalized to one million.

$$\text{Defects Per Million} = \frac{\text{Number of failed circuits}}{\text{Total circuits examined}} \times 10^6 \tag{2.7}$$

The characterization of the reliability of a set of devices can also be given in terms of mean time to failure (MTTF) or failures in time rate (FIT rate). They represent the expected failure rate of semiconductors and other electronic devices during their lifetime operation and can be statistically projected from the results of accelerated test procedures [20], [21], [22].

Mean Time To Failure. Corresponds to the mean time of operation before a failure [23]. MTTF is calculated by the Equation 2.8 [22]:

$$\text{MTTF} = \frac{\text{Device Hours}}{\text{Number of Failures}} \tag{2.8}$$

Failures In Time rate. Represents the failure rate of a circuit during its lifetime, normalized to one billion device hours. 1 FIT stands for one failure observed during a billion device hours [20]. FIT rate and MTTF are connected via the mathematical formulation [23]:

$$\text{FIT}_{\text{rate}} = \frac{10^9}{\text{MTTF}} \tag{2.9}$$

A common practice of the test engineers is the acceleration of the test procedure that enables the simulation of long-lasting operating times. The test procedure can be accelerated by increasing the temperature of the test environment, because most failure mechanisms are temperature activated [21]. In this way, shorter tests can be conducted that simulate many years of device testing under normal temperature conditions. It is therefore important to be able to measure the effect of this testing acceleration. If a device is stressed at a high temperature T_2, for time t_2, the equivalent time t_1 which would cause the device the same level of stress at a lower temperature T_1 can be estimated by the *acceleration factor*, defined as [21]:

$$A = \frac{t_1}{t_2} \tag{2.10}$$

that is, the ratio of the time at the lower temperature, to that at a higher. The *acceleration factor* A can be estimated from the Equation 2.11 [21]:

$$\text{Acceleration factor} = e^{\left[\frac{E_A}{k} \times \left(\frac{1}{T_1} - \frac{1}{T_2}\right)\right]} \tag{2.11}$$

where E_A is the activation energy of the failure mechanism and k is the Boltzmann constant=8.6×10^{-5} eV/K. The value of activation energy can be found from experiments and is usually nearly 1.0 eV [21].

According to relevant work [21], [22], when the failure rate f_2 at temperature T_2 is known, the failure rate at the temperature T_1 can be found by: $f_1 = A \times f_2$. In this way, the failure rates under normal operating conditions can be estimated by testing circuit samples for shorter time windows and at higher temperatures.

2.5 Existing Aging Simulators

In order to understand the advantages of the simulation framework that will be discussed in the next Chapter, other simulation approaches that model the BTI degradation will be presented briefly, along with their distinguishing features. After all, the degradation of FET device characteristics because of aging, has become a critical reliability metric in downscaled technologies hence reliability analysis can be found on many simulating tools and approaches [3], [24], [25], [26]. The software support that models BTI is crucial in order to enable designers to detect reliability failures and estimate the reliability metrics of ICs.

Looking through a variety of simulation approaches, one can find the industry standard tool called Virtuoso UltraSim [3]. Virtuoso is a design suite developed by Cadence Design Systems and the UltraSim package can simulate the effects of NBTI and HCI, independently or together. This tool performs short simulations in order to estimate the aging effect and perform a reliability analysis over a netlist. UltraSim has a built-in support for reliability simulations meaning that both simulations regarding the "fresh" and the "aged" netlist are performed independently and "Fresh and degraded waveforms are output to output waveform files during each simulation" [3]. In Virtuoso, the aged parameters of the devices' modelcards are calculated by interpolation or regression as can be seen in Figure 2.4.

A piece of software developed by Mentor Graphics that can be used for reliability analysis is ELDO [24]. ELDO is a SPICE-like simulator with addons that enable the aging analysis of a circuit. In the

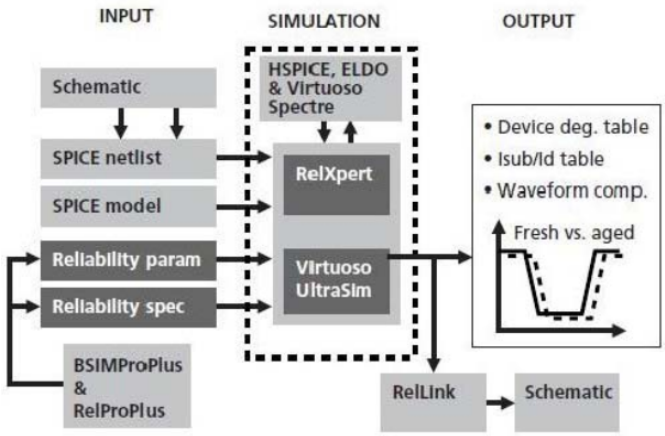

Figure 2.3: Reliability simulation flowchart using UltraSim [3].

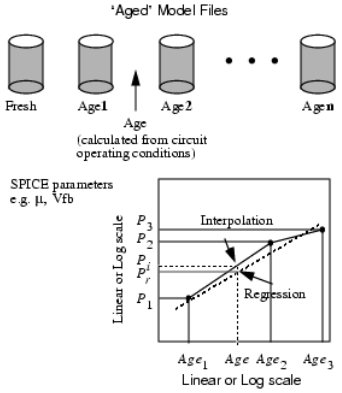

Figure 2.4: Calculating the Aged Model parameters with UltraSim as shown in [3].

ELDO environment, after the stress of each device is set for a transient time, the degradation of the devices for long time spans is calculated. The degradation results are then used to update the model parameters in the model card that can be user defined. The aging analysis in this tool can be achieved for a number of time intervals and at every time step, these model parameters can be updated. This estimation of the aging factor is based on extrapolation and specifically on either an exponential or linear calculation.

Other approaches that are also based on the RD model exist. However, some of them [26] fail to effectively capture the recovery phase leading to over-estimation of the BTI effect and pessimistic results concerning the yield inspection and other reliability metrics. In this work it is clearly mentioned that the model " does not include quantitative results considering an ac recovery" [26]. On the contrary, an RD-based model that can capture both stress and relaxation phases is [25]. The degradation in this model is calculated based on the operating conditions and the signal probability during the stress phase. In this approach, the threshold voltage shift is estimated from the Equation 2.2 but the N_{it} computations "are applicable only to deterministic waveforms" [25]. Considering the fact that predicting the exact nature of a random waveforms is impossible, this approach introduces a method that transforms an input signal of a random nature to an equivalent periodic deterministic waveform. In this work, the modified values of V_{th} of all the devices are computed and then a SPICE simulation is executed, to estimate the NBTI-induced delay.

From the above, it can be concluded that the majority of commercial tools and modeling approaches are based on the RD model and may fail to accurately capture the aging effect in downscaled devices of modern technologies. A second handicap that some of these simulators like UltraSim or ELDO have, is that they focus only on NBTI, while there is little consideration regarding the effects of the PBTI. So, a discussion of a simulation approach that is based on the atomistic theory rather than the RD theory seems a very timely and relevant contribution.

2.6 Conclusions

The BTI-induced degradation, although observed in devices of older technologies with large dimensions, has proven to be a reliability threat of major importance in downscaled technologies. The first theory that tried to explicate and develop the physical mechanisms of this effect was the reaction-diffusion theory. This theory supported that the gradual breaking and annealing of Si-H bonds at the Si/oxide substrate was the cause of BTI. The breaking of the bonds led to carriers being trapped in the broken bonds and cause a threshold voltage shift. Whenever no stress was applied to the gate of the device, hydrogen diffused back and annealed some broken bonds and in this way, partially recovered the V_{th} shift.

However, the reaction-diffusion model was proven inadequate to capture this aging effect in modern, small-dimension devices. The stochastic properties of the BTI observed in experiments were explained by a new atomistic theory. This model treats defects as switching oxide traps that change their state between occupied and unoccupied by capturing and emitting charge in a stochastic way. Another strong feature of this model is its workload-dependency. Hence, a discussion over a simulation framework based on this model is an apt direction for this thesis.

Next, the definitions of time-dependent functional yield, defects per million, failures in time and mean time to failure were reviewed as they constitute common metrics used in reliability analysis. Finally some widely-used industry tools as well as some simulation approaches that aim at modeling the BTI degradation and can help in the reliability analysis were discussed. All of them follow the RD model hence they seem to be of inferior accuracy, in comparison to an atomistic model. Thus, a computationally feasible tool for atomistic reliability analysis constitutes the major motivation behind the current thesis.

CHAPTER 3

Yield Inspection

3.1 Introduction

This Chapter presents an instantiation of a yield analysis methodology that targets the entire lifetime of realistic ICs. First, existing tools that are reused in the current instantiation are displayed and are briefly discussed. These tools have been in place as a result of previous work [27], however many of them have been extended and optimized in the context of this thesis. After the presentation of the existing components, we present the resulting instantiation and illustrate its functionality in a detailed way.

The importance of the `propagate` tool in this framework is underlined. This tool is responsible for estimating the duty factors of the internal nodes of the netlist and stands in contrast to other existing tools in frameworks for simulating BTI degradation. The verification of this component is also examined and the accuracy of the tool is compared to the hypothesis of a fixed duty factor for all nodes that can be seen in other works [28].

Lastly, after the definitions of functional yield, MTTF, FIT rate and DPM were reviewed in the previous Chapter, the mathematical formulation connecting these metrics is underlined.

3.2 Existing Components

In the previous Chapter, the insufficiencies of existing simulation tools based on the atomistic model were mentioned and the significance of a simulation framework that can accurately and effectively capture the principles of the atomistic theory was highlighted. A major handicap of these simulation approaches is the usage of a SPICE signal activity format that is very detailed but comes at a cost of inhibitive processing times. What previous approaches suffer from, is the absence of a signal waveform representation that would enable a precise yet efficient BTI simulation [8]. Recently, a novel waveform representation has been proposed, called Compact Digital Waveform (CDW). This signal representation "exploits optimally the accuracy versus complexity trade-off" and can enable BTI simulations at reduced CPU times [8].

- The tool that implements the idea of the CDW and has been adapted in the discussed framework is called `wavapprox` and a detailed description of its function can be found in previous work [27]. The key feature of this tool is "the grouping of consecutive signal regions that feature a similar f and α figures and occupy a duration of ΔT" [8]. So the percentage of the signal activity

belonging in this time interval (that is characterized by same values of f and α) is represented by a single point with coordinates (f, α, ΔT). Thus, given the SPICE waveform, the initial signal is scanned and the CDW points are produced. Moreover, the compression of the initial signal waveform to a CDW representation is done via selected accuracy, achieved by the user-defined error margins ϵ_f , ϵ_α.

In the context of this thesis, an alteration of this tool has been performed, in order for it to account for dc signals. Specifically, the original version of the tool estimated the duty factor of an ac SPICE signal representation focusing on the alterations of the voltage value of the signal during its duration. The new version of this tool calculates also the duty factor of dc signals were no alteration of the voltage value is observed. The new algorithm compares the constant voltage value of a dc signal to a user-defined limit and iff the voltage value of this signal is higher than the limit, then the duty factor of the relative signal is set to 1. In this simulation framework the limit is set equal to $\frac{V_{dd}}{2}$.

- Another tool that is used in this simulation framework is `pcestimate`. This tool is responsible for calculating the capture probability of every single defect of all the devices of a netlist. Hence, at each point of the CDW representation of a signal, the following Equation is performed, that calculates the P_c of a defect [8]:

$$P_c = \left(P_{c0} + \frac{b}{a} \right) e^{an} - \frac{b}{a} \tag{3.1}$$

  The parameters a and b are functions of α and f, so it is clear that the inputs in this tool are the time constants, the initial capture probability and the α and f that represent the gate workload and provide workload dependency in the simulation.

- The `flatten` tool is used in order to flatten the netlist. It should be noted that commercial tools allow an internal flattened representation. However, given that custom observability of the netlist is preferable during the development stage, we choose to flatten it with a dedicated tool.

- After the netlist is flattened, a tool named `tzinit` is used in order to capture V_{th0} variability. After all, in a population of realistic devices, not all of them have the same threshold voltage but the V_{th0} is distributed near a mean value [15]. Hence, this tool assumes 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 σ^2" [2]. This V_{th0} variability is captured by using the **delvto** command [29], which is available in typical SPICE solvers.

- A tool that populates each device of a netlist with defects and creates initial conditions for the BTI defects is essential. The `definit` tool places a number of defects in all FETs of the netlist. The number of the defects of a device is taken randomly from a Poisson distribution according to the atomistic model with a mean of $N = L_G W_G N_{ot}$ [2]. Hence, devices with larger dimensions carry a higher number of defects. Moreover, the impact of defects on the V_{th} (the ΔV_{th} of every defect) is again taken randomly from an exponential distribution with a mean value of $\eta = \frac{\eta_{p,n} w l}{L_G W_G}$ where w and l are scaling factors of the dimensions of the device and p,n concern the type of the device [2]. The `definit` tool also initializes the capture probability P_{c0} and prints the time

Existing Components	
`flatten`	flattens the netlist
`tzinit`	introduces time-zero variability
`definit`	populates with defects all devices
`wavapprox`	produces a CDW signal representation
`propagate`	propagates the α to all nodes of the netlist
`pcestimate`	calculates the capture probability of a defect
`dvfs`	introduces dvfs technique to the simulation

Table 3.1: The components of the simulation framework

constants of every defect.

Hence, the `definit` tool populates with defects all the devices of the netlist while the time constants are widely distributed along the logarithmic scale. This leads to a separation of the defects according to their time scales. This separation is necessary because in this simulation framework not all defects can be accurately simulated. Specifically, the P_c of the defects with time constants comparable to the bit stream's period $1/f_{avg}$ cannot be estimated by the CDW simulation [8]. Thus, they constitute the first team of defects called "fast" defects. Simulating the BTI degradation imposed by the group of "fast" defects is out of the scope of this thesis because significant work on this matter already exists [14], [30], [31] . Defects with time constants much higher than the ones mentioned earlier are accurately modeled with the proposed flow and compose the second team of the defects, named "slow" defects [8].

- It must be noted that dynamic voltage scaling (dvfs) is a widely-used technique for mitigating the BTI-induced degradation on ICs. Therefore, a `dvfs` tool that changes the bias conditions of the netlist throughout the simulation is also necessary.

- BTI simulation is workload dependent, meaning that the signal activity -the workload- of every gate of a device has to be evaluated. In other words, propagating the CDW signal activity, or the duty factor more specifically, (as frequency and time duration remain constant throughout the circuit) from all primitive nodes to every gate node of the netlist is necessary. The `propagate` tool is used in this simulation framework for this work and the accuracy of this tool will be examined.

3.3 SPICE-free Signal Activity Propagation

As stated earlier, the `propagate` tool can calculate the duty factors of all the internal nodes of a netlist, given only the α of the primitive nodes. It is important to note that the algorithm of this tool is based on estimating the duty factor on the device level while other tools calculate the signal probability at the gate level. According to the algorithm of this tool, the devices of the netlist are checked and depending on the signal activity of the gate node, the α is propagated between the connections of source and drain. In case there is no signal activity in the gate node and the duty factor equals 0, then the device works as an open switch that does not propagate the signal between the connections of source

and drain. The respective algorithm has been presented thoroughly in previous work [27], hence only extensions and optimizations will be discussed in the current thesis.

For this simulation framework, an extension of this tool is suggested. The improved algorithm checks if all the devices are part of a transmission-gate, that has both source nodes of the complementary devices on the one side and the drain nodes of the two devices on the other. After scanning all the transistors of the netlist and a PMOS device is found to share the same drain and the same source with an NMOS, then these devices are considered to compose a transmission-gate. Hence, in this optimized version of the tool, devices that are found to compose a transmission-gate have a different signal activity propagation profile, as can be seen in Figure 3.1.

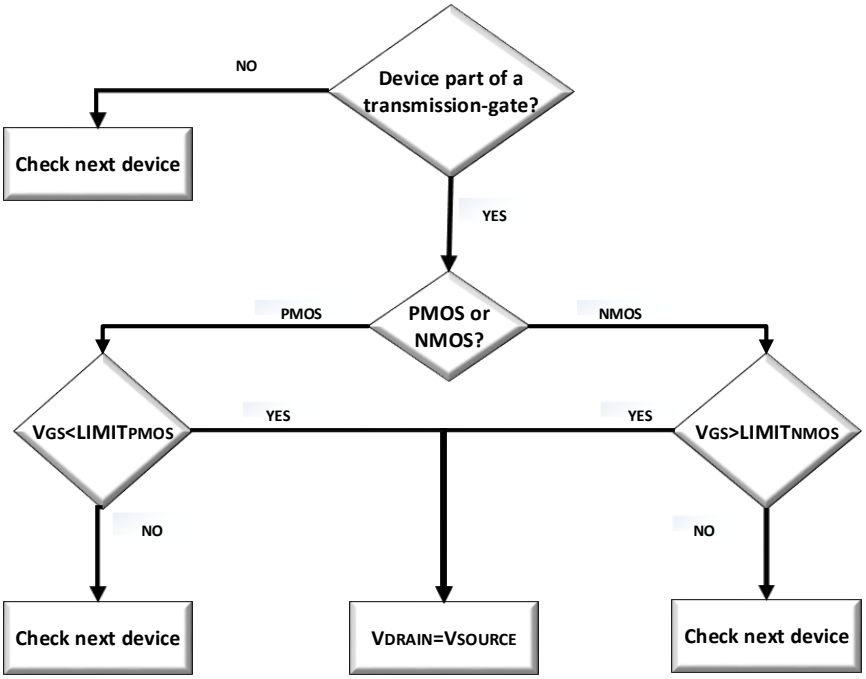

Figure 3.1: Flowchart showing the algorithm of the extension of the propagate tool

In order to achieve that, the algorithm checks the gate voltage of the devices that are part of a transmission-gate. In case the device is a PMOS, and the duty factor is below the limit of 0.95, the transistor is considered to propagate the voltage activity (meaning the duty factor α) from its source to its drain. If the value is greater than this limit, the NMOS device of the same transmission-gate is examined. In case the duty factor of the gate node of the NMOS is greater than the threshold of 0.05 then the drain of the transistor is set equal to the source's α. In any other case, the transmission-gate is on the off-state. Then, there is no propagation of the signal from the one connection of the transmission-gate to the other. This simple extension of the original propagate algorithm is expected to capture the operation of a transmission-gate, thus provides more accurate results than the original version of the tool.

So as to verify the effectiveness of the extended version of the tool, the α of every node of a memory netlist is estimated and is compared to the CDW representation derived from the transient waveform

of the same node. Hence, the absolute error between the duty factors estimated by the newer version of `propagate` and the duty factors calculated by a SPICE simulation can be defined as :

$$\epsilon_1 = |\alpha_{\text{propagate}} - \alpha_{SPICE}| \tag{3.2}$$

It has to be noticed that, in other BTI-related simulation frameworks, the duty factor of all internal nodes is considered fixed to a constant value [28], [32]. Specifically, we find that "a safe and tight upper bound for circuit delay degradation under worst-case signal probabilities is required for reliable operation. In most practical cases, it can be obtained by assuming WC-K_{aging} of 0.95 for the entire circuit" where K_{aging} is "a scalar from the interval (0, 1), a function of stress probability" [28]. Taking into consideration that the atomistic model is workload dependent, it is expected that such a hypothesis may be oversimplified and could lead to pessimistic results when simulating large inventories consisting of numerous internal nodes. Hence, the efficacy of the `propagate` tool and its advantages over such plain estimations need to be tested. Therefore, the hypothesis of a fixed value at 0.95 is also explored while the absolute error in this occasion is defined as:

$$\epsilon_2 = |\alpha_{0.95} - \alpha_{SPICE}| \tag{3.3}$$

The netlist that will be used for this task, is a subset of an embedded Static Random Access Memory (SRAM) [4] that is shown in Figure 3.2.

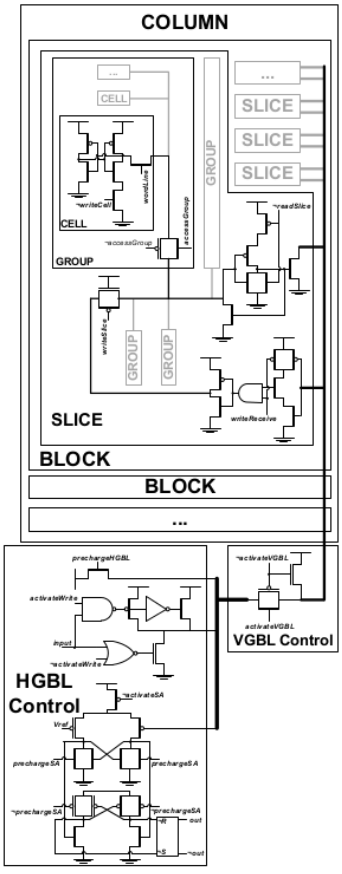

Figure 3.2: Operational path of an embedded SRAM design as presented in [4]

First, from a `createsup` tool a memory trace is created that contains piece-wise linear waveforms of all primary nodes of the netlist. Every memory trace is set to contain the waveforms of the primary nodes for a series of 20 equiprobable operations: the writing of 0 in the memory bit, the writing of 1, the reading of the memory bit and the retaining operation that simply retains the stored bit. By default, the memory trace always starts with the writing of the 0 value in the SRAM cell. Then the rest operations of the memory cell, that will create a single memory trace, are selected randomly.

After the piece-wise linear waveforms (PWL) of the primary nodes of the netlist are created, the `wavapprox` tool is used in order to calculate the duty factors for each of these nodes according to their PWL waveform. A CDW representation of the primary nodes is then available with every waveform being represented by a single CDW point. In order to achieve that, the `wavapprox` tool is set to present the crudest approximation of every signal with $\epsilon_f \to +\infty$ and $\epsilon_\alpha = 1$, thus the duty factor of every node is averaged out from the entire signal duration.

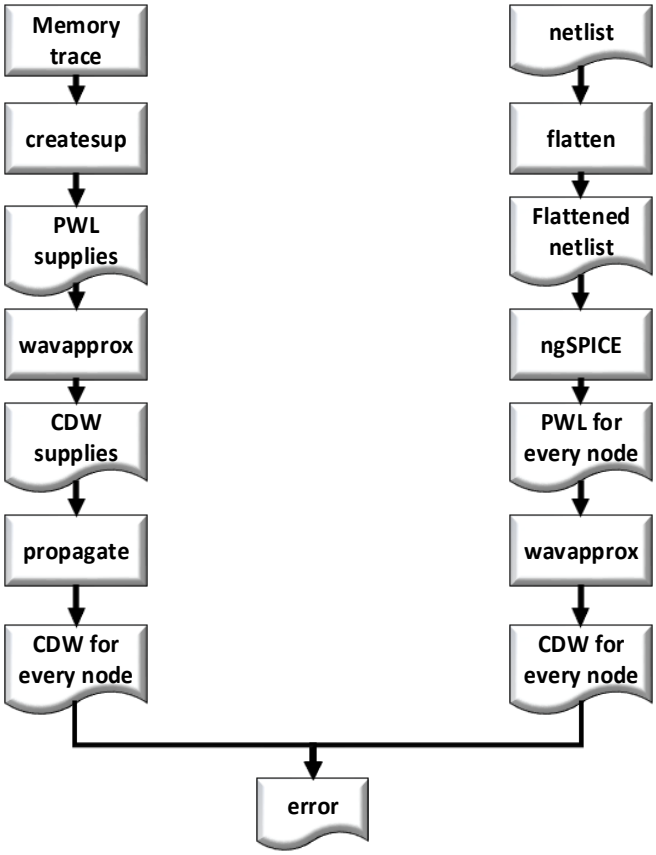

Figure 3.3: Logic diagram for propagate verification

Next, the duty factors of the primary nodes are propagated by using the `propagate` tool and the α of every internal node is estimated. Thus, a file containing the CDW points of all nodes is created. By using the same memory trace and the initial PWL waveforms, a SPICE simulation is completed that calculates the transient waveform of all nodes of the netlist and creates a file for each node. It is generally assumed that a SPICE simulation provides a precise estimation of the signal waveform of all nodes.

After the calculation of the PWL waveforms of all nodes, the `wavapprox` tool is used and the CDW points of all nodes derive. Hence, the propagated duty factors estimated in the first case are compared against the duty factors calculated in the second case by the SPICE simulations and the absolute errors of all nodes of the netlist, defined in Equation 3.2, are estimated.

Following the verification flowchart presented in Figure 3.3 the histograms of the mean value and standard deviation of the absolute errors defined in the Equation 3.2 for 60 memory traces derive. First the accuracy of the extended and the original version of the `propagate` tool are compared and the histograms of the errors for the same 60 memory traces are shown in Figure 3.4.

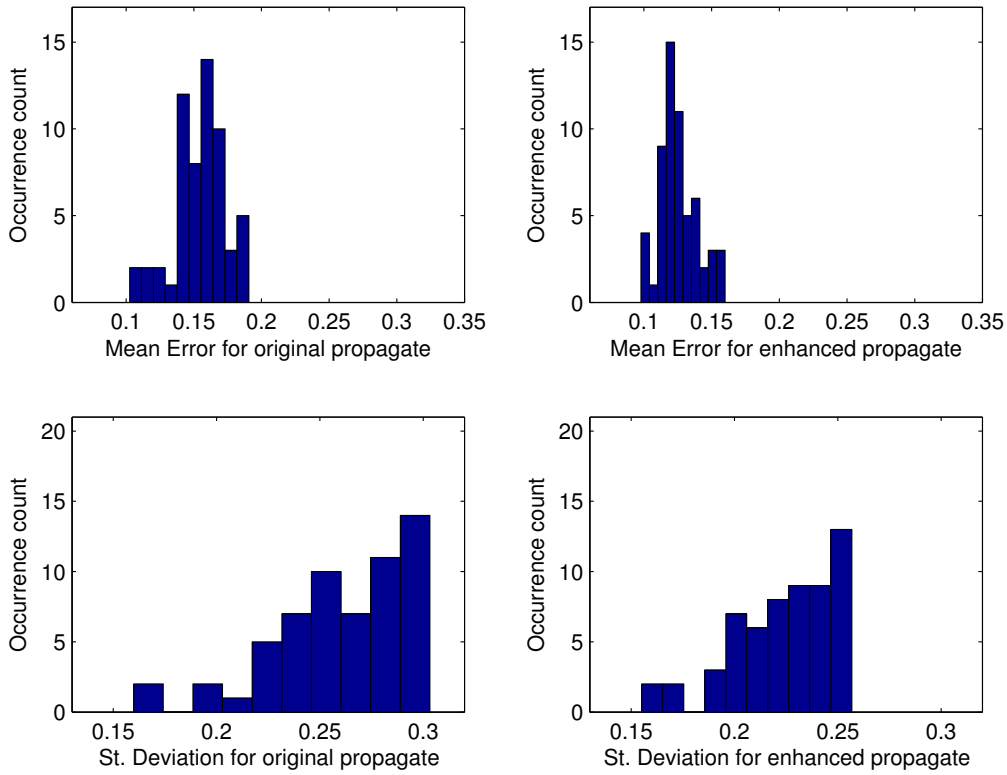

Figure 3.4: Histograms of mean error and st. deviation comparing the original and the extended version of the tool

As it can be seen, the absolute error in the case of the extended version of the tool is less, both in terms of mean value and standard deviation. Therefore, the extended algorithm enhances the original propagate algorithm to provide more accurate estimations. This can be explained by the fact that transmission-gates are fundamental components of an SRAM path thus, by capturing the operation of transmission-gates the accuracy of this SPICE-free signal activity propagation tool is improved.

Next, another 60 memory traces are created and the duty factors of all internal nodes of the circuit are estimated by the enhanced tool. After the absolute errors are calculated, the efficiency of the tool is tested against the hypothesis of a fixed duty factor at 0.95. For the former case the absolute errors are estimated from the Equation 3.2 while for the latter, the absolute errors are calculated from 3.3. The histograms of the mean value and of the standard deviation of the absolute errors are shown in Figure 3.5.

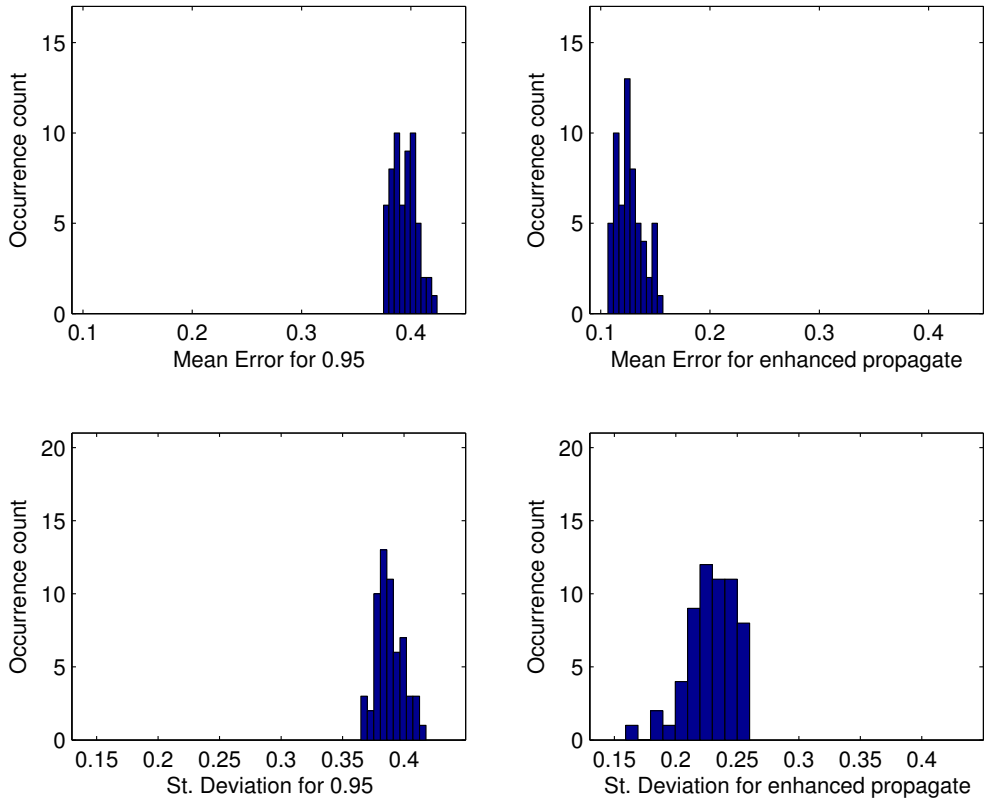

Figure 3.5: Histograms of mean error and st. deviation comparing the 0.95 hypothesis to the extended propagate algorithm

From the graphs of Figure 3.5, it is clear that estimating duty factors by this propagation tool is a more accurate evaluation than the plain hypothesis of $\alpha = 0.95$ for all nodes. In fact, for realistic workload scenarios of the SRAM path, the mean absolute error is nearly tripled in the case of a fixed duty factor at 0.95. This indicates that a fixed duty factor to 0.95 for all nodes, in order to account for the worst-case scenario, is unrealistic and rather pessimistic. Thus, such a hypothesis can lead to overestimations of the BTI-induced degradation.

The efficiency of the `propagate` tool must also be noticed. Obviously, the hypothesis of a constant duty factor for all the nodes of a netlist does not require any computation time, but, as previously demonstrated, is inferior to the estimations of any systematic activity propagation scheme, either using SPICE or the proposed `propagate` tool. A complete SPICE simulation can provide more accurate results however, when large netlists are simulated and for long-lasting simulations, the use of SPICE is computationally prohibitive. Figure 3.6 shows the CPU time needed for the `propagate` estimation compared to the CPU time needed for the SPICE simulation. For the 60 memory traces tested, the mean time on the first case with the `propagate` tool is 0.9 s, while for the SPICE simulator, the mean CPU time equals 23.3 s. Hence, the huge gain in computation time is obvious when using the proposed `propagate` tool.

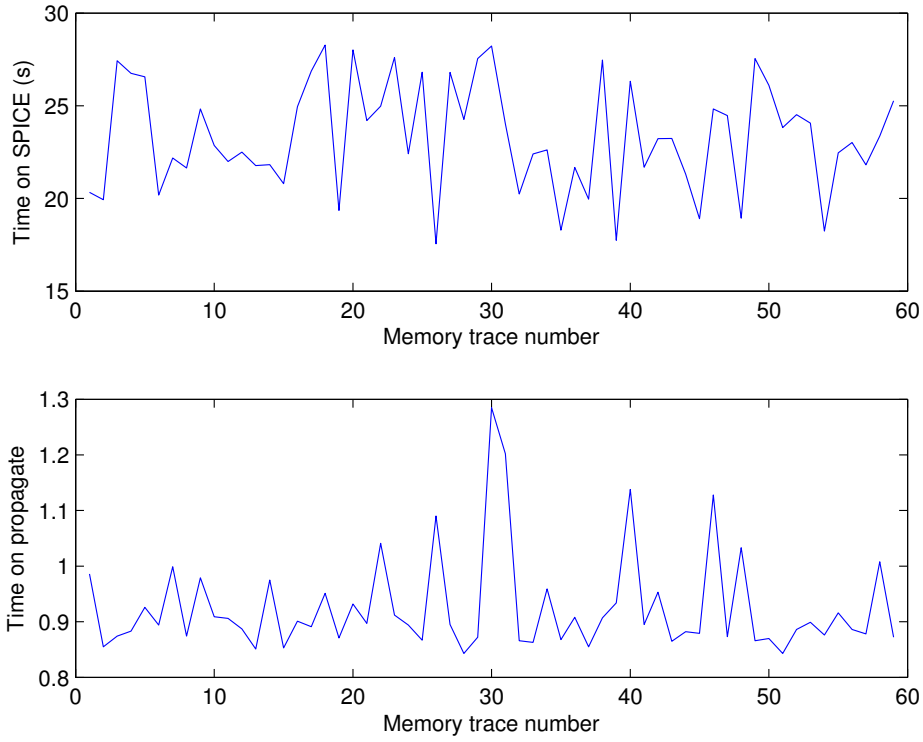

Figure 3.6: Difference in CPU time for SPICE simulation and for propagate

3.4 Combining Existing Tools to a Yield Analysis Tool

After presenting the existing tools and the proposed optimizations, we now combine them into a single framework for yield estimation. The yield of the same netlist (shown in Figure 3.2) used for the verification of the `propagate` tool will be estimated. This circuit contains 63 devices and it supports reading and writing of the cell. A sample of this design is defined as functionally correct if a zero value can be written and then read successfully. 150 Monte Carlo iterations will be simulated with an inspected lifetime of approximately 3 years, which will be split in time intervals. These time windows will be equal and will last $2.5\times\ 10^7$ seconds. The temperature will be constant at 50 $^\circ$ C while it will be assumed that the test-case circuit operates at a frequency $f = 0.88$ GHz.

Firstly, the flattening tool is used in order to print all 63 devices from the subcircuits of the netlist. Then, with the flattened netlist available, ΔV_{th0} variability is estimated via the `tzinit` tool. Hence, a time-zero component of ΔV_{th0} is added to the threshold voltage of the devices that is constant through the device lifetime but varies across the set of the devices. The threshold shift is added to each device of the flattened netlist through the command [29]: **delvto=** ΔV_{th}. The ΔV_{th0} follows the Gaussian distribution which is different for the pFETs and the nFETs.

An initial file of voltage supplies for the SPICE simulation is created. Then, the `definit` tool is used, populating every device with defects. Always according to same atomistic approach [8], the time constants are distributed over a logarithmic scale, between a maximum and a minimum value. As mentioned earlier, defects with time constants comparable to the bit stream's period $1/f_{avg}$ constitute the group of "fast" defects and are not taken into consideration in this thesis. Also, for the Monte-Carlo

Variables for the Simulation Framework	
V_{th0} variability follows a Gaussian distribution $V_{th0} \sim Norm\ (\mu_0, \sigma_0)$	$\mu_0 = 0.397, \sigma_0 = 0.0397$ nMOS $\mu_0 = -0.339, \sigma_0 = 0.0339$ pMOS
Number of Defects is taken from a Poisson distribution with mean N=$N_{ot}WL$	$N_{ot} = 1e + 11$ pMOS $N_{ot} = 6.7e + 10$ nMOS
Time constants are chosen from a logarithmic distribution	$\tau_{range} = 10^{24}$, $\tau_{min} = 10^{-12}$
$V_{th,TD}$ follows an exponential distribution with mean $\eta = \frac{\eta_{p,n} wl}{W_G L_G}$	$\eta_{p,n} = 5e - 3$ V $w \times l = 90$ nm $\times 70$ nm

Table 3.2: Variables for the Simulation Framework

iterations, the definit tool sets the P_{c0} to 0.

After a transient simulation of the netlist takes place, testing the delay (between the writing and reading of the SRAM cell), imposed by V_{th0} variability, the initial source supplies created by the `dvfs` tool are propagated for a defined time-interval to all internal nodes of the circuit. Now that the duty factors, frequency and time duration (which equals the specific time interval) of all the gate nodes of the circuit are identified with a CDW representation of a waveform, the workload of every device is known and via the `pcestimate` tool the capture probability of every single defect of all devices can be estimated. The equation of P_c is a function of the workload of the device, the time constants and the previous value of P_c.

- A `stride` tool is created that compares the value of P_c, calculated by `pcestimate`, with a number r. This number is chosen randomly from the (0, 1) space. Iff $r < P_c$ then the defect is considered occupied. For every defect a new value of r is drawn from the (0, 1) interval. All defects for every device are elaborated and if occupied their state is set to 1 and if unoccupied the state is 0.

- A tool named `aging` is also created that checks the state of all defects. Iff the state value equals 1, the ΔV_{th} of the defect is added to the total threshold shift of the device through the **delvto** command [29]. After the `aging` tool is finished, the aged netlist is created with every device having a ΔV_{th} depending on its occupied defects as claimed by the atomistic theory. This instance of the netlist is again simulated and the delay between the two functions of the circuit is measured.

The sources are propagated for the remaining time intervals until the end of the inspected lifetime. At each time step, following the process demonstrated earlier, the aged netlists are created and then simulated. If a zero value cannot be written and read successfully, the instance is considered as functionally erroneous.

The procedure elaborated in this Section is placed in a loop, simulating a number of 150 samples that constitute a set of Monte Carlo iterations. At each time step, the functional yield of the circuit can be estimated by dividing the number of the circuits that did not fail at that specific time step, to the number of the samples tested. The flowchart for a single iteration of the proposed simulation framework is shown in Figure 3.7.

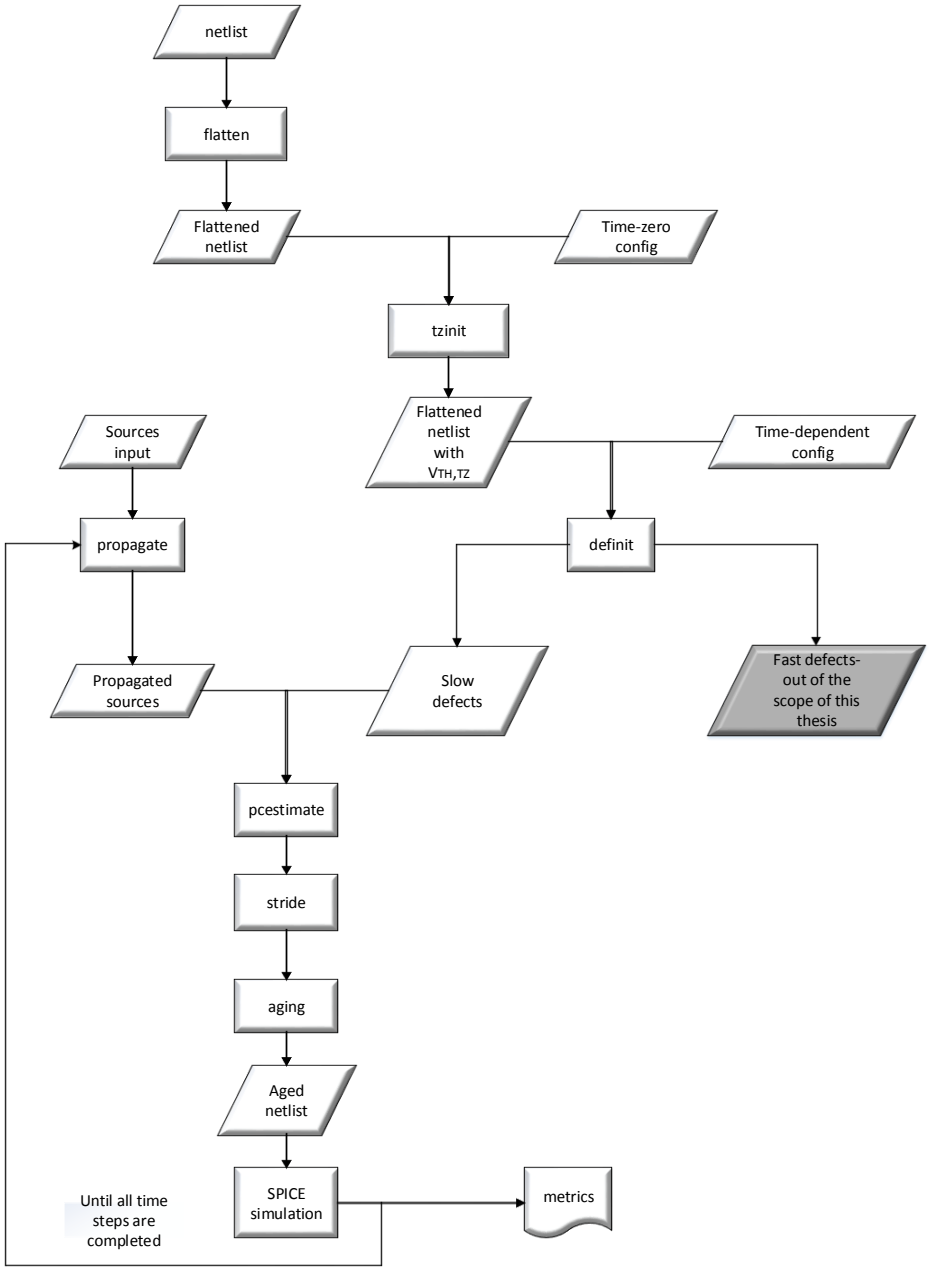

Figure 3.7: Flowchart for a single iteration of the proposed simulation framework

3.5 Yield vs. DPM vs FIT/MTTF

The estimation of the functional yield is rather straightforward. When a failure appears during the simulations, then the respective sample is added to the functional violations. At the end, the functional yield of the circuit can be estimated by dividing the number of circuits that did not fail, to the set of the samples. The number of functionally acceptable circuits is simply the number of samples simulated minus the failed circuits. Hence, the functional yield is estimated by Equation 2.6. Also, considering the definition of defects per million, DPM is estimated by Equation 2.7.

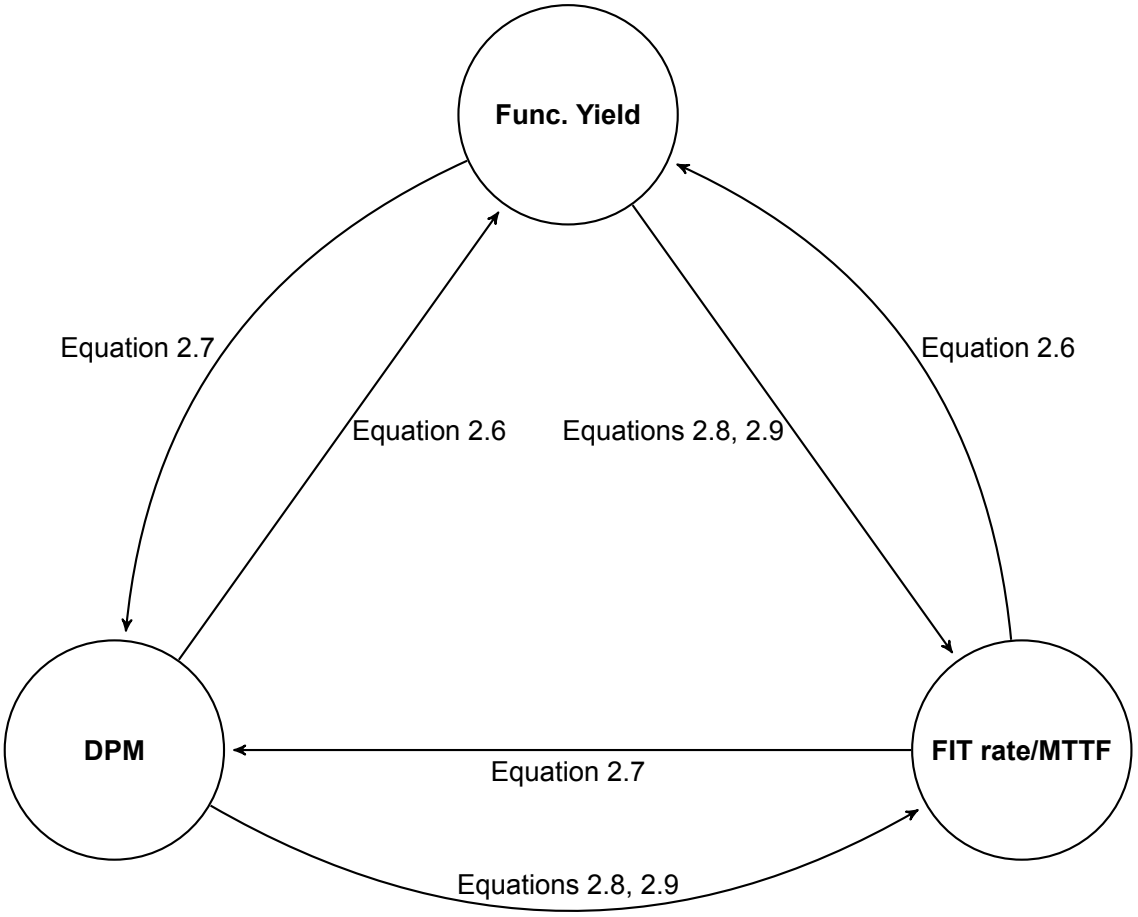

Figure 3.8: Diagram showing the interdependability of Functional Yield, DPM, FIT rate and MTTF

In order to estimate the MTTF and FIT rate, some mathematical transformations need to follow. Firstly, the point of MTTF (in hours) is calculated by 2.8 of the previous Chapter. In this Equation, the Device Hours represent the equivalent device hours of the test procedure, which in the case of elevated test temperatures are estimated by [21]:

$$EDH = A \times CDH \tag{3.4}$$

A is the *acceleration factor* and is given by Equation 2.11 and CDH represents the cumulative device hours and can be calculated by [21]:

$$CDH = \text{Number of Devices Tested} \times \text{Test time (in hours)} \tag{3.5}$$

After the MTTF estimation, the point of FIT rate can be estimated from the Equation 2.9

It must be noted that the FIT rate and the MTTF estimations depend on the number of failures that occurred during the test procedure and on the cumulative device hours. As can be seen from Equation 3.5 longer time tests require fewer testing samples. On the contrary, if simulations for long time windows are infeasible or computationally prohibitive, then in order to calculate these metrics a large number of devices needs to be tested.

3.6 Conclusion

This Chapter discussed a simulation framework that is capable of capturing BTI degradation in large inventories and for long time spans. First, the existing tools that compose this framework were briefly reviewed. The `tzinit` that introduces V_{th0} variability in the simulation, the `definit` tool, that populates every device with defects, the `wavapprox` tool that compresses the signal representation and the `propagate` tool,that estimates the duty factor of all internal nodes of the netlist, are the most significant.

Then, a verification of the accuracy and the efficiency of the `propagate` tool were demonstrated. Specifically, the duty factors estimated by this tool were compared against the duty factors calculated by a SPICE simulation for a number of 60 iterations of a netlist. The netlist that was used, is a representative path of an embedded SRAM. Each iteration consisted of 20 operations of the netlist and the results showed that, indeed, the propagation algorithm constitutes an efficient and accurate tool for an overall estimation of the duty factors of all nodes of a circuit for long time spans, in which a SPICE simulation is computationally prohibitive.

Afterwards, the synthesizing of the simulation framework from its components was presented along with the constants that will be used for the simulations. These constants concern the mean and sigma values for Gaussian distributions, the defect densities for P-MOS and N-MOS transistors and constants that regard time scales. Finally, the inter-dependability between MTTF, FIT rate, DPM and functional yield was described and a unified formulation connected these reliability metrics.

CHAPTER 4

Quantitative Results

4.1 Introduction

This Chapter presents the quantitative results that regard the monitoring of the BTI-induced degradation over the circuit's lifetime. The time inspection will last approximately 3 years (10^8 seconds) while the netlist simulated is a part of an embedded Static Random Access Memory (SRAM), as shown in Figure 3.2. The simulation framework was presented in the previous Chapter, along with the mathematical forms that connect the reliability metrics .

For these simulations, all time steps last 2.5×10^7 seconds. The functional yield at each instance of the simulation will be estimated over a set of 150 Monte Carlo iterations. Then, through the formulation introduced in the previous Chapter, the metrics of DPM, MTTF and FIT rate will also be calculated. Finally, the last section of this Chapter lists a series of concluding remarks that regard the simulation results.

4.2 Functional Yield, DPM, FIT rate and MTTF estimations

Two cases of different V_{dd} configuration will be simulated. For the first case, the voltage level of the netlist will be fixed thus, simulating a normal V_{dd} profile. For the second case, the voltage level will dynamically increase during the circuit's operation, corresponding to a dvs operation.

4.2.1 Circuit Simulation under Fixed Supply Voltage

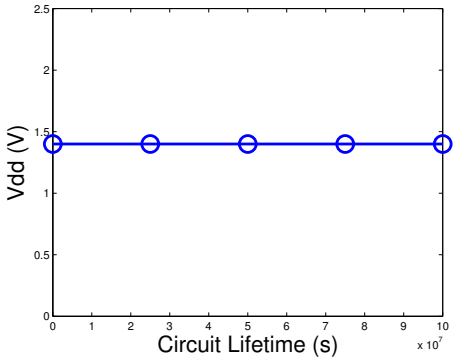

Figure 4.1: Graph showing the V_{dd} configuration for the case of fixed supply voltage

The netlist is stressed under a bias of 1.4 Volts for all time steps and at each time step, a transient SPICE simulation is performed to test the functionality of the sample. An instance of a sample is considered functionally correct, iff the 0 bit can be written on the SRAM cell and then read successfully. Figure 4.2 shows an instantiation of a functionally correct sample.

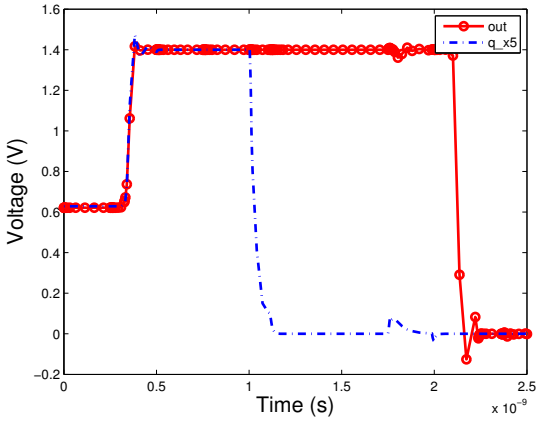

Figure 4.2: Graph showing an instantiation of a functionally correct sample

The two nodes examined represent the node of the SRAM cell and the output node of the circuit. Specifically, “q_X5” corresponds to the node of the cell of the memory path, where the 0 bit is stored, while ”out“ represents the output node, where the stored bit is propagated after a read operation is performed. As it can be seen in Figure 4.2, first the zero bit is stored in the SRAM cell and then, the bit is read and propagated to the output of the SRAM netlist.

On the contrary, examples of failed circuits are shown in Figure 4.3. On the left Graph, it is clear that although the 0 bit is written in the SRAM cell, it cannot be read correctly and the 1 bit is propagated to the output node instead. The right Graph shows an instantiation of a sample where the 0 bit cannot be properly written on the SRAM cell. It must be underlined that, in our case, such failures are only caused by threshold voltage shifts in the devices of the netlist due to V_{th0} variation and BTI degradation.

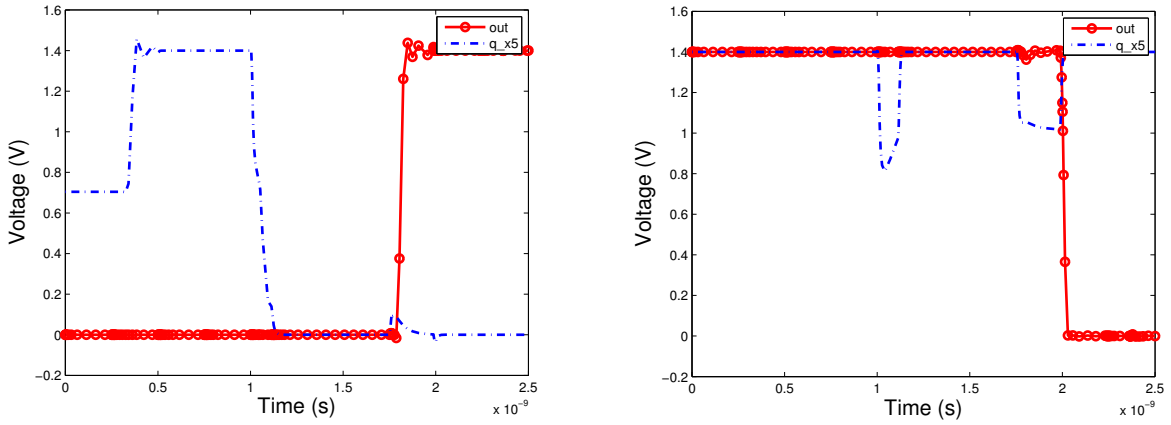

Figure 4.3: Graphs showing examples of functional violations

After the functional yield is estimated for all time steps, the metrics of MTTF, FIT rate and DPM are calculated through the formulation presented in the previous Chapter, according to Figure 3.5. All

estimations are performed for a 95% confidence interval (CI) [22]. It has to be noted that the temperature was fixed for this lifetime inspection at 50 ° C hence, there was no temperature acceleration in these tests. The Upper and Lower limits of the functional yield are estimated by the formulation found in relevant work [17].

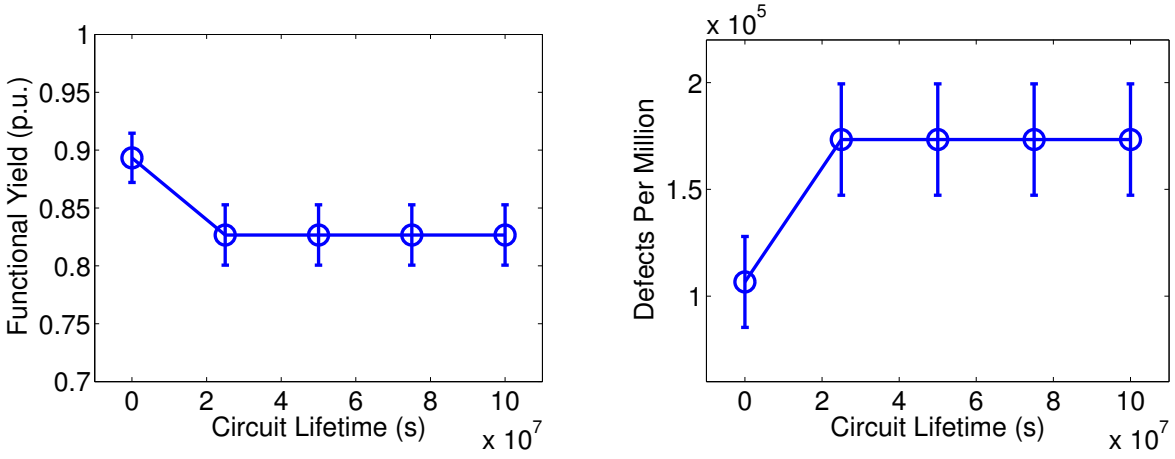

Figure 4.4: Graphs showing the Functional Yield and DPM estimations

From the Graph in Figure 4.4 it is clear that the functional yield drops considerably at the time-zero instance, due to V_{th0} variation. The metric is reduced significantly at the next time step, when BTI degradation manifests. However, for the rest of the inspected lifetime, the functional yield remains at the same level hence, it can be implied that the rate of the degradation rather slows down. In the same spirit, a considerable amount of failed circuit is estimated at the time zero instance, which is attributed only to V_{th0} variation. For the next time instances, the DPM value increases significantly due to the imposed BTI degradation and remains rather constant during the rest time inspection.

Keeping into consideration that the time constants of the defects are distributed along the logarithmic time scale, this constant value of the functional yield and of the DPM, for a linear time inspection, is expected. However, during the first time window, from 0 to 2.5×10^7 seconds, the yield is reduced significantly because of BTI. Therefore it can be implied that BTI degradation manifests heavily during the early lifetime and results to an increase of failed circuits and defects per million and to a decrease of the functional yield. Nevertheless, the rate of BTI degradation fairly slows down for the rest of the operating time.

Lastly, before discussing the FIT rate and MTTF calculations, few things need to be underlined. In the previous Chapter and in the relevant Section, it was specifically mentioned that the cumulative device hours (CDH) are the product of the number of circuits tested and the time duration of a test. Hence, in order to estimate the FIT rate and MTTF metrics of a set of samples and for a given test duration, shorter testing times require a higher number of samples to be tested, in contrast to long-lasting tests that demand a lower number of samples. This is shown from the Equation 3.5. Thus, in order to compare the FIT rate and the MTTF in 4 different time instances, the CDH values need to be equal. In order to achieve that, for the 2.5×10^7 seconds instance, 150 samples were tested. For the second instance at 5×10^7 seconds, the first 75 Monte Carlo iterations were taken into account for the estimations. In the same spirit, for the next instances a number of 50 and 37 samples respectively were tested -the first 50 and 37 of the overall 150 simulations-. These assumptions set the CDH equal for

all four instances. Moreover, in order to estimate Upper and Lower limits of MTTF and FIT rate, the formulation from previous work [20],[33] had to be considered.

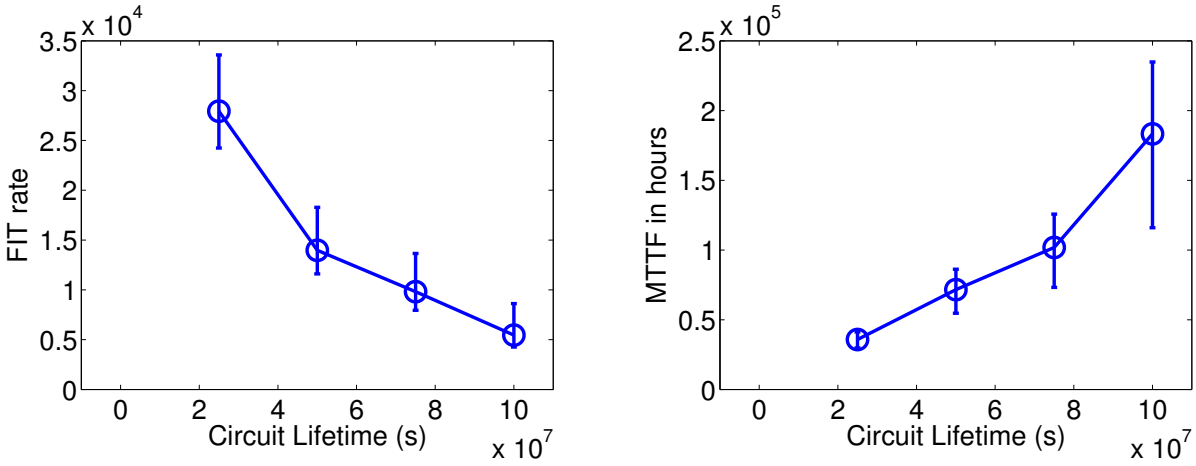

Figure 4.5: Graphs showing the FIT rate and MTTF estimations

From the Graphs of Figure 4.5, it is clear that the MTTF value is slightly increasing along the time instances, while the FIT rate is slowly decreasing. Taking into account that longer simulation times correspond to fewer testing samples, it can be concluded that simulating a small number of samples for long durations can results to higher estimations of the MTTF. On the contrary, testing a larger set of samples for shorter time windows can lead to pessimistic estimations regarding the MTTF and FIT rate. Considering the fact that BTI strongly manifests at an early lifetime and degrades the circuit, such estimations regarding the MTTF and FIT rate can be predicted.

4.2.2 Circuit Simulation under Dynamic Voltage Scaling

A widely used mitigation technique for BTI degradation is dynamic voltage scaling (dvs). Previous works [34], [35] suggest that instead of using constant voltage supplies over the entire circuit's lifetime, dynamically increasing the voltage level of the IC can mitigate the aging effect. Hence, in these simulations, the reliability metrics of the SRAM circuit under BTI stress will be estimated using dynamic voltage scaling.

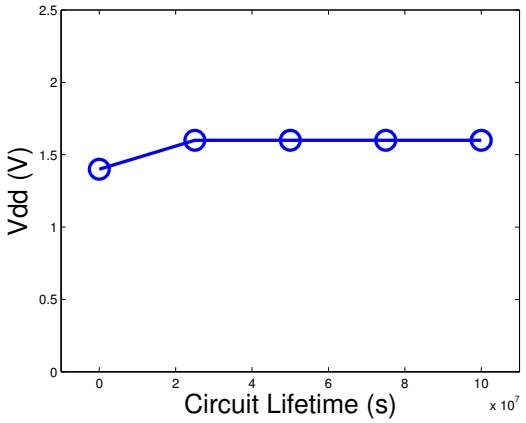

Figure 4.6: Graph showing the V_{dd} configuration for the case of dvs

The temperature was fixed at 50 ° C and the inspected lifetime, that lasted nearly 3 years, was again split in 4 equal time intervals of 2.5×10^7 s. each. However, as can be seen from Figure 4.6, the voltage supplies are fairly increased at the first time step and remain constant throughout the rest simulation time.

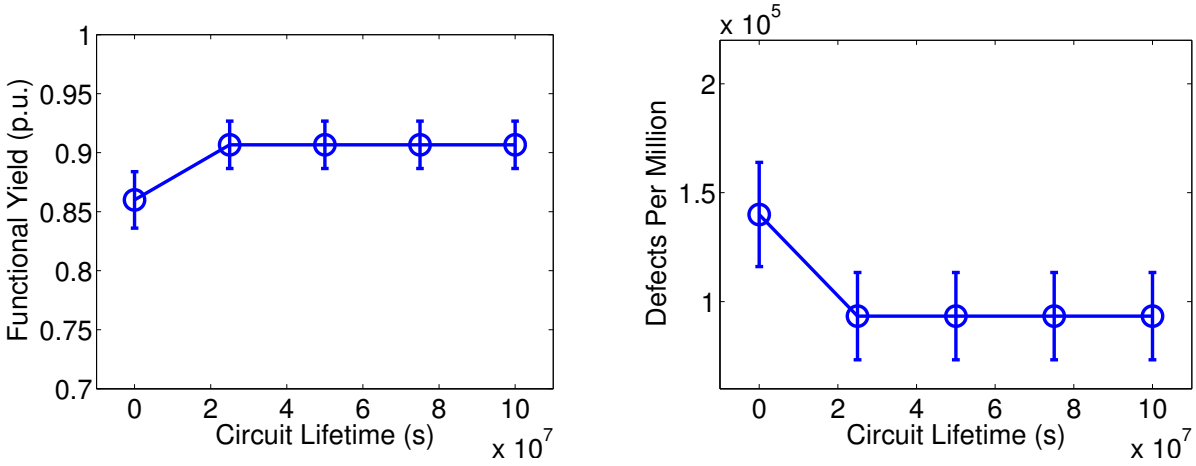

Figure 4.7: Graphs showing the Functional Yield and DPM estimations for the dvs case

From Figure 4.7 it can be concluded that dvs considerably counteracts the BTI degradation. Specifically, although an increase in voltage supplies of the netlist is expected to deteriorate the aging effect, this dynamic V_{dd} scaling results in an significant increase of the functional yield, mitigating BTI. From this Figure it can also be implied that the DPM of the circuit for the 4 time steps is less than the DPM estimated at the time zero instance. Thus, BTI degradation, which is expected to manifest during the circuit's operation at those 4 instances is partially mitigated.

Comparing the Graphs of Figure 4.7 corresponding to the dvs technique to the Graphs from Figure 4.4, we can see that under fixed V_{dd} the functional yield of the circuit drops at the last four instances due to BTI degradation while DPM increases. However, for the dvs case, the functional yield increases during the last four instances while the DPM value is significantly reduced. Hence, it can be implied that dynamic voltage scaling can enhance the circuit's operation and increase the circuit's yield.

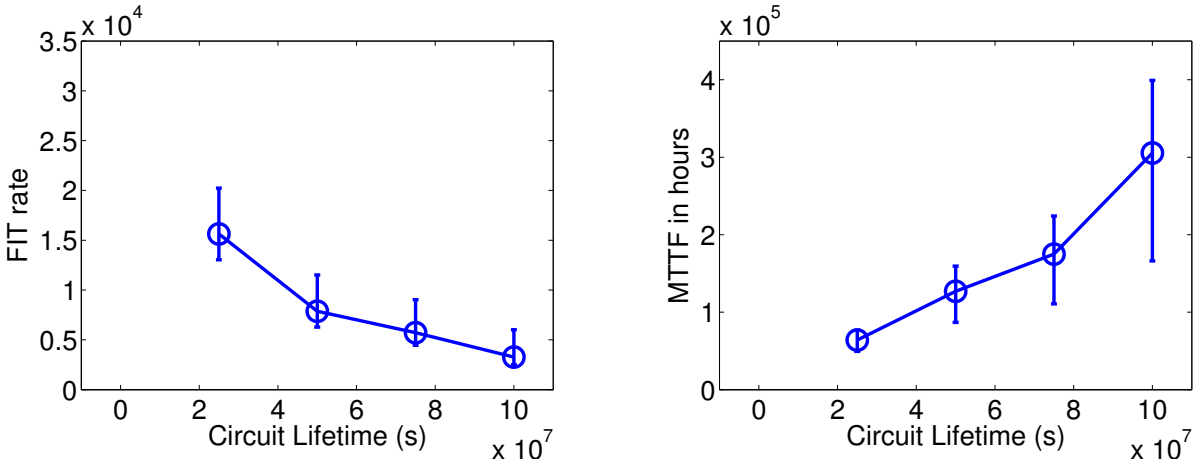

Figure 4.8: Graphs showing the FIT rate and MTTF estimations for the dvs case

The Graphs of Figure 4.8 again show the MTTF estimation to increase over longer-lasting simulations while FIT rate is decreasing. Hence, these Graphs indicate once more the logarithmic degradation imposed by BTI. Nevertheless, the overall increase of the MTTF estimations and the shift of the estimated FIT rates to lower values advocate in favor of the dvs as a mitigation technique for BTI. Specifically, comparing the FIT rate Graph in Figures 4.7 and 4.4, it is clear that regarding the first time instance at 2.5×10^7 seconds, the FIT rate estimated when using dvs is significantly lower than the FIT rate estimated under fixed V_{dd}. While comparing the next time instances, this overall decrease of FIT rate in the case of dvs becomes less obvious because fewer samples are tested. Regardless, the FIT rate drops when using a dynamic V_{dd} configuration while the estimated MTTF increases considerably.

4.3 Comments

In this Chapter the quantitative results of the estimations of functional yield, DPM, MTTF and FIT rate of the circuit shown in Figure 3.2 under BTI degradation were presented. It can be concluded that the V_{th0} variability results in a decrease of the functional yield of the netlist hence a small fraction of these circuits are expected to fail due to the time zero threshold voltage variation. After stressing the devices and due to the BTI manifestation, the functional yield of the circuit drops significantly at an early stage of the lifetime and the number of failed circuits increases. After this sudden decrease of the functional yield and the increase of the DPM, these metrics remain constant throughout the rest of the lifetime inspection. However, BTI is expected to manifest at a logarithmic time scale, due to the distribution of the time constants of the defects over a logarithmic scale.

This logarithmic degradation imposed by BTI leads to pessimistic results of the FIT rate and MTTF when the test duration is short. On the contrary, long-lasting tests of a smaller set of samples can lead to more realistic results. This is because after a certain time, the number of failures due to BTI-imposed degradation do not increase significantly, but rather remain constant. Hence, in order to estimate the overall FIT rate and MTTF of a circuit due to BTI, long-lasting test durations over the entire lifetime of the circuit should be preferred.

Moreover, it can be concluded that dynamic voltage scaling can partially mitigate the BTI effect and lead to an increase of the functional yield of the circuit. Although, BTI degradation deteriorates from this increase of the bias stress, it is counteracted by the operating enhancement dvs provides. The number of failures due to BTI decrease considerably which results to an overall rise of the MTTF, stretching the expected operating time before failure.

CHAPTER 5

Conclusions

5.1 Overall

The BTI effect is the aging phenomenon that results in an absolute increase of the threshold voltage of FET transistors. This shift of the threshold voltage depends on the temperature and voltage stress and leads to a degradation of the drain current and the transconductance. The downscaling of device dimensions, has aggravated this degradation, increasing the frequency of system failures. Thus, the BTI effect has become a serious reliability issue in modern integrated circuits, which is why explaining and accurately modeling this phenomenon is necessary. Over the last 30 years many industrial tools have been developed in order to capture BTI and mitigate the imposed degradation. On the one hand, some tools can provide efficient results but their estimations seem to be crude when simulating small-dimensioned devices. On the other hand, other approaches that accurately simulate the BTI degradation require a prohibitive CPU processing time, rendering them inadequate for long-lasting tests.

The purpose of this thesis was to discuss a simulation framework that is able to accurately and efficiently simulate the BTI effect in complex circuits and over extended strides of circuit lifetime. Through this simulation framework the functional yield of a part of an embedded SRAM memory under BTI stress was estimated, along with some other important reliability metrics, such as defects per million, mean time to failure and failures in time. This simulation framework, was composed by some existing tools that were found in previous work, while some of these tools were enhanced, in order to achieve more accurate estimations.

The first part of this thesis elaborated on the two existing models that try to explain the BTI effect: the RD model and the atomistic model. The RD model claims that the aging effect originates from the breaking of the Si-H bonds at the substrate and from creation of new bonds between the free Si atoms and minority carriers. The atomistic model is more recent and is based on the stochastic nature of defects created in the substrate. Each defect is characterized by some unique time constants that define whether the defect is occupied and adds to the threshold shift or not. This atomistic model seems to capture the BTI effect in a more accurate way.

Some existing tools that can perform reliability analysis and focus on simulating the BTI effect were also mentioned. Most of them follow the Reaction-Diffusion approach hence, discussing BTI modeling tools that are based on the atomistic model is a timely contribution of this thesis. In addition, the definitions of functional yield, defects per million, failures in time and mean time to failure, metrics that are essential in reliability analysis, were reviewed.

Secondly, a simulation framework that follows the atomistic model was discussed. Some of the existing tools composing this model were presented and the function of each one of them was briefly explained. The `wavapprox` component, that transforms a SPICE waveform to a CDW signal representation, the `definit` tool, that populates all devices with defects and the `propagate` tool, that estimates the duty factor of all the nodes of a netlist are the most important. For this simulation framework, some of these tools were improved while the extended algorithm in the `propagate` component was thoroughly enhanced. The superiority of the newer version over the older version of this tool was proven as well with histograms showing the mean absolute error between the two versions of the tool and a thorough SPICE estimation. Next, the instantiation of the proposed simulation framework took place, by combining existing and enhanced components into a single flow. For the tool to estimate the reliability metrics discussed in the previous Chapter, the mathematical formulation connecting them was also presented.

In the last part of this thesis, a part of an embedded SRAM memory under BTI stress was simulated by the respective framework and the quantitative results of the estimation of the functional Yield, DPM, FIT rate and MTTF of the circuit were demonstrated. A number of 150 samples of the netlist were simulated while the samples were simulated for roughly three years of operation. As it can be seen from the graphs, the BTI effect manifests at an early stage and does not deteriorate significantly after a certain time step. The main reason behind this is that the defects responsible for the threshold voltage shift have time scales distributed across a logarithmic time scale thus, the changing of the states of the defects occurs at a logarithmic time scale as well. Moreover the dvs technique can considerably mitigate BTI and leads to an increase of the circuit's functional yield.

5.2 Future Work

After discussing the instantiation of the proposed simulation framework, few steps have to be considered in order to enhance and improve this simulating tool. The current simulation framework efficiently estimates reliability metrics of a circuit under BTI stress, assuming of course that circuits are based on complementary MOS (CMOS) technology. CMOS is the dominant technology of IC nowadays nevertheless, this trend is expected to change in the near future [36]. Hence, improving this simulation framework to estimate reliability metrics and capture BTI on circuits of different technologies might be useful.

Moreover, reliability concerns on downscaled devices are not solely attributed to BTI degradation but to other gate oxide failure modes as well [23]. The HCI effect is another important degradation mode that has to be taken into account when performing a reliability analysis [6]. The gate oxide wearout is also a serious reliability threat that can lead to increased current noise and cause system failure [23].

In the following Section, two topics for improving the proposed simulation framework are discussed: updating the simulation framework to capture BTI on FinFET technology and enhancing this simulating tool to account for both BTI and HCI effects.

5.2.1 Improving the Simulation Framework to Capture BTI on FinFET Technology

The demand for large integration of transistors has led industry to follow a downscaling trend for the creation of devices. This resulted in a decrease of the device dimensions to nearly 20nm. However,

further shrinking the transistors has become rather difficult because MOSFETs with dimensions lower than this limit would suffer from short channel effects and increased leakage currents [37]. Hence, in order to meet the demand for larger integration and the need for reduced power per area product, the FinFET technology has been proposed that would overcome the difficulties limiting the device downscaling [38].

The idea behind this technology is that the gate of the device would be wrapped around the channel in contrast to the planar FET technology that places the gate above the channel. This 3-D built of the gate is expected to result in a better control of the channel and lead to a dramatic decrease in leakage and dynamic power consumption enabling higher integration [37]. In fact EDA companies like Synopsys have already started to develop tools modeling this technology [37] while the industry has already manufactured processors based on this technology [39].

However it is expected that BTI would continue to pose a reliability threat on devices of this technology and specifically, relevant work [40] suggests that the threshold voltage shift due to NBTI degradation would deteriorate in FinFET devices. Hence, the necessity for providing a simulation framework that accurately and efficiently models BTI on FinFET devices is understandable. Thus an extension of this thesis is the enhancement of the simulation framework in order to capture the aging effect on FinFETs and allow reliability simulations on systems composed by such devices.

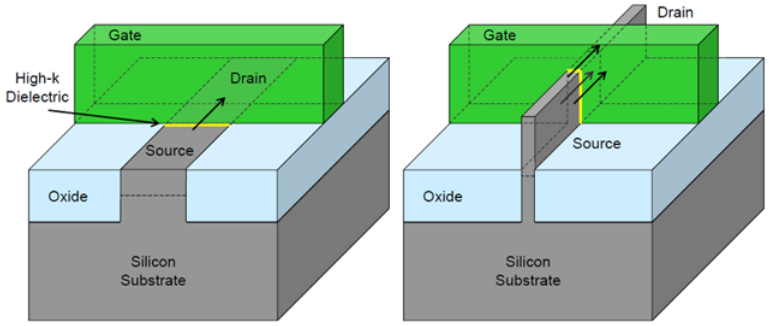

Figure 5.1: A planar device vs a FinFET device as shown in [5]

The atomistic approach that this simulation framework is based on, may be different from an atomistic approach for the FinFET technology thus, the simulation framework may also need to be enhanced, so that it can capture BTI on both planar FET and FinFET devices. Also, an update of the SPICE-free signal activity propagating tool may also be important in order to capture the propagation of the signal activity from one connection of the FinFET device to other and enable long-lasting BTI simulations on a pseudo-transient way, as achieved by this simulation framework.

5.2.2 Enhancing the Simulation Framework to Account for both BTI and HCI

Another important degradation mode that threatens circuit reliability is the hot carrier injection (HCI). This effect has become a dominant reliability issue in FET devices (and especially in NFETs), thus a detailed reliability analysis must account from both BTI and HCI effects [41]. The physical mechanism responsible for the HCI wearout is the trapping of hot electrons in the gate oxide that results in a voltage threshold shift of the device and in a degradation of transconductance and drain current [6].

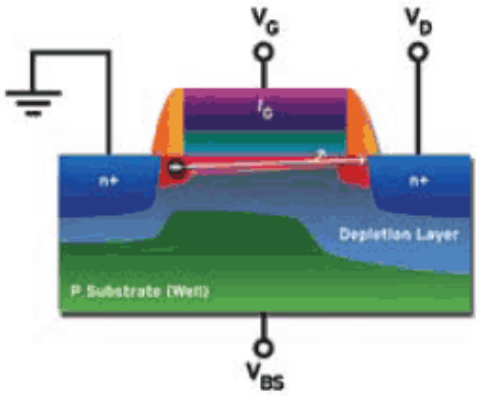

Figure 5.2: The physical mechanism of HCI as presented in [6]

According to the theory, due to the high electric fields near the drain area of the NFET, some electrons gain enough kinetic energy and are accelerated into the gate oxide or even few times to the gate area [6]. The hot electrons that tunnel into the gate are responsible for a leakage gate current, that is estimated to be 10^{-15} A and can even reach a value of 10^{-12} A [42]. Some of the hot electrons that pass into the gate oxide remain trapped thus leading to HCI degradation. This degradation can obviously limit the lifetime operation of an integrated circuit. The mechanism of HCI is shown in Figure 5.2.

Therefore, it seems important to perform reliability tests that focus on both BTI and HCI phenomena. The updating of the simulation framework instantiated in this thesis to account also for the HCI effect is a valid extension. In order to achieve that, new tools must be developed, based on models that accurately and effectively capture the HCI effect. Simulation approaches already exist [41], [3] and are based on the RD theory thus, a tool that simulates both degradation factors and models BTI based on an atomistic approach could be rather useful.

Bibliography

[1] R. Wittmann, "Nbti reliability analysis, reaction-diffusion model." Miniaturization Problems in CMOS Technology: Investigation of Doping Profiles and Reliability www.iue.tuwien.ac.at/phd/wittmann/node10.html.

[2] B. Kaczer, V. V. de Almeida Camargo, M. Toledano-Luque, P. J. Roussel, T. Grasser, F. Catthoor, P. Dobrovolny, P. Zuber, G. Wirth, and G. Groeseneken, "Atomistic approach to variability of bias-temperature instability in circuit simulations," in *Reliability Physics Symposium (IRPS), 2011 IEEE International*, pp. XT.3.1 – XT.3.5, 2011.

[3] C. D. Systems, "Virtuoso ultrasim simulator user guide." White Paper http://lost-contact.mit.edu/afs/rose-hulman.edu/cadence-0910/IC610/doc/UltraSim_User/UltraSim_Chap11.html#64747.

[4] S. Cosemans, W. Dehaene, and F. Catthoor, "A 3.6 pj/access 480 mhz, 128 kb on-chip sram with 850 mhz boost mode in 90 nm cmos with tunable sense amplifiers," *Solid-State Circuits, IEEE*, pp. 2065–2077, July 2009.

[5] D. Kanter, "Intel's 22nm tri-gate transistors." http://www.realworldtech.com/intel-22nm-finfet/, May 2011.

[6] M. S. et al, "Reliabililty-simulation environment tackles lsi design." http://chipdesignmag.com/display.php?articleId=1450.

[7] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," *IEEE, Transactions on Dependable and Secure Computing*, vol. 1, no. 1, pp. 11–33, 2004.

[8] D. Rodopoulos, P. Weckx, M. Noltsis, F. Catthoor, and D. Soudris, "Atomistic pseudo-transient bti simulation with inherent workload memory," *IEEE, Transactions on Device and Materials Reliability*, pp. 704 – 714, June 2014.

[9] D. K. Schroder, "Negative bias temperature instability: What do we understand," *Microelectronics Reliability*, vol. 47, no. 6, pp. 841–852, 2007.

[10] M. A. Alam, "On the reliability of micro-electronic devices: An introductory lecture on negative bias temperature instability," 2005.

[11] B. Paul, K. Kang, H. Kufluoglu, M. Alam, and K. Roy, "Temporal performance degradation under nbti: Estimation and design for improved reliability of nanoscale circuits," in *Design, Automation and Test in Europe 2006. DATE '06*, vol. 1, pp. 1–6, 2006.

[12] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P.-J. Wagner, F. Schanovsky, J. Franco, M. T. Luque, and M. Nelhiebel, "The paradigm shift in understanding the bias temperature instability: From reaction—diffusion to switching oxide traps," *IEEE, Transactions on Electron Devices*, vol. 58, no. 11, pp. 3652–3666, 2011.

[13] K. HUNG, P. K. KO, C. 1lU, and Y. C. CHENG, "Random telegraph noise of deep-submicrometer mosfets," vol. 11, 1990.

[14] D. Rodopoulos, S. B. Mahato, V. V. de Almeida Camargo, B. Kaczer, F. Catthoor, S. Cosemans, G. Groeseneken, A. Papanikolaou, and D. Soudris, "Time and workload dependent device variability in circuit simulations," *IC Design Technology (ICICDT), 2011 IEEE International Conference Review of Scientific Instruments*, pp. 1–4, May 2001.

[15] M. Miranda, "The threat of semiconductor variability." IEEE SPECTRUM, June 2012.

[16] S. R. Welke, B. W. Johnson, and J. H. Aylor, "Reliability modeling of hardware/software systems," *IEEE, Transactions on Reliability*, vol. 44, no. 3, pp. 413–418, 1995.

[17] T. McConaghy, K. Breen, J. Dyck, and A. Gupta, *Variation – Aware Design of Custom Integrated Circuits: A Hands–on Field Guide*. Springer, 2013.

[18] F. N. Najm, N. Menezes, and I. A. Ferzli, "A yield model for integrated circuits and its application to statistical timing analysis," *IEEE, Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, pp. 574 – 591, March 2007.

[19] S. V. Shinde, "Intel–22nm squelch yield analysis and optimization," *International MultiConference of Engineers and Computer Scientists*, vol. II, March 2014.

[20] "Vishay semiconductors, reliability." `http://www.vishay.com/docs/80116/80116.pdf`, February 2002.

[21] Z. Semiconductors, "Reliability handbook." `http://diodes.com/pdfs/handbook.pdf`, June 2004.

[22] E. R. Hnatek, *Practical Reliability Of Electronic Equipment And Products*. October 2002.

[23] S. Mukherjee, *Architecture Design for Soft Errors*. 2008.

[24] "Eldo user's manual." `http://www.engr.uky.edu/~elias/tutorials/Eldo/eldo_ur.pdf`.

[25] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "Nbti-aware synthesis of digital circuits," pp. 370 – 375, 2007.

[26] B. C. Paul, K. Kang, H. Kufluoglu, M. A. Alam, and K. Roy, "Negative bias temperature instability: Estimation and design for improved reliability of nanoscale circuits," *IEEE, Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 743 – 751, April 2007.

[27] D. Rodopoulos, "Optimizing bti/rtn modeling: Technical details and extensions," April 2013.

[28] E. Mintarno, J. Skaf, R. Zheng, J. B. Velamala, Y. Cao, S. Boyd, R. W. Dutton, and S. Mitra, "Self-tuning for maximized lifetime energy-efficiency in the presence of circuit aging," *IEEE Transcactions on Computer-Aided Design for Integrated Circuits and Systems*, vol. 30, pp. 760–773, May 2011.

[29] P. Nenzi and H. Vogt, *Ngspice Users Manual*. January 2013.

[30] L. Grigorios, R. Dimitrios, P. Antonis, and S. Dimitrios, "Hypervised transient spice simulations of large netlists and workloads on multi-processor systems," March 2013.

[31] R. Dimitrios, S. Dimitrios, L. Grigorios, S. Dimitrios, and F. Catthoor, "Understanding timing impact of bti/rtn with massively threaded atomistic transient simulations," March 2014.

[32] D. Zoni and W. Fornaciari, “Sensor-wise methodology to face nbti stress of noc buffers,” 2013.

[33] R. S. Corporation, “Calculating mttf when you have zero failures.” Technical Brief http://microblog.routed.net/wp-content/uploads/2006/08/calculating_mttf_with_zero_failures.pdf.

[34] X. Chen, Y. Wang, Y. Cao, Y. Ma, and H. Yang, “Variation-aware supply voltage assignment for minimizing circuit degradation and leakage,” August 2009.

[35] T.-B. Chan, J. Sartori, P. Gupta, and R. Kumar, “On the efficacy of nbti mitigation techniques,” pp. 1–6, 2011.

[36] S. T., H. J.A., T.-J. King, W. H.-S.P., and B. F, “The end of cmos scaling: toward the introduction of new materials and structural changes to improve mosfet performance,” *Circuits and Devices Magazine, IEEE*, vol. 21, no. 1, pp. 16–16, 2005.

[37] Synopsis, “Finfet technology – understanding and productizing a new transistor.” White Paper http://www.fortronic.it/user/file/A%26VElettronica/tsmc_snps_finfet_wp.pdf, April 2013.

[38] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, J. Bokor, and C. Hu, “Finfet—a self-aligned double-gate mosfet scalable to 20 nm,” *IEEE, TRANSACTIONS ON ELECTRON DEVICES*, vol. 47, pp. 2320 – 2326, December 2000.

[39] A. C. Community, “First tape-out with tsmc’s 16nm finfet and arm’s 64-bit big.little processors,” February 2014.

[40] K. T. Lee, W. Kang, E.-A. Chung, G. Kim, H. Shim, H. Lee, H. Kim, M. Choe, N.-I. Lee, A. Patel, J. Park, and J. Park, “Technology scaling on high-k and metal-gate finfet bti reliability,” pp. 1–4, 2013.

[41] H. Kufluoglu and M. A. Alam, “A unified modeling of nbti and hot carrier injection for mosfet reliability,” 2004.

[42] D. Neamen, *An Introduction to Semiconductor Devices*. 2006.