Ricardo Reis · Yu Cao
Gilson Wirth   *Editors*

Circuit Design for Reliability

Springer

Circuit Design for Reliability

Ricardo Reis • Yu Cao • Gilson Wirth
Editors

Circuit Design for Reliability

Springer

Editors
Ricardo Reis
Instituto de Informática
Universidade Federal do Rio
    Grande do Sul (UFRGS)
Porto Alegre, Rio Grande do Sul, Brazil

Gilson Wirth
Electrical Engineering Department
Universidade Federal do Rio
    Grande do Sul (UFRGS)
Porto Alegre
Rio Grande do Sul
Brazil

Yu Cao
School of ECEE
Arizona State University
Tempe, USA

Printed on acid-free paper

Contents

Chapter 1
Introduction

Ricardo Reis, Yu Cao, and Gilson Wirth

The scaling of CMOS technology to the nanometer regime inevitability increases reliability concerns, profoundly impacting all aspects of circuit performance and posing a fundamental challenge to future IC design. These reliability concerns arise from many different sources, and become more severe with continuous scaling.

VLSI design in the late CMOS era is then driven by an increasing challenge to cope with unreliable components at the device and circuit levels. Early approaches focused only on technology improvement, but ignoring these effects in the design process causes an excessive amount of over-margining, with negative impact in both cost and performance of the final product. At the device and circuit levels, physical understanding, modeling, detection, and successful design techniques for leading reliability mechanisms are vitally important, not only to current robust design practice, but also to the prediction and management of system reliability over the full product life cycle. It is critical to understand, simulate, and mitigate their impact during the stage of physical and circuit design. This book presents physical understanding, modeling and simulation, on-chip characterization, layout solutions, and design techniques that are effective to enhance the reliability of various circuit units.

We start by addressing the physical mechanisms of reliability, focusing on aging and noise. Aging is among the critical reliability issues facing present and future deeply downscaled CMOS devices, and the major concern is the so-called Bias Temperature Instability (BTI). The elementary definitions and experimental

R. Reis • G. Wirth
Electrical Engineering Department, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil

Y. Cao (✉)
School of ECEE, Arizona State University, Tempe, AZ, USA
e-mail: ycao@asu.edu

observations of BTI are first briefly reviewed. Afterwards it is shown how the understanding of gate oxide defect properties can be used to explain BTI and its variability, and possible technological solutions for both nFETs and pFETs are discussed. The charge trapping phenomena that plays a dominant role in BTI also dominates low-frequency noise. A modeling and circuit simulation approach that focuses on operation conditions relevant for digital and analog design, including large signal AC operation, is presented. The modeling and simulation approach aims at helping circuit designers to cope with this reliability issue.

Next, the modeling and characterization is addressed. Any modeling effort needs a solid base to be built on. Models are based on accurate measurement methodologies, which contribute to understanding the mechanisms behind the device degradation. In this book, authors survey characterization methods for the purpose of device modeling that includes the effects of temporal parameter degradation, such as BTI and noise. Particular focus is placed on the approaches that utilize on-chip circuits to enhance the accuracy and fidelity of the measurements. Modeling must focus on real circuit operation conditions, such as Dynamic Voltage Scaling (DVS), or biasing found in mixed signal designs where the stress waveform is much more random. Modeling solutions suitable to allow aging simulation under all dynamic stress conditions are presented.

Variability in electrical parameters, such as variability in threshold voltage, is also known to be a major reliability concern. Among the sources of variability is random dopant fluctuation (RDF). We discuss how RDF can induce significant variation in saturation (on) current and transconductance degradation—two key metrics for benchmark performance of digital and analog integrated circuits.

An analysis of ongoing scaling related trends of integrated circuits is also presented. For this purpose, authors discuss major trends observed in process technology, as well as in memory and logic designs. While newer technology styles, such as FinFETs provide necessary advances to enable further technology scaling, we can observe that there exists also a trend for reduced resilience with ongoing scaling. Our hope is that with the roadmap presented in this book, which concentrates on FinFET-based designs, we are able to sensitize the research community to present and near future problems associated with this technology. Additionally, different architectures for SRAM memories are analyzed, to demonstrate the different design choices available and tradeoffs that have to be made.

Finally, the last chapters present design techniques for reliability. FinFETs have emerged as alternatives to conventional bulk MOSFETs in scaled technologies due to superior gate control of the channel, lower short channel effects and higher scalability. However, width quantization in FinFETs constrains the design space of FinFET-based circuits, especially SRAMs in which transistor sizing is critical for the circuit robustness. The adverse effects of width quantization can be mitigated by appropriate device-circuit co-design of FinFET-based memories. Some of such techniques are described, with an emphasis on the device-circuit interactions associated with each methodology. The impact of different technology

options in FinFETs like gate-underlap, fin orientation, fin height, gate workfunction and independent control of the gates on the stability, power and performance of 6T SRAMs is discussed.

Electromigration is an aging effect that affects metal wires (interconnections). In this book, authors briefly introduce physical foundations of electromigration (EM). We discuss physical parameters affecting EM wire lifetime and we introduce some background related to the existing EM physical simulators. In the work presented in this book for EM physical simulation, authors adopt the atomic concentration balance-based model. We discuss the simulation setup and results. We present VEMA, a variation-aware electromigration (EM) analysis tool for power grid wires. The tool considers process variations caused by the chemical–mechanical polishing (CMP) and edge placement error (EPE). VEMA is a full-chip EM analysis tool. It performs EM lifetime calculation, and analyzes process variation effects on EM reliability and reports variation tolerances of EM-sensitive power grid wires.

Integrated circuits subject to ionizing radiation are sensitive to transient faults caused by the interaction of ionizing particles with the semiconductor. At ground level, the main source of radiation are neutrons that interact with the material generating secondary particles that can ionize the silicon, provoking transient upset in circuits fabricated in nanometer technology. The interaction of the ionizing radiation with the transistor may provoke transient and permanent effects. Different fault tolerance techniques can be applied to FPGAs according to their type of configuration technology, architecture and target operating environment. A chapter is dedicated to present a set of fault mitigation techniques for SRAM, FLASH and ANTIFUSE-based FPGAs. A test methodology to characterize those FPGA under radiation is also presented. Results from neutron-induced faults are presented and discussed.

The relevance of power consumption and leakage currents is also increasing. Although several Design-for-Low-Power and Design-for-Variability options are already available in modern EDA suites, the contrasting nature of the two metrics makes their integration extremely challenging. Most of the approaches used to compensate and/or mitigate circuit variability (e.g., Dynamic Voltage Scaling and Adaptive Body Biasing) are, in fact, intrinsically power inefficient, as they exploit the concept of redundancy, which is known to originate power overhead. In this book, we introduce possible solutions for concurrent leakage minimization and variability compensation. More specifically, authors propose Power-Gating as a mean for simultaneously controlling static power consumption and mitigating the effects induced by two of the most insidious sources of variability, namely, electrical parameter variations and Aging due to BTI.

High-performance clock network design has been a challenge for many years due to the drastically increasing effect of process variability. In addition, tight power budgets have lowered supply voltage levels, which make designs more sensitive to noise. Together, variability and noise present a colossal challenge to clock designers in order to meet timing, yield, and power simultaneously. This chapter discusses the different strategies that designers use to ameliorate variability and noise problems in clock network design.

The project of this book was born during the successful IEEE Circuits and Systems Society Summer School on Physical Design of Reliable Circuits that was held in Porto Alegre, Brazil, from January 12 to 15, 2010. Some chapters of the book are based on the talks realized during the school and other well-known invited authors wrote other chapters.

It is our hope that this book will provide the vitally important physical understanding and design techniques for state of the art and future technologies, ranging from technology modeling, fault detection and analysis, circuit hardening, and reliability management in different circuit units.

Chapter 2
Recent Trends in Bias Temperature Instability

**B. Kaczer, T. Grasser, J. Franco, M. Toledano-Luque, J. Roussel, M. Cho,
E. Simoen, and G. Groeseneken**

Abstract   The paradigm shifts occurring in the past few years in our understanding
of BTI are reviewed. Among the most significant ones is the shift from perceiving
NBTI in terms of the Reaction-Diffusion model to analyzing BTI with the tools
originally developed for describing low-frequency noise. This includes the interpre-
tation of the time, temperature, voltage, and duty cycle dependences of BTI. It is
further demonstrated that a wealth of information about defect properties can be
obtained from deeply-scaled devices, and that this information can allow projection
of variability issues of future deeply downscaled CMOS devices. The chapter is
concluded by showing the most promising technological solutions to alleviate both
PBTI and NBTI.

1   Introduction

Among the critical reliability issues facing present and future deeply downscaled
CMOS devices is the so-called Bias Temperature Instability (BTI). While BTI in
nFET devices was generally ascribed to charge trapping in the high-k portion of
the gate oxide, the interpretation of BTI in pFET devices still generates controversy
[1–7]. This phenomenon in pFET devices has been previously described by
diffusion of hydrogen from and back into substrate/gate oxide interface states [8].
The so-called Reaction–Diffusion model based on this assumption is still popular,
especially in the design community [9], despite being inconsistent with some crucial
observations [10].

In this chapter, the elementary definitions and experimental observations of BTI
are first briefly reviewed. One of the most intriguing properties of BTI—the lack

---

B. Kaczer (✉) • J. Franco • M. Toledano-Luque • J. Roussel • M. Cho
• E. Simoen • G. Groeseneken
imec, Leuven, Belgium
e-mail: kaczer@imec.be

T. Grasser
TU Wien, Wien, Austria

G. Groeseneken
KU, Leuven, Belgium

of characteristic time scale, especially in pFETs—is then argued to point to a *dispersion* in the underlying mechanism [5, 7, 11]. In CMOS technologies, a response on many time scales is typical for low-frequency noise [and its manifestation as Random Telegraph Noise (RTN) in deeply-scaled devices], suggesting that (the principal component of) BTI is in fact caused by the same defects [12, 13]. The link between BTI and low-frequency noise is then further developed—it is shown that many properties of gate oxide defects can be directly extracted from BTI relaxation measurements in deeply-scaled devices [14] and the noise-inspired model capable to fully describe these properties is reviewed. Afterwards it is shown how the understanding of gate oxide defect properties can be used to explain variability of BTI in deeply-scaled technologies, as well as possible technological solutions for both nFETs and pFETs.

2   Brief Overview of BTI

BTI is a consequence of charging of defect states in the gate oxide and at its interface [2]. The defects could be both pre-existing or generated during device operation. The trapped charge results in a shift of the device parameters, such as its threshold voltage V_{th}, channel mobility, transconductance, and subthreshold slope, and generally a decrease of the FET's drive current. The name is derived from the phenomenon being strongly accelerated by *temperature T* and gate *bias V_G*. BTI in n-channel FET devices, which are typically biased in circuits at positive V_G, is referred to as Positive BTI (PBTI), while negative BTI (NBTI) takes place in p-channel FETs. Constant V_G stress bias is often referred to as static, or "DC" BTI, while periodically interrupted V_G stress is called "AC", or dynamic BTI.

3   Static BTI

Figure 2.1 illustrates the typical gradual shift of pFET threshold voltage ΔV_{th} during accelerated stress at elevated T [5]. The stress data are typically measured at several V_G's and ΔV_{th} is extrapolated to 10 years at the circuit operating voltage V_{DD} (or $V_{DD} + 10\ \%$). The extrapolated ΔV_{th} must be below a given value (typically 30 or 50 mV) for the technology to qualify.

This simple extrapolation procedure is, however, complicated by ΔV_{th} decreasing *immediately* after the stress bias is removed, as illustrated in Fig. 2.1 [15]. As will be discussed henceforth, this *recovery*, or *relaxation*, component R typically proceeds simultaneously on many time scales, making it difficult to determine its beginning or end and thus separating it from the final *non-recoverable*, or *permanent*, component P [2, 6]. This ΔV_{th} relaxation is thus a crucial problem for BTI measurement, interpretation, and extrapolation.

Fig. 2.1 Shift in pFET threshold voltage is observed during negative gate bias stress. When the stress bias is removed, a recovery of the effect is seen (note: $V_{th} \sim 0$ V for this device). Inset illustrates the biases applied at FET terminals during the BTI measurement
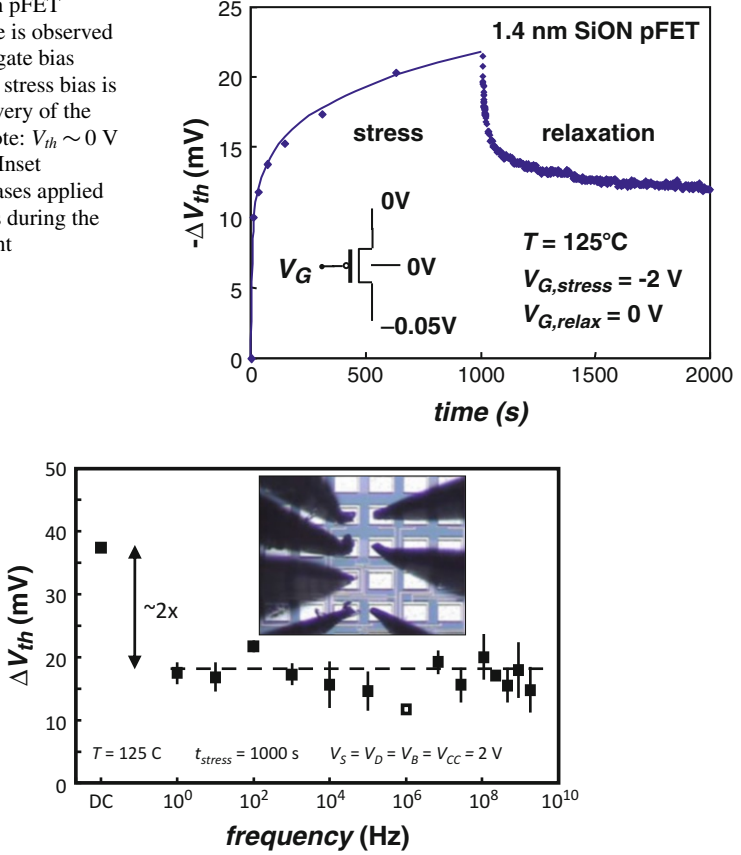
Fig. 2.2 The V_{th} shifts due to 50 % AC unipolar NBTI stress in pFETs are seen independent of frequency in the entire frequency range of 1 Hz–2 GHz. The V_{th} shift of the corresponding DC NBTI stress obtained on an identical device is shown for comparison. Inset: Micrograph of the *on-chip* circuit for the DC and AC BTI measurements consisting of a ring oscillator, a frequency divider, a buffer, a pass-gate-based multiplexer, and the device under test [17]

4 Dynamic BTI

In many CMOS applications, such as logic, the majority of the FETs are constantly switched and thus exposed to *dynamic* stress [16]. Figure 2.2 documents that NBTI is present at frequencies up to the GHz range, i.e., there does not appear to be any "cut-off" time constant of the degradation mechanism above $\sim$1 ns [17]. Furthermore, the AC bias signal reduces BTI with respect to the DC stress. This provides some additional reliability margin, which can be factored in during the application design phase [9].

Fig. 2.3 Total degradation ΔV_{th} after 6000 s of unipolar NBTI stress shows a distinctive dependence on the duty factor DF. In particular, a weak dependence, or a "plateau" between ~10 and ~90 % is observed, complemented by rapid ΔV_{th} increase for the outermost DF values. Data at different relaxation times are shown
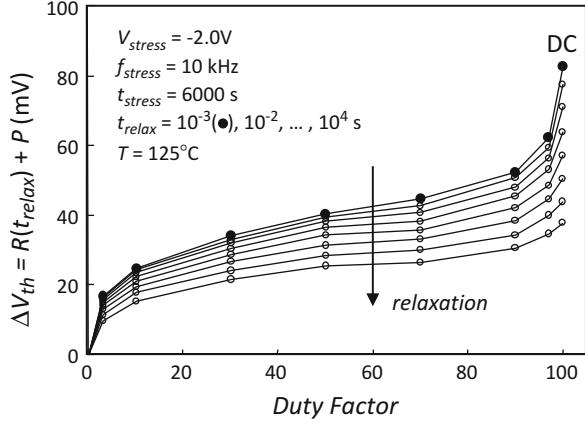
In an arbitrary FET of an arbitrary digital circuit, the average probability of a signal being high can vary between 0 and 100 %. The dependence of BTI on the *duty cycle* (called *duty factor* or *DF* here) thus needs to be studied. A NBTI ΔV_{th}-*DF* dependence with an inflection point around *DF* ~50 %, first reported in [17], is shown in Fig. 2.3 [12].

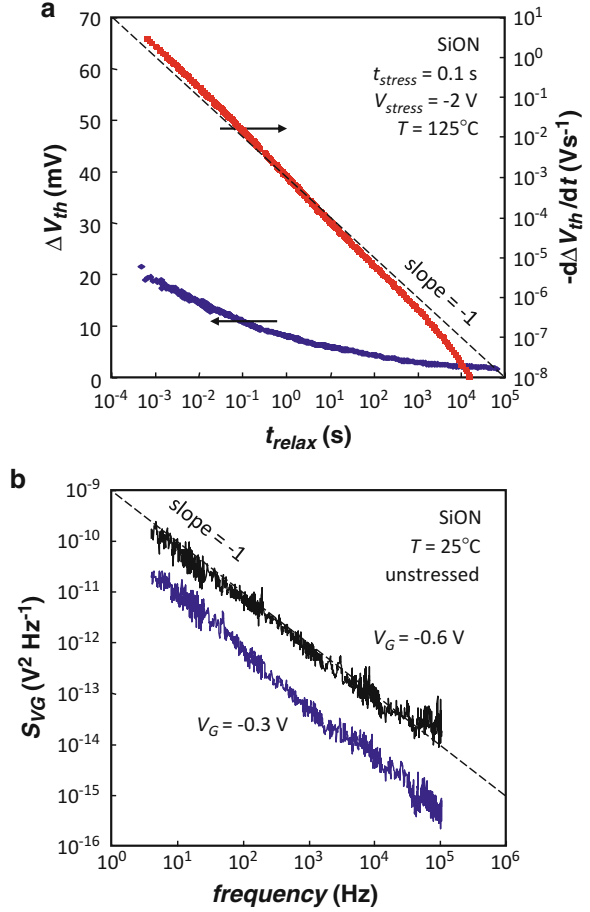## 5   Similarity Between BTI Relaxation and Low-Frequency Noise

Long, log(t)-like behavior of ΔV_{th} without a characteristic time scale is typically observed in both the initial portion of NBTI degradation [13, 18] and the recovery phase. Figure 2.4 illustrates that the *rate of degradation* dΔV_{th}/dt_{relax} [7] extracted from the log(t_{relax})-like ΔV_{th} NBTI relaxation transient after even a *very short*, 0.1 s stress, follows $1/t_{relax}$ for over seven decades. Such behavior is a signature of states with discharging time constants covering as many decades [19].

Incidentally, *superposition of states with widely distributed time scales* is the standard explanation of the 1/f noise spectra [20], which are clearly observed in our pFETs (Fig. 2.4). This obvious similarity leads us to argue that the same states with widely distributed time scales in fact play a fundamental role in both NBTI and noise measurements.

## 6   Semi-quantitative Model for BTI Relaxation

In order to visualize this common property it is beneficial to consider an equivalent circuit representing states with widely distributed time scales [19]. Note that in either NBTI relaxation or 1/f noise measurements, no maximum or minimum "cut-

off" times are typically observed [12]. For the sake of simplicity it is therefore assumed here that the time constants are log-uniformly distributed from times much shorter than the switching time of a pFET to very long, corresponding to the lifetime of a CMOS application. Such states with widely distributed time scales are then represented by "RC" elements in Fig. 2.5 with the total FET ΔV_{th} being proportional to the sum of voltages ("occupancies") on all capacitors. For the sake of simplicity, it is assumed that all RC elements have the same weight and can be partially occupied, which emulates the behavior of a large-area device. Most properties of the recoverable component can be reproduced when the ohmic resistors in Fig. 2.5 are replaced with a *non-linear* component (simulated by two diodes with different parameters, see Fig. 2.5), which emulates different charging (i.e., capture) and discharging (i.e., emission) time constants of each defect [21]. Such a circuit correctly reproduces *DF* (Fig. 2.6, cf. Fig. 2.3) and also the log-like relaxation (Fig. 2.4a) and the log-like initial phase of stress (not shown) [19].
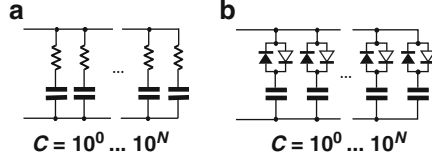
Fig. 2.5 (**a**) An equivalent circuit with exponentially increasing capacitances used to emulate defect states with widely distributed time scales, such as those active in low-frequency noise. (**b**) The same circuit modified to account for charging (i.e., capture) and discharging (i.e., emission) time constants being voltage dependent, represented by asymmetric diodes. The sum of voltages on capacitors is assumed to be proportional to FET ΔV_{th}

Fig. 2.6 The plateau in *DF* dependence of *R* is also qualitatively well reproduced by the equivalent circuit in Fig. 2.5, as is the decrease with increasing relaxation time (cf. Fig. 2.3, which, however, shows the sum of *R* and *P*)
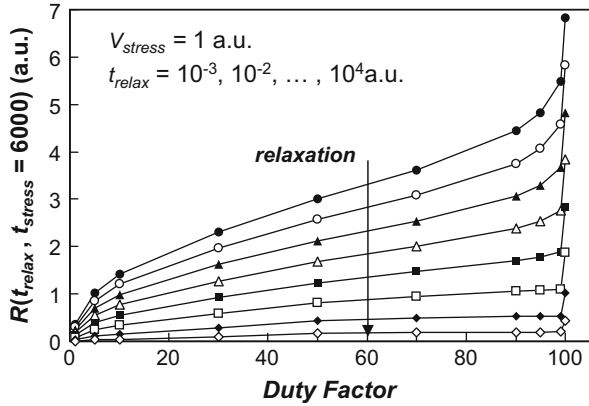
7 Properties of Individual Defects

Figure 2.7 shows two typical ΔV_{th} relaxation transients following positive V_G stress on a single 70×90 nm^2 nMOSFET (i.e., corresponding to PBTI). Conversely to the continuous relaxation curves obtained on large devices, a quantized ΔV_{th} transient is observed in the deeply-scaled devices. In such devices, the relaxation is observed to proceed in discrete voltage steps, with each step corresponding to discharging of a *single* oxide defect [12, 22, 23]. Upon repeated perturbation, each defect shows up in the relaxation trace with a characteristic "fingerprint" consisting of its discharge, or *emission* time, and its voltage step [14].

Figure 2.8 shows the two-dimensional histogram of the heights and the emission times of the steps when the experiment was repeated 70 times at the same stressing and relaxing condition as in Fig. 2.7 [14]. In Fig. 2.8, four clusters are clearly formed that correspond to four active defects in the time window of the experimental setup.

The emission times of each defect are stochastically distributed and follow an exponential distribution. This allows us to determine the average emission time τ_e. The capture time of each trap can be obtained by varying the stress (i.e., charging) time from 240 ms down to 2 ms. The intensity of the cluster decreases with reducing

Fig. 2.7 Characteristic ΔV_{th} transients of a single 70×90 nm^2 1 nm-SiO$_2$/1.8 nm-HfSiO nMOSFET device stressed at 25 °C and $V_G = 2.8$ V for 184 ms. Four discrete drops are observed indicating the existence of four active traps at this stress condition
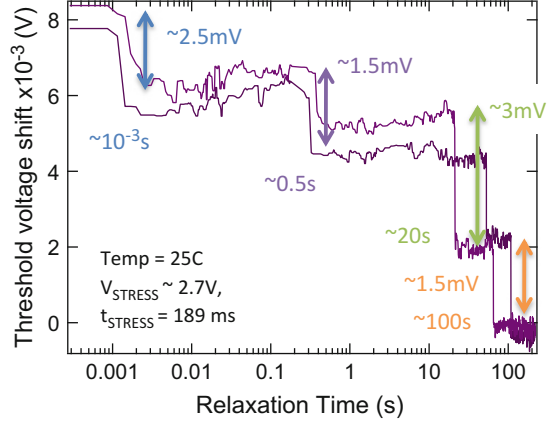
Fig. 2.8 Two-dimensional histograms (TDDS spectra) of the heights and emission times of the steps extracted from 70 ΔV_{th} transients of the particular device of Fig. 2.7 at (**a**) 25 °C and (**b**) 50 °C. Four clusters are formed that shift horizontally to shorter emission times with increasing temperature. Note that trap #3 disappears from the experimental window at 50 °C
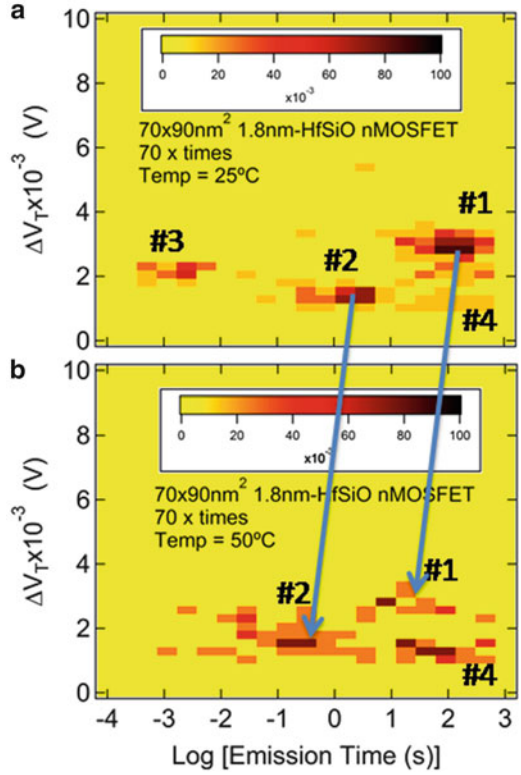
stress time when the characteristic capture time is in the range of the stress time. The fit of the intensity to $P_c = 1 - \exp(-t_{stress}/\tau_c)$ lets us calculate the average capture time τ_c. This technique is known as Time Dependent Defect Spectroscopy (TDDS) [14].

In Fig. 2.8, an identical experiment was repeated at 50 °C on the same device. Note the large horizontal shift of the clusters to shorter emission times with only a 25 °C temperature increase. The Arrhenius plots of the emission and capture times obtained at T from 10 to 50 °C (not shown) provide activation energies of 0.48 eV for emission and 0.25 eV for capture. Similarly thermally activated capture and emission times are also observed in both nFET and pFET (i.e., corresponding to NBTI) with conventional SiO_2 gate oxide [14, 23, 24]. One can therefore conclude for all these cases that *both emission and capture in both electron and hole gate oxide traps are without a doubt thermally activated processes*. This experimental fact is incompatible with direct elastic tunneling theories widely used in different oxide trap characterization techniques and calculations. Consequently, a new model that takes into account this thermal dependence has to be considered.

8 Modeling Properties of Individual Defects

A model of the above-described properties of individual gate oxide defects can be constructed by drawing on the above similarities with low-frequency and Random Telegraph Noise (RTN) [25]. An example of the configuration coordinate diagram of the model is shown in the inset of Fig. 2.9. Four different configurations of the defect are considered [14]. Two of the states are electrically neutral while two of them correspond to the singly positively charged state. In each charge state the defect is represented by a double well, with the first of the two states being the equilibrium state and the other a secondary (metastable) minimum. The time dynamics of the defect can be described by a simple stochastic Markov process. Broadly, transition rates between states involving charge transfer assume (1) tunneling between the substrate and the defect, and (2) nonradiative multiphonon (NMP) theory, which has been often applied to explain RTN [26, 27]. Introduction of the NMP theory naturally explains the temperature dependence of both capture and emission time constants observed in the previous section. The wide distribution of time scales is then readily described by a distribution of the overlaps of the potential wells (i.e., a distribution of "potential barriers") [14].

The crucial extension of the NMP theory is the assumption of the relative position of the potential wells changing with gate bias [14], quite naturally introducing the required strong V_G dependence. As documented in Fig. 2.9, the model successfully describes the bias as well as the temperature dependences of the characteristic time constants. It is also noted that, contrary to techniques for the analysis of RTN, which only allow monitoring the defect behavior in a rather narrow time window, TDDS can be used to study the defects capture and emission times over an extremely wide range.
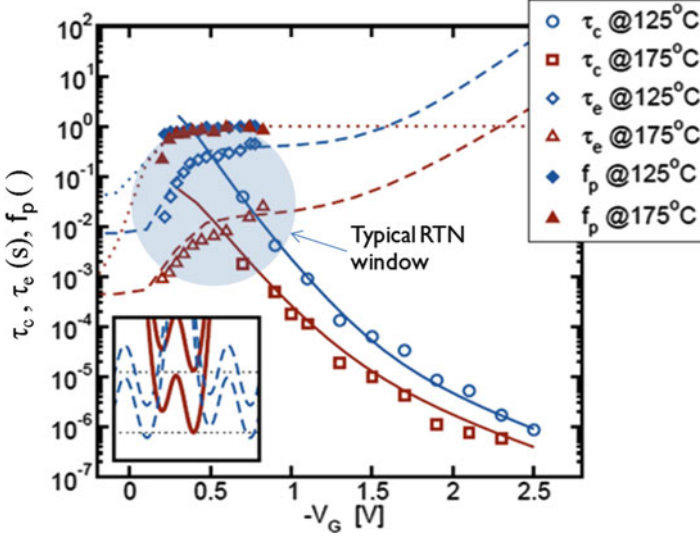
Fig. 2.9 Simulated capture and emission time constants (*lines*) compared with to the experimental TDDS values obtained on SiO$_2$ pFETs (*symbols*) during NBTI stressing at 125 and 175 °C and varying V_G. The experimental occupation probability of the charged state f_p is also indicated. The configuration coordinate diagram is shown in the inset (*dashed line*: neutral defect state; *solid line*: charged state potential)
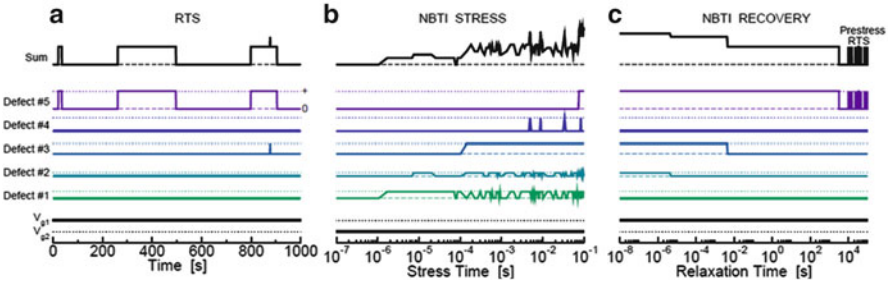
Fig. 2.10 Simulated RTN, stress, and recovery behavior of a nano-scale device using a stochastic solution algorithm of the proposed model. (**a**) At the threshold voltage ($Vg1$), the RTN is dominated by defect #5 with the occasional contribution from defect #3. Defects #1, #2, and #4 remain positively charged within the 'simulation/experimental' window. (**b**) During stress ($Vg2$), the capture times are dramatically reduced by the higher (more negative) gate voltage and the defects #3 and #5 become predominantly positively charged ($\tau c \ll \tau e$). Defects #1, #2, and #4 start producing RTN. (**c**) During recovery (back at $Vg1$), trapped charge is subsequently lost and the dynamic equilibrium behavior is gradually restored

It has been previously argued that the phenomenon called NBTI relaxation in pFET devices is in fact just a different facet of the well-known low-frequency noise in these devices. While the low-frequency noise corresponds to the channel/gate dielectrics system being in the state of dynamic equilibrium, NBTI relaxation corresponds to the perturbed system returning to this equilibrium [28]. Figure 2.10

Fig. 2.11 Histogram of NBTI transient individual step heights measured on 72 devices shows a clear exponential distribution. The average V_{th} shift η corresponding to a single carrier discharge is 4.75 ± 0.30 mV in the pFETs with metallurgic length $L = 35$ nm, width $W = 90$ nm, and HfO$_2$ dielectrics with EOT $= 0.8$ nm
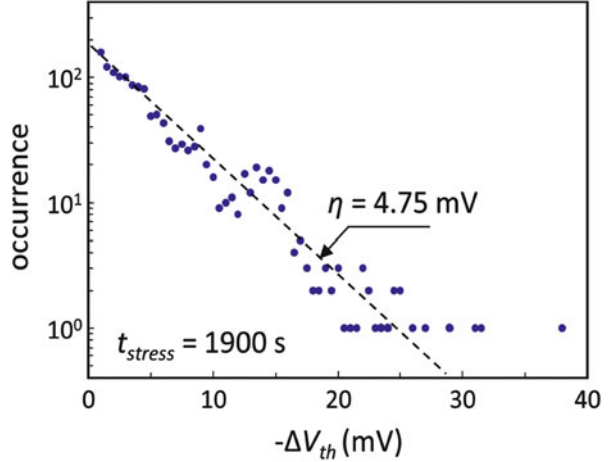
then illustrates this concept on a simulated example of a deeply scaled pFET containing only five active defects [23]. In particular it shows that the same defects can be responsible both for RTN as well as the NBTI relaxation and the (initial phase of) NBTI stress.

9   BTI Distribution in Deeply-Scaled FETs

As CMOS devices scale toward atomic dimensions, device parameters become statistically distributed. Similarly, parameter *shifts* during device operation, once studied in terms of the average value only, will have to be described in terms of their distribution functions. The understanding of the properties of individual defects helps us to explain this distribution. Namely, much like in the case of random telegraph noise (RTN) [29, 30], it is observed the distribution of down-steps ΔV_{th} due to *individual* discharging events to be *exponentially* distributed (Fig. 2.11). The exponential distribution of single-charge ΔV_{th} can be understood if non-uniformities in the pFET channel due to random dopant fluctuations (RDF) are considered [28–30]. A *single* discharging event in many devices routinely exceeded 15 mV, *and in several devices exceeded 30 mV*, the NBTI lifetime criterion presently used by some groups. For comparison, ΔV_{th} of less than 2 mV would be expected based on a simple charge sheet approximation. The large observed step height amplitude is due to the aggressively scaled dimensions of the pFETs used [28, 31].

Assuming the lateral locations of the trapped charges are uncorrelated, the overall ΔV_{th} distribution can be readily expressed as a convolution of individual exponential distributions [28, 31]. An actual population of stressed devices will
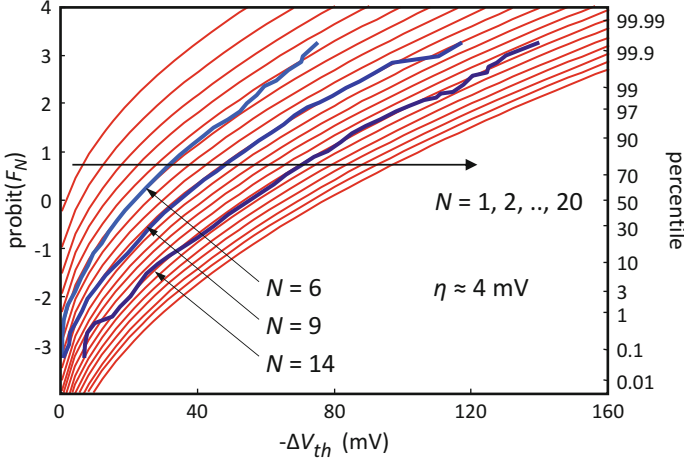
Fig. 2.12 Equation (2.1) in a probit plot rescaled to fit experimental distributions from Fig. 2.10 of Ref. [22], with the corresponding values of N and η readily extracted

consist of devices with a *different* number n of oxide defects in each device, with n being Poisson distributed [12, 22, 28]. The *total* ΔV_{th} distribution can be therefore obtained as

$$F_N(\Delta V_{th}, \eta) = \sum_{n=0}^{\infty} \frac{e^{-N} N^n}{n!} \left[1 - \frac{n\, \Gamma\,(n, \Delta V_{th}/n)}{n!} \right], \qquad (2.1)$$

where N is the mean number of defects in the FET gate oxide and is related to the oxide trap (surface) density N_{ot} as $N = W\,L\,N_{ot}$. The CDF is plotted in Fig. 2.12 for several values of N. For comparison, measured total ΔV_{th} distributions for three different stress times from Ref. [22] are excellently fitted by the derived analytical description.

The advantage of describing the total ΔV_{th} distribution in terms of Eq. (2.1) is its relative simplicity and tangibility of the variables. The analytical description allows, among other things, to calculate NBTI threshold voltage shifts in an unlimited population of devices, a feat practically impossible through device simulations.

10 Technological Solutions

Once the underlying BTI mechanisms are understood, the defect properties can be modified to beneficial ends. Below, two possible technological solutions for both PBTI and NBTI are discussed.
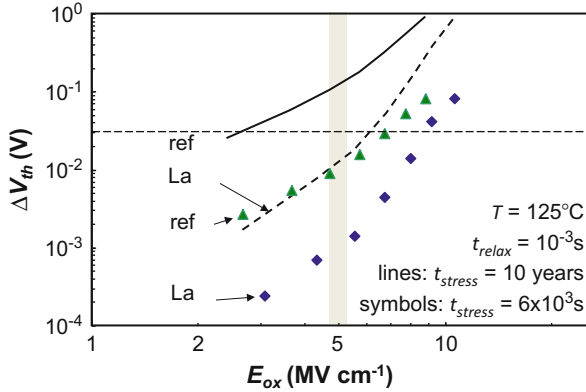
Fig. 2.13 A significant reduction of PBTI threshold voltage shift is observed in planar nFETs with La passivation ("La") over the reference stack ("ref") without passivation. Simplified power-law projection to 10 years shows passivated stack having sufficient reliability ($\Delta V_{th} < 30$ mV) at ~5 MV/cm operating field

11 Improving PBTI with Rare-Earth Incorporation

PBTI was considered a minor problem in technologies based on SiO_2. It arose as a reliability issue when high-k materials were incorporated into the gate stack. However, when rare earths were introduced to adjust the nFET initial threshold voltage, this issue was mitigated, as can be seen in Fig. 2.13. A significant reduction of PBTI is observed in planar nFETs with lanthanum with respect to a lanthanum-free reference [32].

Positive BTI in nFETs with high-k materials like HfO_2 has been linked to oxygen vacancies, which produce a defect level in the upper part of the oxide band gap. Group III elements compensate unpaired electrons around the oxygen vacancy in HfO_2 and the defects are "passivated" by being pushed up toward the conduction band minimum [33]. Such states are not easily accessible to nFET channel electrons, resulting in the significant reduction of negative charge capture in the stack and hence the reduction of PBTI.

12 Improving NBTI in High-Mobility SiGe pFETs

Reduction of gate stack EOT, which is one of the most efficient ways to improve FET performance, enhances NBTI due to increased oxide electric field. As a consequence, 10 year lifetime can be guaranteed for sub-1 nm EOT Si pFETs only at gate overdrive voltages far below the expected operating voltages (Fig. 2.14).

Another way to improve FET performance is the use of high-mobility substrates, such as buried-channel SiGe [34]. Because of the valence band offset between the
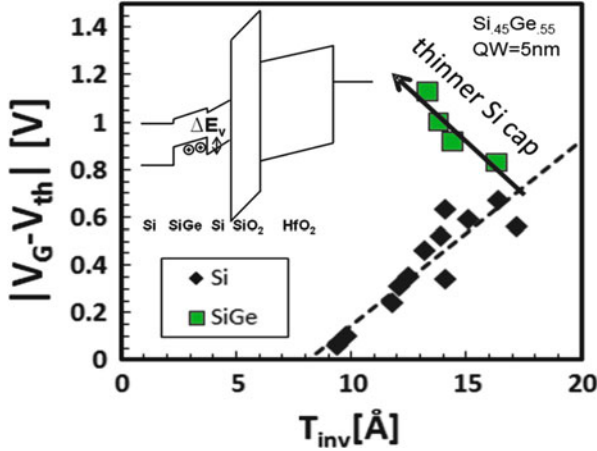
Fig. 2.14 Plot of the operating overdrive voltage $|V_G - V_{th}|$ for 10 year lifetime assuming a 30 mV threshold voltage shift criterion vs. the inversion capacitance equivalent thickness T_{inv}, for Si channel devices with different processing used as a reference, and for SiGe pFETs. For low T_{inv}, Si devices $|V_G - V_{th}|$ is below the expected operating voltage. In contrast to that, optimized SiGe devices show improved lifetime. *Inset*: gate-stack band diagram in inversion. Si cap acts as a barrier ΔE_V for holes

SiGe and the Si cap (see inset of Fig. 2.14), inversion channel holes are confined in the SiGe layer, which therefore acts as a quantum well (QW) for holes. The Si cap lowers the inversion capacitance as compared to the accumulation capacitance. For these devices it is therefore necessary to report the capacitance-equivalent thickness in inversion (T_{inv}, evaluated at $V_G = V_{th} - 0.6$ V) which will be affected by the thickness of the Si cap [35].

As can be seen from Fig. 2.14, SiGe-based device gate stacks significantly increase operating gate overdrive while still guaranteeing 10 year device lifetime and at the moment seem to be the only solution to the NBTI issue for sub-1 nm EOT devices. It has been recently observed that both increasing the Ge content in the channel as well as increasing the SiGe QW thickness reduces NBTI. Most intriguingly, a *reduction* of Si cap thickness also diminishes NBTI [35]. The most likely hypothesis explaining all three trends appears to be the energetic decoupling of the buried channel and the gate oxide defects [36].

Conclusions

In this chapter some of the shifts occurring in the past few years in our understanding of BTI were reviewed. Among the most significant ones is the shift from perceiving NBTI in terms of the Reaction–Diffusion model to analyzing BTI with the tools originally developed for describing low-frequency

(continued)

noise. This includes the interpretation of the time, temperature, voltage, and duty cycle dependences of BTI. It was further demonstrated that a wealth of information about defect properties can be obtained from deeply-scaled devices, and that this information can allow interpretation of variability issues of future deeply downscaled CMOS devices. This theme was complemented by showing the most promising technological solutions to alleviate both PBTI and NBTI.

References

1. J. H. Stathis and S. Zafar, Microelectronics Reliab. **46**, 270 (2006).
2. V. Huard, M. Denais, C. Parthasarathy, *Microelectronics Reliab.* **46**, 1 (2006).
3. D. K. Schroder, Microelectronics Reliab. **47**, 841 (2007).
4. B. Kaczer, R. Degraeve, V. Arkhipov, N. Collaert, G. Groeseneken, M. Goodwin, as discussed at SISC, San Diego, CA, 2006.
5. B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, *Proc. Int. Reliab. Phys. Symp.*, 381 (2005); *Appl. Phys. Lett.* **86**, 143506 (2005).
6. T. Grasser, B. Kaczer, P. Hehenberger, W. Goes, R. O'Connor, H. Reisinger, W. Gustin, and C. Schlunder, *Int. Electron Devices Meeting Tech. Dig.*, 801 (2007).
7. A. Kerber, K. Maitra, A. Majumdar, M. Hargrove, R. J. Carter, and E. A. Cartier, *IEEE T. Electron Dev.* **55**, 3175 (2008).
8. M. A. Alam, *Int. Electron Devices Meeting Tech. Dig.*, 345 (2003).
9. S. V. Kumar, C. H. Kim, S. S. Sapatnekar, *IEEE/ACM International Conference on Computer-Aided Design ICCAD'06*, 493 (2006).
10. T. Grasser, B. Kaczer, W. Goes, H. Reisinger, Th. Aichinger, Ph. Hehenberger, P.-J. Wagner, F. Schanovsky, J. Franco, Ph. Roussel, and M. Nelhiebel, to be presented at *Int. Electron Devices Meeting* 2010.
11. T. Grasser, B. Kaczer, and W. Goes, *Proc. Int. Reliab. Phys. Symp,* 28 (2008).
12. B. Kaczer, T. Grasser, J. Martin-Martinez, E. Simoen, M. Aoulaiche, Ph. J. Roussel, and G. Groeseneken, *Proc. Int. Rel. Phys. Symp.*, 55, 2009.
13. T. Grasser, B. Kaczer, W. Goes, Th. Aichinger, Ph. Hehenberger, and M. Nelhiebel, "A Two-Stage Model for Negative Bias Temperature Instability", *Proc. Int. Reliab. Phys. Symp.*, 2009.
14. T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Goes, and B. Kaczer, *Proc. Int. Rel. Phys. Symp.*, 16 (2010).
15. S. Rangan, N. Mielke, and E. C. C. Yeh, *Int. Electron Devices Meeting Tech. Digest*, 341 (2003).
16. G. Chen, K. Y. Chuah, M. F. Li, D. S. H. Chan, C. H. Ang, J. Z. Zheng, Y. Jin, and D. L. Kwong, *Proc. Int. Reliab. Phys. Symp.*, 196 (2003).
17. R. Fernández, B. Kaczer, A. Nackaerts, S. Demuynck, R. Rodríguez, M. Nafría, G. Groeseneken, *Int. Electron Devices Meeting Tech. Dig..* 1 (2006).
18. H. Reisinger, O. Blank, W. Heinrigs, A. Mühlhoff, W. Gustin, and C. Schlünder, *Proc. Int. Reliab. Phys. Symp.*, 448 (2006).
19. B. Kaczer, T. Grasser, Ph. J. Rousse, J. Martin-Martinez, R. O'Connor, B. J. O'Sullivan, G. Groeseneken, *Proc. Int. Reliab. Phys. Symp.*, 20 (2008).
20. E. Milotti, arXiv:physics/0204033v1.
21. H. Reisinger, T. Grasser, W. Gustin, and C. Schlünder, *Proc. Int. Reliab. Phys. Symp.*, 1 (2010).

22. V. Huard, C. Parthasarathy, C. Guerin, T. Valentin, E. Pion, M. Mammasse, N. Planes and L. Camus, *Proc. Int. Reliab. Phys. Symp,* 289 (2008).
23. T. Grasser, H. Reisinger, W. Goes, Th. Aichinger, Ph. Hehenberger, P.-J. Wagner, M. Nelhiebel, J. Franco, and B. Kaczer, *Int. Electron Devices Meeting Tech. Dig.*, 729 (2009).
24. M. Toledano-Luque, B. Kaczer, Ph. Roussel, M.J. Cho, T. Grasser, and G. Groeseneken, *presented at WoDiM 2010*.
25. M. J. Uren, M. J. Kirton, and S. Collins, *Phys. Rev. B* **37**, 8346 (1988).
26. M. J. Kirton and M. J. Uren, *Adv. Phys.* **38**, 367 (1989).
27. A. Palma et al., *Phys. Rev. B* **56**, 9565 (1997).
28. B. Kaczer, T. Grasser, Ph. J. Roussel, J. Franco, R. Degraeve, L.-A. Ragnarsson, E. Simoen, G. Groeseneken, H. Reisinger, *Proc. Int. Reliab. Phys. Symp.*, 26 (2010).
29. A. Asenov, R. Balasubramaniam, A. R. Brown, J. H. Davies, *IEEE T. Electron Dev.* **50**, 839 (2003).
30. A. Ghetti, C. M. Compagnoni, A. S. Spinelli, A. Visconti, *IEEE T. Electron Dev.* **56**, 1746 (2009).
31. B. Kaczer, Ph. J. Roussel, T. Grasser, and G. Groeseneken, *IEEE Electron Dev. Lett.* **31**, 411 (2010).
32. B. Kaczer, A. Veloso, M. Aoulaiche, and G. Groeseneken, Microelectronic Engineering, **86**(7-9), 1894 (2009).
33. D. Liu and J. Robertson, *Appl. Phys. Lett.* 94, 042904 (2009).
34. N. Collaert, P. Verheyen, K. De Meyer, R. Loo and M. Caymax, *IEEE T. Nanotech* **1**, 190 (2002).
35. J. Franco, B. Kaczer, M. Cho, G. Eneman, G. Groeseneken, and T. Grasser, *Proc. Int. Rel. Phys. Symp.*, 1082 (2010).
36. J. Franco, B. Kaczer, G. Eneman, J. Mitard, A. Stesmans, V. Afanas'ev, T. Kauerauf, Ph. J. Roussel, M. Toledano-Luque, M. Cho, R. Degraeve, T. Grasser, L.-Å. Ragnarsson, L. Witters, J. Tseng, S. Takeoka, W.-E. Wang, T.Y. Hoffmann, and G. Groeseneken, *Int. Electron Devices Meeting* 2010.

Chapter 3
Charge Trapping Phenomena in MOSFETS: From Noise to Bias Temperature Instability

Gilson Wirth and Roberto da Silva

Abstract Charge trapping phenomena is known to be a major reliability concern in modern MOSFETS, dominating low-frequency noise behavior and playing a significant role in aging effects such as Bias Temperature Instability (BTI). In this chapter we address this reliability issue.

We start by discussing MOSFET low-frequency (LF) noise, which is known to be dominated by charge capture and emission by defects (traps) close to the semiconductor–dielectric interface. Standard (LF) noise models used today (e.g. BSIM and PSP) do not properly model noise behavior under large signal excitation. A circuit level modeling and simulation approach, valid at both DC and large signal (AC) biasing, is presented.

The role of charge trapping and de-trapping in BTI (Bias Temperature Instability) is also discussed and modeled.

Mutual relation between the different reliability phenomena (LF noise, BTI and RDF) is also studied. For instance, random dopant fluctuations (RDF) may exacerbate the impact of BTI and LF noise on circuit performance.

1 Introduction

The major goal of this chapter is to compile and critically discuss recent work performed by the authors on modeling of charge trapping and de-trapping phenomena in nanometer scale CMOS devices [1–6, 16–18]. The role of charge trapping and de-trapping in both low-frequency noise and bias temperature instability (BTI) is discussed.
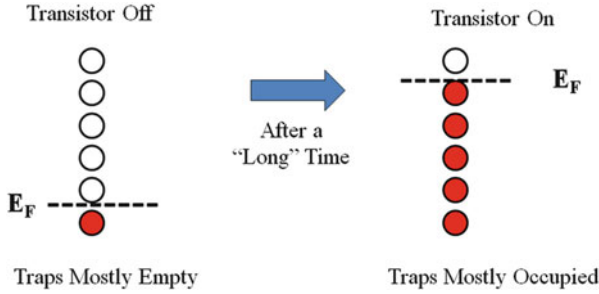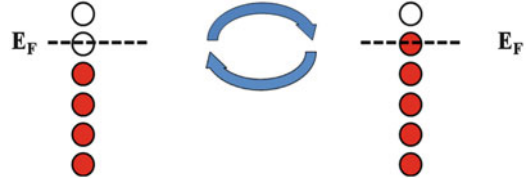

G. Wirth (✉)
Electrical Engineering Department, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil
e-mail: wirth@ece.ufrgs.br

R. da Silva
Physics Institute, UFRGS, Porto Alegre, Brazil

Fig. 3.1 Traps within a few
kT from the Fermi Level
contribute to noise, by
switching their state between
occupied (by a charge carrier)
and empty

Fig. 3.2 If a transistor is abruptly turned on, the Fermi level abruptly changes and trap occupation probability follows the Fermi level (E_F) change. Charge trapping/de-trapping is not an instantaneous event. It is governed by characteristic time constants, and the trap occupation reflects the new Fermi level after an elapsed time. This figure depicts the situation for electron trapping in the channel of a NMOS transistor

Charge trapping and de-trapping at localized states (charge traps) at the interface or in the gate dielectric is a significant reliability issue for CMOS applications. It is known to be a source of low-frequency noise, besides playing a role in Bias Temperature Instability (BTI).

Traps that contribute to noise are the ones that keep switching their state between occupied and empty, as depicted in Fig. 3.1. These are the traps with occupation probability close to 50 %, which means that their capture and emission times are similar.

Traps that contribute to BTI are the ones that have a high probability to stay occupied after a charge trapping event, as depicted in Fig. 3.2. These are the traps with occupation probability close to 100 %, which means that their capture time is much shorter than the emission time.

Figure 3.3 depicts the mechanisms leading to BTI and low-frequency noise.

If a constant bias is applied to a Metal–Insulator–Semiconductor system, only the localized states (traps) with energy close to the Fermi level show significant activity, i.e., change their states between occupied and empty, by capture and subsequent emission of charge carriers from the channel region. This may be considered a steady state condition, where the *occupation probability* of a trap is time independent, i.e., a constant that does not change with time. This mechanism is known to originate low-frequency noise.

However, if the bias point changes abruptly, as is common in e.g. digital CMOS applications, the trap occupation probability abruptly changes. For instance, if the