

Chapter 2

Degradation Mechanisms

2.1 Introduction

In this chapter, a general description of the main MOSFET degradation mechanisms considered in this work is given. The proposed dissertation, while not aiming to give a complete coverage of the wide literature available on the treated topics, is meant to provide the reader with a sufficient basis to follow the discussion of the original experimental work presented in the following chapters.

Most of the chapter is devoted to the Negative Bias Temperature Instability (NBTI) since this degradation mechanism pertains to most of the original experimental work on SiGe and Ge channel devices presented later in the [Chaps. 4 and 5](#). In [Sect. 2.2](#), we present a phenomenological description of NBTI degradation, discussing a first simple empirical model capturing the major experimental observations ([Sect. 2.2.1](#)). A general interpretation of what causes the degradation follows in [Sect. 2.2.2](#). A distinctive and crucial aspect of NBTI is the partial recoverability of the induced degradation, as discussed in [Sect. 2.2.3](#).

The first and probably still most widely used attempt to physically describe NBTI, i.e., the Hydrogen Reaction–Diffusion (R–D) model, is introduced in [Sect. 2.2.4](#). This model was originally proposed back in 1977 [1] and the bulk of NBTI literature appeared to be unanimously supportive of it until about 2006. However, recent experimental observations about the NBTI relaxation ([Sect. 2.2.5](#)) and the NBTI degradation of deeply scaled devices ([Sect. 2.2.6](#)) pose significant doubts about the correctness of the R–D model.

Recent modeling attempts are focused on the central role of hole trapping mechanisms in NBTI degradation. However, the standard Shockley–Read–Hall models for carrier generation-recombination have been proven to be incapable of correctly capturing the properties of hole trapping into bulk oxide defects ([Sect. 2.2.7](#)). A physically more accurate description of the trapping mechanism, including vibrational motion of the atoms at the defect sites, is needed for reproducing the observed trapping behavior ([Sect. 2.2.8](#)). Based on these insights, a defect model compatible with a wide range of NBTI degradation observations

has been recently proposed (Sect. 2.2.9) and it is at the base of the recent NBTI modeling attempts (Sect. 2.2.10).

This detailed overview of the latest trends and controversies in NBTI modeling will be needed and referred to in Chaps. 4 and 5 to understand the improved reliability of Ge-based high-mobility channel devices.

Finally, a general overview of the Hot Carrier effects in scaled transistors (Sect. 2.3) and of the Time-Dependent Dielectric Breakdown (TDDB, Sect. 2.4) is given at the end of this chapter, which will serve as a reference for the discussion of the original experimental results that will be presented in Chap. 6.

2.2 Negative Bias Temperature Instability

The Bias Temperature Instability (BTI) is one of the most serious concerns for the reliability of MOS technologies. The highest impact is observed in p-channel MOSFETs, which during their operation are biased with negative gate-to-source voltages (Negative BTI–NBTI, see Fig. 2.1a). A very distinctive feature of BTI is the partial recoverability of the degradation once the stress conditions (gate bias and operating temperature) are reduced or removed. This aspect significantly complicates BTI studies and it is often considered as the key for understanding the exact physical mechanism behind the device degradation, as we will discuss in detail later in this section.

From an empirical point of view, NBTI induces a degradation of all the main electrical parameters of the stressed device, including:

- an increase of the absolute value of the threshold voltage, V_{th} (see Figs. 2.1b and 2.2);
- a reduction of the device transconductance g_m (see Fig. 2.2b);
- a reduction of the device ON drain current, both in the linear (I_{Dlin}) and saturation (I_{Dsat}) regimes;
- a reduction of the channel mobility (μ_{eff});
- an increase of the sub-threshold swing (SS), i.e., less steep transition from the OFF to the ON regime.

In order to give the impression of the relevance of the NBTI concern, it is worth briefly mentioning some examples of the possible impact on the functionality of different families of MOSFET circuits:

- CMOS logic circuits: reduced circuit switching speed due to increasing charging times of the load capacitances (interconnects or next logic stage) leading eventually to time violation of the synchronous operation;
- Ring Oscillator circuits: reduced and unstable oscillation frequency;
- SRAM cells: increased access time and reduction of the noise margins, eventually causing bit loss (inability to program the desired value or wrong read value);

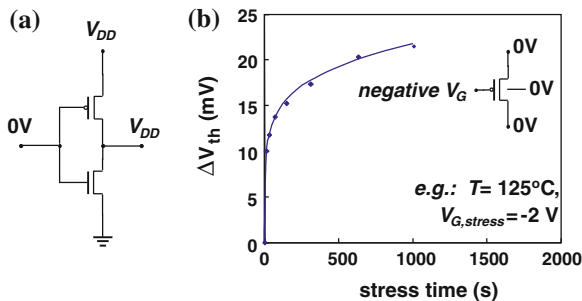


Fig. 2.1 **a** Sketch of a CMOS logic inverter: during normal operation, when the input signal to the gate is ‘0’, the ‘pull-up’ pMOS device experiences a negative $V_{GS} = -V_{DD}$ which, combined with the high operation temperature reached in scaled high performance integrated circuits (e.g., CPU), causes NBTI. **b** Due to NBTI, the device threshold voltage increases (in absolute value) during operation, leading to performance degradation

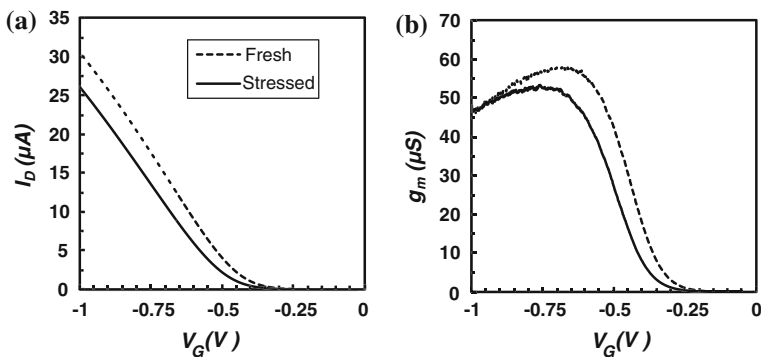


Fig. 2.2 **a** I_D – V_G characteristic of a pMOSFET in the linear regime ($V_D = -50\text{ mV}$) before and after an NBTI stress. On top of the threshold voltage shift, **b** a reduction of the device peak transconductance is clearly visible

- Analog stages: reduced amplifier gain, induced mismatch in current mirrors, increased offset and reduced bandwidth in op-amps.

The effect of NBTI has been reported since the early days of MOS technology [2], but it gained much attention in recent years due to the aggressive scaling of the semiconductor technology. In particular, the following aspects have been found to lead to increasing susceptibility to NBTI:

- the shift to CMOS technology, elevating the importance of p-channel MOSFETs;
- higher gate oxide electric fields due to oxide scaling with limited supply voltage reduction;
- higher operating temperatures due to higher power dissipation;

- introduction of alternative dielectric materials (nitrided oxides, high-k dielectrics).

Due to the inherent complexity of NBTI and its measurement, a general introduction to this phenomenon is first given in this Section. The observed characteristics of NBTI are outlined and the recent understanding of the involved physical mechanisms is discussed.

2.2.1 First-Order Phenomenological Observations

The basic setup for the observation of negative bias temperature degradation of either a MOSFET or a MOS capacitor is schematically depicted in Fig. 2.3. The substrate and (in the case of a MOSFET) the source and drain contacts are grounded, while the gate is negatively biased. These bias conditions are applied at elevated temperatures, typically ranging between 100 and 200 °C, in order to accelerate the device degradation.

The simplest and the most direct methodology for monitoring the NBTI degradation of a transistor consist of a periodical measurement of its I - V characteristics. As schematically shown in Fig. 2.4a, NBTI is clearly visualized as shift of the I_D - V_G characteristics as a function of the stress time. Such shift can be, in a first approximation, described simply as a threshold voltage shift ΔV_{th} (i.e., assuming a parallel shift of the I - V curves). The ΔV_{th} follows a power-law of the stress time (Fig. 2.4b):

$$\Delta V_{th} = A t_{stress}^n, \quad (2.1)$$

with A being a stress-dependent prefactor, and the power-law time exponent n typically ranging between 0.1 and 0.25, as discussed later in this section [3].

The magnitude of the NBTI-induced ΔV_{th} is universally observed to depend on the applied gate stress voltage (the ‘bias’ instability), with higher negative voltages

Fig. 2.3 The basic setup for NBTI stress tests. The gate is negatively biased while the source, the drain, and the substrate contacts are grounded. These conditions are applied at elevated temperatures for a certain duration

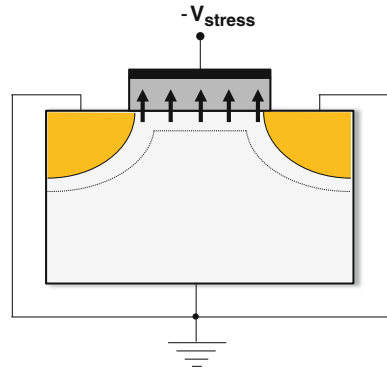
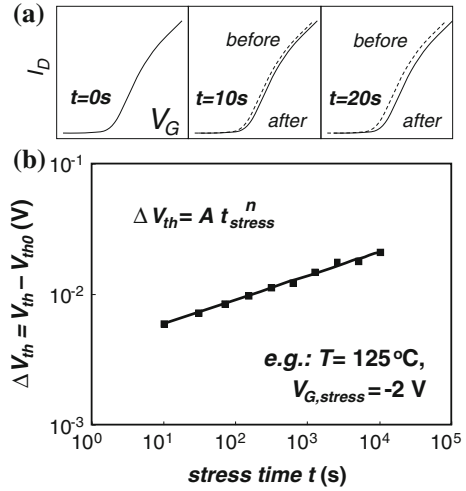


Fig. 2.4 **a** Schematic depiction of the typical I – V characteristic drift during a NBTI stress. In the first approximation, i.e., assuming only a parallel shift of the I_D – V_G curves, the drift can be represented by a threshold voltage shift, ΔV_{th} . **b** The ΔV_{th} appears to follow a power-law dependence of the stress time



causing larger degradation (Fig. 2.5). More precisely, reported studies on MOS structures with varying oxide thickness [4], clearly revealed NBTI degradation to depend on the electric field experienced by the oxide layer (E_{ox}) and not on the absolute value of the gate voltage, as shown in Fig. 2.6.

When studying NBTI on DUTs with a fixed oxide thickness, it is customary to look at the degradation dependence on the gate overdrive voltage $V_{ov} = |V_G - V_{th0}|$, this voltage being directly linked to the oxide electric field ($E_{ox} \approx V_{ov}/t_{ox}$). Figure 2.7 show the typically observed dependence on the applied V_{ov} . The prefactor A of the NBTI time power-law [see Eq. (2.1)] is typically observed to follow itself a power-law of the applied V_{ov} ,

$$A \propto E_{ox}^\gamma \approx \left(\frac{V_{ov}}{t_{ox}} \right)^\gamma, \quad (2.2)$$

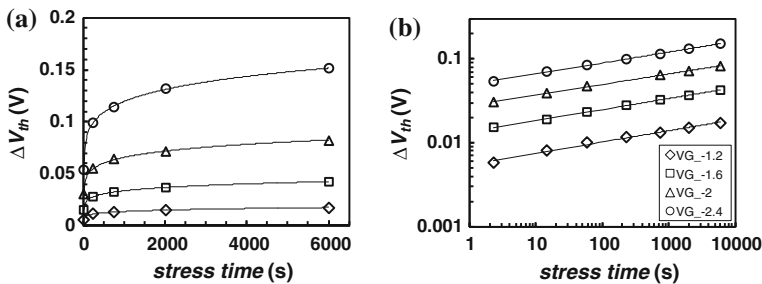


Fig. 2.5 Measured NBTI-induced ΔV_{th} as a function of the stress time for increasing negative gate stress voltages **a** lin–lin scale; **b** same data in a log–log; note the superlinear dependence on the gate stress voltage. Experimental data collected on Si/SiON/poly-Si pMOSFETs, $t_{ox} \approx 1.6$ nm, $T = 125$ °C

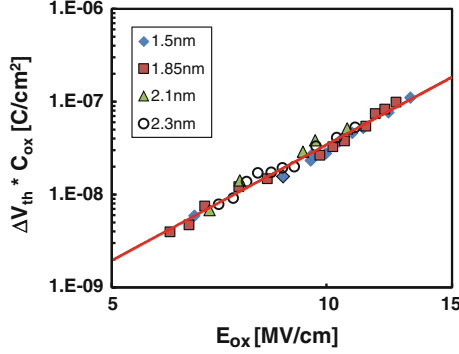


Fig. 2.6 Different gate voltages are needed in order to obtain the same electric field on MOS structures with different oxide thicknesses. However, the NBTI-induced degradation (converted into equivalent charge by multiplying the measured ΔV_{th} times the oxide capacitance) is shown to be independent of the oxide thickness once plotted against the oxide electric field. After [4]

with the exponent γ typically in the range 2.5–3 for Si devices. Conversely, the apparent time exponent n (Fig. 2.27b) is reported to be almost independent of the applied gate bias [3].

The temperature dependence of NBTI-induced ΔV_{th} is generally understood to be an Arrhenius law:

$$\Delta V_{th} \propto \exp\left(-\frac{E_A}{k_B T}\right), \quad (2.3)$$

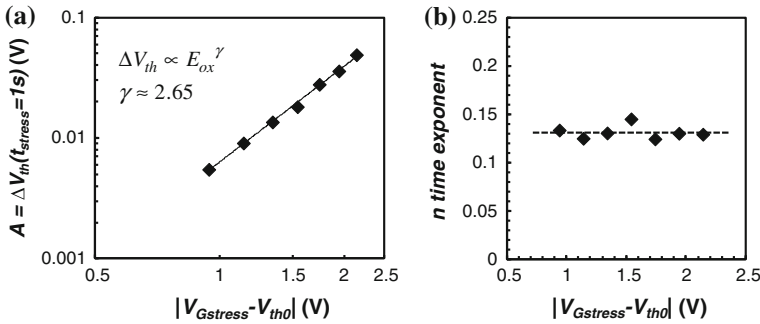
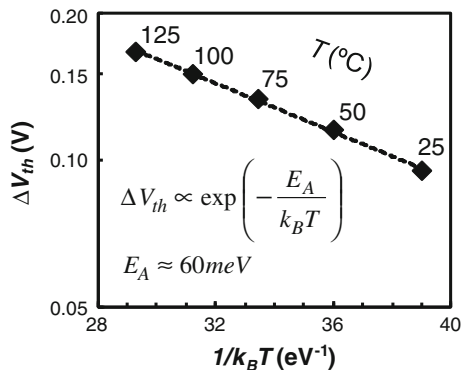


Fig. 2.7 The NBTI dependence on the stress time can be described to the first order as a power-law [Eq. (2.1)]. **a** The power-law prefactor A (i.e., the ΔV_{th} one would measure after 1 s of stress) shows a superlinear dependence on the applied oxide electric field E_{ox} (or, equivalently V_{ov}), often described as a power-law with exponent $\gamma \approx 2.5 \sim 3$ for Si channel devices [Eq. (2.2)]. **b** Conversely, the time exponent n is typically observed to be almost independent of the stress bias. Data are for Si/SiON devices with $t_{ox} = 1.6$ nm

Fig. 2.8 Temperature dependence of NBTI. An Arrhenius law is typically found with activation energy $E_A \approx 60$ meV



where k_B is the Boltzmann constant, T the stress temperature, and E_A the apparent activation energy, typically reported in the range 60–80 meV [3] as shown in Fig. 2.8. Thanks to the wide consensus of this observation, higher test temperatures are often used for accelerated testing and subsequently the measured trends are safely extrapolated to operating conditions.

The dependences of NBTI on stress time, oxide field, and temperature can be summarized into a compact analytic form for the purpose of extrapolation from accelerated tests to realistic operating conditions (i.e., reduced stress conditions but longer device lifetime). Eq. 2.4 represents a typical simple analytic description, corresponding to the observations above discussed:

$$\Delta V_{th} \approx C \exp\left(-\frac{E_A}{k_B T}\right) \left(\frac{|V_G - V_{th0}|}{t_{ox}}\right)^\gamma t_{stress}^n \quad (2.4)$$

A typically used failure criterion for extrapolating the device lifetime under NBTI stress is $\Delta V_{th} = 30$ mV. An empirical description as the one given in Eq. (2.4) can be used to estimate the device lifetime under operating conditions based on accelerated stress tests. However, such extrapolations based on purely empirical formulation involve a significant risk: a small uncertainty in the experimentally extracted parameters (e.g., in the time exponent n) might easily lead to under- or overestimation of the product lifetime by several years. Furthermore, empirical models are typically based on experimental observations apparent when using a given measurement technique which, however, might not capture the real complexity of NBTI, as discussed later.

For a proper lifetime extrapolation, a thorough understanding of the physical mechanisms behind NBTI degradation is therefore necessary. At present, while several physics based models have been proposed, there is still a clear lack of consensus. In the following subsections, after briefly discussing the general interpretation about what causes the device parameter shifts, and after discussing the crucial issue of NBTI relaxation, we will review some of the most relevant models which have been proposed in the literature.

2.2.2 Basic Interpretation: Why does the V_{th} Shift?

Generally, two types of charges can be induced in the dielectric layer of an MOS structure by means of an electrical stress: interface-trapped charge (Q_{it}) and oxide-trapped charge (Q_{ot}). The interface-trapped charge is due to electronic states located at the Si–SiO₂ interface with energy levels within the silicon bandgap. As such, they can capture or emit electrons or holes from both conduction and valence bands. These electronic states arise mainly due to the presence of dangling (incomplete) bonds at the interface. Under the equilibrium condition, the occupancy of the interface states is governed by the position of the Fermi level at the semiconductor/oxide interface. Since their net effect depends on the applied gate voltage, interface states cause a “stretch-out” of the Capacitance–Voltage (C – V) characteristic of the MOS system and a degradation of the subthreshold slope (i.e., a less steep I_D – V_G characteristic below threshold). Moreover, they are observed to cause a reduced carrier mobility at the surface, which results in the reduction of the device transconductance.

The oxide-trapped charge Q_{ot} instead is associated with defects in the bulk of the dielectric layer where electron or holes can be trapped. Oxide states typically have significantly larger trapping/detrapping characteristic time constants as compared to interface states, and therefore their charge state is unable to dynamically follow a gate bias sweep during C – V or I_D – V_G measurements. Hence, the presence of oxide-trapped charge is typically visible as a parallel shift of the MOS characteristics and not as a “stretch-out”.

In their pioneering work on NBTI, Jeppson and Svensson [1] observed already the effect of both an increased Q_{it} and Q_{ot} on the C – V of the MOS sample under stress (Fig. 2.9). The interface state density can be directly monitored during an NBTI stress by means of the Charge Pumping (CP) technique [5], as shown in Fig. 2.10. A significant creation of interface states is clearly visible.

The generation of interface states is typically reported to follow a power-law of the stress time with exponent ~ 0.25 – 0.3 , Fig. 2.11 [3]. Moreover, by comparing MOS stacks with different oxide thicknesses, Huard et al., clearly demonstrated such degradation to depend on the oxide electric field and not on the absolute gate voltage.

Although the generation of new interface states is universally reported in the literature, it has been often noted to be insufficient to explain the total measured V_{th} shift during NBTI. This can be observed, e.g., in Fig. 2.12 where Aoulaiche et al. converted the measured total ΔV_{th} into an equivalent charge density $\Delta N_{eff} = \Delta V_{th} C_{ox} / q$, and compared it with the created interface state density ΔN_{it} as measured by CP. As one can observe, ΔN_{it} contributes only a small part of the total degradation, while the presence of an additional trapped charge component ΔN_{ot} is evident. ΔN_{ot} is found to be dominant for high- k stacks [6].

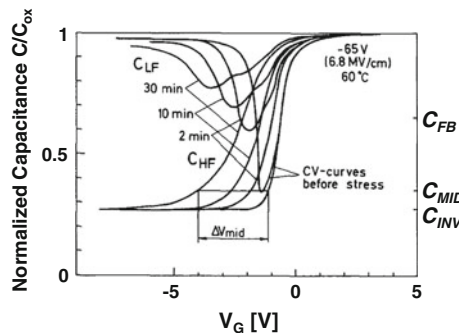
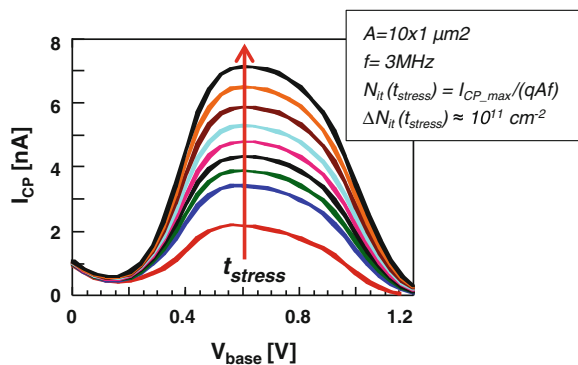


Fig. 2.9 Low- and high-frequency C - V characteristics of a MOS structure monitored during a NBT stress as reported by Jeppson and Svensson in their seminal work [1]. Both a shift and a ‘stretch-out’ of the high-frequency C - V are visible, together with an increase of the depletion capacitance visible in the low-frequency measurements. This observation suggested already the effect of increased both Q_{it} and Q_{ot}

Fig. 2.10 Interface state generation during NBTI, monitored by the charge pumping (base-voltage sweep) technique. A significant generation of interface states is visible ($\Delta N_{it} \sim 1 \times 10^{11} \text{ cm}^{-2}$, in this case). After, e.g., [6]



2.2.3 NBTI Relaxation

The most peculiar characteristic of NBTI degradation is its partial recoverability. Once the stress is removed or diminished, the induced ΔV_{th} is observed to partially recover (Fig. 2.13) [7].

This feature is commonly referred to as “relaxation” or “recovery”. It poses a clear challenge for the measurement of NBTI degradation: any delay between the end of stress and the actual measurement yields an underestimation of the device degradation. Furthermore, the estimation of the NBTI time exponent was observed to strongly depend on the measurement delay (Fig. 2.14) [7]. A range of several NBTI time exponents between ~ 0.1 and ~ 0.25 has been reported in the literature, possibly due to different measurement techniques. Such observation induces a significant uncertainty in device lifetime extrapolations based on purely empirical models as, e.g., the one discussed in the previous subsection [see Eq. (2.4)].

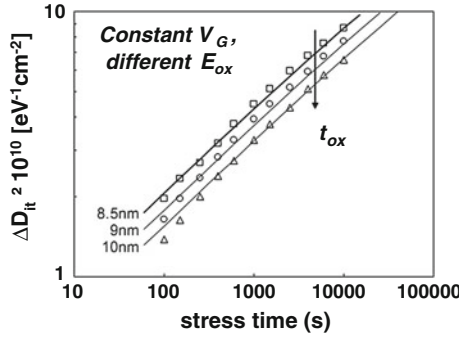


Fig. 2.11 Generated interface state density as a function of NBTI stress time. A power-law dependence with an exponent $n \sim 0.25-0.3$ is visible. Samples with a thicker oxide show reduced interface state generation at constant gate stress voltage, suggesting the degradation to be driven by the oxide electric field, rather than gate voltage. After [3]

Fig. 2.12 The measured ΔV_{th} can be converted into equivalent charge density ($\Delta N_{eff} = \Delta V_{th} C_{ox}/q$). When comparing ΔN_{eff} to the created interface state density ΔN_{it} as monitored by CP, the presence of an additional trapped charge component ΔN_{ot} is evident. After [6]

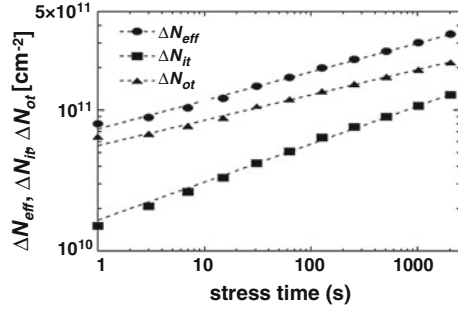
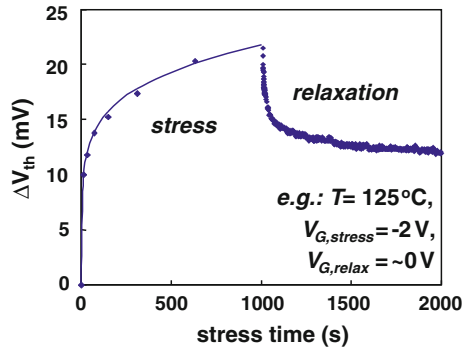


Fig. 2.13 The NBTI-induced ΔV_{th} is observed to partially recover after the stress voltage is removed. After [7]



Moreover, it also complicates the direct comparison of degradation data measured with different instrumentation and by different groups. For empirical cross-comparison of different MOS stacks, it is therefore mandatory to use and report fixed measurement conditions.

Fig. 2.14 The estimation of the time exponent is affected by different measurement delays. After [7]

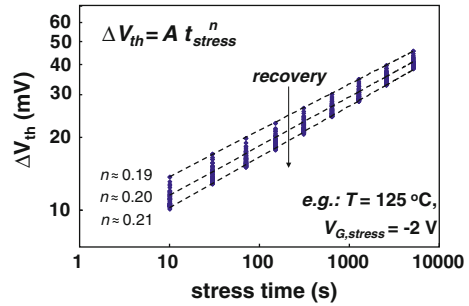
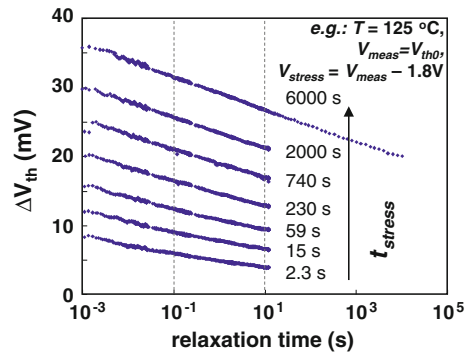


Fig. 2.15 A set of relaxation transients recorded after stress phases of increasing duration. The relaxation transients show a $\log(t)$ trend over the whole experimental time window. After [7]



Due to the significant relaxation observed already one microsecond after removing the stress [8], it is evident that the basic measurement technique consisting of periodical stress interruptions for monitoring the full I - V characteristic of the DUTs (see Fig. 2.4) is not appropriate for proper NBTI assessment. A more appropriate approach consists of using measurement techniques capable of capturing both the degradation and the relaxation behaviors. An overview of such techniques will be given in Chap. 3.

Figure 2.15 reports a set of relaxation transients recorded after NBTI stress periods of increasing duration with one such technique (i.e., the extended Measure-Stress-Measure technique, or eMSM [7]). It is worth noticing already that the NBTI relaxation is observed to proceed with a $\log(t)$ trend over a wide range of timescales (see, e.g., the extended relaxation transients covering ~ 7 decades in time in Fig. 2.15). This is a crucial aspect which any NBTI modeling attempt should be able to properly capture, as will be discussed later on.

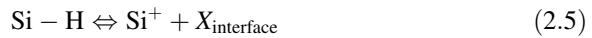
2.2.4 Reaction-Diffusion Model

Although NBTI has been reported since the early days of the MOS technology [2], there is still much controversy about the physical mechanisms behind the

degradation. However, broad agreement has been found that when pMOSFETs are stressed with a negative gate voltage at an elevated temperature, positive charge builds up either at the Si/SiO₂ interface and/or in the gate oxide layer. This charge leads to the above discussed degradation of the transistor parameters.

The reaction–diffusion (R–D) model was originally proposed by Jeppson and Svensson in their pioneering NBTI study in 1977 [1] and further developed during the years by several groups of authors [9, 10, 11]. The basic concept on which the R–D model is based is interface state creation controlled by hydrogen diffusion through the gate oxide.

A significant density of Si dangling bonds is known to be present at the Si/SiO₂ interface after Si oxidation. These defects are electrically deactivated in a following process step consisting of a wafer anneal in a hydrogen-rich ambient (commonly indicated as Forming Gas Anneal, or FGA), with the Si dangling bonds being passivated by H [2]. The R–D model describes the device degradation as a combination of two linked aspects. First, a field-dependent electrochemical reaction involving channel holes at the Si/SiO₂ interface is assumed to break the Si–H bonds at the interface (the ‘Reaction’ step). As a consequence, an electrically active new interface state (N_{it}) and a mobile hydrogen-related species (X) are formed:



Consequently, the model involves the transport of the hydrogen species away from the interface toward the bulk of the dielectric (the ‘Diffusion’ step). Also the reverse process is assumed to be possible, i.e., backward diffusion of hydrogen species toward the interface leading to the repassivation of previously created Si dangling bonds. This reverse process is suggested to be responsible of the NBTI recovery. Figure 2.16 gives a schematic illustration of the R–D model.

The process at the interface is modeled by a rate equation as:

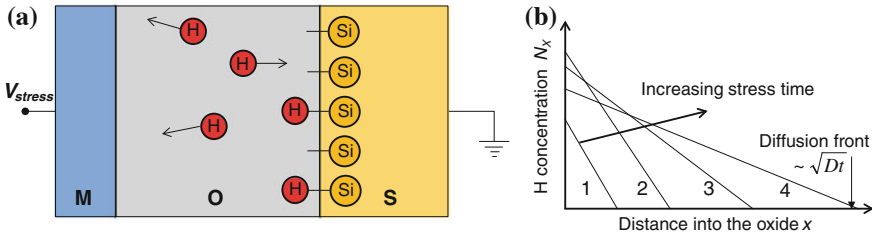


Fig. 2.16 **a** Schematic representation of the reaction–diffusion (R–D) model. Electrically inactive Si–H bonds at the Si/SiO₂ interface are broken and the hydrogen diffuses into the dielectric, leaving behind an electrically active interface trap N_{it} . **b** Sketch of the hydrogen profile diffusing into the oxide during stress

$$\frac{\partial N_{it}(t)}{\partial t} = \underbrace{k_f(N_0 - N_{it}(t))}_{\text{generation}} - \underbrace{k_r N_{it}(t) N_X(0, t)^{1/a}}_{\text{annealing}}, \quad (2.6)$$

where k_f is the interface-trap generation rate (i.e., forward rate) and k_r the interface-trap repassivation rate (i.e., reverse rate). N_0 denotes the initial number of electrically inactive Si–H bonds (the defect ‘precursors’), while $N_X(0, t)$ represents the surface (i.e., at the interface, $x = 0$) concentration of the hydrogen diffusing species. The value of a gives the order of the reaction. In the original publication by Jeppson and Svensson, neutral hydrogen H^0 was proposed which would give $a = 1$. For molecular hydrogen, H_2 , $a = 2$.

The equilibrium of the forward and backward reaction is controlled by the hydrogen density at the interface $N_X(0, t)$. Thus, the transport mechanism of the hydrogen species away from the interface characterizes the degradation mechanism, controlling the device parameter shift. The original reaction–diffusion model describes the transport as a purely diffusive mechanism, which is described by the diffusion equation (Fick’s second law):

$$\frac{\partial N_X(x, t)}{\partial t} = D \nabla^2 N_X(x, t), \quad (2.7)$$

where, D is the diffusivity of the hydrogen species in the dielectric. For each generated interface trap, a hydrogen is assumed to be released, thus:

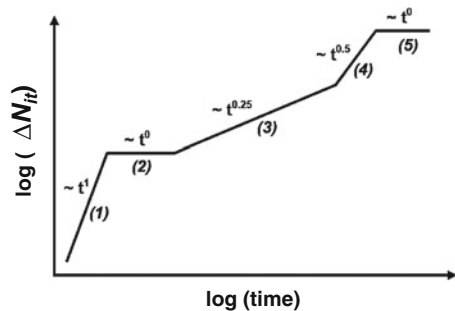
$$N_{it} = \int_x N_X dx \quad (2.8)$$

i.e., the generated interface states are assumed to be equal to the total number of released hydrogen atoms diffusing in the oxide depth (Note: 1D representation).

For the stress phase, the solution of the R–D Eqs. (2.6)–(2.7) predict five different regimes (Fig. 2.17) distinguished by their time exponents n , as discussed in the following.

Regime 1: in the early stage, the amount of free hydrogen is very low, while the amount of already broken Si–H bonds is close to zero. Therefore, the

Fig. 2.17 The classical R–D model results in five different degradation regimes with five different time exponent n . After [12]



kinetic is dominated by the forward reaction rate only, and Eq. (2.7) simplifies into:

$$\frac{\partial N_{it}(t)}{\partial t} \approx k_f N_0, \quad (2.9)$$

with its solution being:

$$N_{it}(t) \approx k_f N_0 t, \quad (2.10)$$

which follows a linear dependence of the stress time (i.e., $n = 1$).

Regime 2: after a while, a considerable amount of hydrogen has been released at the interface and the forward reaction reaches a quasi-equilibrium with the backward reaction:

$$k_f(N_0 - N_{it}) \approx k_r N_{it} N_X. \quad (2.11)$$

Equation (2.11) can be simplified assuming N_0 to be very large with respect to N_{it} :

$$k_f N_0 \approx k_r N_{it} N_X. \quad (2.12)$$

Since the diffusion process has not removed a significant amount of released hydrogen atoms from the interface yet, N_{it} is assumed to be equal to N_X at the interface, and therefore Eq. (2.12) can be solved as:

$$N_{it} \approx \sqrt{\frac{k_f N_0}{k_r}}. \quad (2.13)$$

This equation is not time dependent and therefore $n = 0$ in this regime.

Regime 3: when the diffusion of hydrogen atoms sets in, it acts as the limiting factor of the degradation. Jeppson and Svensson [1] proposed a simplified analytical solution of Eqs. (2.6)–(2.7):

$$N_{it} \approx \sqrt{\frac{k_f N_0}{2k_r}} (Dt)^{0.25}, \quad (2.14)$$

while several groups have proposed numerical solutions resulting in n values close to 0.25 [10, 13]. This regime is assumed to describe the typical lifetime of a device, setting in after ~ 1 s of stress and dominating for several orders of magnitude in time.

Regime 4: when the hydrogen diffusion front through the oxide reaches the gate, the time exponent increases to $n = 0.5$ due to the higher diffusivity of hydrogen in poly-Si

Regime 5: theoretically, the degradation ends (i.e., $n = 0$) when all interface bonds are broken and $N_{it} \approx N_0$

Once the stress is removed, the dissociation of Si–H bonds that forced forward diffusion of the hydrogen away from the interface is stopped, and therefore hydrogen atoms can diffuse back toward the interface to passivate Si dangling bonds. Similarly to the stress phase, the R–D model predicts a multiregime annealing kinetics; however, analogously to the degradation phase, the diffusion-controlled regime is expected to dominate the annealing phase. In the diffusion-limited recovery regime, an approximate solution of the R–D equations is [9, 14]:

$$\frac{N_{\text{it}}(t_s, t_r)}{N_{\text{it}}(t_s, 0)} \approx \frac{1}{1 + (t_r/t_s)^{0.5}}, \quad (2.15)$$

where t_s is the duration of the previous stress phase, and t_r is the recovery time elapsed since the end of the stress phase. Note Eq. (2.15) predicts a universal relaxation (i.e., not depending on the absolute time but only on the ratio t_r/t_s), which moreover does not depend on any model parameter: in the diffusion-limited regime the reverse reaction rate k_r has no influence, while the diffusion coefficient D cancels out of the expression since there are two opposite hydrogen diffusion fronts (one diffusing back toward the interface to passivate Si dangling bonds, and one diffusing away from it, see Fig. 2.16a).

2.2.5 NBTI Relaxation: A Crucial Benchmark for Degradation Models

The R–D model has been shown by several groups to be able to describe the device degradation during constant voltage stress, well capturing the stress time dependence of NBTI. However, when researchers started to look in detail to the NBTI recovery phenomenon, striking contradictions between experimental observation and the R–D model emerged [15]. In particular, the observation of relaxation transients extended over a wide timescale window is the crucial aspect which reveals the inadequacy, or at least the incompleteness of the R–D model, as discussed next.

The observation by ultrafast measurements of the NBTI recovery being already active in the microsecond time scale [8], shows a first significant inconsistency of the pure R–D model. As discussed in the previous Section, the conventional R–D approach predicts universal recovery, meaning that the recovered fraction of the degradation depends only on the ratio of the stress and relaxation times, t_s and t_r . Equation. (2.15) is plotted in Fig. 2.18 for a wide range of t_r/t_s values. The model predicts the recovery (90–10 %) to take place over ~ 4 time decades, while experimentally it is observed to be rather uniform and roughly follow $\log(t_r)$ over the whole experimental time window (see Fig. 2.15); actually a clear saturation of the recovery has been rarely experimentally observed [16]. For $t_r \ll t_s$ the R–D model predicts a negligible recovery, that is in clear contradiction with the experimental data. Even for very long stress times of more than 1,000 s, a

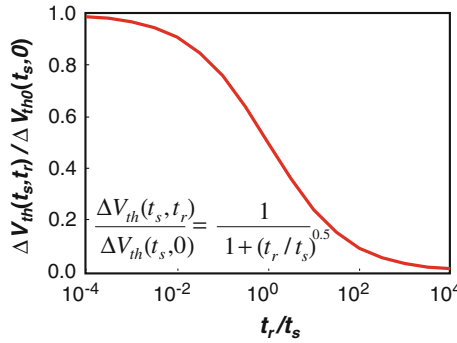


Fig. 2.18 Recovery as predicted by the R–D model. A negligible recovery is expected for $t_r \ll t_s$, while the whole relaxation process is expected to happen in ~ 4 time decades. Both these aspects are in stark contrast with the experimental data (see Fig. 2.15), where a $\log(t_r)$ trend is observed over a wide time window, starting from very short timescales. After [15]

significant recovery of V_{th} is observed in the first milliseconds after stress removal (see Fig. 2.15).

This controversy can be seen also from a different perspective. The R–D model assumes both a diffusion-limited degradation and also a diffusion-limited recovery. When considering a stress time of 1,000 s, a certain amount of hydrogen-related species is assumed to diffuse into the dielectric as shown in Fig. 2.19. The total amount of hydrogen in the dielectric must equal the number of interface traps N_{it} [see Eq. (2.5)]. N_{it} is in turn directly proportional to the shift of the threshold voltage ΔV_{th} . To argue a significant recovery during the first second in the framework of the R–D model, most of the hydrogen must diffuse back to the interface and anneal the dangling silicon bonds within the same time. This would result in a hydrogen profile in the dielectric, as seen on the right hand side of Fig. 2.19, implying that the backward diffusion must be orders of magnitude faster

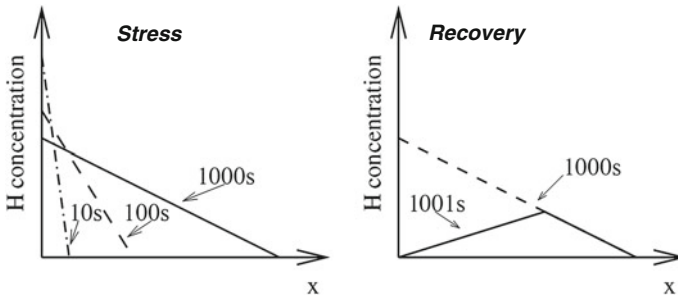


Fig. 2.19 Hydrogen profiles in the dielectric for increasing stress time (up to 1,000 s) (*left*) and after a recovery of $\sim 60\%$ of the degradation, observed 1 s after stress removal. (*right*) More than half of the hydrogen which took 1,000 s to diffuse into the dielectric would need to diffuse back to the interface in ~ 1 s in order to justify the observed recovery. After [17]

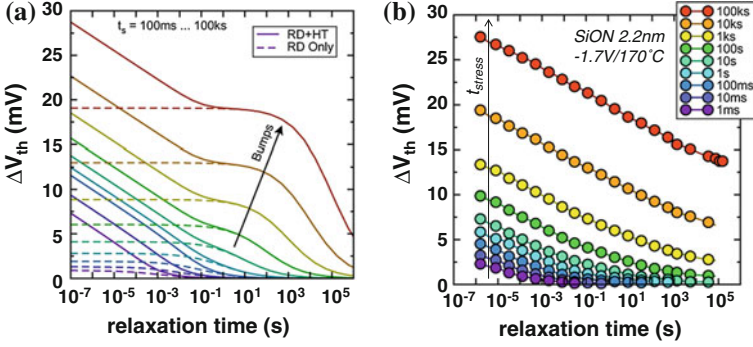
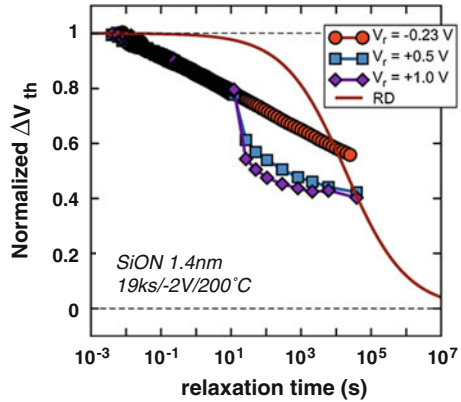


Fig. 2.20 **a** Extended R-D models with hole trapping superposition result in discontinuous recovery traces: a “bump” is unavoidable when R-D recovery takes over from hole detrapping. **b** Such “bumps” are not observed in the experimental traces. After [15]

than the forward diffusion [17]. Such asymmetry between the stress and the relaxation timescales poses other significant doubts about the correctness of diffusion-based models.

The R-D model assumes the interface states to be the only contribution to device parameter shifts. However, hole trapping into preexisting oxide defects is generally accepted to play also a role in the NBTI degradation, and recent studies strongly suggest hole detrapping might be responsible of the $\log(t)$ NBTI relaxation behavior [15]. In the attempt to overcome the above-mentioned issues with the modeling of the recovery, recent development of the R-D theory assumed NBTI to be constituted by a fast hole trapping/detrapping component which saturates within ~ 1 s, followed by the classical hydrogen-diffusion-controlled interface state generation/repassivation which would dominate degradation and recovery at longer times. However, as shown by Grassler et al. [15], the predicted recovery characteristic obtained when coupling two disjunct processes yields unnatural “bumps” not observed experimentally (Fig. 2.20).

Fig. 2.21 Switching to positive gate bias considerably accelerates recovery. Conversely, the R-D models based on neutral H_2 diffusion predict the recovery to be bias-independent. After [15], experimental data from [7]



Moreover, most R–D model implementations ascribe the recovery to the back diffusion of neutral molecular hydrogen H_2 , therefore postulating a bias independence of the recovery. However, experimental recovery has been observed to be strongly bias dependent when V_G is switched toward accumulation (Fig. 2.21) [7]. Such bias dependence has been observed to occur also when the positive bias is applied only beyond the expected saturation time of the hole detrapping component ($t_r > 1$ s) and thus has to be an intrinsic feature of the dominant recoverable component.

Furthermore, the alleged conjunction of two superimposed mechanisms, namely fast hole trapping and hydrogen-diffusion-controlled interface state creation, should be observable also during the stress. Although a transition from a $\log(t)$ degradation to a power-law t^n seems to be observable in fast ΔV_{th} data collected during stress, the two different evolutions show identical dependences on the stress voltage and temperature (Fig. 2.22) [18]. Therefore, a single mechanism or alternatively two tightly coupled ones should be considered responsible for both the regimes. Such observation poses doubts about the correctness of any NBTI model involving multiple independent physical mechanisms in order to capture a larger set of experimental features.

2.2.6 Recent NBTI Observations: Small-Area Devices

A significant boost to the understanding of NBTI degradation has come in the very recent years from the study of deeply scaled devices [19], i.e., devices significantly smaller than those usually used for NBTI testing, with channel area smaller than

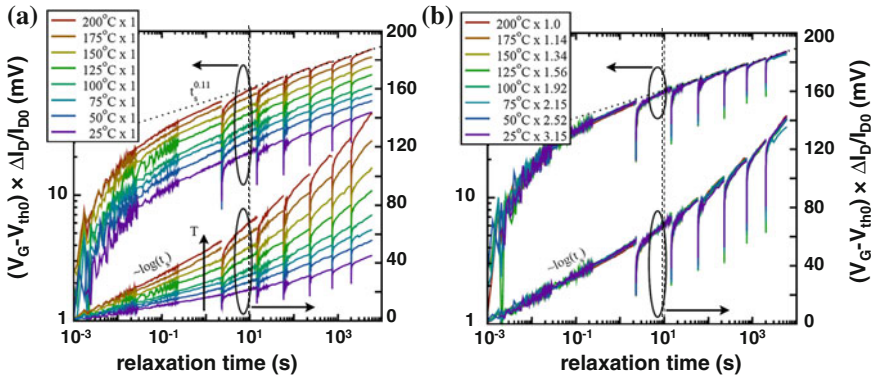


Fig. 2.22 **a** Fast NBTI degradation traces recorded at the stress bias (*Note* the drops are due to the used eMSM technique, see Chap. 3). Two trends are present: first a $\sim \log(t_s)$, followed by a power-law like t_s^n for stress time > 1 s. This observation might be interpreted as two different mechanisms being involved in NBTI. **b** However, the two trends follow perfectly identical dependences on the stress voltage and temperature, revealing that either only one, or two tightly coupled mechanism must be responsible of NBTI. After [18]

$100 \times 100 \text{ nm}^2$. The crucial observation attracting researchers' attention is that NBTI recovery in small-area devices proceeds in discrete steps, conversely to the smooth $\sim \log(t)$ transient typically observed (Fig. 2.23) [20, 21, 22]. Each individual step is interpreted to be caused by the emission of an individual hole previously trapped in an oxide defect.

As pointed out by Kaczer et al. [23], the smooth $\sim \log(t)$ relaxation traces typically measured on large area devices can be modeled as discharging of a large number of defects with widely distributed time scales. Conversely, the discharging of only a handful of oxide defects included in nanoscale devices, each one showing its own characteristic time constant, yields the discrete stepwise relaxation observed experimentally (Fig. 2.24).

This interpretation exposed the existence of a link between BTI and noise phenomena, as Random Telegraph Noise (RTN) and low-frequency $1/f$ noise. In particular, while in RTN experiments only a limited number of defects having capture and emission time constants within the experimental window are visible, many more defects are charged during a BTI stress. In other words, NBTI can be considered as the nonsteady-state response of oxide defects, while RTN represents their quasistationary behavior (Fig. 2.25) [21].

Like in RTN, the charging of each individual defect has an impact, in terms of a measurable discrete ΔV_{th} step height, that can be considerably larger than the simple charge sheet approximation ($\Delta V_{th} = q/C_{ox}$), see Fig. 2.26. This is a consequence of the percolative nature of the channel current conduction in nanoscale devices caused by the random location of discrete dopants in the channel [24]. In this scenario, the impact of each charged defect depends on the underlying channel potential profile: charged defects located in the proximity (i.e., 'on top') of the

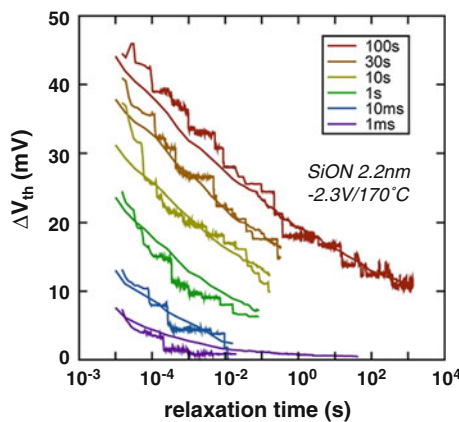


Fig. 2.23 NBTI relaxation transients recorded on a $100 \times 100 \text{ nm}^2$ pMOSFET show discrete ΔV_{th} steps corresponding to the emission of an individual hole previously trapped in the gate oxide. Averaging out the relaxation transients recorded on multiple (25) nominally identical devices yields the typical smooth relaxation transients measured on large area devices, demonstrating that discrete recovery steps are at the heart of the NBTI relaxation phenomenon. After [15]

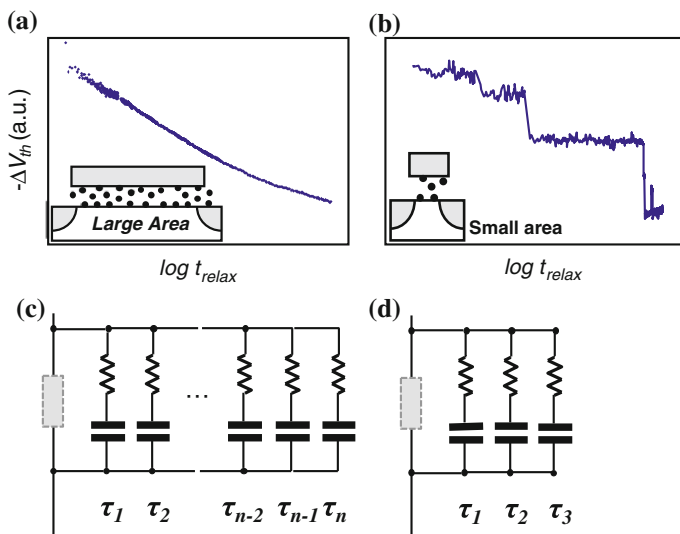


Fig. 2.24 **a** The smooth relaxation traces measured in large area devices, as opposite to **b** the discrete step-like relaxation observed in nanoscale devices. **c** In large area devices, a large number of discharging defects with widely distributed time scales yields the smooth transients [23], while **d** only a limited number of defects are visible in nanoscaled devices

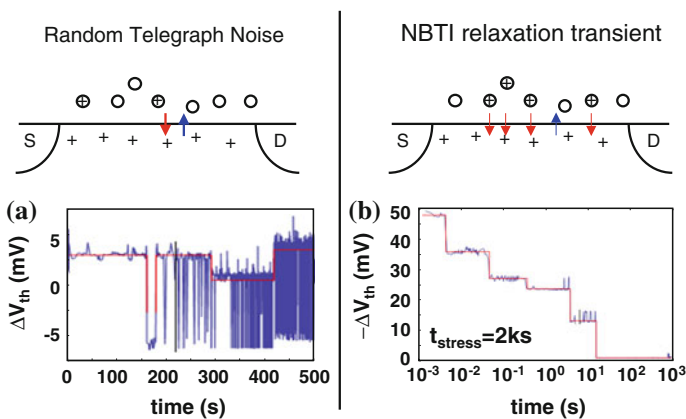


Fig. 2.25 **a** At constant bias, oxide defects are charged by channel carriers and subsequently discharged back into the channel with a wide range of time constants. The system is in dynamic equilibrium, manifested by low-frequency noise or Random Telegraph Noise (RTN) in small devices. **b** Following the perturbation by NBTI stress, excess charged oxide defects gradually discharge and the system is returning to the dynamic equilibrium of (a), resulting in long NBTI transients. After [21]

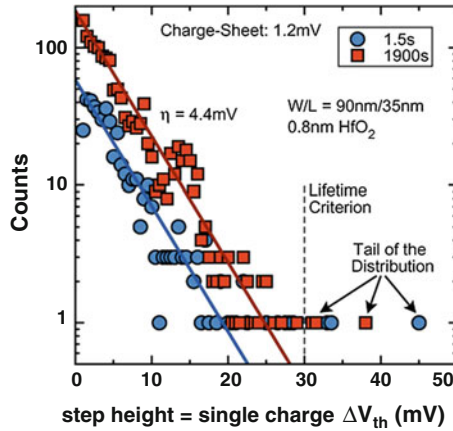


Fig. 2.26 Exponential distribution of discrete ΔV_{th} step for two different stress times (Note that the average value η (≈ 4.4 mV) is significantly higher than the charge sheet approximation (≈ 1.2 mV). Defects from the tail of the distribution can produce $\Delta V_{th} > 30$ mV (i.e. a typical failure criterion). After [21] and [15]

critical spot of a channel percolation path might yield a significant reduction of the channel current (i.e., observed as a large ΔV_{th} step height), as depicted in Fig. 2.27.

The observable magnitude of each single step height is, therefore, a result of the interaction between each charged defect and the underlying channel potential profile caused by the random dopant configuration. As such, the step height constitutes a “signature” which clearly identifies each individual defect. This observation paved the way for the introduction of a novel experimental methodology to study the properties of the defects responsible of BTI and RTN: the time-dependent defect spectroscopy (TDDS, see Fig. 2.28) [26]. It allows the study of the characteristic times of each individual defect as a function of stress/recovery bias conditions and temperature.

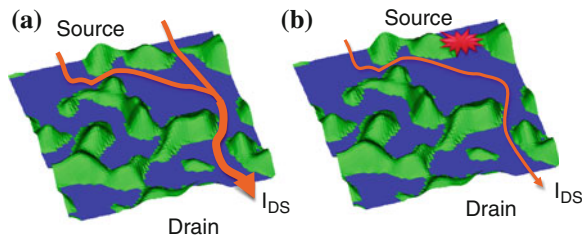


Fig. 2.27 **a** Discrete dopants arranged in random configurations cause a nonuniform potential profile in the channel of nanoscale devices. The carrier conduction proceeds through percolation paths (sketches generated via [25]). **b** In the unlucky case of a charged gate oxide defect located above the constriction point, the percolation path will be blocked off, causing a strong reduction of the current (i.e., a large ΔV_{th} step, see Fig. 2.26)

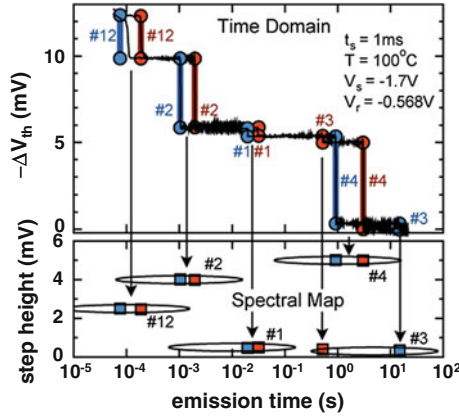


Fig. 2.28 Two NBTI relaxation traces measured on the same nanoscale devices. Each ΔV_{th} step height is an unambiguous fingerprint of the corresponding defect. A TDDS spectral map is created by collecting each step height/emission time pair. With repeated experiments, clearly visible clusters form, which allow for the identification of various defects and for the study of the stochastically distributed capture and emission times. After [26]

By means of the TDDS, several observations crucial for understanding BTI have been made [15]:

- Even for short stress times (e.g., in the ms range), emission events are observed with emission times five orders of magnitude larger than the stress time;

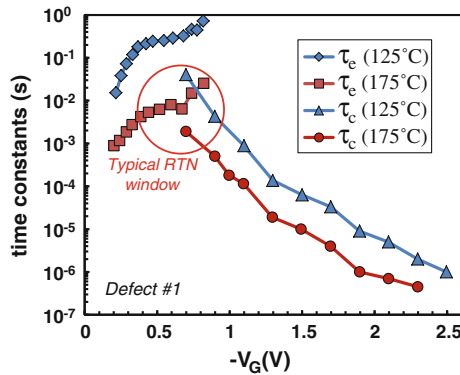


Fig. 2.29 The TDDS technique allows to study defect properties by measuring the characteristic capture and emission times over an extremely wide bias range. For comparison, the typical RTN measurement window is shown (i.e., the window for which the capture and emission time are comparable). Both the capture and emission time constants are observed to be strongly temperature dependent. The capture time constants always show a very strong bias dependence, while the emission time constants can show either a strong or a weak (not shown here, see Fig. 2.41b) bias dependence. After [26]

- Both the capture and emission time constants are strongly temperature dependent (see Fig. 2.29). No defects with temperature-independent capture time constants—hinting at the elastic tunneling process invoked by some extended R-D models—has been observed;
- The capture time constants always show a very strong bias dependence, while the emission time constants can show either a strong or a weak bias dependence (see Fig. 2.29);
- For short stress times, the charging process is fully reversible and therefore appears to be dominated by charge capture in preexisting defect precursors (e.g., oxygen vacancies). At longer stress time, a defect transformation may occur.

These observations about NBTI-related defects are fully compatible with those made about RTN-related defects. Consequently, for a thorough understanding of the BTI degradation mechanisms, the defect charge capture and emission process has to be described by a physics-based model, as discussed in the following sections.

Finally, it is worth noting that a similar information about defect charge capture and emission times as provided by the TDDS experiment (see Fig. 2.28) performed on multiple nanoscale devices, can be obtained from a single large area device with a recently proposed procedure [27]. A Capture-Emission Times (CET) map can be extracted by numerically differentiating the relaxation traces collected for increasing charging times, as shown in Fig. 2.30. The experimentally derived CET map can be

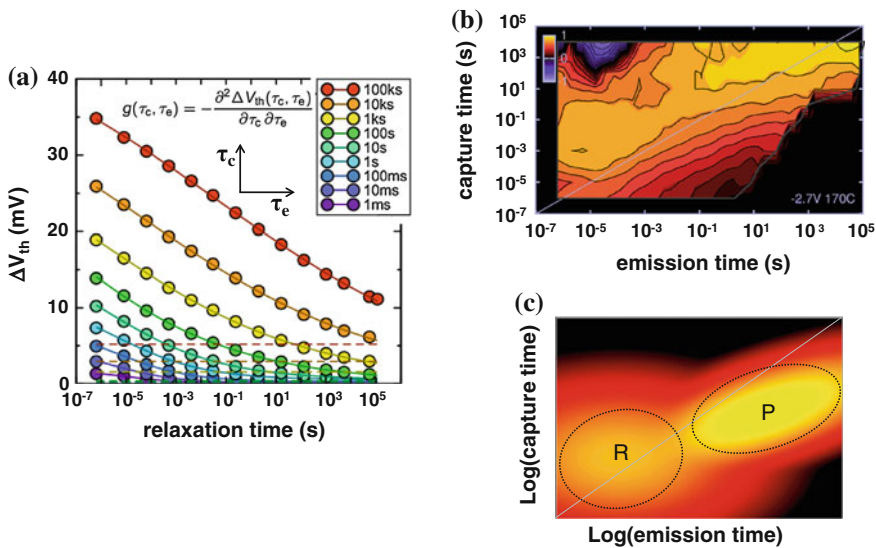


Fig. 2.30 **a** By numerically differentiating the ΔV_{th} relaxation transient measured for finely increasing stress times, **b** a capture-emission times (CET) map can be obtained. **c** Typically, the experimentally obtained CET maps can be analytically modeled by using two bivariate Gaussian distributions, which corresponds to the so-called Recoverable and Permanent components of NBTI [28]

converted into an accurate analytic representation, which typically includes two Gaussian bivariate distributions, one centered at short capture/emission times and the other centered at long capture/emission times [28]. These two components, given their characteristic time windows, resemble the often invoked Recoverable and Permanent components of NBTI, as we will discuss later in this chapter.

2.2.7 Extended Shockley–Read–Hall Trapping Models

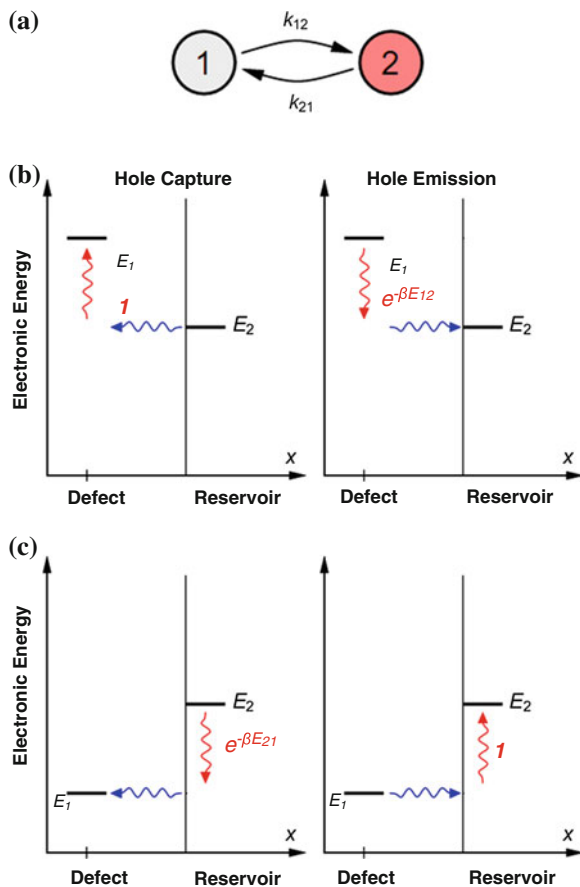
The conventional models for charge trapping and detrapping in preexisting oxide defects consist of an extension of the Shockley–Read–Hall (SRH) theory of carrier Generation–Recombination mechanism [29, 30]. This theory was originally developed for describing the charge exchange process between the conduction and valence bands of a semiconductor and states located within the semiconductor bandgap. The extensions for oxide defects typically involve the convolution of the SRH theory with the WKB approximation for the charge tunneling probability toward defects located within the dielectric layer. This is typically achieved by including an exponential term $\sim \exp(-x/x_0)$ in the capture/emission probability in order to account for defects at different oxide depths. A renowned example of such theory is the one proposed by McWhorter [31] who attributed the wide dispersion in the capture and emission time constants required to explain RTN and 1/f noise to defects located at various depths in the oxide.

In general, a hole trap can be represented by two possible states: a neutral one and a positively charged one. The transition from one state to the other (i.e., capture/emission) is typically described as a stochastic process governed by forward and reverse rates (k_{12} and k_{21} in Fig. 2.31a). The characteristic time constant of the charge exchange is commonly described as proportional to two terms, one accounting for the tunneling probability toward defects located at different depths in the oxide and the other one accounting for the transition toward defect levels located at different electronic energies (curled arrows in Fig. 2.31).

The values assumed by the latter term according to the SRH theory are summarized in Fig. 2.31: for a hole capture in a defect with electronic energy higher than the charge reservoir energy (i.e., Fermi level in the channel) the term is equal to 1 (i.e., there is no energy barrier associated with the transition—the hole ‘*bubbles up*’) while for an hole emission it is equal to the Boltzmann factor $\sim \exp(-\beta\Delta E)$ (where $\beta = 1/k_B T$ and $\Delta E = E_2 - E_1$ is the energy difference between the defect and the charge reservoir). Conversely, for a defect having an electronic energy lower than the charge reservoir energy, the term is given by the Boltzmann factor $\sim \exp(-\beta\Delta E)$ for an hole capture and it is equal to 1 for an hole emission.

As depicted in Fig. 2.32, the electronic defect energy level depends on the oxide electric field (E_{ox}). According to SRH-based models, only oxide defect levels which moved from below to above the Fermi level in the channel (e.g., due to the application of a negative gate voltage during an NBTI stress) change significantly their probability of being charged, while they can be discharged once the

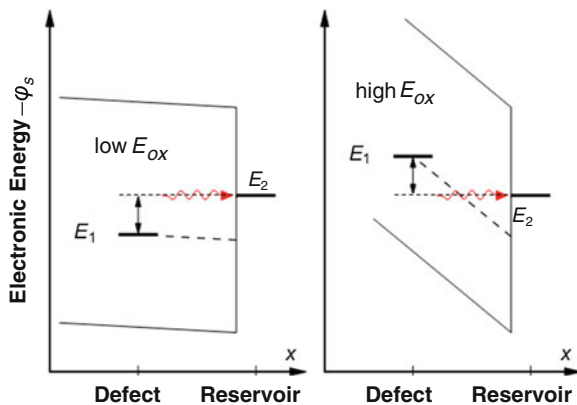
Fig. 2.31 **a** State transition rate diagram for a simple two-state defect. **b, c** The energetic barrier encountered in a model which considers only the electronic energies of the defect to describe a hole capture process (as in SRH-based models). The barrier is determined by the energy difference $\Delta E = E_2 - E_1$ (or E_{21}) which can be **b** negative or **c** positive. **b** When $E_2 < E_1$, no energy barrier has to be overcome for hole capture (i.e., the hole ‘bubbles up’). However, emission of the hole is only possible after sufficient energy has been absorbed from phonons. **c** When $E_2 > E_1$, phonons have to be absorbed for hole capture from E_2 to E_1 , while no energy barrier has to be overcome for hole emission from E_1 to E_2 . After [32]



stress bias is removed. In other words, both the defect charging during stress and discharging after stress are almost independent of the exact defect energy level, but correlate mainly to the general energy location—above or below—with respect to the Fermi level in the channel.

The main issue with the extensions of the SRH theory is that a distribution of characteristic times is obtained only by the spatial tunneling probability. Trapping/detrapping proceeds into the oxide layer with a tunneling front [i.e., the defect charge capture/emission times are correlated with the spatial depth into the oxide, since the tunneling probability is proportional to $\exp(-x/x_0)$] which typically reaches ~ 1 nm in a few ms (Fig. 2.33). In this picture, the capture and emission times are strictly correlated, which is in contrast with the experimental observation highlighted in the previous section. Moreover, modern oxides are simply too thin to support the wide distribution of characteristic times experimentally observed based on tunneling distance alone. Furthermore, it is worth mentioning that the temperature- and bias-dependencies of the capture/emission process as described

Fig. 2.32 Illustration of the impact of the electric field on the energy levels of an oxide defect. At small fields, $E_1 < E_2 = E_v$ (left). Application of a large field moves E_1 above E_2 (right). After [32]



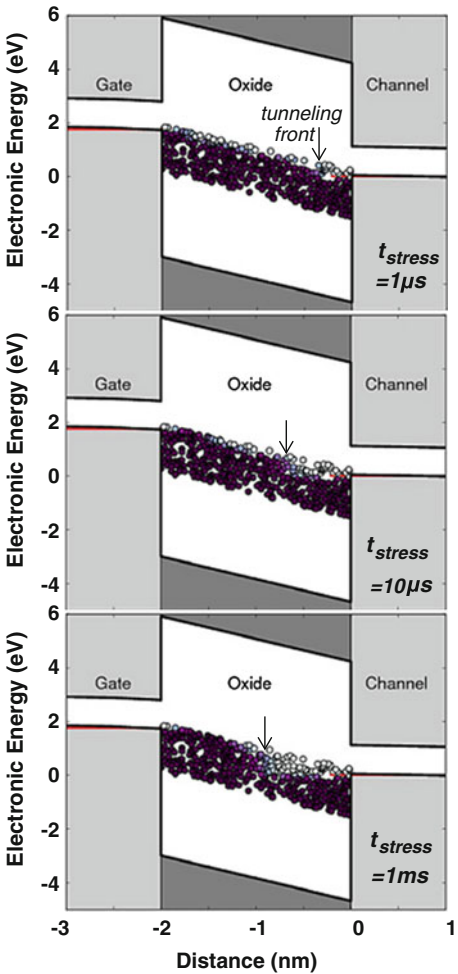
by SRH theories are typically weak compared to the strong dependencies observed experimentally [32]. A more accurate theory is therefore necessary to understand the charge trapping mechanism.

2.2.8 Nonradiative Multiphonon Theory

A fundamental aspect neglected by SRH-based models is the reconfiguration of the atoms at the defect site when its charge state is changed (i.e., that is a structural relaxation). Although the microscopic nature of the charge trapping sites is still not established, the oxygen vacancy/E' center (Fig. 2.34) is the most commonly studied defect in amorphous SiO₂ [33]. Figure 2.35 depicts an oxygen vacancy defect in its neutral and positively charged configurations: the defect charge state determines the exact atomic positions. In particular, upon positively charging the oxygen vacancy, the distance between the silicon atoms increases creating what is known as an E' center. The atomic positions determine the electronic energy level of the state, which is the energy information displayed in a band diagram, as in Fig. 2.32. However, the electronic energy describes only partially the total energy of the system constituted by the atoms at the defect site. The information missing in the band diagram representation is the energy associated with the lattice vibrations. With increasing temperature, the atoms vibrate more vigorously; every displacement from their equilibrium position modifies (i.e., increases) the total energy of the system. This aspect affects the energy barriers associated with charge trapping/detrapping in/from the defect and therefore it strongly influences the charge transition rates.

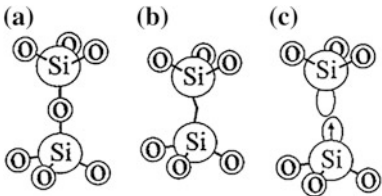
A typically used simplified model for including the vibrational energy in the representation of the defect site is the quantum harmonic oscillator (depicted in Fig. 2.36) [35]. This model is the quantum-mechanical analog of the harmonic oscillator, i.e., a system that, once displaced from its equilibrium position, experiences a restoring force proportional to the displacement. A typical illustrative

Fig. 2.33 Time dependence of the charging/discharging transients according to the SRH model. Prior to stress, it is assumed that all defects are in equilibrium with the Fermi level of the channel and the gate. With increasing stress time (from *top to bottom*), a tunneling front progresses from the interface toward the middle of the device. The situation is analogous during recovery, where the defects are neutralized in a tunneling front progressing from the interface toward the middle of the device. After [32]



example of such system is two masses connected by a spring (also depicted in the sketch of Fig. 2.36). The same energy can be stored into the system by pulling the two masses aside (extension) or pushing them toward each other (contraction). Thanks to this symmetry, the system can be simply represented by a parabola. In the absence of external forces, the system undergoes sinusoidal oscillations about its equilibrium point.

Fig. 2.34 Sketches of **a** a correct bond of two Si atoms to an O atom in SiO_2 ; **b** an oxygen vacancy defect, i.e., a Si-Si bond in SiO_2 ; **c** a positively charged oxygen vacancy, i.e. an E' center



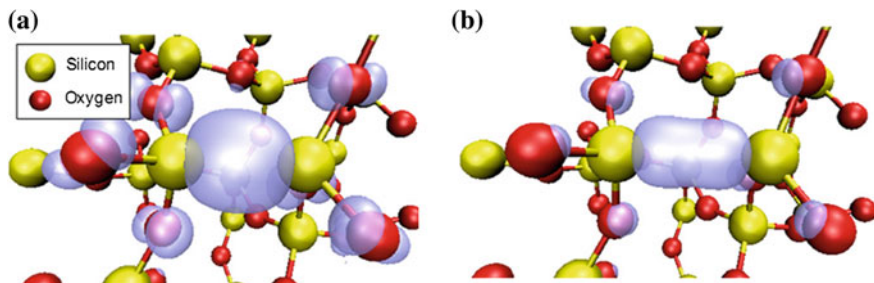


Fig. 2.35 Two charge states of the E' center calculated by density-functional-theory (DFT) [34], **a** neutral and **b** positive. The electron density is shown as blue ‘bubbles’. The atomic equilibrium positions change when the charge state is changed. After [32]

Two different energetic levels, e.g., the valence band in the channel (hole reservoir) and a defect site in the dielectric, can be represented by two parabolas (Fig. 2.37a), while the energy information displayed in a band diagram corresponds only to the equilibrium point (ground state) of the harmonic oscillator. A nonradiative transition (i.e., a transition not involving any external energy exchange by means of a photon emission or absorption) from one state to the other takes place only at the intersection of the two parabolas (i.e. $V_1(q) = V_2(q)$ in Fig. 2.37b).

The nonradiative condition has two immediate consequences: first, the energy barrier associated with a transition from state 1 to state 2 can be significantly different with respect to the energy difference of the two ground states (ε_{21} as compared to $E_2 - E_1$ in Fig. 2.37b) depending on the potential surfaces of the levels (i.e., the parabolas). Second, there is an energy barrier $\varepsilon_{21} - (E_2 - E_1)$ involved also in the transition from state 2 to state 1: this aspect represents a significant modification to

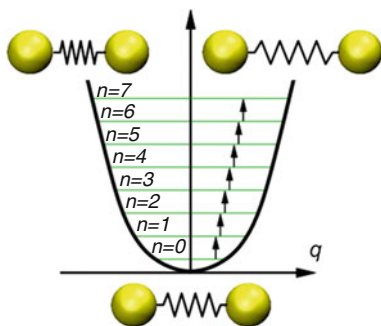


Fig. 2.36 A conceptual sketch of the harmonic oscillator. Once displaced from its equilibrium position, the system experiences a restoring force proportional to the displacement (as two masses connected by a spring). Every displacement from the equilibrium increases the total energy of the system, in discrete energy steps for the quantum version of the model (quantum harmonic oscillator). Thanks to the symmetry between extension or contraction the energy of the system can be represented as a parabola

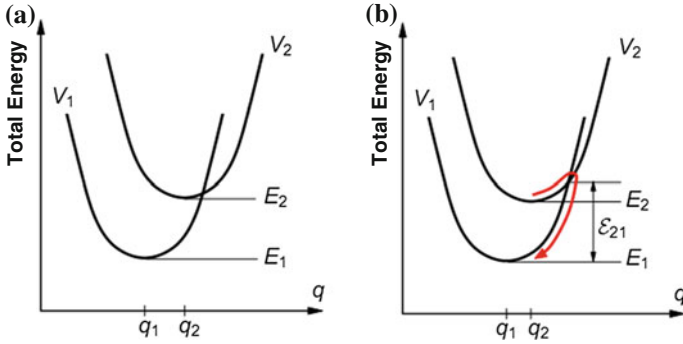


Fig. 2.37 Two different energetic levels represented by two parabolas V_1 and V_2 . The electronic energy information displayed in a band diagram corresponds only to the ground state of the harmonic oscillator. A nonradiative transition can take place only at the intersection of the two parabolas. The necessary energy has to be supplied by phonons. After [32]

SRH-based charge trapping models, which assume no energy barrier to be involved in one of the two possible transitions (see Fig. 2.31). The necessary energy is supplied by phonons.

Let us now consider, within the framework of the Nonradiative Multiphonon (NMP) model, the charge exchange process between allowed energy levels in the oxide bandgap (i.e. defects) and the energy bands in the semiconductor channel (i.e. a charge reservoir), depicted in Fig. 2.38a. An electron is bound at the defect site in the neutral state 1. During the transition to state 2, the electron is transferred to the reservoir (i.e., the channel of the transistor). The potential energy of the electron in state 1 is given by E_1 , while in state 2 it is equal to E_2 . With the energy barrier ϵ_{12} higher than ϵ_{21} , the transition rate from state 1–2 is lower than its reverse rate: as a consequence it is more probable that the electron will stay in state 1 most of the time.

The ground state energy level of the neutral defect increases linearly with the applied electric field (Fig. 2.38a), similarly to the SRH model. However, a superlinear field dependence of the transition rates is predicted by the NMP model since both the forward and backward barriers ϵ_{12} and ϵ_{21} are affected. Moreover, while in the SRH model the barrier E_{21} , and thus the field dependence, enters either the capture or the emission time constant, in the NMP model the bias dependence is equally shared (with different signs) between the time constants.

In conclusion, for a sufficiently accurate modeling of the charge trapping all the different contributions to the transition probability, namely the tunneling probability (WKB), the electronic energy, and the vibrational energy need to be taken into account, with the latter aspect being of fundamental importance for capturing the correct temperature and bias dependence of the trapping process.

A final consideration is that due to the amorphous nature of the dielectric, each defect is expected to be different from the others. Therefore, different parabolas describe the potential surface of the vibrational motion of each defect. As a consequence, the charge trapping in some defects can be significantly more

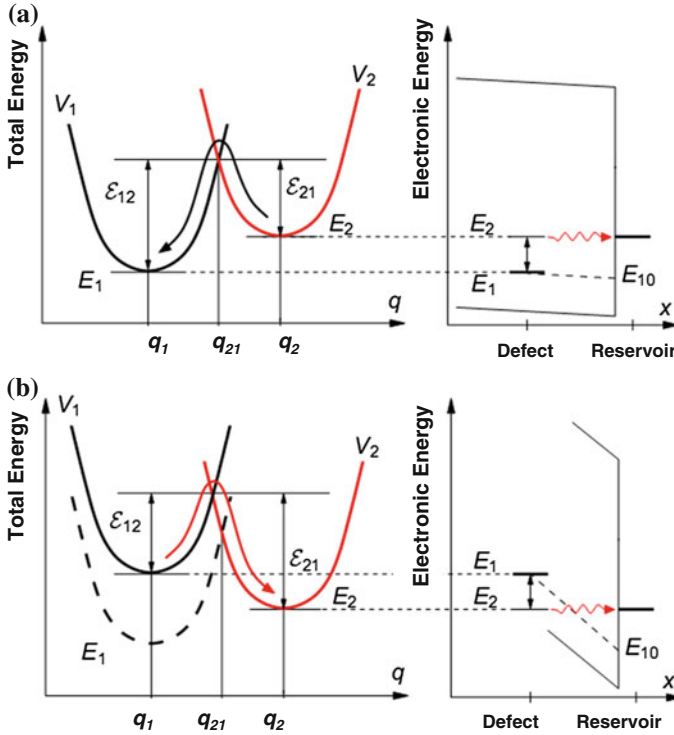


Fig. 2.38 The field dependence of the NMP transition rates is a consequence of the electrostatic shift of the defect level. A superlinear dependence on the electric field is obtained by the modulation of both the forward and reverse barriers. States 1 and 2 are indicated with indices 1 and 2 respectively. After [32]

probable than in other ones. This aspect has an important implication: a tunneling front from the channel toward the bulk of the oxide is not well defined anymore, i.e., capture and emission times can be uncorrelated with the defect depth into the oxide. This is a fundamental aspect for correct NBTI modeling.

2.2.9 A Defect Model for RTN and NBTI

Based on the standard model for E' centers developed in the context of irradiation damage [36, 37], Grassler et al. have suggested a defect model consistent with the experimental TDDS data [26], see Fig. 2.39. The essence of the model is that, for a given structural configuration of the defect site, the charge state of the defect can repeatedly change depending on its energetic level with respect to the charge reservoir. Such defects are known as switching traps. Depending on the defect potential energy surfaces, which are different for each defect in an amorphous

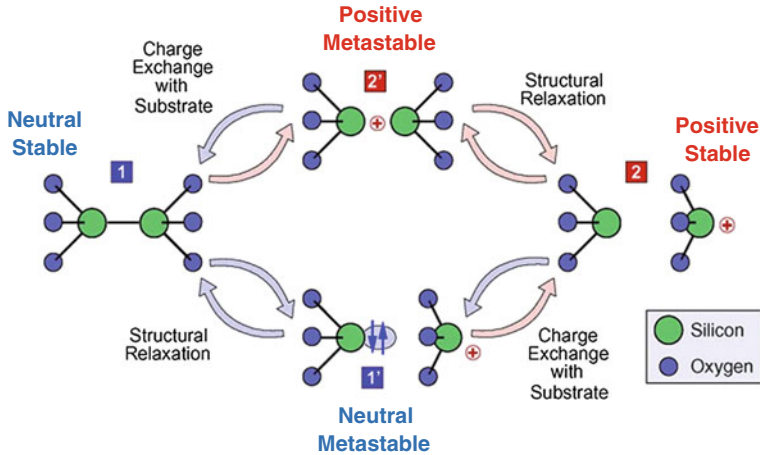


Fig. 2.39 Four states of the defect as extracted from the experimental data [26]. The defect is modeled around reported properties of the E' center. After [15]

oxide, the defect will show a different behavior when monitored under constant (RTN) or dynamic bias conditions (BTI).

In order to capture various anomalies observed by TDDS, this extended model introduces the possibility of having metastable states, represented by a second

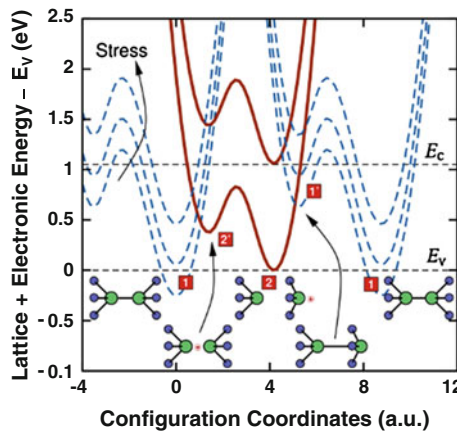


Fig. 2.40 Coordinate diagram showing the total energy potentials corresponding to the four defect configurations. The energy is given relative to the valence band edge E_v . In states 1 and 1', the electron is at the defect site, that is, the electronic contribution to the total energy shifts the potential up with increasing NBT stress (more negative V_G). In state 2' and 2, the electron is either (approximately) at the valence band or conduction band edge (i.e., the defect site is positively charged). The relative position of these potentials (solid curves) with respect to E_v does not change with bias. After [26]

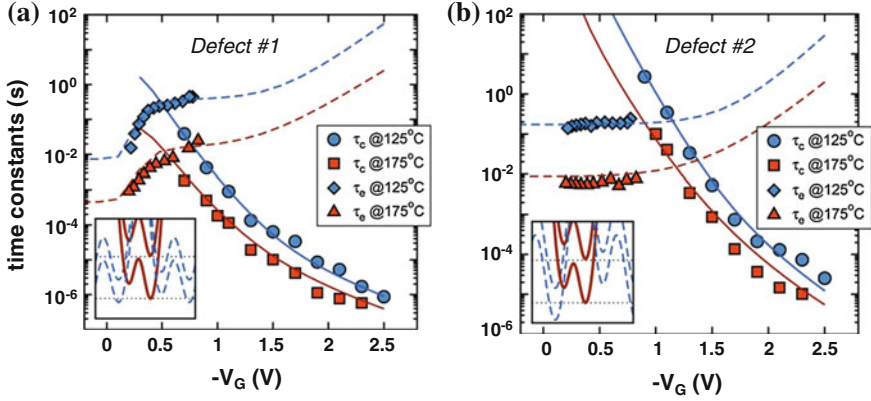


Fig. 2.41 Simulated capture and emission time constants for two different defect configurations, calibrated on experimental data. The energy diagrams shown in the insets determine the switching trap behavior. The strong temperature and bias dependences of the capture times are reproduced, while depending on the defect configuration, both weak and strong bias dependences of the emission times are possible, as observed experimentally. After [15]

parabola merged with the main defect potential energy surface (Fig. 2.40). With this model, Grasser et al. were able to explain the large difference between capture and emission times as well as the extremely strong bias dependence observed experimentally (see Fig. 2.41). Moreover, it is worth mentioning that other experimentally observed anomalies (out of the scope of this chapter) were successfully reproduced by the proposed model, including defect disappearing from the TDDS spectral maps for random amount of time, or defects producing only a temporary RTN after having been positively charged [38], (see Fig. 2.42).

2.2.10 A Two-Stage Model for NBTI

In order to explain their NBTI experimental data, several groups [3, 7] have proposed in the past to separate the NBTI degradation into a recoverable and a permanent component, R and P . As shown in the previous sections (see, e.g., Fig. 2.30a), R clearly dominates over the whole experimental windows, from the microseconds regime and lasting up to days. As a consequence, the characterization of P remains difficult, since its estimation tends to be close to the last measured ΔV_{th} . Huard et al. proposed that the P component might be the contribution to the total ΔV_{th} which is visible in the charge pumping data (see, e.g., Fig. 2.12) [3]. However, Rangan et al. showed a strong recovery of the charge pumping current, while a complete annealing of NBTI degradation was obtained at high temperatures (~ 300 °C) [39]. These results suggested that the recovery of the P component is possible in a thermally activated process, just like the recovery

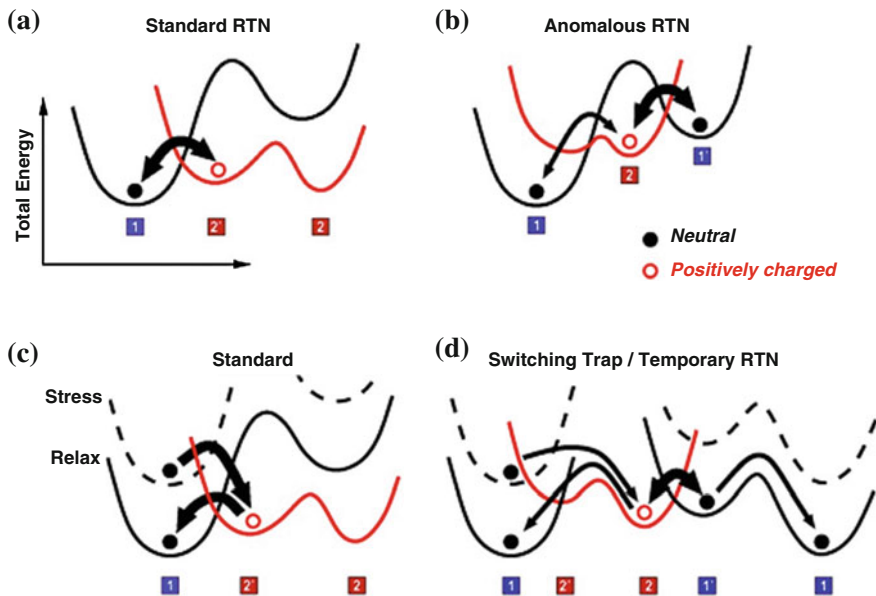


Fig. 2.42 Schematic configuration coordinate diagram model for a general defect [26]. In an amorphous oxide, each defect will have different surface potentials determining the NMP process. Depending on the bias conditions, the relative position of the potentials changes, causing a strong field dependence of the transition rates (*thicker arrows* indicate larger transition probabilities). The general defect model including the metastable states as proposed by Grasser et al. can capture both standard observations and experimental anomalies: **a** standard RTN, switching between the neutral and the positively charged defect state; **b** anomalous RTN: a larger transition rate is observed when the defect switches from its positively charged state (2) to its neutral metastable state (I'), while a reduced transition rate is observed when the switching occurs between the positively charged state and the neutral stable state (I), causing the defect to disappear from the TDDS spectral maps for random amount of time; **c** standard NBTI: large transition rate from neutral to positive state during stress and large rate of the reverse transition during relaxation; **d** temporary RTN: defects producing RTN only after having been positively charged. After [15]

of R . As a consequence, the separation of NBTI degradation into R and P appears arbitrary. Two possibilities are still open: either R and P correspond to two different microscopic defects, or they might be a consequence of the wide distribution of time constants of a single or two tightly coupled components (see Fig. 2.22).

In favor of the latter option, Grasser et al. have recently observed that oxide defects can also contribute to the charge pumping current [16]. In order to explain the larger time constants associated with P , the defect might undergo a transformation which could be still fully reversible. In such a scenario, the expression ‘permanent component’ would assume the meaning of a fraction of degradation with time constants outside of the typical measurement windows (see Fig. 2.30).

The same group of authors has recently proposed a two-stage model (Fig. 2.43) for NBTI which is able to capture a large number of the experimental features [40]. The degradation is assumed to proceed in two coupled stages. For the first stage, a

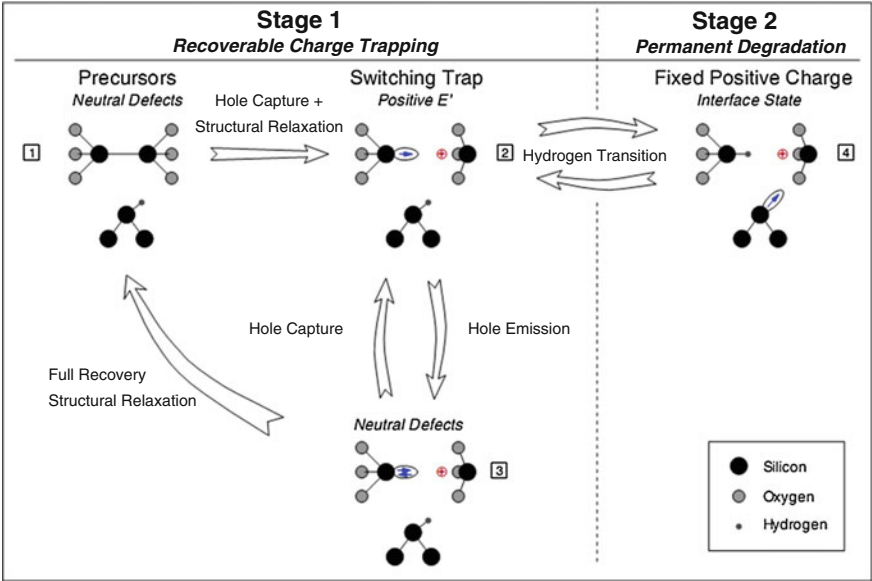


Fig. 2.43 The two-stage model for a switching oxide trap coupled to the creation of a dangling bond at the interface as suggested by Grasser et al. [18, 40]. When the E' center is positively charged (in state 2), the hydrogen passivating a silicon dangling bond at the interface can move to the E' center, thereby effectively locking in the positive charge (state 4). The charge state of the thereby created dangling bond depends on the position of the Fermi level

simplified version of the defect model shown in Fig. 2.39 (i.e., without the metastable positive state 2') was used. The two-stage model assumes that a charge is trapped in an E' center, creating a switching trap. Once the charge is trapped in the defect, the defect might transform into a more permanent state P via hydrogen exchange with a P_b center (i.e. a Si–H bond) at the interface.

Fig. 2.44 The two-stage model proposed by Grasser et al. [40] has been shown to capture a number of crucial features of NBTI. E.g. the asymmetry between stress and recovery, with recovery taking much longer than degradation is properly reproduced by the model

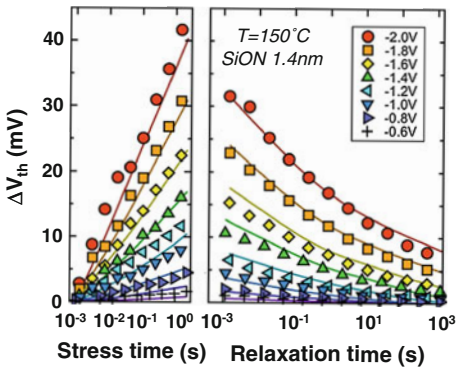
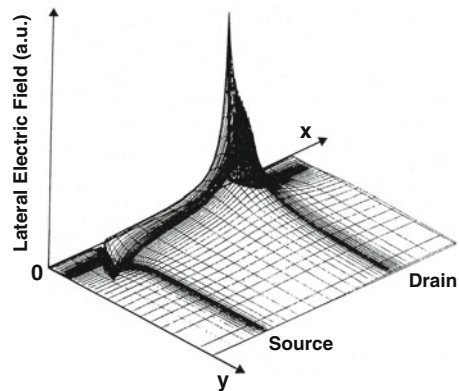


Fig. 2.45 Simulated lateral electric field peak at the drain side of a MOSFET operated in saturation mode. After [41]



This model was shown to capture the crucial NBTI aspects, including the asymmetry between stress and relaxation (see Fig. 2.44), the acceleration of recovery at positive bias or temperature switches (not shown).

2.3 Hot Carriers

Hot carriers (HC) are charge carriers that gain a high kinetic energy from being accelerated by a high electric field. As such, their energy is not in thermal equilibrium with the semiconductor lattice (thus the definition of ‘hot’ carriers). Some of these energetic carriers can generate electron–hole pairs by impact ionization or can get injected into the gate oxide where they can become trapped, generate new interface states, or reach the gate and contribute to the gate current.

As discussed in Chap. 1, with the aggressive scaling of the MOSFET technology, the electric fields involved are ever increasing. In particular, the scaling of the channel length toward nanoscale dimensions has caused the lateral electric field in the channel at $V_{DS} = V_{DD}$ to reach the ~ 1 MV/cm range. Figure 2.45 depicts the lateral electric field peak reached in the pinch-off region at the drain side of a MOSFET operated in saturation (i.e., $V_{DS} > V_{GS} - V_{th}$). Such electric field peak enhances the probability of hot carrier generation, as discussed later on. Several approaches have been introduced over the years to limit the generation of hot carriers by engineering field reducing regions close to the device junctions (e.g., the Lightly Doped Drain, LDD). However, as we will discuss in Chap. 6, some of the solutions introduced to limit other short channel effects (SCE), such as V_{th} roll-off and punch-through leakage current, can have a detrimental effect on hot carrier generation (e.g., halo counter doping for abrupt junctions).

2.3.1 Hot Carrier Typology

Four different kinds of hot carrier generation are possible [42]:

- Channel Hot Carriers (CHC) are heated up in the channel near the drain side of a MOSFET operated at $V_{GS} = V_{DS}$. As schematically shown in Fig. 2.46a, carriers flowing from source to drain without suffering any energy losing collision (“lucky electrons”) might get sufficient energy to surmount the Si/SiO₂ barrier and get injected into the gate oxide.
- Drain Avalanche Hot Carriers (DAHC) are due to the high electric field near the drain region promoting avalanche multiplication. The carriers gain enough energy while flowing in the channel so that they produce electron–hole pairs by impact ionization, which in turn produces further electron–hole pairs, resulting in an avalanche process (Fig. 2.46b). These avalanche hot electrons and hot holes are injected into the gate oxide. This condition is mostly observed at $|V_{DS}| > |V_{GS}| > |V_{th}|$. It is often considered as the most severe device degradation as both holes and electrons might get injected into the gate oxide.
- Secondary Generated Hot Carriers (SGHC) are the secondary minority carriers originating from secondary impact ionization of the substrate current. It occurs

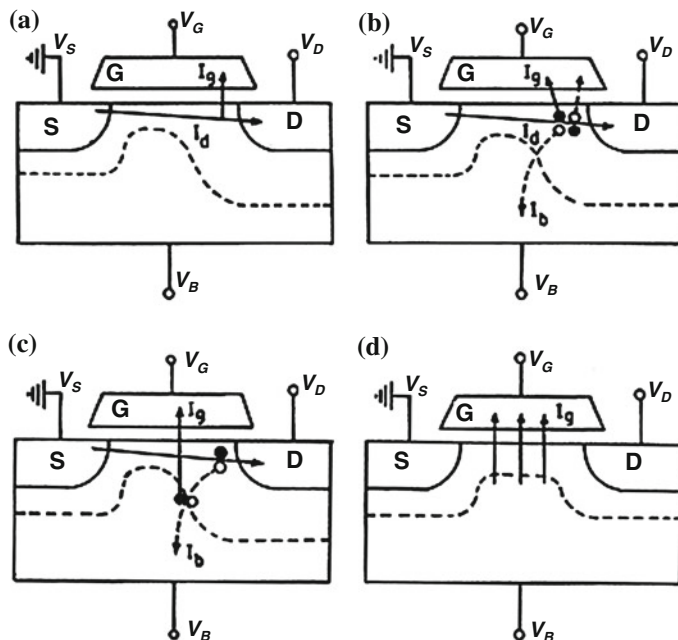


Fig. 2.46 Sketches of the four hot carrier generation modes: **a** Channel Hot Carriers (CHC); **b** Drain Avalanche Hot Carriers (DAHC); **c** Secondary Generated Hot Carriers (SGHC); and **d** Substrate Hot Carriers (SHC). After [42]

when the substrate current, produced by avalanche effect near the drain, generates further electron–hole pairs (Fig. 2.46c).

- Substrate Hot Carriers (SHC) are thermally generated or injected carriers (i.e., by means of an external p–n junction injector structure adjacent to the MOSFET channel). SHC injection occurs when a large body bias is applied, independently of the drain bias (Fig. 2.46d). The carriers gain energy from the high field in the surface depletion region. This mechanism is often exploited for hot carrier reliability accelerated tests. Its advantage is that the energetic carriers at the interface are uniformly distributed along the channel, in contrast to the other generation mechanisms, whereas the hot carrier effects are localized near the drain end of the channel. Moreover, more control over the stress conditions is possible by independently adjusting several parameters: the oxide field is controlled by the gate voltage; the potential drop in the substrate and thus the energy of the injected carriers are controlled by the body bias; while the carrier availability is controlled by the injected current level. This particular methodology will be used in a study presented in Chap. 6.

Typically, HC effects have been studied mostly in nMOS devices, with the degradation in pMOSFETs being limited thanks to the higher barrier for holes at the Si–SiO₂ interface and to the lower efficiency of holes in generating electron–hole pairs. However, the use of small bandgap semiconductors for high-mobility channel pMOSFET (e.g., the Ge-based technologies studied in this chapter) is expected to favor the impact ionization process. As a consequence, HC might constitute a severe reliability threat for novel pMOSFET technologies.

2.3.2 First-Order Modeling of Channel hot Carrier Generation

The probability that a channel carrier is accelerated to a given energy Φ has been described as an exponential distribution (e.g., [42]):

$$f(\Phi) = \exp\left(-\frac{\Phi}{q\lambda E}\right), \quad (2.16)$$

where E is the accelerating electric field and λ is the carrier mean free path (in the order of a few nm for electrons in Si at room temperature). It is worth to note here that, contrary to BTI, reducing hot carrier degradation has been sometimes reported for increasing stress temperature in non-aggressively scaled technologies. This observation has been ascribed to enhanced collision probability reducing the carrier mean free path. Eq. (2.16) is schematically plotted in Fig. 2.47.

The source current provides a direct measurement of the total number of carriers available to potentially become “hot”. Carriers with energy larger than the semiconductor bandgap might generate an electron–hole pair by impact ionization,

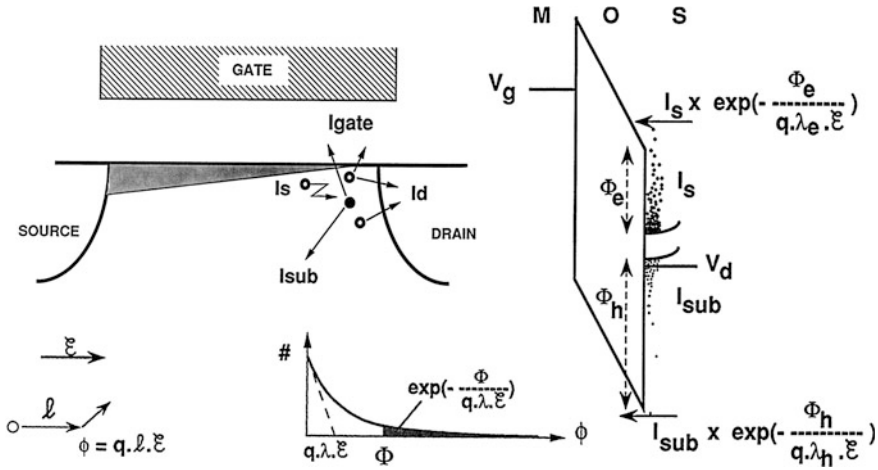


Fig. 2.47 Sketch depicting HC generation at the drain end of a MOSFET operated in saturation. The carrier energy follows an exponential distribution. In the presence of a higher electric field, more highly energetic carriers are expected to be found. The temperature-dependent carrier mean free path also affects the carrier energy distribution. Carriers having a sufficient energy (larger than the semiconductor bandgap) can generate electron-hole pairs by impact ionization, while very highly energetic carriers can directly overcome the barriers between the semiconductor and the insulating layer. After [42]

while carriers with an energy higher than the barrier between the semiconductor and the oxide can be directly injected into the dielectric layer and reach the gate through transport through the oxide conduction or valence bands. The substrate current can be used as a monitor of the impact ionization, while the ratio of the substrate and the source currents (often called the multiplication factor $M = I_{sub}/I_s$) can be used as an indirect measurement of the electric field generating hot carriers through Eq. (2.16).

The hot carrier-induced gate current has been described with the “lucky electron” model. The term “lucky” pertains to an electron that survives several phenomena before being injected toward the gate. In order to calculate the probability of an electron reaching the gate, several probabilities have to be convoluted:

- the probability of acceleration of a carrier to a given energy;
- the probability of quasi-elastic scattering, i.e., hot carrier redirection towards the interface without losing energy;
- probability of collision-free travel of the electron to the point of maximum potential barrier;
- probability of scattering in the oxide image potential well.

For scaled V_{DD} technologies, refinement to the lucky electron model have been proposed to explain why hot carrier effects are still present when qV_{DD} is lower than the energy barriers toward the gate oxide or even lower than the impact ionization barrier (i.e., energy bandgap) in the substrate. Energy-redistribution

mechanisms have been introduced, including electron–phonon [43] and electron–electron interactions [44], to support the possibility of an extended tail of the carrier energy distribution. Furthermore, Multi Vibrational Excitation (MVE) model of Si–H bond breaking at the interface have been introduced to reproduce HC degradation in scaled CMOS nodes, where a direct temperature dependence similar to BTI has been observed. [45].

2.3.3 Hot Carrier Degradation

The injection of hot carriers into the gate oxide of MOSFETs causes charge trapping in the oxide- and interface-trap generation. As we have discussed for NBTI in Sect. 2.2, the presence of interfacial and bulk charge in the gate oxide affects the I – V characteristics of the device, inducing a shift of the threshold voltage, of the subthreshold slope, and of the transconductance. These parameter shifts can be used as degradation monitors.

It is, however, important to note that for hot carriers generated by lateral field in the channel, the degradation is localized at the drain side where the electric field peaks. Such localized degradation induces an asymmetric shift in the MOSFET I – V characteristic depending on the direction of the current flow in the channel, as illustrated in Fig. 2.48.

To properly interpret the I – V characteristic degradation under nonuniform carrier injection, nonuniform spatial distributions of interface state density and trapped-oxide charge need to be considered. Due to this complication in the

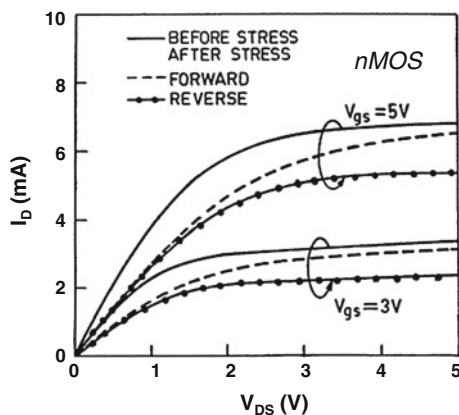


Fig. 2.48 I_D – V_D characteristics of a nMOSFET device before and after a HC stress. In the saturation regime ($V_{DS} > V_{GS}$), the shift of the characteristic depends on the direction of the current flow in the channel. Maximum sensitivity to the induced degradation is observed when the damaged region is located at the source side of the pinched-off channel (i.e., reversed S/D current flow). After [46]

interpretation of the degradation in the drain current characteristics, other characterization techniques, such as modified lateral-profiling charge pumping are often used [47].

In order to estimate the device time-to-failure (τ) under Hot Carrier stress, several device parameter shifts have been alternatively used in literature as failure criteria, including fixed threshold voltage shift ΔV_{th} , drain current reduction in the linear regime ΔI_{Dlin} , transconductance degradation Δg_m , or charge pumping current increase ΔI_{cp} [48]. Moreover, several acceleration models have been proposed in order to estimate the device lifetime at operating conditions from accelerated stress tests. Two of the most commonly used acceleration models were proposed by Takeda et al. [49], and by Hu et al. [50] (Fig. 2.49). The first empirical one assumes an exponential dependence of τ on the drain stress voltage as:

$$\tau \propto \exp\left(\frac{1}{V_D}\right), \quad (2.17)$$

while the second one assumes the cumulative charge flowed through the channel before the device failure (i.e. $\tau \times I_D$ [C]) to follow a power-law of the current multiplication factor ($M = I_{sub}/I_S$):

$$\tau \times I_D \propto \left(\frac{I_{sub}}{I_S}\right)^{-m}. \quad (2.18)$$

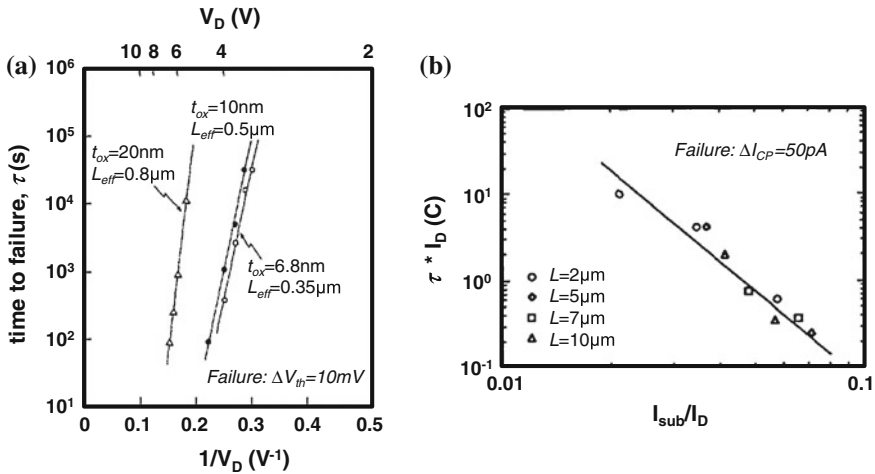


Fig. 2.49 **a** HC acceleration model proposed by Takeda et al.: the device time-to-failure follows an exponential dependence on $1/V_D$. After [49]. **b** Acceleration model proposed by Hu et al. [50]: the cumulative charge flowed in the channel before the device failure follows a power-law of the multiplication factor. After [48]

2.4 Time-Dependent Dielectric Breakdown

The dielectric breakdown phenomenon is the sudden loss of insulating properties of the dielectric layer caused by an electrical stress [51]. Although this degradation mechanism is out of the main scope of this work, in the following a limited treatment is given in order to provide the reader with sufficient information to understand the experimental data presented in Sect. 6.6.2.

The oxide breakdown is commonly assumed to be caused by the formation of a conducting path between the gate and the semiconductor substrate due to charge injection. This failure mechanism constitutes a serious reliability concern for the MOSFET technology due to continuous down scaling of the oxide thickness, which has been the first among the other device dimensions to reach the nanometer scale.

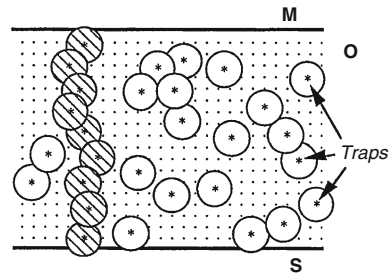
Time-Dependent Dielectric Breakdown (TDDB) failures are commonly divided into two categories: extrinsic and intrinsic. The former involves a breakdown occurring early in the life of the device and is related to the presence of weak spots in the oxide layer arising from a poor processing.

However, even defect-free oxides can undergo an *intrinsic* breakdown during the device operation due to an electrical stress-induced generation of oxide traps. The mechanism leading to intrinsic breakdown is generally understood as follows: charge injection into the oxide generates oxide traps which provide additional trap-assisted leakage paths and therefore cause a gradual increase of the gate current (Stress-Induced Leakage Current, or SILC). Traps are generated randomly in the oxide until a conduction path from the substrate to the gate is formed through the traps (see Fig. 2.50) causing an abrupt increase of the gate current.

One of the most widely used TDDB tests (used also in this work, for an experiment presented in Chap. 6) consists in applying a constant voltage to the gate of an MOS structure while measuring the gate leakage current versus time (I - t traces, see Fig. 2.51). A breakdown is easily visualized as a sudden increase of several orders of magnitude of the gate leakage current (see Fig. 2.51a). However, upon a closer inspection (see Fig. 2.51b) several phenomena are visible, namely:

- a gradual increase of the leakage current in a first stage (SILC);
- the first discrete step in the gate leakage current, corresponding to the formation of the first percolation path through the oxide ('Soft Breakdown' or SBD);

Fig. 2.50 Sketch of the commonly considered percolation model for TDDB: due to charge injection, traps are generated randomly in the oxide until a conduction path is formed. After [51]



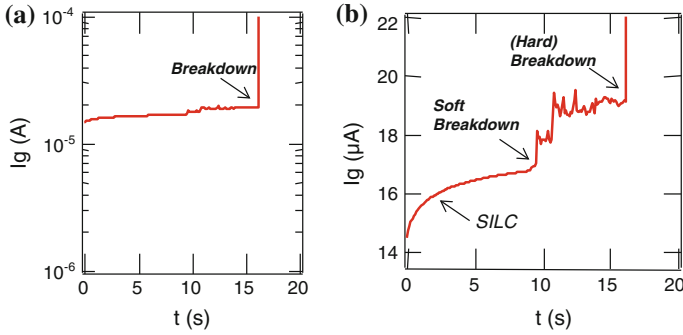


Fig. 2.51 **a** Constant voltage stress (CVS) methodology for TDDDB assessment. The MOS structures are biased at a constant stress voltage while the gate leakage current is measured as a function of the stress time. The breakdown is visible as a sudden significant increase of the current. **b** Zoom-in: a gradual increase of the leakage (SILC) is typically visible at short stress time, followed by the first discrete increase (*Soft Breakdown*) and later on by the complete loss of the insulating properties (*Hard Breakdown*) [52]

- a ‘wear-out’ phase, during which the localized current flow through the conduction path causes a thermal runaway which leads to:
- a sudden complete loss of the insulating properties of the oxide (i.e., the *real* oxide breakdown, often labeled as ‘Hard Breakdown’ or HBD).

A schematic representation of these stages and the corresponding gate leakage characteristics is given in Fig. 2.52.

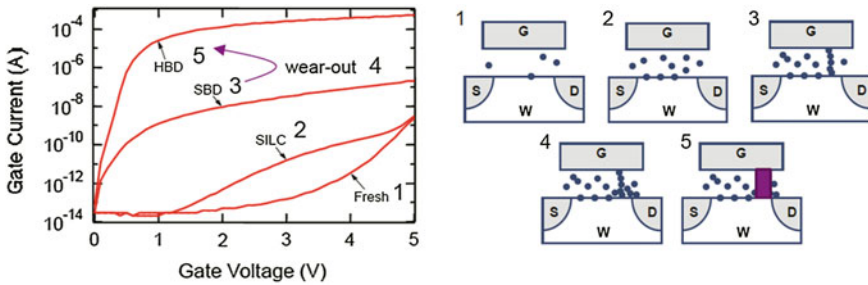


Fig. 2.52 Schematic representation of the different stages of the breakdown mechanism. Due to charge injection traps are generated randomly in the oxide, contributing additional trap-assisted leakage paths: the gate current gradually increases (SILC, stage ‘2’) as compared to the leakage measured on the fresh device (stage ‘1’). When a conduction path from the substrate to the gate is formed through the traps, the gate leakage increases abruptly (SBD, stage ‘3’). In the following “wear-out” phase (stage ‘4’), the localized current flow through the conduction path causes a thermal runaway which leads to a sudden complete loss of the insulating properties of the oxide (HBD, stage ‘5’). After [53]

Figure 2.53a shows a histogram of the measured time-to-breakdown for a population of large area (2.4 cm^2) MOS capacitors: both the extrinsic and intrinsic failure modes are visible, with a small part of the device population failing at very short stress times. However, the extrinsic failure mode is only visible when large area devices are used, i.e., maximizing the probability of having a locally defective region in several devices.

The time-to-breakdown in a device population is a statistically distributed quantity, obeying Weibull statistics. The cumulative density function of the Weibull distribution is:

$$F(t) = 1 - \exp \left[- \left(\frac{t}{\eta} \right)^\beta \right], \quad (2.19)$$

which can be rewritten as:

$$\ln[-\ln[1 - F(t)]] = \beta \ln(t) - \beta \ln(\eta) \quad (2.20)$$

Equation (2.20) describes the Weibull plot of measured times to breakdown in a device population, where represents the time at which 63 % of the devices have already shown a breakdown. As such, η is often marked as the time-to-breakdown t_{BD} . However, t_{BD} can be easily rescaled to lower percentiles (typically 0.01 % device failing) for process qualification purposes.

Fig. 2.53 **a** Histogram of a typically measured time-to-breakdown dataset: the extrinsic failure mode at short stress times concerns a small part of the device population, while the main part of the population fails later on (intrinsic mode). **b** Weibull plot of the measured t_{BD} values: the extrinsic and intrinsic failure modes are visible as two different slopes. After [54]

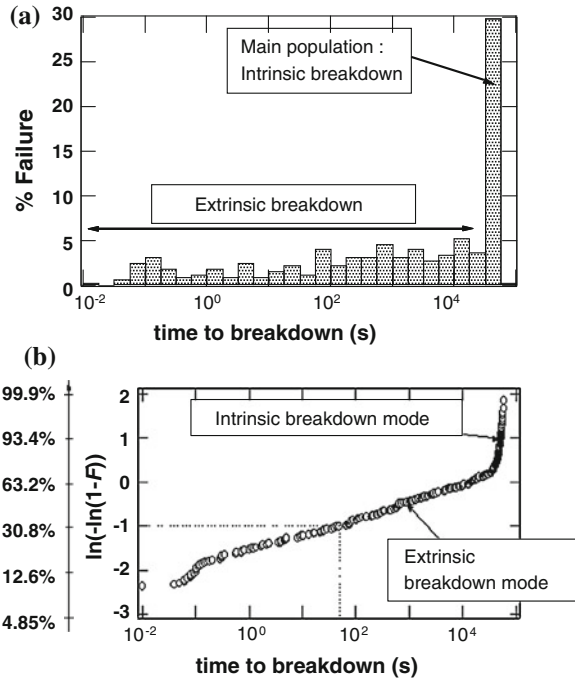
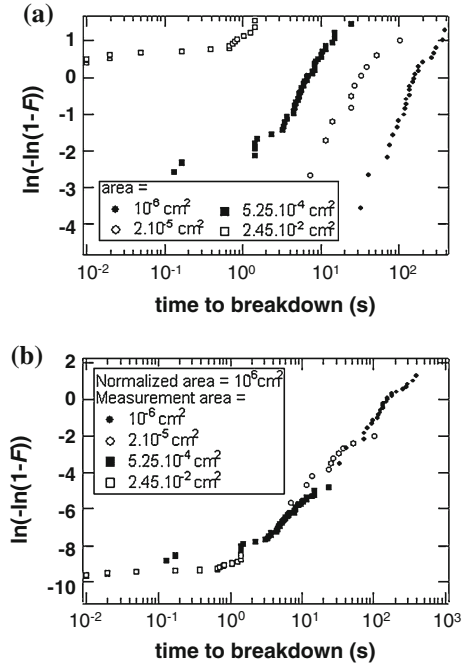


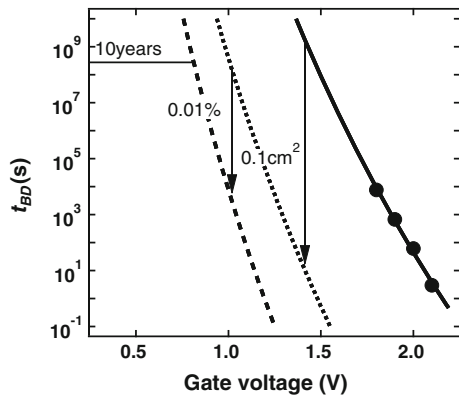
Fig. 2.54 a Time-to-breakdown distributions measured on device sets with different areas, plotted in a Weibull plot. Larger area devices show reduced average t_{BD} . **b** Rescaling the distributions to a single area yields a single distribution where the Weibull parameters (η and β) can be fitted. After [51]



In Fig. 2.53b, a measured time-to-breakdown dataset is shown in a cumulative density plot with a Weibull probit axis: two slopes appear, corresponding to the two failure modes.

The average time-to-failure for devices with different areas can be understood considering the large area device as a series reliability system made of devices with smaller area. With TDDB being a weak spot phenomenon (i.e., the first failing element causes a failure of the whole system), the time-to-failure scales inversely

Fig. 2.55 Average t_{BD} ($=\eta$) estimated for different gate stress voltage and temperature. Using an acceleration model ($\sim V_G^m$ in the example), the maximum allowed gate voltage for 10 year reliable operation is extrapolated. Area- and percentile-scaling are also shown



with the device area (see Fig. 2.54a). The distribution measured on device sets with different areas can be rescaled using the relation:

$$\ln[-\ln[1 - F_{A1}(t)]] = \ln\left(\frac{A_1}{A_2}\right) + \ln[-\ln[1 - F_{A2}(t)]] \quad (2.21)$$

Thanks to this property, multiple TDDB distributions measured on several device areas are rescaled onto a single distribution, which is then used to obtain a more robust estimation of the Weibull parameters (see Fig. 2.54b). Moreover, the area-scaling property allows to reduce test time using large area DUTs and then rescaling the estimated time-to-failure to realistic device areas. This approach is particularly useful for low stress voltage (close to operating voltage) tests.

A typical TDDB accelerated testing methodology includes the statistical evaluation of the average time-to-breakdown t_{BD} for different gate stress voltages and/or stress temperatures (Fig. 2.55), which can be subsequently rescaled to operating conditions using an acceleration model. Several laws for the bias stress have been proposed in the literature over the years, including exponential and power-law dependencies on $-1/E_{ox}$, $-E_{ox}$, and $-V_G$. For contemporary ultrathin oxides, a power-law of V_G is typically used [54].

2.5 Summary of this Chapter

A general description of the MOSFET degradation mechanisms considered in this work has been given in this chapter, and it will serve as a basis for the discussion of the original results presented in this work.

In the first part of the chapter, we have discussed in detail the NBTI mechanism which pertains to most of the original experimental work on SiGe and Ge channel devices presented later in Chaps. 4 and 5. A phenomenological description of the NBTI degradation and its distinctive relaxation process has been given first. Next, we have reviewed the classic Hydrogen Reaction–Diffusion (R–D) model and we have discussed how recent experimental observations about NBTI pose significant doubts about the correctness of the R–D model.

We have therefore given an overview of a recent modeling attempt focused on the central role of hole trapping in NBTI. The proponents of this approach pointed out that the standard Shockley–Read–Hall theory does not correctly capture the properties of hole trapping into bulk oxide defects. On the contrary, the Nonradiative Multiphonon Theory (NMP) which includes the energetic description of the vibrational motion of the atoms at the defect sites is capable of correctly reproducing the observed trapping behavior. As we have discussed, these insights have recently led to an oxide defect model compatible with a wide range of NBTI observations.

In the second part of this chapter, we have given a general overview of the Hot Carrier degradation mechanism and of the Time-Dependent Dielectric Breakdown.

This overview will serve as a reference for the discussion of the experimental results presented in [Chap. 6](#).

References

1. K.O. Jeppson, C.M. Svensson, Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices. *J. Appl. Phys.* **48**(5), 2004–2014 (1977)
2. B.E. Deal, M. Sklar, A.S. Grove, E.H. Snow, Characteristics of the surface-state charge (Q_{ss}) of thermally oxidized silicon. *J. Electrochem. Soc.* **114**(3), 266–274 (1967)
3. V. Huard, M. Denais, C. Parthasarathy, NBTI degradation: from physical mechanism to modeling. *Microelect. Reliab.* **46**(1), 1–23 (2006)
4. S. Tsujikawa et al., Negative bias temperature instability of pMOSFETs with Ultra-thin SiON gate dielectrics, in *IEEE Proceedings of IRPS* 2003, pp. 183–188
5. G. Groeseneken, H.E. Maes, N. Beltran, R.F. De Keersmaecker, A reliable approach to charge pumping measurements in MOS transistors. *IEEE Trans. Electron Dev.* **31**(1), 42–53 (1984)
6. M. Aoulaiche et al., Contribution of fast and slow states to negative bias temperature instabilities in Hf_xSi_(1-x)ON/TaN based pMOSFETs. *Microelect. Eng.* **80**, 134–137 (2005)
7. B. Kaczer et al., Ubiquitous relaxation in BTI stressing—new evaluation and insights, in *IEEE Proceedings of IRPS* (2008), pp. 20–27
8. H. Reisinger et al., Analysis of NBTI degradation- and recovery-behavior based on ultra-fast V_{th}-measurements. in *IEEE Proceedings of IRPS* (2006), pp. 448–453
9. M.A. Alam, A critical examination of the mechanics of dynamic NBTI for pMOSFETs, in *IEEE Proceedings IEDM* (2003), pp. 345–348
10. S. Chakravarthi, A.T. Krishnan, V. Reddy, C. F. Machala, S. Krishnan, A comprehensive framework for predictive modeling of negative bias temperature instability, in *IEEE Proceedings of IRPS* (2004), pp. 273–282
11. S. Mahapatra et al., Negative bias temperature instability in CMOS devices. *Microelect. Eng.* **80**, 114–121 (2005)
12. H. Kufluoglu and M.A. Alam, A geometrical unification of the theories of NBTI and HCI time-exponents and its implications for ultra-scaled planar and surrounded-gate MOSFETs, in *IEEE Proceedings of IEDM* (2004), pp. 113–116
13. M.A. Alam, S. Mahapatra, A comprehensive Model of PMOS NBTI degradation. *Microelect. Reliab.* **45**(71–81), 71–81 (2005)
14. T. Grasser, W. Goes, V. Sverdlov, and B. Kaczer, “The universality of NBTI relaxation and its implications for modeling and characterization,” in *IEEE Proc. IRPS*, pp. 268–280, 2007
15. T. Grasser et al., The paradigm shift in understanding the bias temperature instability: from reaction-diffusion to switching oxide traps. *IEEE Trans. Electron Dev.* **58**(11), 3652–3666, (2011)
16. T. Grasser et al., The ‘permanent’ component of NBTI: composition and annealing, in *IEEE Proceedings of IRPS* (2011), pp. 605–613
17. C. Shen et al., Characterization and Physical Origin of Fast V_{th} Transient in NBTI of pMOSFETs with SiON Dielectric, in *IEEE Proceedings of IEDM* (2006), pp. 1–4
18. T. Grasser and B. Kaczer, “Evidence that two couple mechanisms are responsible for Negative Bias Temperature Instability. *IEEE Trans. Electron Dev.* **56**(5), 1056–1062 (2009)
19. T. Grasser et al., Recent advances in understanding the bias temperature instability, in *IEEE Proceedings of IEDM* (2010), pp. 82–85
20. V. Huard et al., NBTI Degradation: from transistor to SRAM arrays, in *IEEE Proceedings of IRPS* (2008), pp. 289–300

21. B. Kaczer et al., Origin of NBTI variability in deeply Scaled pFETs, in *IEEE Proceedings of IRPS* (2010), pp. 26–32
22. H. Reisinger, T. Grasser, W. Gustin, C. Schlünder, The statistical analysis of individual defects constituting NBTI and its implications for modeling DC- and AC-stress, in *IEEE Proceedings of IRPS* (2010), pp. 7–15
23. B. Kaczer et al., NBTI from the perspective of defect states with widely distributed time scales, in *IEEE Proceedings of IRPS* (2009), pp. 55–60
24. A. Asenov, R. Balasubramaniam, A.R. Brown, J.H. Davies, RTS Amplitude in Decananometer MOSFETs: 3-D simulation study. *IEEE Trans. Electron Dev.* **50**(3), 839–845 (2003)
25. <http://www.ibiblio.org/e-notes/Perc/contour.htm>
26. T. Grasser et al., The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability, in *IEEE Proceedings of IRPS* (2010), pp. 16–25
27. H. Reisinger et al., Understanding and modeling AC NBTI, in *Proceedings of IRPS* (2011), pp. 597–604
28. T. Grasser et al., Analytic modeling of the bias temperature instability using capture and emission time maps, in *Proceedings of IEDM* (2011), pp. 618–621
29. R.N. Hall, Electron-hole recombination in Germanium, *Phys. Rev.* **87**, 387 (1952)
30. W. Shockley, W.T. Read, Statistics of the recombinations of holes and electrons. *Phys. Rev.* **87**, 835–842 (1952)
31. A.L. McWhorter, 1/f noise and germanium surface properties, in *Semiconductor Surface Physics*. pp. 207–228 (1957)
32. T. Grasser, Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities. *Microelect. Reliab.* **52**, 39–70 (2012)
33. P.M. Lenahan, Atomic scale defects involved in mos reliability problems. *Microelect. Eng.* **69**, 173–181 (2003)
34. F. Schanovsky, W. Goes, T. Grasser, Multiphonon hole trapping from first principles. *J. Vac. Sci. Technol. B* **29**(1), 01A2011–01A215 (2011)
35. T. Grasser, Charge trapping in oxides—from RTN to NBTI, *IEEE IRPS Tutorial* (2011)
36. J.F. Conley Jr., P.M. Lenahan, A.J. Lelis, T.R. Oldham, Electron spin resonance evidence that E'_{γ} centers can behave as switching oxide traps. *IEEE Trans. Nucl. Sci.* **42**(6), 1744–1749 (1995)
37. A.J. Lelis, T.R. Oldham, Time dependence of switching oxide traps. *IEEE Trans. Nucl. Sci.* **41**(6), 1835–1843 (1994)
38. M. Uren, M.J. Kirton, S. Collins, Anomalous telegraph noise in small-area silicon metal-oxide-semiconductor field-effect transistors. *Phys. Rev. B*, **37**(14), 8346–8350 (1988)
39. S. Rangan, N. Mielke, E.C.C. Yeh, Universal recovery behavior of negative bias temperature instability, in *IEEE Proceedings of IEDM* (2003), pp. 341–344
40. T. Grasser et al., A two-stage model for negative bias temperature instability, in *IEEE Proceedings of IRPS* (2009), pp. 33–44
41. K.R. Hofman, C. Werner, W. Weber, G. Dorda, Hot-Electron and hole-emission effects in Short n-channel MOSFETs. *IEEE Trans. Electron Dev.* **32**(3), 691–699 (1985)
42. N. Arora, MOSFET modeling for VLSI simulation: theory and practice. (World Scientific Publishing Co. Pte. Ltd., 1992)
43. A. Lacaita, Why the effective temperature of the hot electron tail approaches the lattice temperature. *App. Phys. Lett.* **59**(13), 1623–1625 (1991)
44. M.V. Fischetti and S.E. Laux, Monte Carlo study of sub-band-gap impact ionization in small Silicon field-effect transistors. in *IEEE Proceedings of IEDM* (1995), pp. 305–308
45. A. Bravaix et al., Hot-Carrier acceleration factors for low power management in DC-AC stressed 40 nm NMOS node at high temperature, in *IEEE Proceedings of IRPS* (2009), pp. 531–548
46. T.-C. Ong, P.K. Ko, C. Hu, Hot-carrier current modeling and device degradation in surface-channel p-MOSFETs. *IEEE Trans. Electron Dev.* **37**, 1658–1666 (1990)

47. M.G. Ancona, N.S. Saks, D. McCarthy, Lateral distribution of hot-carrier-induced interface traps in MOSFETs. *IEEE Trans. Electron Dev.* **35**(12), 2221–2228 (1988)
48. R. Bellens, P. Heremans, G. Groesenekn, H.E. Maes, On the channel length dependence of the Hot-Carrier degradation of n-channel MOSFETs. *IEEE Electron Dev. Lett.* **10**(12), 553–555 (1989)
49. E. Takeda, N. Suzuki, An empirical model for device degradation due to Hot-Carrier injection. *IEEE Electron Dev. Lett.* **4**(4), 111–113 (1983)
50. C. Hu et al., Hot-electron-induced MOSFET degradation—model, monitor, and improvement. *IEEE Trans. Electron Dev.* **32**(2), 375–385 (1985)
51. R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, H.E. Maes, A consistent model for the thickness dependence of intrinsic breakdown in ultra-thin oxides. in *IEEE Proceedings of IEDM* (1995), pp. 863–866
52. T. Kauerauf et al., Methodologies for sub-1 nm EOT evaluation, in *IEEE Proceedings of IRPS* (2011), pp. 2A.2.1–2A.2.10
53. B. Kaczer, FEOL reliability: BTI and TDDB in high-k/metal gate, finFET and Ge-based technologies, *IEEE IEDM Short Course* **27**, 356–359 (2010)
54. R. Degraeve et al., A new model for the field dependence of intrinsic and extrinsic time-dependent dielectric breakdown. *IEEE Trans. Electron Dev.* **45**(2), 472–481, (1998)

Reliability of High Mobility SiGe Channel MOSFETs for
Future CMOS Applications

Franco, J.; Kaczer, B.; Groeseneken, G.

2014, XIX, 187 p. 219 illus., Hardcover

ISBN: 978-94-007-7662-3