

14

Toward Engineering Modeling of Negative Bias Temperature Instability

Tibor Grasser, Wolfgang Goes, and Ben Kaczer

CONTENTS

14.1	Introduction	400
14.2	Interface States	401
14.3	Oxide Charges	404
14.4	Measurement Issues	405
14.5	Characterization of Relaxation.....	405
14.5.1	Functional Form of the Universal Relaxation Function	407
14.6	Characterization of MSM Data	408
14.6.1	Influence of Measurement-Delay on the Power-Law Parameters.....	410
14.7	Modeling of NBTI.....	411
14.7.1	Reaction-Diffusion Models	413
14.7.1.1	Standard RD Model	413
14.7.1.2	Pre-RD Regime	415
14.7.1.3	Relaxation as Predicted by the RD Model.....	416
14.7.2	Extended Classical RD Models	417
14.7.2.1	Two-Region RD Model.....	418
14.7.2.2	Two-Interface RD Model.....	418
14.7.2.3	Explicit H-H ₂ Conversion RD Model	419
14.7.2.4	Vanderbilt Model	420
14.7.3	Final Notes on RD Models	421
14.7.4	Dispersive NBTI Models.....	421
14.7.4.1	Reaction-Dispersive-Diffusion (RDD) Models	421
14.7.4.2	Dispersive Pre-RD Regime.....	425
14.7.4.3	Relaxation as Predicted by the RDD Models.....	426
14.7.4.4	Dispersive-Rate Coefficients	427
14.7.4.5	Simple Dispersive Hole Trapping Model	429
14.7.4.6	Detailed Dispersive Hole Trapping Model	431
14.7.5	Multiple Mechanisms	432
14.8	Conclusions.....	432
	References.....	433

14.1 Introduction

Negative bias temperature instability (NBTI) has been known for 40 years [1] and is attracting an ever growing industrial and scientific attention as one of the most important reliability issues in modern complementary metal-oxide semiconductor (CMOS) technology. It affects mostly p-metal-oxide-semiconductor field-effect transistors (pMOSFETs) at elevated temperatures with a large negative voltage applied to the gate. While the typical NBT setup requires the other terminals to be grounded, an application of a larger voltage at the drain creates interesting mixed patterns with hot-carrier degradation (HCI) and a large voltage at the bulk contact can be used to investigate the dependence of NBTI on hot or cold holes. Altogether, as a result of this stress condition, a shift in the threshold voltage is observed [2,3]. In addition to this threshold voltage shift, other crucial transistor parameters degrade as well, such as the drain current, the transconductance, the subthreshold slope, the gate capacitance, and the mobility [2,3].

The evolution of the threshold voltage during stress is commonly described by a power law of the form:

$$\Delta V_{th}(t) = A(T, E_{ox}) t^n, \quad (14.1)$$

with the prefactor A strongly depending on the temperature and the electric field. The actual dependencies of the power-law exponent n are still not fully clarified with some groups [4,5] reporting a temperature- and technology-independent value around $n \approx 0.15$, while recent publications show considerably smaller values [6,7]. Alternatively, some groups have reported a log-like dependency [6,8,9], for instance of the form:

$$\Delta V_{th}(t) = A(T, E_{ox}) \log(1 + t/\tau), \quad (14.2)$$

at least at early times. A typical scenario is depicted in [Figure 14.1](#) where the same data are shown once on a lin-log and on a log-log plot. Depending on the accuracy of the initial threshold voltage determination or, in that example, the initial drain current in the linear regime, different interpretations seem possible [6].

The detailed microscopic physics behind NBTI are not yet fully understood [10–14] but the creation of interface states seems to be a universally acknowledged feature of NBTI [2,15]. A growing number of recent publications, however, attribute at least a part of the degradation to positive charge generation in the oxide bulk [11,13,16,17]. Possible positive charges that have been suggested include holes trapped in either preexisting traps [11,16] or in traps generated by the hydrogen species released during the creation of the interface states [13].

Other potential contributions to a threshold voltage shift like mobile charges are commonly assumed to be negligible for NBTI [2] and the total threshold voltage shift is thus given by

$$\Delta V_{th}(t) = -\frac{\Delta Q_{it}(t) + \Delta Q_{ox}(t)}{C_{ox}}, \quad (14.3)$$

with ΔQ_{it} and ΔQ_{ox} being the effective charges due to interface and oxide states and C_{ox} the gate capacitance per area.

The fundamental problem in the context of NBTI is given by the fact that the degradation created during the stress phase begins to recover immediately once the stress is removed. This makes the classic measurement technique where the stress is interrupted during the

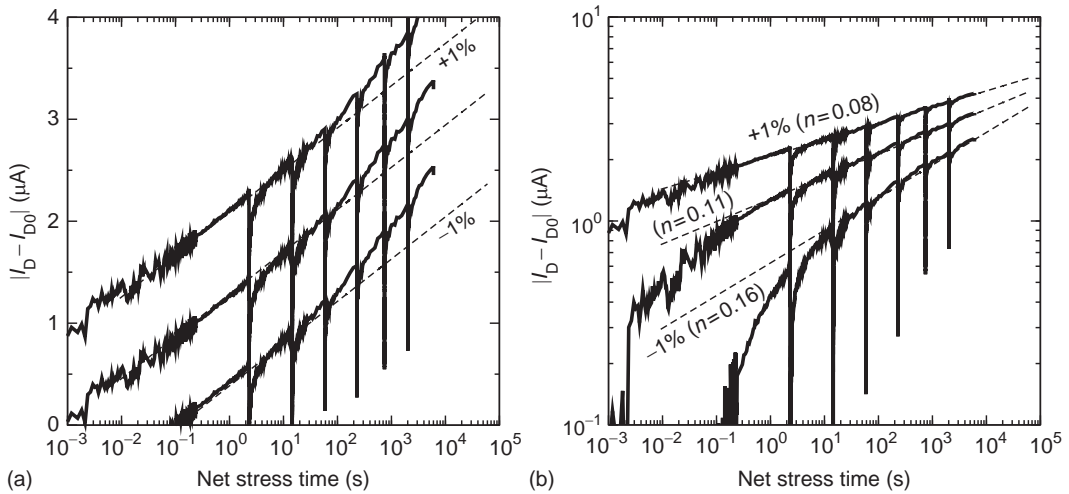


FIGURE 14.1

(a) Degradation of the drain current in the linear regime during stress ($V_G = V_G^{\text{stress}}$). The measurement is periodically interrupted to record the drain current around $V_G = V_{\text{th}}$. The initial degradation appears to be linear on a lin-log scale and any uncertainty in the initial drain current results in a shift of the whole curve. (b): Same data on a log-log plot. When the drain current measured at $t_s = 1$ ms is directly used for I_{D0} , a power-law exponent of $n=0.11$ is obtained. An uncertainty in I_{D0} of $\pm 1\%$ changes the slope to 0.08 and 0.16, thereby making the interpretation of the data extremely difficult.

extraction of the threshold voltage problematic [9,18]. In particular, the value of the extracted power-law exponent depends significantly on the delay introduced during the measurement [5,14,19]. Experimental results obtained with delayed measurements show a linear increase of the exponent with temperature [5,8,14] with values around 0.2–0.3. In contrast, temperature-independent exponents in the range 0.07–0.2 have been extracted from recent delay-free measurements [4,6,20].

Of particular interest is the question related to the origin of this extremely fast relaxation [9,14,21]. While some authors assume that hole trapping is negligible and both degradation as well as relaxation are determined by the temporal change of the interface state density and an associated back- and forth-diffusion of hydrogen [5], others acknowledge at least partial importance of trapped charges [6,8,13,22]. In the latter case it has been assumed that trapped charges either form the fast component of NBTI relaxation superimposed onto some interface defect relaxation [6,22] or are solely responsible for any recovery while created interface defects do not recover at all [8,13].

14.2 Interface States

The most commonly and earliest reported effect related to NBTI is the creation of defects at the fundamentally important Si/SiO₂ interface. These interface states are often assumed to be P_b centers [23–25] which are known to have electrically active levels within the silicon band gap. In particular, for industrially relevant samples with (100) surfaces, two variants of P_b centers have been identified [24], the P_{b0} and the P_{b1} center. Both defects are silicon dangling bonds, with the Si atom backbonded to three other Si atoms [26]. While some researchers argue that only the P_{b0} is electrically active [27], others have observed additional electrically active peaks which were claimed to originate from P_{b1} centers [26].

A recent study suggests that in nitrated oxides the role of the P_b center is taken over by K centers, which are silicon dangling bonds backbonded to three nitrogen atoms [28]. K centers are located inside the nitrated oxide, rather than at the interface as P_b centers. As such, a model relating NBTI to K centers could be different from the available theories, a question open to future research.

P_b centers are present in a considerable number at every Si/SiO₂ interface with a concentration in the order of 10^{12} cm^{-2} . During device fabrication these defects have to be passivated through some sort of hydrogen anneal [2], thereby reducing the electrically active trap levels to a value below 10^{10} cm^{-2} . The electrically active trap levels are of amphoteric nature, meaning that each interface state can accommodate two electrons. Possible transitions are from the positive to the neutral state (+/0), which appears as a donor-like trap level in the lower half of the silicon band gap, and the neutral to negative charge state (0/−) which is commonly assumed to act as an acceptor-like trap level in the upper half of the band gap.

Although the P_b H bonds obtained after the passivation step are relatively stable, they can be broken at elevated temperatures and higher electric fields, thus reactivating the electrically active trap levels. In our analysis, we will denote the time-dependent density of interface states as $N_{it}(t) = [P_b^\bullet]$. Depending on the trap occupancy, the initial value of $N_{it0} = N_{it}(t_0)$ is inherently visible in the reference threshold voltage $V_{th}(t_0)$ and the change in the density of interface states is given through $\Delta N_{it}(t) = N_{it}(t) - N_{it0}$. It is normally assumed that charging and discharging of these interface states is very fast, and consequently that the positive charge in these interface states immediately follows the Fermi-level via

$$\Delta Q_{it}(t) = q \int \Delta D_{it}(E_t, t) f_{it}(E_F, E_t, t) dE_t. \quad (14.4)$$

Here, ΔD_{it} is the time-dependent density of interface states in the units of $\text{cm}^{-2} \text{ eV}^{-1}$, which is by a still to be quantified relation [29] directly linked to $\Delta N_{it}(t)$, and $f_{it}(E_t)$ their occupancy with holes. In addition to the exponential band-tail states of a passivated Si/SiO₂ interface, the P_b centers create Gaussian peaks in the Si band gap where the broadening is probably linked to the disorder at the interface [30]. As an example, the measured concentration of P_{b0} centers as obtained by Ragnarsson and Lundgren [31] is shown in Figure 14.2 for an initially unpassivated interface and after a short hydrogen passivation step. This may correspond to the inverse process occurring during NBT stress, that is, the relaxation part which we have argued to be of fundamental importance for the understanding of NBTI [32]. The measurement data can be nicely fitted by two Gaussian peaks or by using a Fermi-derivative function [33] (which can be analytically integrated):

$$g_P(E_t, E_P, \sigma) = \frac{1}{\sigma} \frac{\exp\left(\frac{E_P - E_t}{\sigma}\right)}{\left(1 + \exp\left(\frac{E_P - E_t}{\sigma}\right)\right)^2} \quad (14.5)$$

as

$$\Delta D_{it}(E_t) = N_{it}(g_P(E_t, E_{P1}, \sigma_1) + g_P(E_t, E_{P2}, \sigma_2)). \quad (14.6)$$

Note that in order to fit the data of Ragnarsson and Lundgren, the variances of the two peaks have to evolve differently in time, with the acceptor-like peaks becoming narrower sooner (Figure 14.2). In contrast, other groups have reported a similar time evolution of both peaks [30]. This disorder-induced broadening of the electrical active levels is

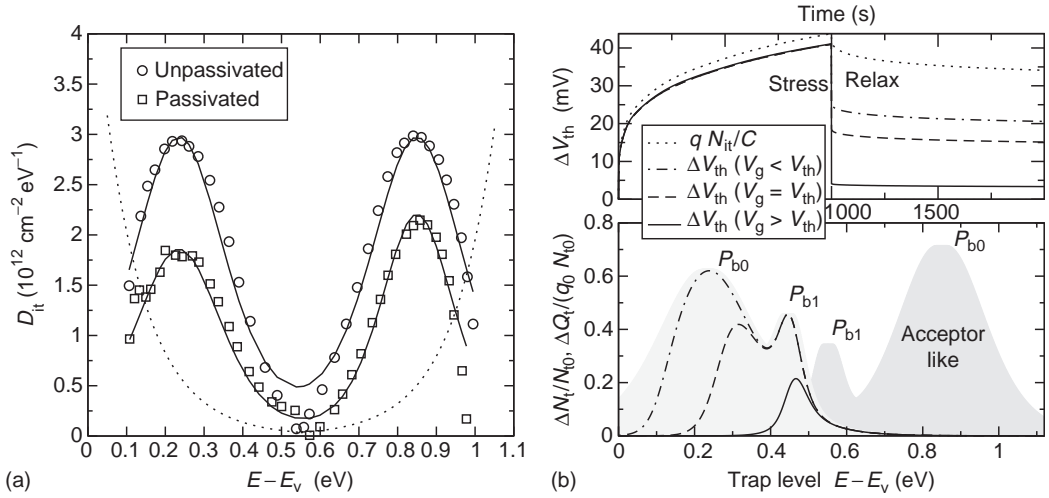


FIGURE 14.2

(a) Measured density of interface states supposedly related to P_{b0} centers before and after a short hydrogen passivation process [31]. The symbols are the measurement data, the solid lines give the analytic fit, while the exponential band-tail states are schematically represented by the dotted lines. The donor-like peak is located $\approx 0.24 \text{ eV}$ above the valence band edge, while the acceptor-like peak is at $\approx 0.85 \text{ eV}$. Note that the variance of the unpassivated sample is $\sigma_1 = \sigma_2 = 0.085 \text{ eV}^2$ while after the passivation step one obtains $\sigma_1 = 0.074 \text{ eV}^2$ and $\sigma_2 = 0.062 \text{ eV}^2$, a fact to be included into a model. (b) Influence of the interface state occupancy on the observed threshold voltage shift using on-the-fly measurements. During stress, nearly all interface traps are positively charged. When a different gate voltage is used during relaxation, only a fraction of the traps are visible which must be separated from the real relaxation. Schematically shown is the density-of-states typically associated with P_{b0} and P_{b1} centers [26]. (From Grasser, T., et al., *Proc. IRPS*, 268, 2007. With permission.)

suspected to be closely related to a disorder-induced Gaussian variation of the binding energies of the Si-H bonds at the interface [30,34,35].

Nevertheless, during NBT stress, the Fermi-level E_F is close to the valence band edge and $f_{it}(E_t) \approx 1$ throughout the silicon band gap. Thus, under the assumption that P_b centers introduce states only within the silicon band gap, see Ref. [13] for a different interpretation, all newly generated interface states ΔN_{it} are positively charged and one obtains $\Delta Q_{it}(t) \approx q \Delta N_{it}(t)$, independently of the exact form of the density-of-states. This is the usual assumption employed for instance in the widely used reaction-diffusion (RD) model and quite reasonable during the stress phase. However, in order to measure the degradation, the stress is often interrupted and the various forms of degradation are assessed using different possible techniques. Regardless of the actual measurement technique employed, be it a complete or partial $I_D V_G$ sweep, single point V_{th} determination [9,14], ultrafast pulse $I_D V_G$ [16], capacitance-voltage (CV), DCIV [36], or charge-pumping (CP) [37] measurements, the trap occupancy changes significantly because a different fraction of the traps is charged during stress and measurement. Furthermore, this Fermi-level dependence causes a change in the subthreshold slope during $I_D V_G$ measurements and humps in the CV characteristics, in contrast to constant shifts induced by fixed positive charges, see Figure 14.3 for a qualitative description.

Alternatively, in the model of Zafar [13], a different interpretation is introduced. Zafar assumed that a large number of dangling bonds always exists but that only a fraction can be observed in electrical measurements, while the majority is too close to the band-edges to contribute. During NBT stress the total number of interface states is increased and only this increase is visible during measurements. To properly account for this partial contribution

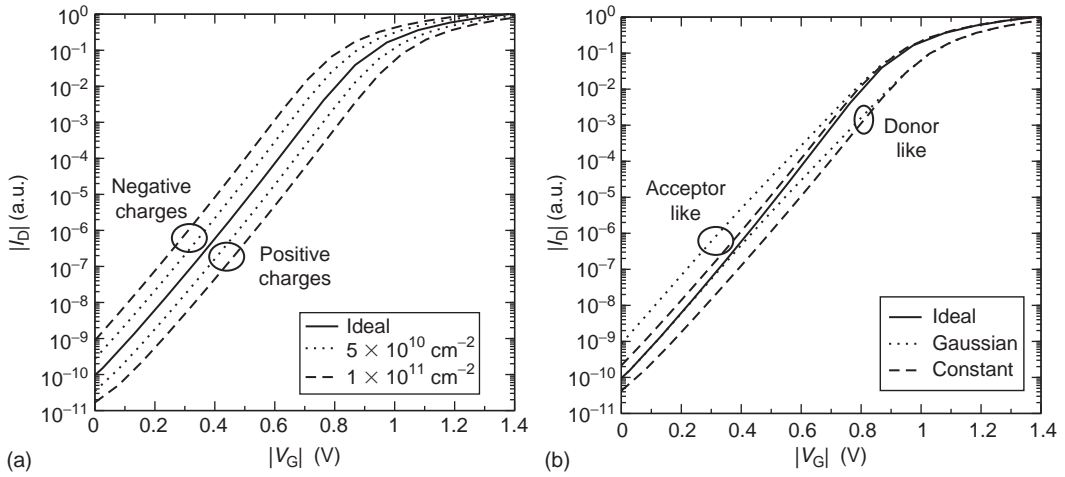


FIGURE 14.3

(a) Simulated influence of fixed charges on the $I_D V_G$ characteristics of a PMOS transistor. (b) Simulated influence of interface states on the $I_D V_G$ characteristics of a PMOS transistor. Depending on the form of the trap density-of-states, a different shift in ΔV_{th} and a different change in the subthreshold slope are obtained.

of the generated interface states to the observable threshold voltage shift, the occupancy of the interface states as a function of the Fermi-level position has to be introduced.

14.3 Oxide Charges

On the top of generated interface defects, charge may be stored in existing or newly created oxide traps. Although some of these traps may still be considered fast, they are more difficult to charge and discharge, that is, have larger time constants than interface states due to their location inside the oxide bulk. It has also been suggested that holes trapped in energetically deep levels give rise to practically permanent charge contributions which can only be neutralized through the application of unusual bias conditions [38]. Altogether, the occupancy of oxide traps cannot follow the Fermi-level immediately and $\Delta Q_{ox}(t)$ will be governed by different dynamics. The contribution of the oxide charges to the threshold voltage shift is formally written as

$$\Delta Q_{ox}(t) = q \iint \Delta D_{ox}(x, E_t, t) f_{ox}(x, E_t, t) (1 - x/t_{ox}) dx dE_t, \quad (14.7)$$

where

ΔD_{ox} is the spatially and energy-dependent density-of-states in the oxide

f_{ox} is the hole occupancy of these traps

t_{ox} is the oxide thickness

Note that the issue of whether oxide charges are important during NBTI or not is currently one of the most debated ones [5,11,13,16]. Also, the question whether ΔD_{ox} consists mainly of preexisting traps [11,16] or traps that are created during stress [13] remains to be answered.

14.4 Measurement Issues

The understanding and characterization of NBTI is considerably hampered by the difficulties arising during measurement. Currently, two techniques are often used to characterize NBTI: the classic measurement/stress/measurement (MSM) technique, which is handicapped by undesired relaxation, and on-the-fly (OTF) measurements, which avoid any relaxation by maintaining a high stress level throughout the measurement and directly monitor the drain current in the linear regime, ΔI_{Dlin} . Since the threshold voltage shift ΔV_{th} is more suitable to study the creation of charges, ΔI_{Dlin} has to be converted to ΔV_{th} which is commonly done using the simple SPICE Level-1 model [8] or an empirical formalism [39]. The applicability of the OTF technique is particularly troublesome when one switches from stress to relaxation. When V_G is left at V_G^{relax} , the interface trap occupancy is considerably lower than during the stress phase [29], resulting in spurious additional relaxation (Figure 14.2). Conversely [4], when V_G is brought back to V_G^{stress} , one faces the opposite problem one is trying to avoid during the stress phase, since now additional uncontrolled stress is introduced during the measurement cycles. Even more important is the fact that the initial value of I_{Dlin} is extremely difficult to determine as it is already obtained at the stress voltage. Conventionally, the time required for this is in the milliseconds range where already significant degradation can be observed [6] but any uncertainty in I_{D0} modifies the time exponent (the slope) of ΔV_{th} on a log-log plot in a somewhat arbitrary manner, see also Figure 14.1. This may render many results obtained by the OTF technique questionable.

In contrast, the MSM technique probes the interface under comparable conditions during both the stress and relaxation phase. In addition, the voltage applied to the gate can be kept close to the threshold voltage where only negligible degradation can be expected. However, as has been pointed out [7,32,40], it is probably very difficult to minimize the measurement delay in such a way that the true degradation is observed.

14.5 Characterization of Relaxation

In order to properly understand and characterize NBTI it is mandatory to take a close look at the relaxation behavior. Particularly noteworthy are the long tails of logarithmic-like nature that may cover more than 12 decades in time [9,14,42,43]. In order to formalize the description, we use the term $S(t_s) = \Delta V_{\text{th}}(t_s)$ for the real degradation accumulated during the stress phase. As soon as the stress voltage is removed, relaxation sets in as a function of the accumulated stress time t_s and the relaxation time $t_r = t - t_s$. In the following, we will assume that the accumulated degradation $S(t_s)$ consists of a recoverable component $R(t_s)$ and a permanent component [8,38] $P(t_s)$ as

$$S(t_s) = R(t_s) + P(t_s). \quad (14.8)$$

As the recoverable component depends on the recovery time t_r , any measurement conducted with a certain delay $t_r = t_M$ observes only

$$S_M(t_s, t_M) = R_M(t_s, t_M) + P(t_s) \leq S(t_s), \quad (14.9)$$

with the subscript M indicating quantities observed in a measurement. Of course, $S_M(t_s, 0) = S(t_s)$ and $R_M(t_s, 0) = R(t_s)$ hold.

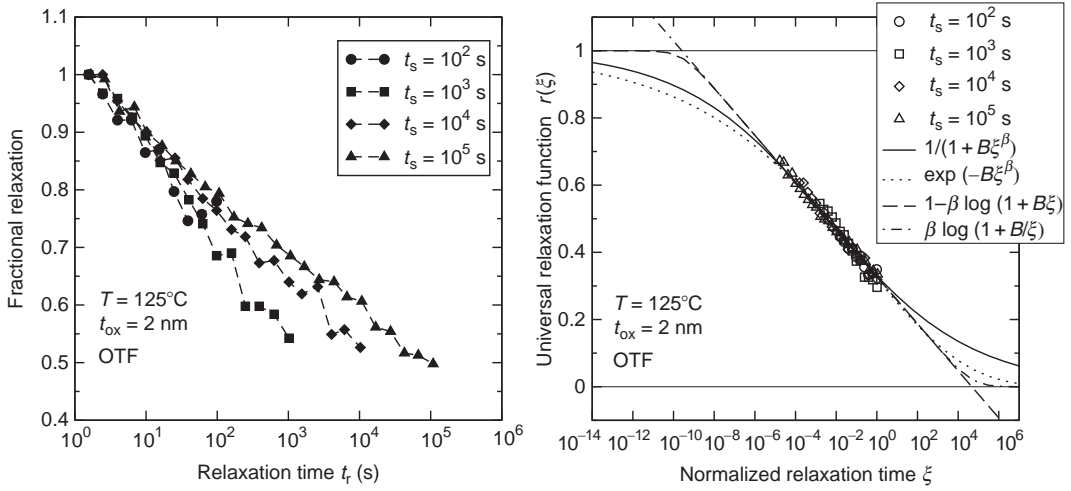


FIGURE 14.4

Demonstration of universal recovery for the OTF data of Denais et al [41]. The left figure shows a conventional view of the fractional recovery as a function of the relaxation time t_r . Apparently, data obtained after longer stress times seem to relax more slowly than data obtained at shorter times. The right figure, on the other hand demonstrates the universality of relaxation when the relaxation data are normalized to the last stress value and plotted over the ratio $\xi = t_r/t_s$ [41]. Also shown are some possible empirical expressions which can be fit to the data. (From Grasser, T., et al., *Proc. IRPS*, 268, 2007. With permission.)

Due to the onset of recovery which may occur at timescales possibly shorter than nano- or even picoseconds [7,9], a rigorous characterization of the relaxation phase is extremely challenging [4,21,40,41]. Typically, the relaxation data $R(t_s, t_r)$ recorded at different stress times t_s have been normalized to the first measurement point t_M as

$$r_i(t_s, t_r) = \frac{S_M(t_s, t_r)}{S_M(t_s, t_M)}, \quad (14.10)$$

giving the fractional measurable recovery, and aligned as a function of the absolute relaxation time t_r [9,14,21,40], see Figure 14.4. The functional form of the relaxation remains elusive in such a plot.

Instead, it has been demonstrated that it is highly advantageous to study the recoverable component in its universal representation [32] which is based on the observation that all individual relaxation curves obtained at different stress times $t_{s,i}$ can be represented by a single universal curve when [41]

- Relaxation data are normalized to the last stress value $S(t_{s,i}) = S_M(t_{s,i}, 0)$ rather than the first measurement point $S_M(t_{s,i}, t_M)$
- Relaxation time t_r is normalized to the last stress time $t_{s,i}$ as $\xi = t_r/t_{s,i}$

The above results in the definition of the universal relaxation function as [32]:

$$r(\xi) = \frac{R_M(t_s, t_r)}{R(t_s)} = \frac{S_M(t_s, t_r) - P(t_s)}{S(t_s) - P(t_s)}, \quad (14.11)$$

which is a function of ξ only. For the special case of a negligible permanent component, note the relationship between the universal recovery function and the fractional recovery given by $r_i(t_s, t_r) = r(\xi)/r(\xi_M)$ with $\xi_M = t_M/t_s$. This concept is visualized in Figure 14.5.

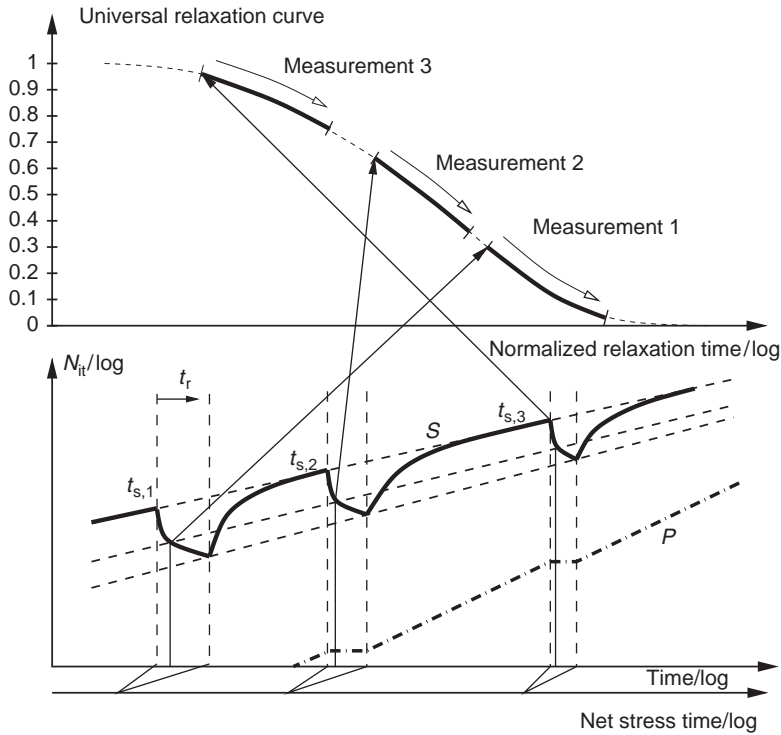


FIGURE 14.5

Schematic view of universal relaxation. The stress is interrupted three times to record relaxation data on the relative time scale $t_r = t - t_s$. For stress intervals considerably larger than the relaxation intervals, the device forgets the interruption. Note how for larger stress times the relaxation data move to smaller normalized relaxation times $\xi = t_r/t_s$ and how the relative recovery becomes less significant. Also indicated is a possible permanent/slowly relaxing component P .

Because the bulk of relaxation data available in the literature do not allow a definite identification of the permanent component, which requires very detailed data [7], we will consider the permanent component to be negligible in the following and assume $P(t_s) = 0$. Recent studies have shown how more detailed data allow for a clear identification of a permanent component and the extension of the method presented here [7,43].

14.5.1 Functional Form of the Universal Relaxation Function

Lacking a universally accepted and valid theory for NBTI, the exact form of the universal relaxation function $r(\xi)$ remains illusive at this point and empirical functions have to be used. So far, excellent results have been obtained with the power-law-like expression:

$$r(\xi) = (1 + B\xi^\beta)^{-1}, \quad (14.12)$$

where the parameters B and β are in the range $B \approx 0.3-3$ and $\beta \approx 0.15-0.2$ for most of the data available. Of particular interest is the relaxation predicted by the RD model which is well described by Equation 14.12 using $B = 1$ and $\beta = 1/2$. However, we have to point out that the available data are not conclusive at the time being, making alternative expressions such as the logarithmic dependence suggested by Denais et al. [41]:

$$r(\xi) = 1 - \beta \log(1 + B\xi), \quad (14.13)$$

or the frequently used stretched-exponential [44]:

$$r(\xi) = \exp(-B\xi^\beta). \quad (14.14)$$

viable alternatives as well [32]. We remark that Equation 14.13 is nonphysical at larger times, making a reformulation mandatory.

One difficulty in determining the correct choice of the empirical function is the fact that relaxation may occur over more than 12 decades in time [9]. As the delay times in conventional measurements are around 1 ms and relaxation data are not normally recorded for $t_r > 10^5$ s, only eight decades in time is commonly available. By employing fast measurements which start at $t_r = 1 \mu\text{s}$ 12 decades has been reported [42]. Interestingly, the measurement data available behave logarithm-like over most of the recorded regime and excellent fits with Equations 14.12 through 14.14 can be obtained [32]. Only more detailed relaxation data and a solid theoretical description will allow to differentiate between possible expressions which differ mostly in the behavior at extremely short and long times. This is illustrated in Figure 14.4 where possible empirical expressions for the universal relaxation function are compared. All expressions can be fit to the measurement data and give fits of practically the same accuracy. However, they result in different extrapolations for large and small relaxation times, the consequences of which need to be carefully investigated.

14.6 Characterization of MSM Data

Although more delicate to apply, universal relaxation is of particular interest for data obtained by the MSM technique. For the normalization needed in Equation 14.11 one has to keep in mind that the value of $S(t_s) = R_M(t_s, 0)$ is essentially unknown, one only knows $R_M(t_s, t_M)$ determined at the first measurement point available after a short but probably nonnegligible relaxation period t_M . However, making use of the universal relaxation expression 14.11 and assuming for the time being that $r(\xi)$ is known, $S(t_s) = R_M(t_s, 0)$ can be obtained as

$$S(t_s) = R(t_s) = \frac{R_M(t_s, t_M)}{r(t_M/t_s)}. \quad (14.15)$$

Inserting the above into the universal relaxation relation 14.11 we obtain

$$\frac{r(\xi)}{r(\xi_M)} = \frac{R_M(t_s, t_r)}{R_M(t_s, t_M)}. \quad (14.16)$$

From Equation 14.16 the as of yet unknown parameters B and β can be easily determined from a measured sequence of relaxation data $R(t_{s,i}, t_r)$ obtained after N stress intervals $t_{s,i}$ by minimizing for instance

$$\varepsilon_t = \sum_{i=1}^N \int \left(\frac{r(t_r/t_{s,i})}{r(t_M/t_{s,i})} - \frac{S_M(t_{s,i}, t_r)}{S_M(t_{s,i}, t_M)} \right)^2 d \log(\xi_i). \quad (14.17)$$

Note that the parameter extraction is independent of the functional form of R and that the final form of R is directly related to the measurement data through the universal relaxation relation as $R(t_s) = R_M(t_s, t_M)/r(t_M/t_s)$ [32].

Naturally, in contrast to data obtained by OTF measurements where $R_M(t_s, 0)$ is known, the analytical expression determines the final value of $R_M(t_s, 0)$ through the extrapolation given by Equation 14.16. This results in a floating behavior of $r(\xi_M)$ which reflects the uncertainty of this approach [32].

A particularly intriguing feature of Equation 14.16 is that it can be applied to a whole sequence of stress and relaxation sequences as typically encountered during MSM measurements. This is because during MSM sequences the duration of the stress intervals usually grows exponentially while the measurement interval t_M is short and of constant duration. This implies that after a certain stress time, which we determined empirically to be of the order $t_s \gtrsim 10 \times t_M$, the relaxation during the measurement does not significantly alter the degradation at the end of each stress phase, meaning that the degradation relaxed during each measurement interval is mostly restored during the next stress phase. This is in agreement with the reports of Reisinger et al. [9] who report that “the sample completely forgets the effect of the interruption” provided the stress phase following the interruption is by a factor of 100 longer than the interruption.

Consequently, Equation 14.15 holds for every stress point t_s , where t_s is now the accumulated net stress time. The applicability of the procedure outlined above to the detailed relaxation data published by Reisinger et al. [9] and for the IMEC data otherwise published in Ref. [14] is outlined in Figure 14.6. For the IMEC data the universality is also shown at three different temperatures, 50°C, 125°C, and 200°C.

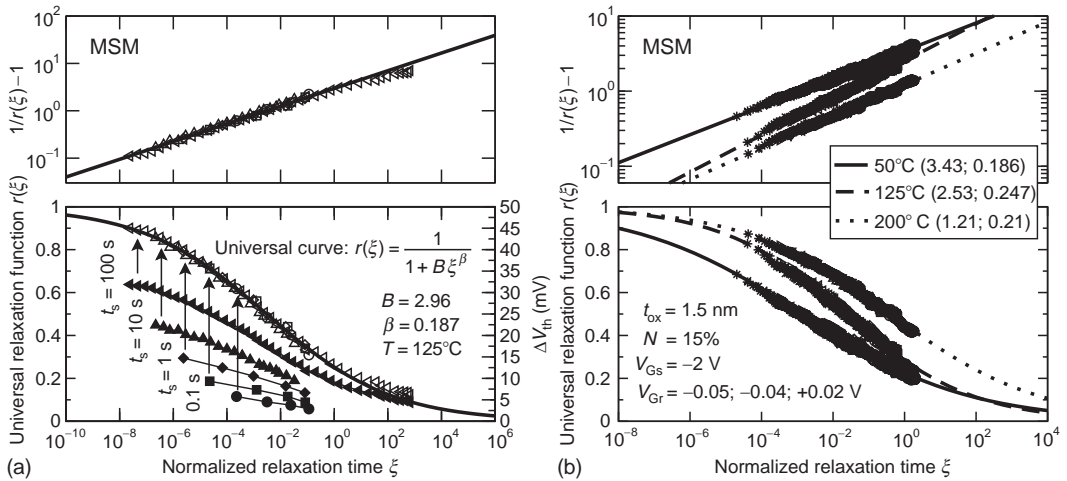


FIGURE 14.6

(a) Application of universal relaxation to the fast MSM data obtained by Reisinger et al. [9]. Depending on the choice of the universal relaxation function, the individual data points can be mapped onto the respective universal curve, in this case Equation 14.12. Note the linear behavior of $1/r - 1$ shown in the upper plot. The slight deviation for $\xi > 10^2$ is introduced by a permanent component $P(t_s)$, see Ref. [43]. (b) Same as (a) but with data from IMEC [14]. Relaxation data of three devices stressed in a single MSM sequence were recorded at 10 different stress times in the interval 10^{-10} – 10^4 s at three different temperatures. The values of B and β (given in parenthesis) depend on the temperature, β even in a nonmonotonic manner which may indicate the existence of two different processes with different temperature dependencies. (From Reisinger, H. et al., *Proc. IRPS*, 2006.)

Universal relaxation thus results in the interesting possibility to reconstruct the true (undelayed) measurement curve from delayed data sets. This suggests a novel measurement technique:

- (1A) Determine $R_M(t_s, t_M)$ using a single delay time and add a long relaxation period at the end. In case a permanent component is present, multiple devices can be subjected to different stress intervals for an accurate determination of the time-dependence of P [7].
- (1B) Alternatively, one may determine $R_M(t_s, t_M)$ using different delay times. This approach probably only works for stress cases where negligible permanent degradation is created [7].
- (2) From that data determine B and β .
- (3) Finally, calculate the true degradation using Equation 14.15.

Variant A, where B and β have been obtained from detailed relaxation data, has already been demonstrated in Figure 14.6. However, the method also works for MSM data obtained with different delay times where no relaxation data are available (Variant B). In that case the parameters A , n , B , and β can be directly extracted through fitting of Equation 14.18. This is demonstrated in Figure 14.7 for the data published by Li et al. [45]. Again, the extracted parameter values agree very well with the cases where we had access to the full relaxation data.

14.6.1 Influence of Measurement-Delay on the Power-Law Parameters

Next, we show that the universal relaxation expression naturally connects individual stress curves obtained using the MSM technique with different delay times. For simplicity, we assume that the true degradation behavior follows a power law as $S(t_s) = At_s^n$ and that the universal relaxation is given by Equation 14.12. Due to the measurement delay one observes instead of the power law

$$S_M(t_s, t_M) = S(t_s)r(t_s, t_M) = \frac{At_s^n}{1 + B(t_M/t_s)^\beta}. \quad (14.18)$$

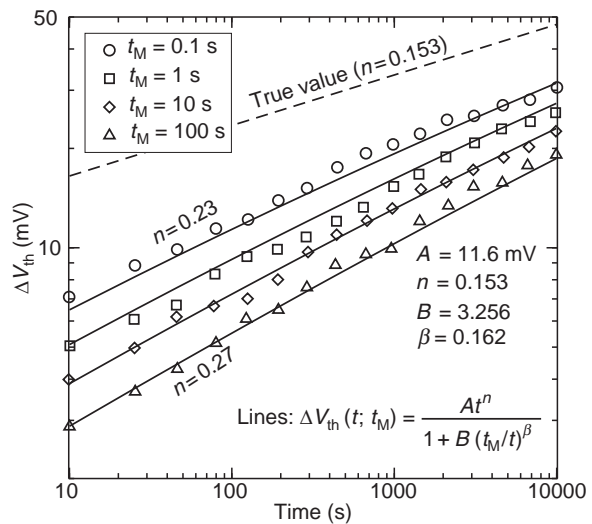


FIGURE 14.7

Reconstruction of the true degradation from MSM data obtained by Li et al. [45] with four different delay times without the knowledge of the detailed relaxation behavior. Again, a corrected slope of about $n \approx 0.15$ is obtained. Note that even at $t_s = 10^4$ s the lines do not merge and the impact of the delay is still clearly visible. (From Grasser, T., et al., *Proc. IRPS*, 268, 2007. With permission.)

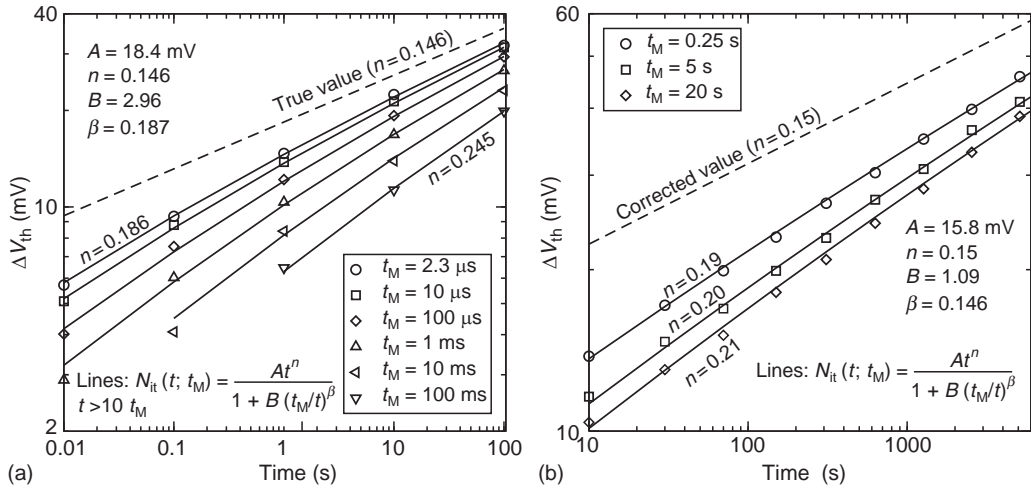


FIGURE 14.8

Comparison of the analytic model for MSM measurements based on the universal relaxation to the data of Reisinger et al. [9] (a) and Kaczer et al. [14] (b). Excellent accuracy of the analytic model is obtained for all available delay times. In addition, the true NBTI degradation can be recovered by extrapolating to $t_M = 0$ s. (From Grasser, T., et al., *Proc. IRPS*, 268, 2007. With permission.)

Equation 14.18 is validated against the Infineon and IMEC data in Figure 14.8 where the parameters B and β are given by the universal relaxation law. The analytic expression 14.18 exactly reproduces the delayed measurement results for various delay times t_M and thereby convincingly confirms our assumptions stated above.

As a consequence of the measurement delay, the observed power-law exponent n_M will be time-dependent and given through Equation 14.18 as

$$n_M(t_s, t_M) = n - r'(t_M/t_s) \frac{t_M/t_s}{r} = n + \frac{\beta B}{B + (t_s/t_M)^\beta}, \quad (14.19)$$

with $r'(\xi) = \partial r(\xi)/\partial \xi$.

It is worthwhile to point out that although many groups report a constant measured power-law exponent over three to four decades which varies as a function of the temperature and delay time, this can of course only be approximately correct. The fact that all curves obtained with different delay times have to merge at larger times, makes a time-dependent slope a necessity. However, depending on the actual values of B and β this time-dependence will be more-or-less visible in a log-log plot. In general, the smaller β , the less visible the time-dependence will be. A comparison of measured power-law exponents as a function of the delay time t_M and temperature is given in Figure 14.9. Most of the data show an apparently constant power-law exponent (within the measurement accuracy) over three to four decades. Clearly, the measured power-law exponents, and consequently B and/or β (see Table 14.1), depend on temperature, on the particular technology, and/or the measurement technique.

14.7 Modeling of NBTI

As has been detailed in the previous sections, the fundamental dilemma encountered during the development of NBTI models is the question of what exactly should be modeled. While conventional models for semiconductor device simulation can rely on a

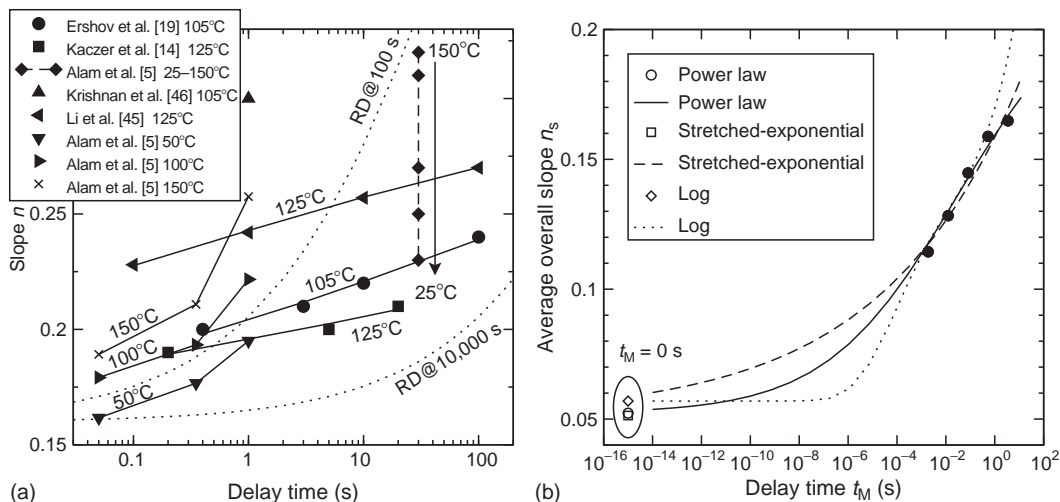


FIGURE 14.9

(a) Influence of the measurement delay on the measured slope as reported by various groups [5,10,14,19,45,46]. The solid lines are given by a fit to Equation 14.19 using the parameters in Table 14.1. Note that the strong temperature-dependence of the reported slopes and that the slopes were found to be constant over three to four decades in many measurements. Clearly, there is a large spread in the measurement data indicating a technology dependence. The dotted lines show the slopes predicted by the RD model at $t_s = 100$ s and $t_s = 10,000$ s. Note that the RD slope changes considerably within two decades, is per construction temperature independent, and cannot be adjusted to the technology. (b) Observed slope in a delayed measurement as a function of the measurement delay [7]. The symbols are the measurement data while the lines give the extrapolation to the true slope using our algorithm. Note that according to the power-law and stretched-exponential model only for delay times in the picosecond range the true slope could be measured. (From Grasser, T., et al., *Proc. IRPS*, 268, 2007. With permission.)

rather robust set of measurement data which need to be captured by the models, the modeling of NBTI has been plagued by a still open definition of what NBTI degradation actually is. The question of whether the model should result in a power law or a logarithmic evolution of the threshold voltage, whether it should predict a temperature-dependent slope, or whether it should relax universally is of fundamental importance to any modeling attempt.

In contrast to previous model validation attempts that have focused almost exclusively on the stress phase, we put a special emphasis on the analysis of the model prediction

TABLE 14.1

Parameters for Equation 14.19 Used to Fit the Data in Figure 14.9 Assuming $t_s = 1000$ s

Reference	T	N	B	β
Ershov et al. [19]	105	0.15 (fixed)	1.49	0.179
Kaczor et al. [14]	125	0.15 (fixed)	1.29	0.136
Li et al. [45]	125	0.15 (fixed)	4.08	0.163
Alam et al. [5]	50	0.155	4.79	0.611
Alam et al. [5]	100	0.177	40.23	0.973
Alam et al. [5]	150	0.186	102.2	1.048

Source: From Grasser, et al., *Proc. IRPS*, 2007, With permission.

Note: The fit was obtained using a fixed $n = 0.15$ with a simple least-square algorithm. However, in order to fit the data of Ref. [5], which are somewhat different from the other sources considered in this study, n had to be included as a free parameter. Interestingly, this results in a significant temperature-dependence of the zero-delay slope, well described by a linear relationship for n reported in Refs. [8,14] for delayed measurements. Keep in mind that these values should be taken with care, since they were extracted by a fit to three or four rather inaccurate slope values using two/three free parameters. The inaccuracy of the slope values is a result of both the measurement uncertainty as well as the time-dependence of the slope.

during the relaxation phase. Thereby two features are of interest, namely the large distribution of timescales and the universal behavior. For the analysis, the models under consideration have been implemented into a partial-differential-equation solver and solved numerically in order to rule out any uncertainties related to approximate analytic expressions. Since most of the published NBTI models can be derived from a generalized RD formalism [7], a short review of the assumptions employed in this model is given.

14.7.1 Reaction-Diffusion Models

RD-like models consist basically of an electrochemical reaction at the semiconductor–oxide interface which is coupled to a transport equation in the oxide bulk. We remark that the questions whether the depassivation process is field-driven [10,47], why holes at the interface are required and how they influence the reaction [48], and in which charge state, neutral or positive, the created trap and the released hydrogen species are, are highly controversial and are put aside for the moment. Nevertheless, for the discussion of the basic properties it is instructive to write the electrochemical reaction at the interface, which creates a dangling bond Si^\bullet from a passivated interface defect SiH , as



Thereby we differentiate between hydrogen in a conduction/mobile state, H_c , and trapped hydrogen [49], H_t . Such a distinction is important, since in dispersive transport models most hydrogen becomes trapped quickly and might not be available for the reverse reaction. We also note that a large background concentration of hydrogen may exist in the vicinity of the interface, possibly in the order of 10^{19} cm^{-3} [50], which, if assumed to be freely available, could dominate the reverse reaction and completely compensate the forward reaction in a standard RD model.

It has been claimed that the binding energies of the Si–H bonds display a Gaussian broadening [8,35]. Previously published dispersive NBTI models consider either a dispersion in the forward rate [8] or a dispersion in the transport properties [13,14,51]. Models based on these assumptions will be discussed in [Section 14.7.4](#). In particular, the variations in the energy barrier for the reverse reaction is important for the investigation of dispersive transport.

14.7.1.1 Standard RD Model

In the standard RD formulation the dissociation barrier is considered to be single valued (dispersion-free) and $\text{H}_t = 0$, meaning that all released hydrogen remains in the conduction state. The kinetic equation describing the interface reaction is commonly assumed to be of the form [46,52,53]:

$$\frac{\partial N_{it}}{\partial t} = k_f(N_0 - N_{it}) - k_r N_{it} \text{H}_{it}^{1/a}, \quad (14.21)$$

where $N_{it} = [\text{Si}^\bullet]$ is the interface state density, $N_0 = [\text{SiH}]_0$ is the initial density of passivated interface defects, H_{it} is the hydrogen concentration at the semiconductor–oxide interface, k_f and k_r are the temperature and possibly field-dependent rate coefficients, while a gives the order of the reaction (1 for H^0 and H^+ , 2 for H_2 , assuming an instantaneous conversion of H^0 to H_2 , cf. Ref. [5,46,54]). In our context it is important to recall that the usual assumptions are that $N_{it0} = N_{it}(t_0) = 0$ at the beginning of the stress period and that all generated N_{it} contribute equally to the threshold voltage shift. A somewhat unappreciated feature of the RD equations is, as will be shown in [Section 14.7.4.2](#), that by allowing for a larger number of initial interface defects, a completely different behavior is obtained.

Hydrogen motion is assumed to be controlled by conventional drift–diffusion [53]:

$$\frac{\partial H_c}{\partial t} = -\nabla \cdot F_c + G_c, \quad (14.22)$$

$$F_c = -D_c \left(\nabla H_c - Z \frac{E_{ox}}{V_T} H_c \right), \quad (14.23)$$

with the (possibly unrealistic) assumption of a negligible initial hydrogen concentration, $H_c(x,0)=0$. Hydrogen transport is postulated to occur on a single energy level, which will be referred to as the conduction state, with H_c , D_c , and G_c the hydrogen concentration, diffusivity, and generation rate in the conduction state, F_c the particle flux, Z the charge state of the particle, $V_T = k_B T_L / q$ the thermal voltage, T_L the lattice temperature, and E_{ox} the electric field inside the oxide.

The generation rate G_c is given by the interface reaction and reads for the usually considered one-dimensional problem:

$$G_c(x,t) = \frac{1}{a} \frac{\partial N_{it}}{\partial t} \delta(x) \quad (14.24)$$

with the interface assumed to be located at $x=0$.

For the calculation of the time-dependent density of interface states, N_{it} , Equations 14.21 and 14.22 can be solved numerically on an arbitrary geometry. Although the solution of the RD model depends on the underlying geometry [55], it is commonly assumed that NBTI is a one-dimensional problem. In particular, for some special cases analytical approximations can be given [46,56,57] which are helpful for understanding the basic kinetics.

Depending on the parameter values and boundary conditions, different phases are observed which are shown in Figure 14.10 for the three most commonly used species

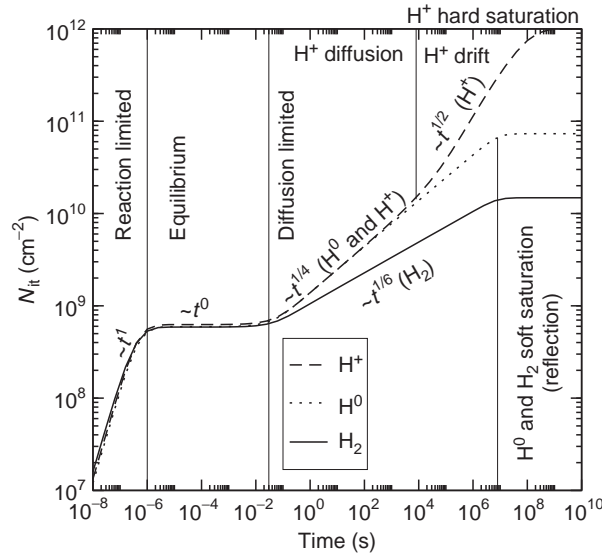


FIGURE 14.10

Five phases of the standard RD model obtained from a numerical solution of Equations 14.21 and 14.22 on a 2 nm oxide, using the parameters $N_0 = 10^{12} \text{ cm}^{-2}$ and $N_{it0} = 0$. Shown are the results for the three species H^+ , H^0 , and H_2 . The time exponent $n = 1$ is the signature of the reaction-limited phase while $n = 1/4 \dots 1/2$, $n = 1/4$, and $n = 1/6$ are observed for the three species in the diffusion-limited phase. At the beginning of the diffusion-limited phase H^+ behaves like H^0 . Furthermore, in the nonself-consistent simulation, where the feedback of the charges on the field distribution is neglected, H^+ does not show a soft saturation since all hydrogen is pulled away from the interface. (From Grasser, T., et al., *Trans. Dev. Mater. Reliab.*, 8(1), 79, 2008. With permission.)

H^+ , H^0 , and H_2 : (1) The reaction-dominated regime with an exponent $n=1$, where the reverse rate is negligible due to the lack of available H_{it} . (2) Depending on the parameter values, a transition regime where $\partial N_{it}/\partial t=0$ which gives an exponent $n=0$. (3) The quasiequilibrium regime where $\partial N_{it}/\partial t$ is much smaller than the generation and passivation terms. This is assumed to be the dominant regime and displays the characteristic time exponent depending on the created species. (4) A saturation regime which could for instance be a soft saturation due to a reflecting boundary condition or a hard saturation resulting from the depassivation of all passivated interface states [58].

The RD model assumes the quasiequilibrium of the interface reaction ($\partial N_{it}/\partial t \approx 0$) to be the dominant regime [10,13,14]. Consequently, we obtain from Equation 14.21 together with the assumption $\Delta N_{it}(t) \gg N_{it0}$ the standard RD model as

$$\Delta N_{it}(t) = A_{RD} C^{1/(1+a)}(t), \quad (14.25)$$

with the species-dependent prefactor

$$A_{RD} = \left(a \left(\frac{k_f}{k_r} \Delta N_{it, \max} \right)^a \right)^{1/(1+a)} \quad (14.26)$$

and the maximum value of ΔN_{it} given by $\Delta N_{it, \max} = N_0 - N_{it0}$. For nondispersive transport $C(t) = D_c E_{ox} t / V_T$ for the proton case while $C(t) \approx \sqrt{D_c t}$ for the neutral species H^0 and H_2 [10,59]. This results in the well-known exponents 1/2, 1/4, and 1/6 for proton, atomic, and molecular hydrogen transport, respectively. These exponents do not depend on temperature nor is it possible to include process dependencies. We recall that such an exponent of 1/2 obtained for H^+ transport is not observed experimentally which led researchers to discard the possibility of drifting protons.

14.7.1.2 Pre-RD Regime

Interestingly, by allowing a relatively large initial concentration of interface states N_{it0} and by assuming $\Delta N_{it}(t) \ll N_{it0}$, a completely different solution is obtained [59],

$$\Delta N_{it}(t) = A_P C(t), \quad (14.27)$$

with the prefactor

$$A_P = a \left(\frac{k_f}{k_r} \frac{\Delta N_{it, \max}}{N_{it0}} \right)^a = A_{RD}^{1+a} N_{it0}^{-a}. \quad (14.28)$$

This regime is termed pre-RD regime [59], because for intermediate concentrations of N_{it0} the number of created interface states ΔN_{it} will eventually become larger than N_{it0} , changing the overall behavior to that of the standard RD model. This is demonstrated in Figure 14.11 for the H_2 -RD model.

In the pre-RD regime the exponents have the values 1, 1/2, and 1/2 for proton, atomic, and molecular hydrogen transport, respectively. Note that these exponents do not depend on the kinetic exponent a as in the standard RD model. For classic drift-diffusion, these resulting exponents are not compatible with measurements. However, as has been shown [13,59], the introduction of dispersive transport can bring the exponents within the observed ranges.

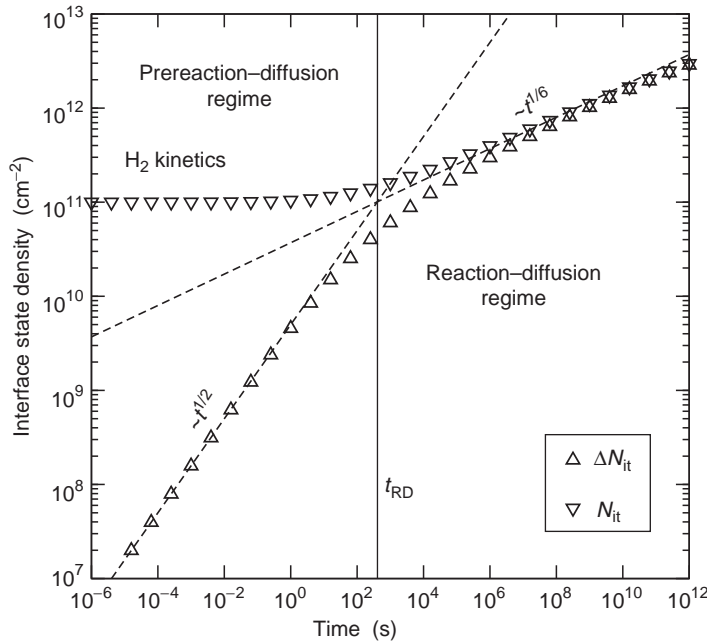


FIGURE 14.11

Two different regimes for a medium number of initial interface defects $N_{it0} = 10^{11} \text{ cm}^{-2}$. The transition between the pre-RD regime and the standard RD regime can be clearly observed. A value of $N_0 = 10^{13} \text{ cm}^{-2}$ was used in the simulations. (From Grasser, T., et al., *Trans. Dev. Mater. Reliab.*, 8(1), 79, 2008. With permission.)

The assumption $\Delta N_{it}(t) \ll N_{it0}$ has originally been introduced by Zafar [13]. This is based on the (actually mandatory) notion that the occupancy of the interface states depends on the position of the Fermi-level and that not all interface states are electrically active. In this context N_0 is now the maximum number of hydrogen binding sites rather than the maximum number of electrically observable interfaces states in a completely depassivated sample.

14.7.1.3 Relaxation as Predicted by the RD Model

As soon as the stress condition is removed, the forward-rate of the RD model is assumed to be negligible. Just like during the stress phase, the reaction is in quasiequilibrium, resulting in the left-hand side of Equation 14.21 to become very small. With $k_f \approx 0$, the actual values of k_r and D_c become irrelevant, except for a very short and insignificant reaction-limited initial phase. In addition, the species type has no influence on the relaxation and the overall behavior is again diffusion-limited. Consequently, the RD model predicts a universal relaxation practically independent of the species (H and H₂) as

$$r(\xi) = 1/(1 + \xi^{1/2}). \quad (14.29)$$

This analytic expression is compared to the numerical results for both species in Figure 14.12. Also shown is the measurement data of Reisinger et al. [9]. It is worthwhile to realize that the relaxation predicted by the RD model does not depend on any model parameters. Consequently, it must be clearly emphasized that since the relaxation predicted by the RD

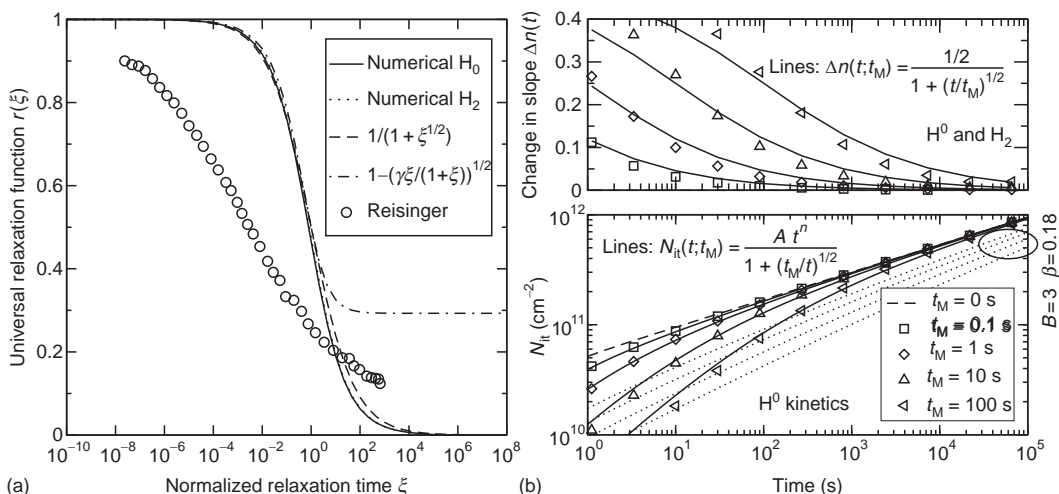


FIGURE 14.12

(a) Comparison of two analytic expressions for the RD relaxation behavior with numerical results obtained for H and H_2 kinetics. The power-law-like expression is accurate for all relaxation times and will be used as reference throughout this work. Also shown is the analytic expression derived in Ref. [56]. Due to the lack of parameters there is no way to fit the measurement data with the RD model. (b) Influence of the measurement delay t_M as predicted by the RD model. Comparison of the analytic model (lines) with the numerical solution (symbols) proves the excellent accuracy of the analytic model for $t > t_M$. Note that the RD model predicts a very small influence of delay for longer stress times, in contrast to Figure 14.8. For the sake of comparison, a more realistic influence of the measurement delay is given by the dotted lines, obtained with typical parameter values $B=3$ and $\beta=0.18$. There the individual curves obtained with different delay times remain clearly separated even after $t_s > 10^3$ s. (From Grasser, T., et al., *Proc. IRPS*, 268, 2007. With permission.)

model cannot be made to depend on gate bias, temperature, and process conditions, in stark contradiction to Figure 14.9, there is no way to fit the available measurement data. In particular, $\beta=1/2$ is much larger than observed experimentally, leading to a relaxation which is too slow in the beginning and too fast in the end. This is clearly visible in Figure 14.12 where most of the relaxation occurs within three to four decades whereas the measurements show relaxation over 12 decades. Consequences of this erroneous relaxation prediction are a heavily time-dependent but temperature-independent slope in the RD simulated delayed measurements, and a vanishing influence of the delay on the measurement result for $t_s \gtrsim 10 \times t_M$ (Figure 14.12), in contradiction to measurements [14,19], see also Ref. [6]. The only way to move the relaxation curve to shorter relaxation times is to bring the forward reaction into the quasisaturation regime where hydrogen has already piled up considerably in the oxide (assuming for instance a reflecting boundary condition). However, in addition to the fact that this behavior is not universal, the slope during the stress phase approaches zero.

14.7.2 Extended Classical RD Models

As the standard form of the RD model has been found to have also limitations during the stress phase [4,54,60], extended versions have been introduced. However, the question of whether these extended models are better able to describe the relaxation behavior has so far only been qualitatively assessed and a rigorous statement is missing. This will be done in the following.

14.7.2.1 Two-Region RD Model

First, it has been noted that due to the extremely thin oxides used in modern CMOS technology, the diffusing hydrogen species may quickly reach the oxide/poly interface [4]. As a consequence, the degradation will be dominated by the presumably slower diffusion in the poly gate. We will discuss two variants of RD models extended to account for such a situation. The first variant assumes the oxide/poly interface to be a perfect transmitter. At short times the oxide will be filled with H_2 . At later times, the overall hydrogen diffusion is dominated by the slower diffusion inside the polygate and the model behaves just like the standard H_2 -RD model. One might suspect that the hydrogen stored inside the oxide, where the diffusivity has been assumed to be larger, modifies the relaxation behavior. Under certain conditions this is indeed the case, with undesired properties, though, as shown in Figure 14.13. For large stress times, most hydrogen is stored in the poly and the model predicts the same relaxation as the RD model. Thus, in order to see the influence of the two regions we have to look at shorter stress times, in our particular case $t_s = 10$ s and $t_s = 100$ s, where the population in both regions is of the same order of magnitude. However, as shown in Figure 14.13, the shape of the relaxation curve does not agree with measurement data. Furthermore, the shape depends on the ratio of both reservoir occupancies, which changes with time and consequently results in a nonuniversal relaxation. We also remark that the assumptions underlying this model are in contradiction to a study which did not show a dependence of NBTI on the gate material [61].

14.7.2.2 Two-Interface RD Model

Next, we discuss a two-interface model which can be considered a refined form of the two-region model. It assumes that atomic hydrogen is released from the silicon/oxide interface which then diffuses through the thin oxide and depassivates defects at the oxide/poly interface [4,62]. The creation of defects at the opposite interface is supported by SILC

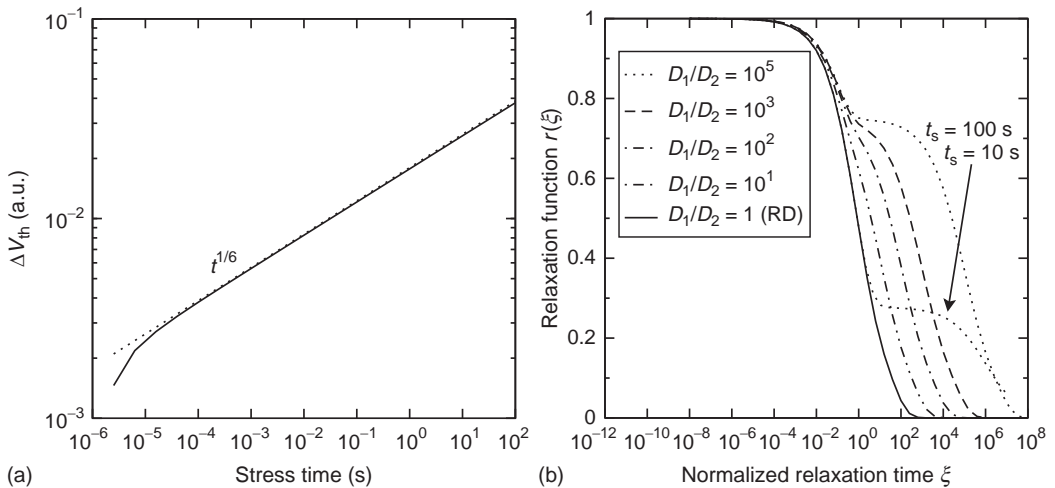


FIGURE 14.13

Numerical simulation of a generalized RD model with two different diffusion coefficients in the oxide and poly. For this particular set of parameters and the small stress time required to bring out this effects, no difference is visible during the stress phase (a), while the relaxation behavior slows down and displays nonuniversal humps (b). (From Grasser, T., et al., *Proc. IRPS*, 268, 2007. With permission.)

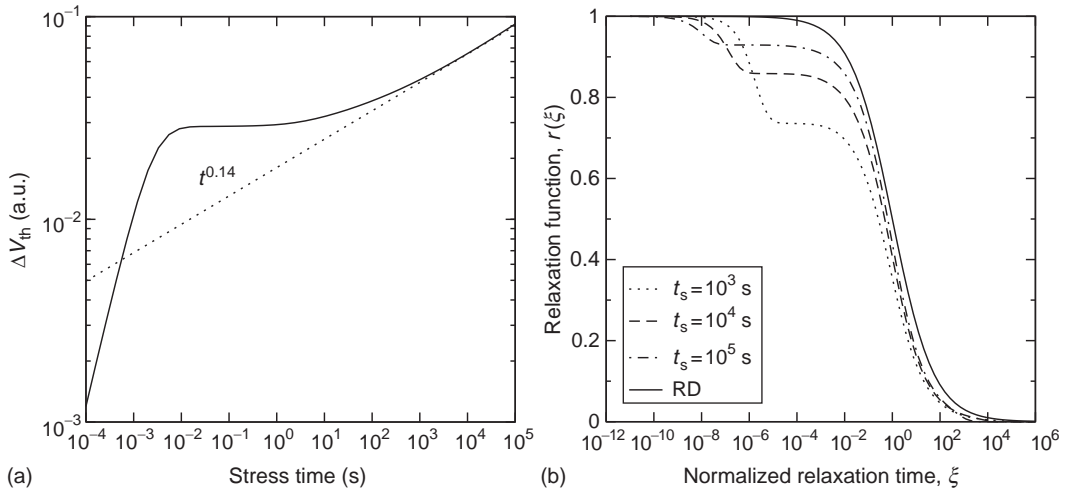


FIGURE 14.14

Numerical simulation of the two-interface RD model stress (a) and relaxation (b) behavior. Although faster relaxation than with the standard RD model is possible, the relaxation is not universal and the shape of the relaxation curve starkly disagrees with measurement data. In order to obtain a visible influence on the relaxation behavior, the hydrogen stored in both regions has to be of the same order of magnitude, which considerably lowers the power-law slope during the stress phase. (From Grasser, T., et al., *Proc. IRPS*, 268, 2007. With permission.)

measurements [4]. The hydrogen from the oxide and the released hydrogen at the oxide/poly interface diffuse as H_2 through the poly and result in an overall power-law exponent of $1/6$ at large times. It has been suggested that such a two-interface model may predict a faster recovery compared to the standard RD model [62]. For this to be the case, the amount of fast hydrogen stored in the oxide must be of the same order of magnitude compared to the slow hydrogen stored in the poly. As in the case of the two-region RD model, it is again possible to modify the relaxation behavior to a certain extent, see Figure 14.14. In this case the relaxation can be made faster than with the standard RD model because the fast hydrogen concentration inside the oxide is saturated, resulting in a shift to smaller normalized relaxation times ξ on the universal plot. However, just as with the two-region model, the resulting relaxation is not universal, as the ratio of these two hydrogen storage areas changes with time, see Figure 14.14.

14.7.2.3 Explicit $H-H_2$ Conversion RD Model

Another variant of the classic RD model aims at improving the model prediction at early times [54]. This is based on the suggestions that measurements might display a power-law exponent of $1/3$ during the initial stress phase [5,40], which is incompatible with the standard RD model. This has been explained by an extended RD model which explicitly accounts for the dimerization of H into H_2 [5,54],

$$\frac{\partial[H]}{\partial t} = k_{H_2}[H_2] - k_H[H]^2, \quad (14.30)$$

rather than assuming an instantaneous transition, in addition to the diffusion of both hydrogen species. Depending on the values of k_{H_2} and k_H , either pure H or H_2 kinetics can be observed. In addition, a regime with the aforementioned transitional power-law exponent of $1/3$, which eventually changes to $1/6$, is possible. Since recent measurements

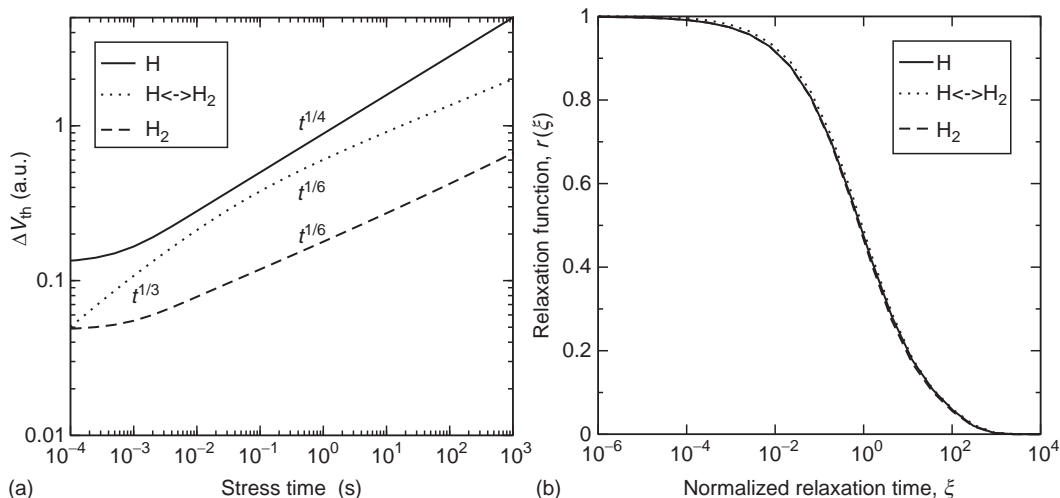


FIGURE 14.15

Numerical simulation of a generalized RD model with explicit H to H₂ conversion. (a) Depending on the choice of parameters, the model gives power-law exponents known from the H and the H₂ models during the stress phase, in addition to a transitional region with $n = 1/3$. (b) Since the model remains within the boundaries set by the pure H⁰ and H₂ model (which are equal during relaxation), the overall relaxation behavior cannot be influenced by any of the available parameters. (From Grasser, T., et al., *Proc. IRPS*, 268, 2007. With permission.)

give a long-term exponent closer to 1/6 than to 1/4, the parameters have to be chosen in such a way that the total amount of stored [H₂] is much larger than [H]. One might conclude from this that the two distinct reservoirs of H and H₂ may allow for a modified relaxation behavior. However, this is not the case for the simple reason that the model stays within the limits set by pure H and H₂ behavior, just as during the stress phase. Since the relaxation of both species is practically equivalent, no influence on the relaxation behavior is obtained from such a model, see Figure 14.15.

14.7.2.4 Vanderbilt Model

By employing first-principles calculations, Tsetseris et al. [48] investigated the electrochemical reaction (Equation 14.20), which is one of the foundations of the RD model. They found an activation energy of about 2.4 eV, in agreement with measurement data [63]. Such a barrier is way too large to allow the bond to be broken during typical bias temperature conditions. Although the presence of holes lowers the activation energy to values around 2.1 eV, this value is still too high for a relevant contribution. Consequently, they suggested an alternative reaction triggered by protons supplied from the semiconductor bulk



Provided H⁺ is readily supplied from the bulk, the differential equations resulting from Equation 14.31 combined with the standard diffusion Equation 14.22 are from a mathematical point of view equivalent to the equations resulting from the standard RD model for atomic hydrogen diffusion [62]. Consequently, the model predicts a slope of 1/4 and the same relaxation as the RD model and can therefore not be used in this form to explain NBTI.

14.7.3 Final Notes on RD Models

We have shown that irrespective of the extensions applied to the RD model, the recovery behavior observed during measurement cannot be described with the published RD variants in their present form. The fact that some OTF measurements and the corrected MSM measurements give exponents of around $n=0.15$, which is close to the value predicted by the H_2 -based RD model ($n=1/6$), should not let one arrive at the conclusion that the RD model is consequently reasonable. In particular, we think one has to be extremely cautious with a point of view that the RD model correctly covers the stress part while only the relaxation part needs to be refined. The point to make here is that the $1/6$ exponent during the RD stress phase is a result of a delicate interplay between the forward and backward reaction [5]. Without the backward reaction, which dominates the time evolution by inserting the diffusion-limited component into the RD model, the forward reaction alone would result in $n=1$. It is only during relaxation, where the forward rate is suppressed, that the poor performance of the RD reverse reaction becomes visible. Consequently, we do not see any reason to believe the very same reverse reaction to be valid during the stress phase to constructively change the reaction-limited exponent of $n=1$ to the proposed diffusion-limited value of $n=1/6$.

14.7.4 Dispersive NBTI Models

It has been clearly shown in the previous sections that the RD model predicts 80% of the relaxation to occur within three to four decades, while in reality relaxation is observed to span more than 12 decades [9,14,21]. This indicates some form of dispersion in the underlying physical mechanism(s). Various forms of dispersion have already been introduced into NBTI models based on either (1) diffusion [12–14], (2) hole tunneling from/into states in the oxide [16], and (3) reaction rates at the interface [11,64]. The models suggested to capture these mechanisms will be benchmarked in the following using the universality as a metric.

14.7.4.1 Reaction-Dispersive-Diffusion (RDD) Models

First, we consider generalized RD models based on dispersive transport of the hydrogen species [14,65]. These models are obtained by replacing the classic drift-diffusion Equation 14.22 in the RD model by its dispersive counterpart. Several different variants of dispersive transport equations have been used for the formulation of NBTI models [10,13,14]. It can be shown that the transport models themselves give practically identical results and that the differences in the final model prediction can be traced back to different assumptions used for the boundary and initial conditions [59].

These differences are best studied using the dispersive multiple-trapping (MT) transport equations [66–68]. In the MT model the total hydrogen concentration H consists of hydrogen in the conduction states H_c and trapped hydrogen as

$$H(\mathbf{x},t) = H_c(\mathbf{x},t) + \int \rho(\mathbf{x},E_t,t)dE_t, \quad (14.32)$$

with $\rho(\mathbf{x},E_t,t)$ being the trapped hydrogen density ($\text{cm}^{-3} \text{ eV}^{-1}$) at the trap level E_t . Transport is governed by the continuity equation and the corresponding flux relation:

$$\frac{\partial H}{\partial t} = -\nabla \cdot \mathbf{F}_c + G_c. \quad (14.33)$$

Note that in contrast to Equation 14.22 the time derivative of the total hydrogen concentration is used in Equation 14.33, which also accounts for the exchange of particles with the trap levels. The occupancy of the trap levels is governed by balance equations which have to be solved for each trap level

$$\frac{\partial \rho(E_t)}{\partial t} = \frac{\nu}{N_c} (g(E_t) - \rho(E_t)) H_c - \nu \exp\left(-\frac{E_c - E_t}{k_B T_L}\right) \rho(E_t), \quad (14.34)$$

with ν being the attempt frequency, N_c the effective density-of-states in the conduction band, and E_c the conduction band edge. An exponential trap density-of-states is commonly used [14,68]:

$$g(E_t) = \frac{N_t}{E_0} \exp\left(-\frac{E_c - E_t}{E_0}\right), \quad (14.35)$$

which, in this particular context, results in a power law [14,67] for the time-dependence of ΔN_{it} . It is also worth recalling that the transport will only be dispersive for $E_0 > k_B T_L$, that is, for sufficiently deep trap distributions [68].

As the MT equations are rather complex and can in general only be solved numerically, simplified equations have been derived by Arkhipov and Rudenko [68,69]. Their approximate solution relies on the existence of the demarcation energy,

$$E_d(t) = E_c - k_B T_L \log(\nu t), \quad (14.36)$$

separating shallow from deep traps and was derived to describe the broadening of an initial particle distribution in the conduction band, see Figure 14.16. This is not the case during NBT stress, however, where we have to deal with a continuous generation of particles at the interface during the stress phase. An extended model suitable for NBTI

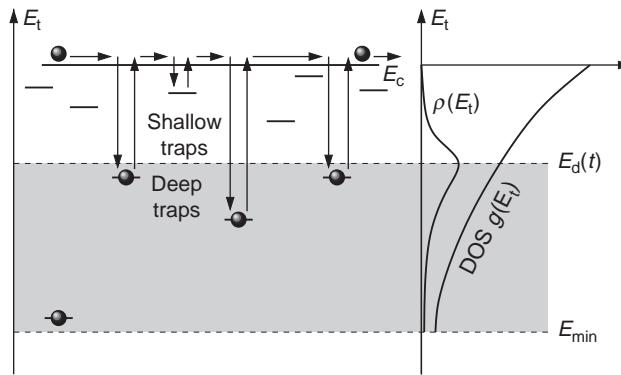


FIGURE 14.16

Schematic illustration of MT dispersive transport. Particles from the conduction band fall into the traps and are thermally reemitted into the conduction band. Reemission is more likely for shallow traps. The time-dependent demarcation energy E_d separates shallow from deep traps. With time, the demarcation energy becomes more negative, until the bottom of the trap distribution is reached ($E_d \rightarrow E_{\min}$) and equilibrium is obtained. Before equilibrium, the motion of the particle packet slows down with time. Note how the individual trap levels, which microscopically correspond to the different energy levels of hydrogen in an amorphous material, are approximated by a macroscopic density-of-states. (From Grassler, T., et al., *Trans. Dev. Mater. Reliab.*, 8(1), 79, 2008. With permission.)

has been derived in Ref. [59] which, in the highly nonequilibrium regime, describes the overall motion as

$$H(\mathbf{x},t) - H_0(\mathbf{x}) = -\nabla \cdot F(\mathbf{x},t) + \int_{t_0}^t G_c(\mathbf{x},t')dt', \quad (14.37)$$

with the effective flux of the total concentration of the species H given by

$$F(\mathbf{x},t) = -D_c\tau(t)\left(\nabla H - Z\frac{E_{\text{ox}}}{V_T}H\right). \quad (14.38)$$

Note that Equation 14.37 is basically the time integral of Equation 14.33 and thus does not contain a time derivative anymore. This is a consequence of the fact that the dynamics of the system can be incorporated solely into $\tau(t)$, which directly depends on the hydrogen trap density-of-states and the demarcation energy as

$$\frac{1}{\tau(t)} = \frac{\nu}{N_c} \int_{-\infty}^{E_d(t)} g(E_t)dE_t. \quad (14.39)$$

This is a characteristic feature of any adiabatic process, where the time-dependence of the whole system is determined by the slowest process, in our case, the thermal equilibration of hydrogen [70]. For the particular case of an exponential trap density-of-states as in Equation 14.35, τ can be evaluated in closed form to

$$\tau(t) = \frac{1}{\nu} \frac{N_c}{N_t} (\nu t)^\alpha, \quad (14.40)$$

with the dispersion parameter

$$\alpha = k_B T_L / E_0. \quad (14.41)$$

It is also worth pointing out that by introducing a time-dependent diffusivity $D(t) = D_c\tau(t)$, under certain circumstances a link to empirical dispersive transport models can be established [59].

Of particular interest for the derivation of NBTI models is the concentration of the mobile hydrogen H_c , which is directly linked to the total hydrogen concentration H as

$$H_c(\mathbf{x},t) = \frac{\partial \tau(t)H(\mathbf{x},t)}{\partial t}. \quad (14.42)$$

This relation will be used for the formulation of the different NBTI boundary conditions.

In order to obtain an NBTI model, the dispersive transport equation has to be coupled to the electrochemical reaction assumed to take place at the interface. For the present analysis we remark that the macroscopic hydrogen trap density-of-states is derived for an amorphous bulk material and is unlikely to be valid close to an interface. In that context, the physical mechanisms justifying the conduction band concept in conjunction with hydrogen hopping next to the interface need to be evaluated and justified. Published dispersive NBTI models, however, are based on the validity of this concept, and the different interpretations explain the discrepancies in these models. As in the RD model, the kinetic equation

describing the interface reaction is assumed to be of the form Equation 14.21. Also, the interface reaction is assumed to be in quasiequilibrium.

A crucial question for the formulation of dispersive NBTI models is how to link it to the interfacial hydrogen concentration H_{it} . In the following, we will consider two different models. The first assumes that H_{it} is given by the total hydrogen concentration H . Thus, in the RD regime one obtains for neutral particles [14,59]:

$$\Delta N_{it}(t) = A_{RD} \left(\frac{D_c}{\nu} \frac{N_c}{N_t} \right)^{1/(2+2a)} (\nu t)^{\alpha/(2+2a)}, \quad (14.43)$$

with the same prefactor A_{RD} as in the RD model given through Equation 14.26. For atomic hydrogen ($a=1$) the exponent is given through $n = \alpha/4$ while molecular hydrogen ($a=2$) gives $n = \alpha/6$ (Figure 14.17).

For the proton, one can show that [14,59]

$$\Delta N_{it}(t) = A_{RD} \left(\frac{D_c}{\nu} \frac{N_c}{N_t} \frac{E_{ox}}{V_T} \right)^{1/2} (\nu t)^{\alpha/2} \quad (14.44)$$

holds. Note that the numerical solution for H^+ may contain a transitional regime with $n = \alpha/4$, where the diffusive component still dominates.

Since α equals 1 in the diffusive limit and 0 in the extremely dispersive case, Equations 14.44 and 14.43 imply that with dispersive transport an exponent smaller than the RD exponents of 1/2, 1/4, and 1/6 can be obtained. Also, for increasing trap density N_t the total amount of degradation decreases.

A qualitative explanation for the reduction of the exponent can be given by noting that dispersive transport results in most particles being trapped close to the interface, yielding a steeper profile compared to classic diffusion. As all hydrogen is available for the reverse rate in Equation 14.21, the net interface state generation is suppressed, resulting in a smaller exponent.

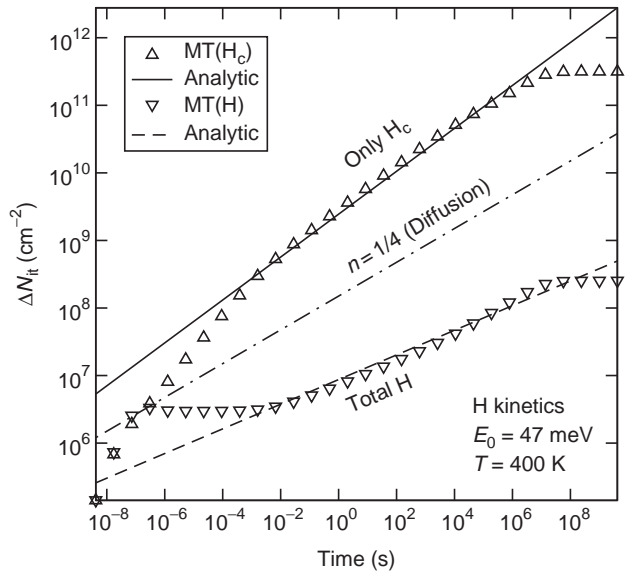


FIGURE 14.17

Interface state density as a function of the boundary condition calculated numerically by solving the MT equations in comparison to the analytic expressions (Equations 14.43 and 14.46) for $N_{it0}=0$. Good agreement between the numerical and analytical solution is obtained for both boundary conditions. (From Grasser, T., et al., *Trans. Dev. Mater. Reliab.*, 8(1), 79, 2008. With permission.)

Since the dispersion parameter α depends linearly on the temperature, a linear temperature-dependence of the exponent is obtained as [14]:

$$n_1 = \frac{\alpha}{2 + 2a} = \frac{k_B T_L}{2E_0(1 + a)}. \quad (14.45)$$

This is consistent with experimental results obtained with delayed measurements [8,14].

The previous model was based on the assumption that all hydrogen, mobile and trapped, can participate in the NBTI reverse rate. In contrast, if we now assume that only the mobile hydrogen can participate in the reverse rate, that is, $H_{it} = H_c(0)$, which appears to be the more natural boundary condition for the MT model [71], one obtains for neutral particles

$$\Delta N_{it}(t) = A_{RD} \left(\frac{D_c}{\nu} \frac{N_t}{N_c} \right)^{1/(2+2a)} \left(\frac{1+a}{1+a\alpha/2} \right)^{1/(1+a)} (\nu t)^{(1-\alpha/2)/(1+a)}. \quad (14.46)$$

For atomic hydrogen, the exponent $n = 1/2 - \alpha/4$ is obtained while H_2 results in $n = 1/3 - \alpha/6$. Hence, for increased dispersion the exponents become now larger than their RD equivalents. Furthermore, when the trap density is increased, the degradation increases. This is in agreement with the previously stated result that the inclusion of traps into a standard RD model increases the exponent [29,72].

Interestingly, for H^+ one obtains [59]:

$$\Delta N_{it}(t) = \left(2 \frac{k_f N_0}{k_r} D_c \frac{E_{ox}}{V_T} \right)^{1/2} t^{1/2}, \quad (14.47)$$

with an exponent $n = 1/2$. This is equal to the result obtained by the standard RD model [59].

Again, qualitatively, in this model the newly released hydrogen quickly falls into the traps, but for times larger than $1/\nu$ most hydrogen resides in deep traps and is therefore not as easily available for the reverse rate in Equation 14.21. This suppresses the reverse reaction and consequently enhances the net interface state generation and results in a larger exponent.

In contrast to the total hydrogen boundary condition, now the exponent decreases with increasing temperature through

$$n_2 = \frac{1 - \alpha/2}{1 + a} = \frac{2E_0 - k_B T_L}{2E_0(1 + a)} = \frac{1}{1 + a} - n_1. \quad (14.48)$$

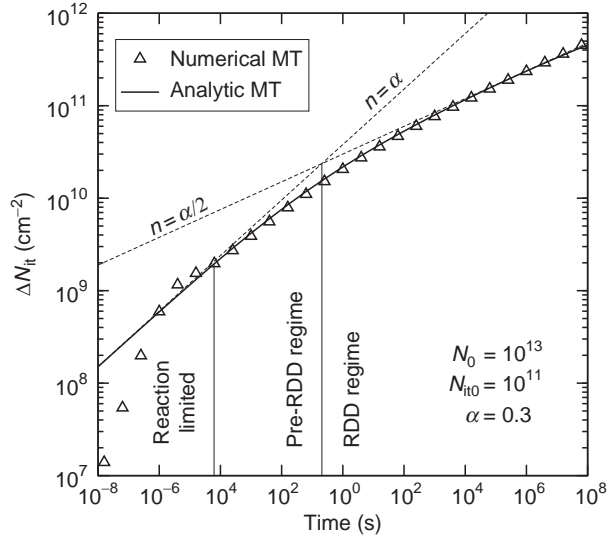
This is in contradiction to currently available observations [8,14,20]. Note, however, that this particular temperature-dependence is a consequence of the exponential trap density-of-states and a hardly noticeable temperature-dependence has been reported [29] using a Gaussian distribution on top of the exponential density-of-states.

14.7.4.2 Dispersive Pre-RD Regime

For the case that a large initial concentration of interface states is allowed, the pre-RD result (Equation 14.27) can be directly transferred to the dispersive case and one obtains for the total-hydrogen-boundary-condition [59]:

FIGURE 14.18

Example simulation showing the transition between the pre-RDD regime and the RDD regime which could be used to explain a different initial exponent compared to the long-time exponent. Shown are the numerical and analytical solutions of the MT equations. Note that the analytic solution is only valid after the reaction-limited phase. (From Grasser, T., et al., *Trans. Dev. Mater. Reliab.*, 8(1), 79, 2008. With permission.)



$$\Delta N_{it}(t) = A_P \left(\frac{D_c}{\nu} \frac{N_c}{N_t} \right)^{1/2} (\nu t)^{\alpha/2}. \quad (14.49)$$

For the proton one obtains

$$\Delta N_{it}(t) = A_P \left(\frac{D_c}{\nu} \frac{N_c}{N_t} \frac{E_{ox}}{V_T} \right) (\nu t)^{\alpha}. \quad (14.50)$$

As before, the exponents $n = \alpha$, $\alpha/2$, and $\alpha/2$ for H^+ , H^0 , and H_2 reduce to their pre-RD equivalents 1, $1/2$, and $1/2$ for $\alpha = 1$. Also, the exponent increases linearly with temperature similarly to Equation 14.45, and the same compatibility to measurements is given.

Interestingly, it can be shown that Equation 14.50 is equivalent to the Zafar model, which also has a slope $n = \alpha$ rather than $n = \alpha/2$ as obtained in the RD regime [59]. For an intermediate concentration of interface states, the transition between the pre-RD and the RD regime is shown in Figure 14.18.

14.7.4.3 Relaxation as Predicted by the RDD Models

A previous analysis of the relaxation behavior predicted by dispersive transport equations [14] was based on various assumptions (such as pulse-like excitation [65], uncertainties in the boundary conditions [65], and a neglected history of previously trapped hydrogen atoms during relaxation) which led to only approximative solutions. As it turned out, a more rigorous analytic derivation is rather involved. An approximation for $\xi < 1$ ($t_r < t_{sr}$ as normally encountered during typical MSM measurements), is given by Equation 14.11 with B and β depending on the boundary condition and the dispersion coefficient α . Interestingly, for $\xi > 1$ the behavior changes and different values for B and β have to be used (cf. Figure 14.19).

In order to avoid any uncertainties inherent in approximate analytical solutions, we numerically solve the full time-dependent multiple trapping model [66] to allow for an accurate description of both the stress and the relaxation phase. The results shown in Figure 14.19 display a much broader range of possible relaxation characteristics compared to classic diffusion. Nevertheless, the dispersive transport models in their present form are

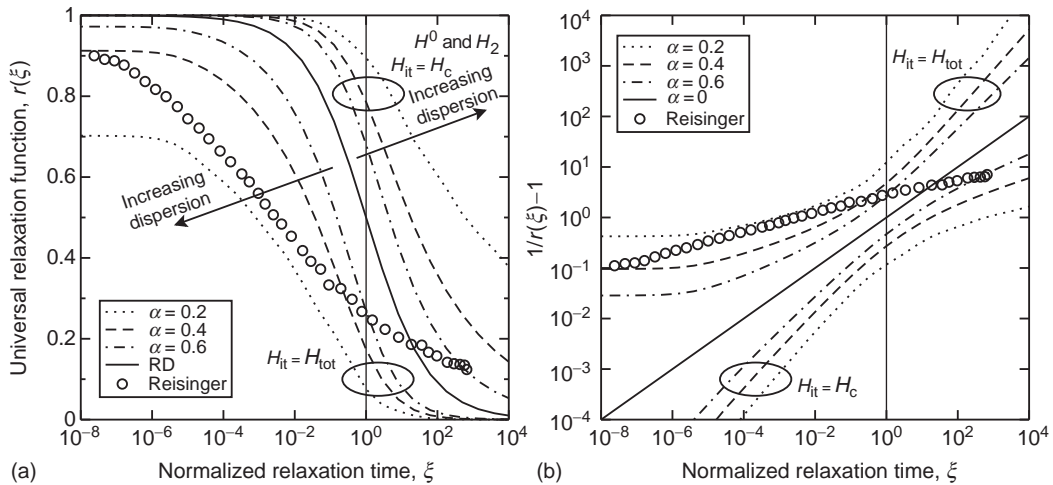


FIGURE 14.19

Relaxation as predicted by the full numerical solution of the dispersive transport models for various values of the dispersion parameter α . (a) Shows the universal relaxation function $r(\xi)$ while (b) shows $1/r(\xi) - 1$ which should be close to a straight line according to Equation 14.12. Also shown are the data from Reisinger as a reference. The model with $H_{it} = H_{tot}$ always stays below the diffusive (RD) limit, while the model with $H_{it} = H_c$ always stays above. The diffusive limit is like a watershed which cannot be crossed by either model. Also note that the $H_{it} = H_{tot}$ model appears to have a limit different from unity for $\xi \rightarrow 0$, which is a result of an extremely fast relaxation triggered by the hydrogen stored right at the interface. Note that this component is not universal. (From Grasser, T., et al., *Proc. IRPS*, 268, 2007. With permission.)

not able to fully explain the experimentally observed relaxation on their own. They might, however, be combined with a hole trapping model in order to account for a slow component during relaxation. Note that the standard RD model is an unlikely candidate for this slow component since any contribution would be negligible for large relaxation times.

Also note that the $H_{it} = H_{tot}$ model appears to have a limit different from unity for $r(\xi \rightarrow 0)$. This is a result of the extremely fast relaxation triggered by the hydrogen stored directly at the interface. The exact shape of this initial hump (not shown) depends on the stress time and the width of the interfacial layer, thereby rendering this model nonuniversal.

14.7.4.4 Dispersive-Rate Coefficients

Next we consider reaction-limited models using a dispersion in the rate coefficients [11,64]. This is based on the observation of Stesmans et al. [35] who could best describe the dissociation kinetics of hydrogen-passivated P_b centers at the interface using first-order kinetics and a Gaussian distribution of interface states. A similar observation was made regarding the passivation of P_{b0} and P_{b1} interface defects [34]. Huard et al. [11] base their permanent component on such a dispersive forward rate, assuming that the generated interface states do not relax at all, or at least not at shorter and medium relaxation times [73].

The model derivation uses the RD interface reaction given in Equation 14.21. In contrast to the RD model, however, it is assumed that the generated interface states are permanent. Thus, k_r can be set to zero and Equation 14.21 has the solution:

$$N_{it} = N_0(1 - \exp(-k_f(E_d)t)). \quad (14.51)$$

By invoking the Arrhenius law for the rate coefficient k_f , one obtains the forward rate as a function of the dissociation energy E_d as

$$k_f(E_d) = k_{f0} \exp\left(-\frac{E_d}{k_B T_L}\right). \quad (14.52)$$

Assuming a distribution of dissociation energies given by the Fermi-derivative function (Equation 14.5), which, in contrast to a Gaussian distribution, allows for a closed form solution, one obtains by summing the individual contributions

$$\frac{\Delta N_{it}}{N_0} = \int g_P(E_d, E_{dm}) (1 - \exp(-k_f(E_d)t)) dE_d. \quad (14.53)$$

This integral can be approximated by realizing that $N_{it}(E_d)$ is close to unity below $E^*(t) = k_B T_L \ln(k_{f0}t)$ and zero otherwise. One can thus approximately write

$$\frac{\Delta N_{it}}{N_0} \approx \int_0^{E^*(t)} g_P(E_d) dE_d = \frac{1}{1 + \left(\frac{t}{\tau}\right)^{-\alpha}}, \quad (14.54)$$

with $\tau = k_{f0}^{-1} \exp(E_{dm}(E_{ox})/k_B T_L)$ and $\alpha = k_B T_L / \sigma_f$. Note the similarity with the relaxation expression (Equation 14.12) and the correspondence between E^* and the demarcation energy in the dispersive multiple trapping equations. The median dissociation energy E_{dm} was assumed to depend on the oxide electric field in order to accommodate for the reported field dependence. For short stress times, the above simplifies to a power law

$$\Delta N_{it}(t) = \Delta N_{it, \max} \left(\frac{t}{\tau}\right)^\alpha. \quad (14.55)$$

Again, as with the dispersive transport model, a temperature-dependent slope is obtained. The analytic solution (Equation 14.54) is compared to the numerical solution in [Figure 14.20](#), where excellent accuracy is obtained for $\sigma_f > 0.12$, which corresponds to $\alpha < 0.21$, and is thus well within the required regime.

In the above model the backward rate was assumed to be negligible, resulting in an unrecoverable degradation of ΔN_{it} . In order to generalize this model to allow at least for some recovery, one has to account for the reverse rate in Equation 14.21. In contrast to the RD model, however, where the diffusion of the hydrogen species eventually limits the reverse rate, it is now assumed that hydrogen at the interface is readily available. Formally, this may be done by setting H_{it} constant in Equation 14.21, equivalent to a large background hydrogen concentration. The solution of Equation 14.21 with $k_r \neq 0$ is readily obtained as

$$N_{it}(t, E_d, E_a) = N_0 \frac{k_d(E_d)}{k_d(E_d) + k_a(E_a)H_{it}} (1 - \exp(-k_d(E_d)t - k_a(E_a)H_{it}t)), \quad (14.56)$$

with the overall time evolution of N_{it} given through

$$N_{it}(t) = \int dE_d \int dE_a N_{it}(t, E_d, E_a) g_P(E_d, E_{dm}) g_P(E_a, E_{am}). \quad (14.57)$$

A numerical solution of Equation 14.57 is given in [Figure 14.20](#) for varying parameters σ_f and σ_r . Obviously, the introduced reverse rate strongly influences the stress phase.

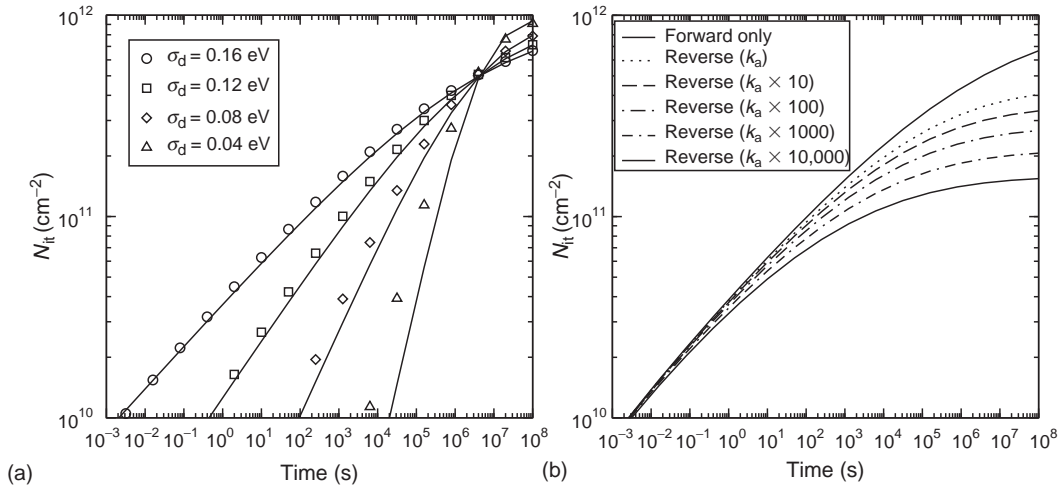


FIGURE 14.20

(a) Comparison of the analytic result of Equation 14.54 (lines) with the numerical solution of Equation 14.53 (symbols). Good agreement is obtained for larger dispersion σ_f . (b) As soon as the reverse rate is taken into account, an additional curvature in the slope is introduced which increases with the reverse rate.

During the relaxation phase an apparently very flexible model behavior is observed and the model can be nicely fit to a single relaxation curve. Unfortunately, however, the excellent fit during the single relaxation phase adversely affects the quasi-power-law exponent during the stress phase which reduces to very small values ($n \approx 0.03$). Furthermore, the model does not scale universally as demonstrated in Figure 14.21.

14.7.4.5 Simple Dispersive Hole Trapping Model

In addition to the creation of interface states, trapped charges have been made responsible for the observed threshold voltage shift during NBT stress. In particular, it has been argued

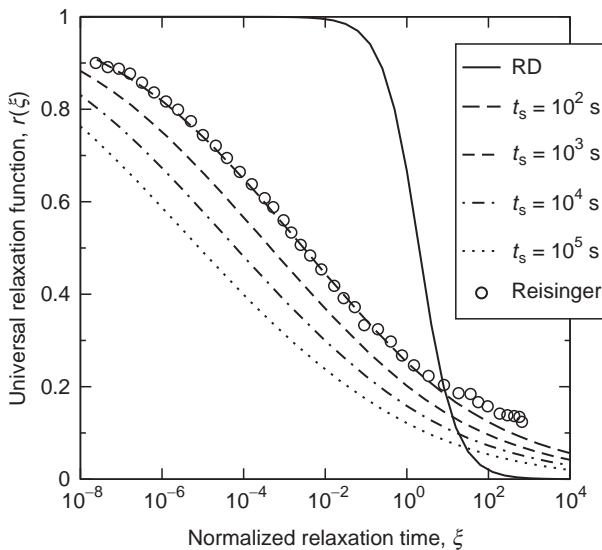


FIGURE 14.21

The dispersive rate model can be fit to an individual relaxation curve, $t_s = 100$ s in this case, but does not scale universally. In addition, the calibrated model gives a rather strong curvature during the stress phase with a too small power-law exponent $n \approx 0.03$, the variance of the rate coefficients had to be set to a value considerably larger than reported ($\sigma_A = 0.211$ eV and $\sigma_D = 0.264$ eV, compare Refs. [34,35]), and in general the model cannot be fit to both the stress and relaxation phase.

that these charges are responsible for the fast component observed both during stress and relaxation [11,16]. A simple phenomenological hole trapping model has been used by Yang et al. [16] based on a broad distribution of trapping times. It is assumed that hole trapping occurs over a broad spectrum of capture and release times following first-order kinetics

$$\frac{\partial p(t, \tau_c, \tau_e)}{\partial t} = \frac{1}{\tau_c} (N_{ot} - p) - \frac{p}{\tau_e}. \quad (14.58)$$

Here, p is the hole concentration of a trap with capture and emissions times τ_c and τ_e , while N_{ot} is the trap density. In order to fit their measurement data, Yang et al. coupled the two time-constants via $\tau_e = k\tau_c$ and employed different capture and release times during stress and relaxation. The overall time evolution of all trapped charges is obtained by weighing all contributions using a probability density function in a manner similar to the rate-limited model (Equation 14.53):

$$p(t) = \int d\tau p(t, \tau) f(\tau). \quad (14.59)$$

The probability density function for the relaxation times was assumed to be given by a log normal distribution:

$$f(\tau) = \frac{1}{\sqrt{2\pi\tau\sigma}} \exp\left(-\frac{1}{2} \left(\frac{\log(\tau) - \mu}{\sigma}\right)^2\right). \quad (14.60)$$

The numerical solution of Equation 14.59 is shown in Figure 14.22. Although a fast component can indeed be formed and a single relaxation curve can be nicely fit, no universality is observed. Note that this was to be expected due to the mathematical similarity of Equation 14.59 with Equation 14.53. We remark that this is not in disagreement with the good agreement to the measurement data reported in Ref. [16], but possibly a consequence of the narrow range of stress times employed in that study.

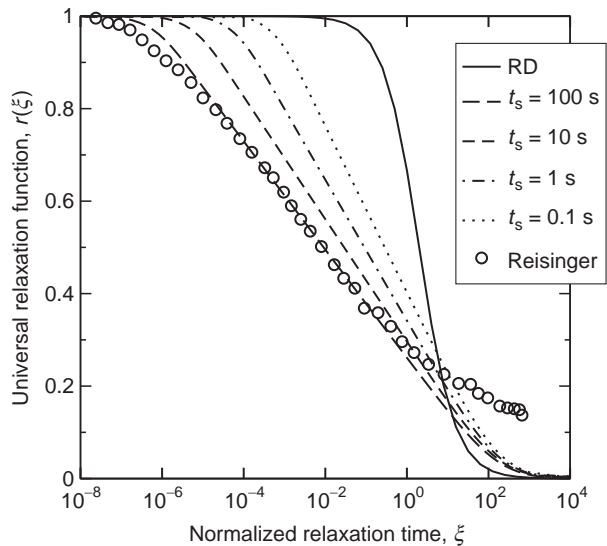


FIGURE 14.22

(a) Simple dispersive hole trapping model used in Ref. [16] can be fit to an individual relaxation curve but does not scale universally and gives a very small slope during the stress phase. (From Grasser, T., et al., *Proc. IRPS*, 268, 2007. With permission.)

14.7.4.6 Detailed Dispersive Hole Trapping Model

A more detailed hole trapping model has been derived by Tewksbury [74] based on a number of possible transitions from conduction, valence, and interface states into bulk oxide traps. Its use for NBTI has been suggested by Huard et al. [75] to cover the recoverable part of the degradation. For the following discussion we limit ourselves to the component of the model which results from charge transfer from an interface state into an oxide trap and back to the interface state, the other suggested mechanisms behave similarly [74] and follow analogously. During stress, the threshold voltage shift due to trapped bulk charge accumulated via transfer from the interface states can be given as $S(t_s) \approx A \ln(t_s/\tau_{0s})$ while the absolute relaxation is given by $R_M(t_s, t_r) \approx A \ln(t_s \tau_{0r}/(t_r \tau_{0s}))$ and depends (at least in this approximate form) universally on t_r/t_s . However, using the previous two relations, the relative relaxation function is given by

$$r(t_s, t_r) \approx 1 - \ln\left(\frac{t_r}{\tau_{0r}}\right) \ln^{-1}\left(\frac{t_s}{\tau_{0s}}\right), \quad (14.61)$$

which cannot be written as a function of t_s/t_r and is consequently not universal in our sense. The full numerical solution of the Tewksbury model is given in Figure 14.23 together with an excellent fit for a single relaxation curve. However, in order to obtain such a fit, the logarithmic behavior of the hole trapping component results in a slope close to zero during the stress phase. Also shown in Figure 14.23 is a permanent component modeled by a numerical solution of a dispersive forward rate only, as suggested by Huard et al. [75]. Note however, that after a certain stress time the degradation will be dominated by ΔQ_{it} and the observed relaxation given only through ΔQ_{ox} will be minimal. This is also not compatible with the data at hand where even at large stress times considerable relaxation can be observed.

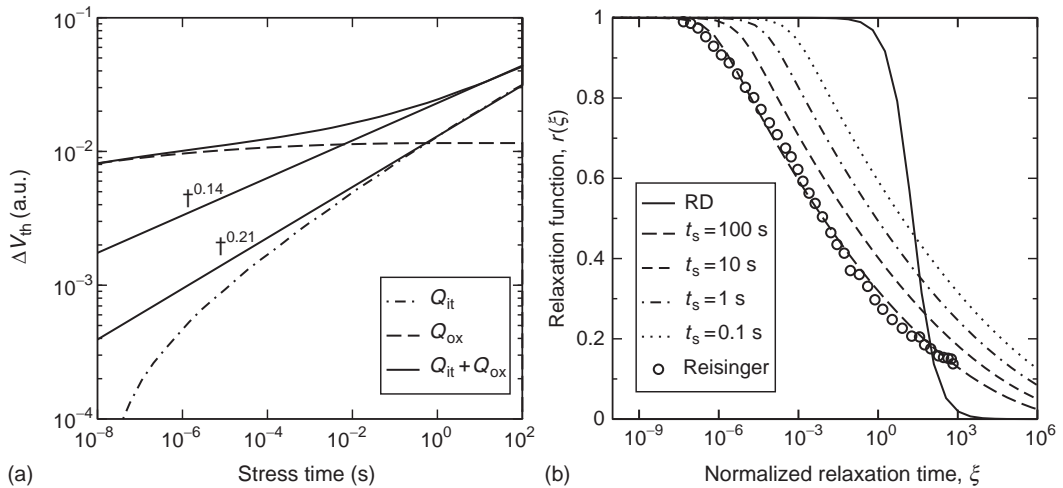


FIGURE 14.23

(a) Behavior during stress as predicted by the Tewksbury model on top of the dispersive rate (Equation 14.54). (b) Relaxation predicted by the Tewksbury model. The model can be fit to the data for an individual relaxation curve, here again at 100 s, but does not scale universally. Also, the excellent fit comes at the price of a very small power-law exponent at early times during the stress phase, in contradiction to Figure 14.4 of Ref. [9]. We were not able to fit both the stress and relaxation phases with the same set of parameters.

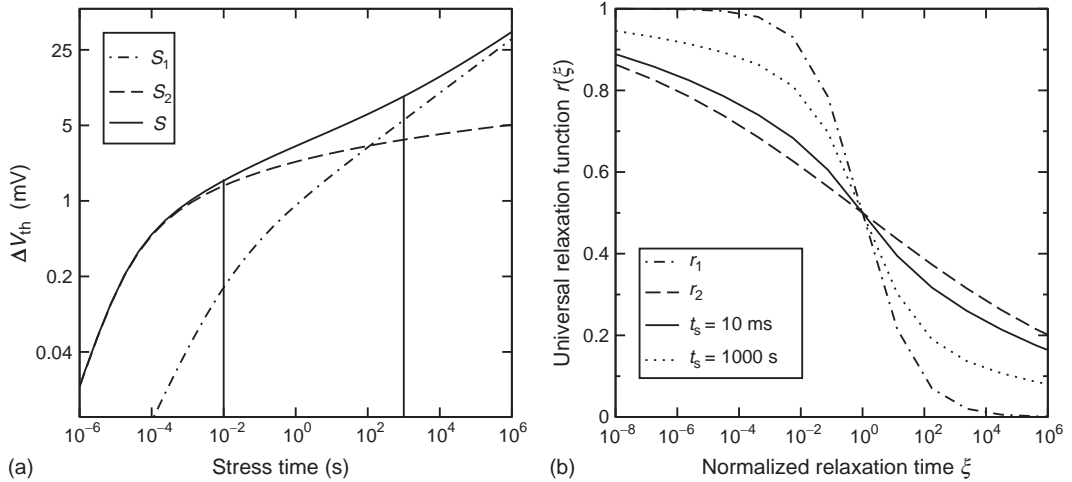


FIGURE 14.24

Schematic illustration of the contribution of two processes to the overall NBTI stress and relaxation characteristics. Process 1 could be the RD model while process 2 might be a simple hole-trapping process. If both processes are of the same order of magnitude, as is frequently assumed, the relaxation recorded at short stress times will be considerably different from a relaxation characteristic recorded at later times, in contrast to measurement data which are universal for usually employed stress times.

14.7.5 Multiple Mechanisms

Since none of the studied mechanisms can fully capture the universal relaxation, we have considered various combinations in our numerical framework. In order to obtain a universal behavior, some points need to be considered. Consider the case that the total observed threshold voltage shift is the result of two independent mechanisms, that is, $S = S_1 + S_2$. During relaxation one observes $R = S_1 r_1 + S_2 r_2$ and the normalized relaxation function is given by $r = \rho r_1 + (1 - \rho) r_2$ with $\rho = S_1 / (S_1 + S_2)$. If the two degradation mechanisms progress differently with time, ρ will be a function of t_s and r cannot be universal, see Figure 14.24. We thus conclude that for the relaxation to be universal, the two mechanisms need to be tightly coupled, that is, $S_1 / S_2 = \text{const}$, or at least roughly constant within the range of measured ξ and within the measurement tolerance. Alternatively, both mechanisms could relax equally, $r_1 \approx r_2$, or one process could be dominant for the range of recorded stress times. Finally, one process could be permanent or slowly relaxing, forming the permanent component identified in Ref. [59] which would have to be subtracted from the total relaxation data in order to study the universally recovering component.

14.8 Conclusions

We have thoroughly analyzed the relaxation of NBT stress-induced degradation using data from various groups. The observed universal relaxation behavior has been quantified and modeled using possible empirical expressions. It has been demonstrated that data obtained via conventional MSM sequences can be analytically described as a function of the delay introduced during the measurement. In particular, this analytic expression allows one to reconstruct a corrected degradation curve. Using this corrected curve, it might be possible to more accurately estimate the time-to-failure.

We have then used the relaxation behavior and in particular the universality as a benchmark for existing NBTI models. There we have found that none of the existing models are capable of reproducing both the stress and the relaxation phase with the same set of parameters. While the classic RD model scales universally, it predicts relaxation to occur mainly during three to four decades, in contradiction to detailed relaxation measurements available in literature which span more than 12 decades. No improvement could be found in extended RD models using two regions, a second interface, or an explicit transition from atomic to molecular hydrogen. Models based on an extension of the RD model with dispersive transport somewhat improve on the situation but are still not able to cover the whole relaxation regime. Other dispersive models, like dispersive forward and backward rates or dispersive hole-trapping models allow one to fit an individual relaxation curve only but are not universal. In addition, we were not able to describe both the stress and relaxation phase with the same set of parameters. This indicates a significant gap in our current understanding of NBTI.

We particularly wish to point out that it is of utmost importance not to consider the inaccuracies of existing models during the relaxation phase of secondary importance compared to the stress phase. The reason for this is only partially related to the frequently quoted fact that continuous DC stress is rarely observed in a circuit and that duty-cycle dependent corrections have to be applied. The more important point we want to make here is that during the stress phase the relaxation mechanism in existing models always interacts with the degradation mechanism, dominating the overall time behavior during the stress phase. It is only during the relaxation phase, where the degradation mechanism is more or less absent, that the relaxation mechanism can be studied in full detail, despite the difficulties arising during measurements. We therefore stipulate that a more complete NBTI model needs to focus on the relaxation phase first before attempting to cover the stress phase as well.

It is also important to stress that the observed discrepancies in the available models with measurement data do not necessarily indicate that the physical processes involved in NBTI, predominantly hole trapping and interface state generation, have been wrongly identified. The main finding of this study is that NBTI relaxation, and consequently the stress phase, are strongly influenced by physical mechanisms that are not yet fully understood and require a refined set of models.

References

1. Miura, Y. and Matukura, Y., Investigation of silicon–silicon dioxide interface using MOS structure, *Jpn. J. Appl. Phys.*, 5, 180, 1966.
2. Schroder, D.K. and Babcock, J.A., Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing, *J. Appl. Phys.*, 94(1), 1–18, 2003.
3. Schroder, D.K., Negative bias temperature instability: What do we understand? *Microelectron. Reliab.*, 47(6), 841–852, 2007.
4. Krishnan, A.T. et al., Material dependence of hydrogen diffusion: Implications for NBTI degradation, in *Proc. IEDM*, Washington, Washington D.C., 2005, pp. 688–691.
5. Alam, M.A. et al., A comprehensive model for PMOS NBTI degradation: Recent progress, *Microelectron. Reliab.*, 47(6), 853–862, 2007.
6. Shen, C. et al., Characterization and physical origin of fast V_{th} transient in NBTI of p-MOSFETs with SiON dielectric, in *Proc. IEDM*, San Francisco, CA, 2006, pp. 333–336.
7. Grassler, T. and Kaczer, B., Negative bias temperature instability: Recoverable versus permanent degradation, in *Proc. ESSDERC*, Munich, Germany, 2007, pp. 127–130.

8. Huard, V., Denais, M., and Parthasarathy, C., NBTI degradation: From physical mechanisms to modelling, *Microelectron. Reliab.*, 46(1), 1–23, 2006.
9. Reisinger, H. et al., Analysis of NBTI degradation- and recovery-behavior based on ultra fast V_{th} -measurements, in *Proc. IRPS*, San Jose, CA, 2006, pp. 448–453.
10. Alam, M.A. and Mahapatra S., A comprehensive model of PMOS NBTI degradation, *Microelectron. Reliab.*, 45(1), 71–81, 2005.
11. Huard, V. et al., A thorough investigation of MOSFETs NBTI degradation, *Microelectron. Reliab.*, 45(1), 83–98, 2005.
12. Houssa, M., Modelling negative bias temperature instabilities in advanced p-MOSFETs, *Microelectron. Reliab.*, 45(1), 3–12, 2005.
13. Zafar, S., Statistical mechanics based model for negative bias temperature instability induced degradation, *J. Appl. Phys.*, 97(10), 1–9, 2005.
14. Kaczer, B. et al., Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification, in *Proc. IRPS*, San Jose, CA, 2005, pp. 381–387.
15. Stathis, J.H. and Zafar, S., The negative bias temperature instability in MOS devices: A review, *Microelectron. Reliab.*, 46(2–4), 270–286, 2006.
16. Yang, T. et al., Fast DNBTI components in p-MOSFET with SiON dielectric, *IEEE Electron Dev. Lett.*, 26(11), 826–828, 2005.
17. Mahapatra, S. et al., On the physical mechanism of NBTI in silicon oxynitride p-MOSFETs: Can differences in insulator processing conditions resolve the interface trap generation versus hole trapping controversy? in *Proc. IRPS*, Phoenix, AZ, 2007, pp. 1–9.
18. Denais, M. et al., Interface trap generation and hole trapping under NBTI and PBTI in advanced CMOS technology with a 2-nm gate oxide, *IEEE Trans. Dev. Mater. Reliab.*, 4(4), 715–722, 2004.
19. Ershov, M. et al., Transient effects and characterization methodology of negative bias temperature instability in pMOS transistors, in *Proc. IRPS*, Dallas, TX, 2003, pp. 606–607.
20. Varghese, D. et al., On the dispersive versus Arrhenius temperature activation of NBTI time evolution in plasma nitrided gate oxides: Measurements, theory, and implications, in *Proc. IEDM*, Washington, D.C., Dec. 2005, pp. 1–4.
21. Denais, M. et al., On-the-fly characterization of NBTI in ultra-thin gate oxide p-MOSFETs, in *Proc. IEDM*, San Francisco, CA, 2004, pp. 109–112.
22. Li, M.F. et al., Dynamic bias-temperature instability in ultrathin SiO₂ and HfO₂ metal-oxide-semiconductor field effect transistors and its impact on device lifetime, *Jpn. J. Appl. Phys.*, 43(11B), 7807–7814, 2004.
23. Poindexter, E.H. et al., Electronic traps and P_b centers at the Si/SiO₂ interface: Band-gap energy distribution, *J. Appl. Phys.*, 56(10), 2844–2849, 1984.
24. Lenahan, P.M. and Conley, J.F. Jr., What can electron paramagnetic resonance tell us about the Si/SiO₂ system? *J. Vac. Sci. Technol. B*, 16(4), 2134–2153, 1998.
25. Campbell, J.P. et al., Direct observation of the structure of defect centers involved in the negative bias temperature instability, *Appl. Phys. Lett.*, 87(20), 1–3, 2005.
26. Campbell, J.P. et al., Observations of NBTI-induced atomic-scale defects, *IEEE Trans. Dev. Mater. Reliab.*, 6(2), 117–122, 2006.
27. Stesmans, A. and Afanas'ev, V.V., Electrical activity of interfacial paramagnetic defects in thermal (100) Si/SiO₂, *Phys. Rev. B*, 57(16), 10030–10034, 1998.
28. Campbell, J.P. et al., Density of states and structure of NBTI-induced defects in plasma-nitrided pMOSFETs, in *Proc. IRPS*, Phoenix, AZ, 2007, pp. 503–510.
29. Grasser, T. et al., TCAD modeling of negative bias temperature instability, in *Proc. SISPAD*, Monterey, USA, Sept. 2006, pp. 330–333.
30. Cartier, E. and Stathis, J.H., Hot-electron induced passivation of silicon dangling bonds at the Si (111)/SiO₂ interface, *Appl. Phys. Lett.*, 69(1), 103–105, 1996.
31. Ragnarsson, L.-A. and Lundgren, P., Electrical characterization of P_b centers in (100)Si/SiO₂ structures: The influence of surface potential on passivation during post metallization anneal, *J. Appl. Phys.*, 88(2), 938–942, 2000.

32. Grasser, T. et al., The universality of NBTI relaxation and its implications for modeling and characterization, in *Proc. IRPS*, Phoenix, AZ, 2007, pp. 268–280.
33. Haggag, A. et al., High-performance chip reliability from short-time-tests, in *Proc. IRPS*, Orlando, FL, 2001, pp. 271–279.
34. Stesmans, A., Passivation of P_{b0} and P_{b1} interface defects in thermal (100) Si/SiO₂ with molecular hydrogen, *Appl. Phys. Lett.*, 68(15), 2076–2078, 1996.
35. Stesmans, A., Dissociation kinetics of hydrogen-passivated P_b defects at the (111)Si/SiO₂ interface, *Phys. Rev. B*, 61(12), 8393–8403, 2000.
36. Neugroschel, A. et al., Direct-current measurements of oxide and interface traps on oxidized silicon, *IEEE Trans. Electron. Dev.*, 42(9), 1657–1662, 1995.
37. Groeseneken, G. et al., A reliable approach to charge-pumping measurements in MOS transistors, *IEEE Trans. Electron. Dev.*, 31(1), 42–53, 1984.
38. Zhang, J.F. et al., Hole traps in silicon dioxides—Part I: Properties, *IEEE Trans. Electron. Dev.*, 51(8), 1267–1273, 2004.
39. Parthasarathy, C.R. et al., New insights into recovery characteristics post NBTI stress, in *Proc. IRPS*, San Jose, CA, 2006, pp. 471–477.
40. Rangan, S., Mielke, N., and Yeh, E.C.C., Universal recovery behavior of negative bias temperature instability, in *Proc. IEDM*, Washington, Washington D.C., 2003, pp. 341–344.
41. Denais, M. et al., Paradigm shift for NBTI characterization in ultra-scaled CMOS technologies, in *Proc. IRPS*, San Jose, CA, 2006, pp. 735–736.
42. Reisinger, H. et al., A comparison of very fast to very slow components in degradation and recovery due to NBTI and bulk hole trapping to existing physical models, *IEEE Trans. Dev. Mater. Reliab.*, 7(1), 119–129, 2007.
43. Grasser, T. et al., Simultaneous extraction of recoverable and permanent components contributing to bias-temperature instability, in *Proc. IEDM*, 2007, pp. 801–804.
44. Kakalios, J., Street, R.A., and Jackson, W.B., Stretched-exponential relaxation arising from dispersive diffusion of hydrogen in amorphous silicon, *Phys. Rev. Lett.*, 59(9), 1037–1040, 1987.
45. Li, J.-S. et al., Effects of delay time and AC factors on negative bias temperature instability of PMOSFETs, in *IIRW Final Rep.*, Lake Tahoe, CA, 2006, pp. 16–19.
46. Krishnan, A.T. et al., Negative bias temperature instability mechanism: The role of molecular hydrogen, *Appl. Phys. Lett.*, 88(15), 1–3, 2006.
47. Ushio, J., Okuyama, Y., and Maruizumi, T., Electric-field dependence of negative-bias temperature instability, *J. Appl. Phys.*, 97(8), 1–3, 2005.
48. Tsetseris, L. et al., Physical mechanisms of negative-bias temperature instability, *Appl. Phys. Lett.*, 86(14), 1–3, 2005.
49. Van De Walle, C.G., Stretched-exponential relaxation modeled without invoking statistical distributions, *Phys. Rev. B*, 53(17), 11292–11295, 1996.
50. Nickel, N.H., Yin, A., and Fonash, S.J., Influence of hydrogen and oxygen plasma treatments on grain-boundary defects in polycrystalline silicon, *Appl. Phys. Lett.*, 65(24), 3099–3101, 1994.
51. Houssa, M., Reaction-dispersive proton transport model for negative bias temperature instabilities, *Appl. Phys. Lett.*, 86(9), 1–3, 2005.
52. Ogawa, S., Shimaya, M., and Shiono, N., Interface-trap generation at ultrathin SiO₂ (4 nm–6 nm)–Si interfaces during negative-bias temperature aging, *J. Appl. Phys.*, 77(3), 1137–1148, 1995.
53. Ogawa, S. and Shiono, N., Generalized diffusion-reaction model for the low-field charge build up instability at the Si/SiO₂ interface, *Phys. Rev. B*, 51(7), 4218–4230, 1995.
54. Islam, A.E. et al., Temperature dependence of the negative bias temperature instability in the framework of dispersive transport, *Appl. Phys. Lett.*, 90(1), 083505-1–083505-3, 2007.
55. Kufluoglu, H. and Alam, M.A., Theory of interface-trap-induced NBTI degradation for reduced cross section MOSFETs, *IEEE Trans. Electron Dev.*, 53(5), 1120–1130, 2006.
56. Alam, M.A., A critical examination of the mechanics of dynamic NBTI for p-MOSFETs, in *Proc. IEDM*, Washington, Washington D.C., 2003, pp. 345–348.
57. Yang, J.B. et al., Analytical reaction-diffusion model and the modeling of nitrogen-enhanced negative bias temperature instability, *Appl. Phys. Lett.*, 88(17), 1–3, 2006.

58. Alam, M.A. and Kufluoglu, H., On quasi-saturation of negative bias temperature degradation, *ECS Trans.*, 1(1), 139–145, 2005.
59. Grasser, T., Gös, W., and Kaczer, B., Dispersive transport and negative bias temperature instability: Boundary conditions, initial conditions, and transport models, *IEEE Trans. Dev. Mater. Reliab.*, 8(1), 79–97, 2008.
60. Alam, M.A., NBTI: A simple view of a complex phenomena, in *Proc. IRPS*, San Jose, CA, 2006, (Tutorial).
61. Zafar, S. et al., A comparative study of NBTI and PBTI (charge trapping) in SiO₂/HfO₂ stacks with FUSI, TiN, Re gates), in *Proc. VLSI Symp.*, Honolulu, HI, 2006, pp. 23–25.
62. Chakravarthi, S. et al., Probing negative bias temperature instability using a continuum numerical framework: Physics to real world operation, *Microelectron. Reliab.*, 47(6), 863–872, 2007.
63. Stathis, J.H., Dissociation kinetics of hydrogen-passivated (100) Si/SiO₂ interface defects, *J. Appl. Phys.*, 77(12), 6205–6207, 1995.
64. Houssa, M. et al., Modeling negative bias temperature instabilities in hole channel metal-oxide-semiconductor field effect transistors with ultrathin gate oxide layers, *J. Appl. Phys.*, 95(5), 2786–2791, 2004.
65. Grasser, T., Gös, W., and Kaczer, B., Modeling of dispersive transport in the context of negative bias temperature instability, in *IIRW Final Rep.*, Lake Tahoe, CA, 2006, pp. 5–10.
66. Noolandi, J., Multiple-trapping model of anomalous transit-time dispersion in *a*-Se, *Phys. Rev. B*, 16(10), 4466–4473, 1977.
67. Orenstein, J., Kastner, M.A., and Vaninov, V., Transient photoconductivity and photo-induced optical absorption in amorphous semiconductors., *Philos. Mag. B*, 46(1), 23–62, 1982.
68. Arkhipov, V.I. and Rudenko, A.I., Drift and diffusion in materials with traps, *Philos. Mag. B*, 45(2), 189–207, 1982.
69. Rudenko, A.I. and Arkhipov, V.I., Drift and diffusion in materials with traps: Quasi-equilibrium transport regime, *Philos. Mag. B*, 45(2), 177–187, 1982.
70. Arkhipov, V.I. and Bäessler, H., An adiabatic model of dispersive hopping transport, *Philos. Mag. B*, 68(4), 425–435, 1993.
71. Kaczer, B. et al., Temperature dependence of the negative bias temperature instability in the framework of dispersive transport, *Appl. Phys. Lett.*, 86(14), 1–3, 2005.
72. Chakravarthi, S. et al., A comprehensive framework for predictive modeling of negative bias temperature instability, in *Proc. IRPS*, Phoenix, AZ, 2004, pp. 273–282.
73. Huard, V., Private communications.
74. Tewksbury, T.L. and Lee, H.-S., Characterization, modeling, and minimization of transient threshold voltage shifts in MOSFETs, *IEEE J. Solid State Circ.*, 29(3), 239–252, 1994.
75. Huard, V. et al., Physical modeling of negative bias temperature instabilities for predictive extrapolation, in *Proc. IRPS*, San Jose, CA, 2006, pp. 733–734.