A Two-Stage Model for Negative Bias Temperature Instability

T. Grasser*, B. Kaczer°, W. Goes*, Th. Aichinger†, Ph. Hehenberger◇, and M. Nelhiebel‡
* Christian Doppler Laboratory for TCAD at the ◇Institute for Microelectronics, TU Wien, Austria
° IMEC, Leuven, Belgium † KAI,Villach, Austria ‡ Infineon Technologies, Villach, Austria

***Abstract*— Based on the established properties of the most commonly observed defect in amorphous oxides, the E' center, we suggest a coupled two-stage model to explain the negative bias temperature instability. We show that a full model that includes the creation of E' centers from their neutral oxygen vacancy precursors and their ability to be repeatedly charged and discharged prior to total annealing is required to describe the first stage of degradation. In the second stage a positively charged E' center can trigger the depassivation of P_b centers at the Si/SiO_2 interface or K_N centers in oxynitrides to create an unpassivated silicon dangling bond. We evaluate the new model to experimental data obtained from three vastly different technologies (thick SiO_2, SiON, and HK) and obtain very promising results.**

I. Introduction

Many recent publications dealing with the negative bias temperature instability (NBTI) have suggested the existence of a recoverable component observed on top of a slowly recovering or even permanent component [1–4]. Often, the *recoverable component is attributed to hole trapping* while the *permanent component is explained by the creation of interface states* [1]. We have recently pointed out a *serious problem* with the interpretation that *two independent components* result in the overall degradation observed during NBT stress [4, 5]. This is because these two components should have a different voltage and temperature acceleration, allowing for their separation by the application of a suitably chosen combination of stress temperatures and voltages. Quite to the contrary, however, we have observed [4–7] that NBTI data at various stress and relaxation times, broad ranges of stress voltages and temperatures, and a broad range of technologies (ultra-thin SiON, high-k, ultra-thick SiO_2) can often be made to overlap via multiplication by a suitably chosen scaling factor. This *very broad scalability cannot be explained by any established model* [4] but implies that *NBTI is either due to a single mechanism* (which then would have to be able to explain both the recoverable and the permanent contributions) or due to *two tightly coupled mechanisms*. We suggest a model that captures this behavior during both stress and recovery and explain why previous modeling attempts fail in doing so.

II. Experimental Observations

We characterized SiON pMOSFETs (EOT = 1.4 nm) using the extended MSM scheme (eMSM) [3, 8], which acquires data in alternating on-the-fly (OTF) and relaxation sequences using a wide range of stress voltages ($-0.6\,\mathrm{V}$ … $-2.0\,\mathrm{V}$) and stress temperatures ($25\,^\circ\mathrm{C}\ldots200\,^\circ\mathrm{C}$). The recorded OTF degradation in I_{Dlin} was converted to $\Delta V_{\mathrm{th}}^{\mathrm{OTF}}$ using the simple expression $\Delta V_{\mathrm{th}}^{\mathrm{OTF}} \approx (V_{\mathrm{G}} - V_{\mathrm{th0}})(I_{\mathrm{D}} - I_{\mathrm{D0}})/I_{\mathrm{D0}}$ [9, 10]. It is now understood that $\Delta V_{\mathrm{th}}^{\mathrm{OTF}}$ is contaminated by mobility variations and the error in the initial drain current already determined at stress level, $I_{\mathrm{D0}} = I_{\mathrm{D}}(t_0)$ [11–13]. Correction schemes for these errors have been suggested [13, 14], but are still open to rigorous justification. During model development, the impact of I_{D0} can be easily accounted for by subtracting the simulated $\Delta V_{\mathrm{th}}(t_0)$ from the overall simulation result. The impact of the mobility variation, however, remains unclear at the moment.

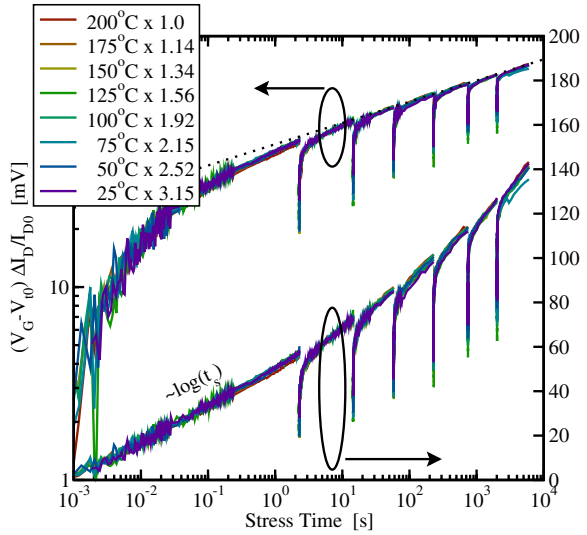


Fig. 1: Degradation of the drain current collected during consecutive stress sequences. The stress temperature was varied at a fixed stress voltage of $-2\,\mathrm{V}$. The unscaled data initially follows a logarithmic time dependence while the long-time data may be approximated by a power-law with the exponent $n = 0.11$. The slope of the initial log shows clear temperature activation while the power-law exponent is roughly temperature-independent. Multiplication of each data set with a constant value results in a *nearly perfect overlap*. The scaling factors are independent of stress time, indicating that the initial log and the long-time power-law behavior are due to a related process.

Nonetheless, we will take the uncorrected $\Delta V_{\mathrm{th}}^{\mathrm{OTF}}$ as an indicator for the overall degradation and assume that $\Delta V_{\mathrm{th}}^{\mathrm{OTF}}$ contains the correct information regarding the time dynamics but may potentially be affected by an unknown error in amplitude [13].

As has been observed previously [1, 4, 6, 15, 16], for short stress times ($t_s \lesssim 1\,\mathrm{s}$, depending on the stress condition) the initial degradation phase is well described by a logarithmic time-dependence (see Fig. 1 for a typical example),

$$\Delta V_{\mathrm{th}}^{\mathrm{OTF}}(t_s) = B_s(F,T)\log_{10}(t_s/t_0), \tag{1}$$

with t_0 being the delay of the first measurement point. The pre-factor during stress, B_s, gives the increase in $\Delta V_{\mathrm{th}}^{\mathrm{OTF}}$ in volts per decade in time and depends on the stress field F and temperature T.

The recovery phase, on the other hand, may be fit by [1, 17–19]

$$\Delta V_{\mathrm{th}}(t_s,t_r) = B_r(F,T)\log_{10}(1+t_s/t_r) + P(t_s), \tag{2}$$

where P is a roughly permanent contribution depending on the stress time only and is possibly due to interface states [17]. Again, the pre-factor B_r gives the recovery rate in volts per decade in time.

Before developing our model, we briefly summarize some key experimental observations obtained by the eMSM technique, which provide important insights regarding the dynamics of NBTI. In particular, they can be used to effectively rule out a number of alternative models as will be shown later.

978-1-4244-2889-2/09/$25.00 ©2009 IEEE

IEEE CFP09RPS-CDR 47th Annual International Reliability Physics Symposium, Montreal, 2009

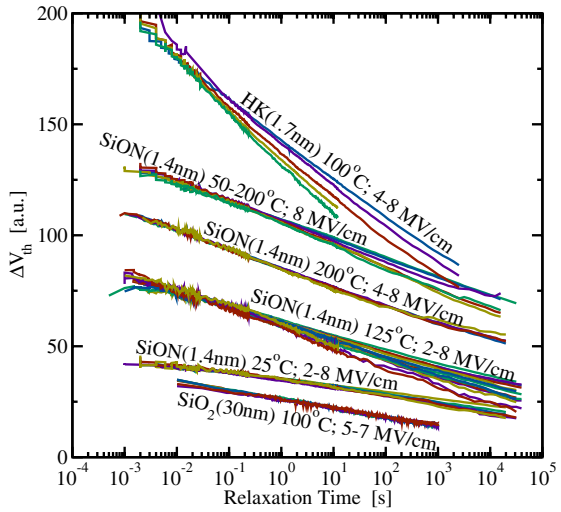


Fig. 2: Recovery of the threshold voltage shift at different temperatures, voltages, and stress times, for three completely different gate stacks, ultra-thick SiO_2, ultra-thin SiON, and HK. A similar scalability as during stress is observed during recovery for all technologies. All data sets show good temperature and voltage scalability and thus do not support the idea that NBTI is a consequence of two independent mechanisms.

A. Scalability

The data recorded using the eMSM scheme can often be made to overlap by multiplication with a suitably chosen scaling factor. While the initial behavior up to about 1 s is well approximated by a logarithmic time dependence and the long-term data approximately follows a power-law, *the same scaling factor can be used in both regimes*, see Fig. 1. A similar scalability is observed during recovery, see Fig. 2 which, to first-order, follows a logarithmic time dependence. We remark that there is an excellent correlation between the scaling factors required during stress and recovery [6].

B. Bias and Temperature Dependence of Stress

Following the previously suggested idea that the initial degradation is dominated by hole trapping while the long-term degradation is due to interface state creation, we focus on the initial degradation phase to obtain some more information on this alleged 'hole trapping' component. We thus subjected devices to short stresses of 2 s and let them recover for about 3000 s, see Fig. 3. For what is typically considered NBTI stress in this oxide thickness range ($V_{\text{stress}} \leq -1\,\text{V}$), the pre-factor B_{s} can be roughly approximated as $B_{\text{s}}(F,T) \approx B_{\text{s},0}\, T^{\Theta} F^2$, with $\Theta \approx 2$. We remark that the power-law temperature dependence is also well approximated by an Arrhenius law with $E_{\text{A}} \sim 70\,\text{meV}$.

C. Asymmetry Between Stress and Relaxation

It has been long understood that recovery takes substantially longer than the time used to build up the degradation. This is shown in Fig. 4 for a typical stress temperature of 125 °C. Using the extracted pre-factors B_{s} (fit to a log in the range 1 ms…1 s) and B_{r} (fit to a log in the range 1 ms…100 ms) we observe that the ratio $B_{\text{s}}/B_{\text{r}}$ is about 2.5, independent of temperature and voltage. For example, if one observes during degradation a rate of 10 mV/dec, recovery will proceed with only about 4 mV/dec. The exact value of $B_{\text{s}}/B_{\text{r}}$ depends on the mobility error in B_{s} but this has no bearing on the fact that stress and recovery are asymmetric. Although this asymmetry may look quite innocent at a first glance, it turns out to be a considerable challenge for any modeling attempt.

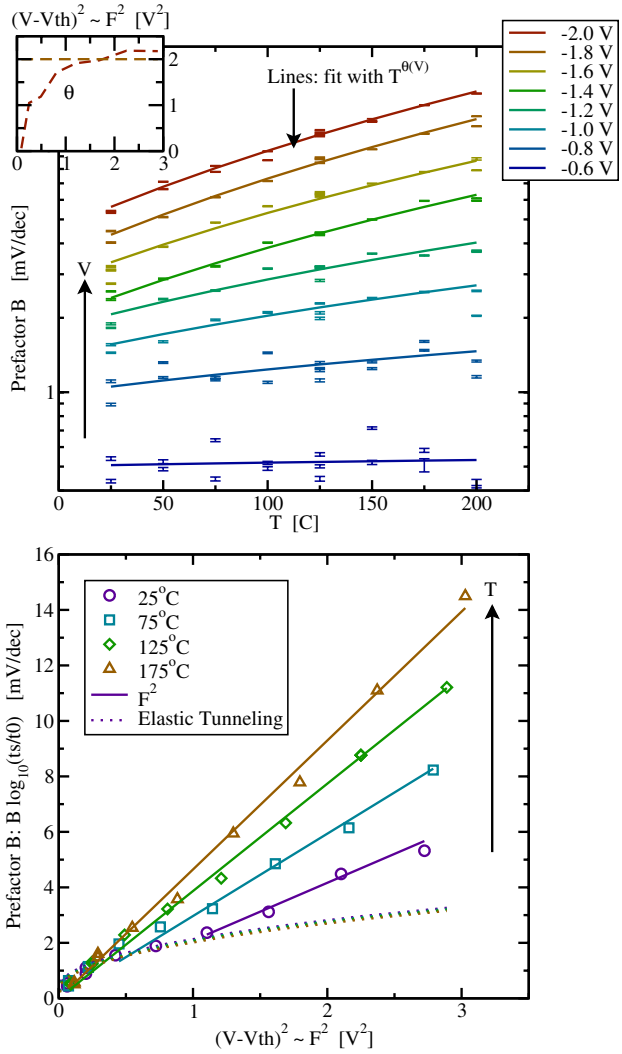


Fig. 3: PMOSFETs with EOT = 1.4 nm are subjected to various stress voltages and temperatures for 1 s. The degradation follows $B\log(t_{\text{s}}/t_0)$, with $t_0 = 1\,\text{ms}$ being the first OTF measurement point. Contrary to the predication of elastic tunneling theory, the prefactor B strongly depends on temperature (T^2) once a critical oxide field is reached (**top**). Also, the field dependence can be well approximated by F^2, with $F \sim V - V_{\text{th}}$ (**bottom**). We conclude from the fact that the data are scalable in the whole bias and voltage regime even for larger stress times [6] where defect creation becomes important ($t_{\text{s}} = 10^5\,\text{s}$), that we are dealing with a single, two-stage process. We rule out elastic tunneling due to its lacking temperature dependence.

D. Bias Dependence of Recovery

Recovery has been shown to depend on the bias voltage applied during recovery [1, 5, 8, 19]. In particular, application of positive bias accelerates recovery. This has been explained by the bias-dependence of H^+ drift [8] or the bias-dependence of hole-detrapping via valence band and interface states [1]. However, just as the asymmetry, this is also challenging to reproduce correctly by a quantitative model.

III. Previous Modeling Attempts

Although it has been widely acknowledged that both interface states and trapped holes can contribute to NBTI, the exact details on how this should occur are highly controversial. The vastly different microscopic explanations that have been suggested will be quickly summarized in the following.

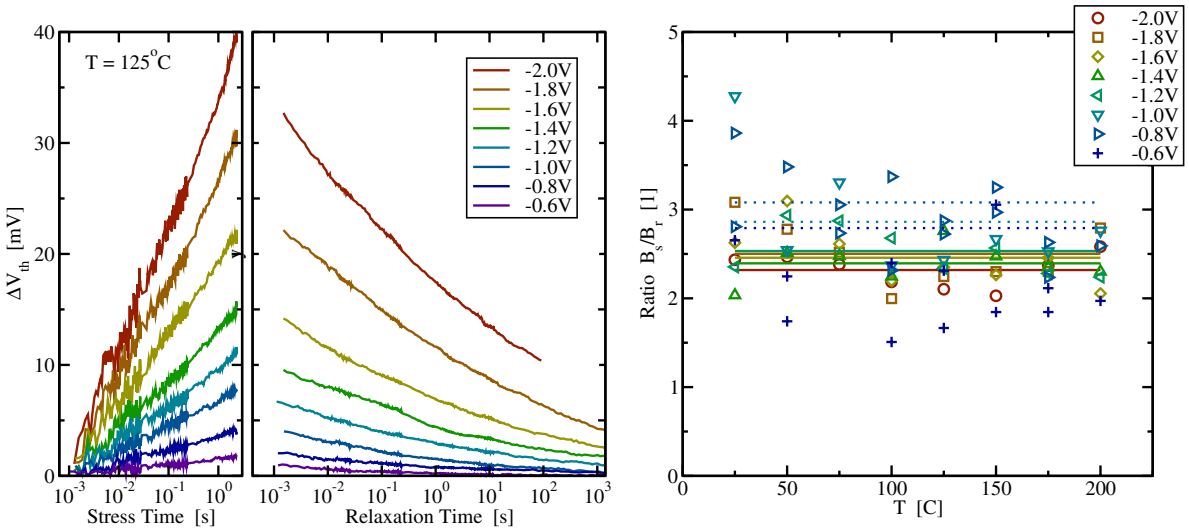


Fig. 4: Left: Asymmetry of stress and recovery measured at 125 °C for 8 different stress voltages. The recovery is considerably slower than the degradation. **Right**: The complete data set using 8 different stress temperatures and voltages. When the data is fitted to $B_s \log(t_s)$ and $B_r \log(t_r)$, the ratio B_s/B_r is roughly independent of the stress voltage and temperature and about 2.5, particularly for data recorded with higher stress conditions which decreases the noise. The dotted lines are averages over all temperatures for each stress voltage.

A. *Interface States Only*

In the simplest case NBTI would be due to a single mechanism only. One prominent example is the reaction-diffusion (RD) theory which claims that only interface states are responsible. However, it has been clearly shown that the RD theory and the various extensions proposed over the years can only explain constant bias stresses while they fail to explain the dynamics (recovery, bias dependence of recovery, duty-factor dependence, etc.) [2, 4, 15]. Furthermore, published data on the recovery of charge-pumping (CP) signals seem to indicate that interface states show only marginal recovery compared to the recovery of ΔV_{th} [1, 20].

Data acquired by the recently suggested on-the-fly charge-pumping technique (OFIT), which tries to minimize the measurement-induced recovery by using the CP base-level as a stress voltage [21], suggested that also interface states can show fast recovery. Consequently, a single mechanism model based on interface states only appeared feasible and we have recently suggested such a model [5]. The model is derived using a suitable generalization of dispersive bond breaking already used previously for the creation of interface states [1, 22]. Although the resulting triple-well model can reproduce complicated stress/relaxation sequences with very good accuracy, it requires a large and (as it now seems) nonphysical variance in the hydrogen binding energies in order to reproduce data recorded in a larger temperature and voltage range such as used in this study. Furthermore, our theoretical and experimental study of the OFIT technique suggests that fast recovery of interface states may be an artifact of the method and requires correction [23]. We conclude from this that the microscopic assumptions underlying the triple-well model are likely not correct. Nonetheless, the mathematical structure of the model can be retained in the following.

B. *Hole Trapping Models*

Hole trapping is often modeled using elastic tunneling into pre-existing traps located at various distances away from the interface. Depending on the distance, an exponentially increasing time constant is obtained, which, at least in thicker oxides, could explain the large spread of time constants observed in NBTI. However, elastic hole trapping is to first-order temperature independent and linearly dependent on the stress field [24]. Furthermore, the model predicts $B_s \sim B_r$, that is, symmetric degradation and recovery behavior. As such, this is incompatible with our data.

C. *Combination of Hole Trapping and Interface State Creation*

Two variants of combined hole trapping and interface state creation are commonly used: *(i)* initial hole trapping which quickly saturates (within about 1 s) and long term degradation dominated by interface state creation according to the RD theory [25]. We rule out this variant based on the various known shortcomings of RD theory [2, 4, 15]. *(ii)* Trapped holes are responsible for the recoverable part of the degradation while interface states form a permanent contribution [1]. We also rule out this variant due to the shortcomings of elastic tunneling and the lack of scalability [4]. *(iii)* Alternatively, it has been speculated that holes can be trapped into newly created defects [26, 27]. Although the details behind the responsible process have only been schematically outlined, they appear broadly consistent with our own observations and the detailed physical model suggested in the following.

IV. Hole Trapping Process

Since the corrected OFIT data do not indicate fast recovery of interface states [23] and elastic hole tunneling also cannot explain the data, we have to look for an alternative explanation. Of particular interest are hole trapping models that have been applied since the 1970s in attempts to understand $1/f$-noise and thermally stimulated currents at semiconductor surfaces [28, 29]. Just like the triple-well model, these models are also based on a dispersion of activation energies but require very large variances in order to reproduce for instance the $1/f$ behavior. In these models it is assumed that holes can be captured via a (*thermally activated*) multiphonon emission (MPE) process into deep near-interfacial states/border traps [24, 29–31], for instance into oxygen vacancies (E' centers) [29]. The MPE process differs from the conventionally invoked elastic tunneling process, notably due to its temperature activation and the larger time constants resulting therefrom [30].

Unfortunately, a trapping model based on the MPE process would still have a linear field dependence in contradiction to our data.

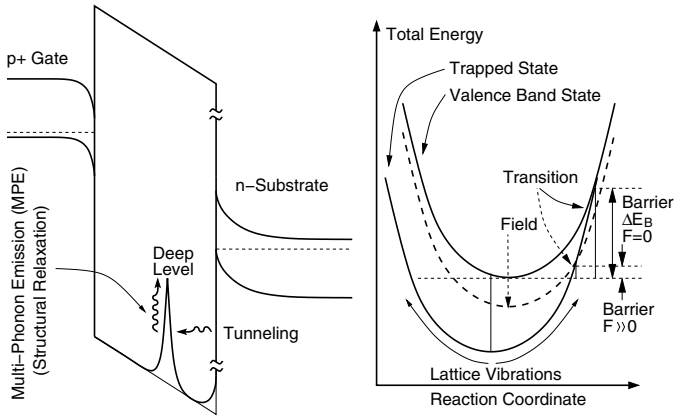


Fig. 5: The multiphonon-field-assisted tunneling (MPFAT) process used to explain the experimental data: elastic tunneling into deep states is only allowed when the excess energy of holes can be released via a multiphonon emission process during structural relaxation. The probability for a thermionic transition over the barrier ΔE_B has been estimated as $\exp(-\beta\Delta E_B)$ using 1D reaction-coordinate calculations [30, 33], with $\beta = 1/k_B T$. Application of an electric field shifts the total energy of the valence band state (dashed line), increasing the transition probability by $\exp(F^2/F_c^2)$ [32, 33].

In order to resolve this issue, we have to recall that the MPE mechanism is derived under the assumption of negligible electric fields. This assumption is definitely violated in the case of NBTI. An extension of MPE to the large electric field case has already been developed for the emission of particles from deep traps and has become known as *multiphonon-field-assisted tunneling (MPFAT)* [32, 33]. The signatures of this mechanism are its $\exp(F^2/F_c^2)$ field dependence (note that only the logarithm of the enhancement factor enters B) and a considerable temperature activation (from the MPE process). The MPFAT process is schematically illustrated in Fig. 5: a hole can either be in the valence band or in a trapped state. These two states are represented by the two solid parabola which give the total energy of the system. The vibrational modes can be approximated using a simple oscillator model and at the intersection point a transition can occur. The intersection point determining the barrier ΔE_B is dramatically lowered by the application of an electric field, resulting in an enhancement of $\exp(F^2/F_c^2)$.

V. Properties of the E' Center

In order to develop an accurate microscopic model for hole trapping in the context of NBTI, we summarize the most important features collected in a long line of studies on oxide defects providing a solid basis for our NBTI model.

- The most likely microscopic candidate for the 'trapped hole' are defects from the E' center family, most notably the E'_γ center [34, 35]. An E'_γ center is thought to be created when a hole is trapped in the precursor structure, which is commonly assumed to be a neutral oxygen vacancy (a Si–Si dimer inside the oxide).
- The energy level for hole trapping is roughly located at about 1 eV below the Si valence band [34], see Fig. 6.
- Once the silicon bond is broken, the distance between the two silicon atoms increases into a new equilibrium position, which requires a large-range structural relaxation of the surrounding lattice (10 Å [36]), and an E'_γ center is obtained. The E'_γ is visible in ESR (electron spin resonance) when positively charged, that is, right after hole capture.
- An important peculiarity about the E'_γ center is that it can be repeatedly charged and discharged. The corresponding energy levels lie within the silicon bandgap [34]. The idea behind this cyclability is that once the hole is emitted (that is, an electron is captured), the bond between the two silicon atoms does not fully reform but the defect remains in a dipole state which can easily lose an electron again. The fact that the E' center can act as a switching trap has been suggested by Lelis *et al.* [37] based on electrical measurements who then suggested the Harry-Diamond-Laboratories (HDL) model. This was later confirmed by ESR studies [38] and theoretical calculations [39], see Fig. 7 for a schematic representation. In [40] it has been suggested that in order to create a stable E' configuration from an oxygen vacancy, the doubly positive configuration could be important.
- Only after having been in its electrically neutral state for a while, the structure relaxes again to the initial dimer configuration and the defect is completely healed.
- The E'_γ is often considered a donor-like defect [34], that is, it is either neutral (ESR inactive) or positively charged (ESR active).
- The oxygen vacancy can also act as an electron trap, with a trap level close to the silicon conduction band [39, 40].
- Due to the amorphous nature of the interfacial layer, a considerable broadening of the energy-levels is to be expected. For instance, theoretical calculations give a spread of about 1.5 eV for the precursor level and a Si–Si bond-length variation from 2.3 Å to 2.7 Å [36].

The above summary bears some important and interesting consequences regarding our understanding of NBTI:

- So far, hole trapping has been mainly considered as being into pre-existing traps which rapidly fill but are not related to the actual NBTI mechanism, a somewhat parasitic component which has to be removed to get to the heart of NBTI [10].
- Using CV measurements, NBTI has been shown to be due to donor-like defects [10, 41, 42]. Models based on the RD theory claim that NBTI is dominantly due to the creation of interface states, which are P_b centers. However, P_b centers are amphotheric, that is, donor-like only in the lower half of the silicon band-gap but acceptor-like in the upper half [43]. The E'_γ center, on the other-hand, is a donor-like defect [34].
- NBTI recovery is strongly bias-dependent, in particular when the gate voltage is moved from inversion into accumulation [8]. This is intuitively consistent with carrier trapping [27].
- Describing hole trapping via the known properties of the E' center *promotes the positive oxide charge component* from a purely parasitic component to the *central contributor* to NBTI. Indeed, as will be shown in the following, some key experimental features of NBTI which are incompatible with the P_b center and simple hole trapping models follow directly from the properties of the E' center.
- The amount of positive charge visible and thus contributing to ΔV_{th} will depend on the position of the Fermi-level, that is, the gate voltage at which the degradation is monitored. For example, during OTF experiments, the Fermi-level is below the valence band edge and most defects will be positively charged (state 2 in Fig. 7) and thus visible. During recovery, the Fermi-level is moved towards mid-gap and a smaller fraction of the defects will be positive and visible. This occupancy effect will manifest itself as a change in the subthreshold slope, and is consistent with recent experimental results [16, 27, 44, 45].
- Since full annealing of oxide defects is only possible when the defect is neutral (state 3), defects having an electrically higher trapping level will show a slower recovery rate. Indeed, as will be shown, this is fully compatible with what is observed during

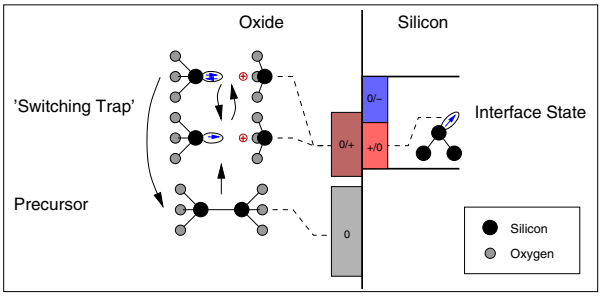


Fig. 6: Electronic energy-levels required in the HDL model and in our two-stage model: the neutral precursors lie about 1 eV below the silicon valence band edge. The E'_γ levels are assumed to be inside the silicon bandgap, while for simplicity both the charged and the uncharged level are assumed to be roughly identical. Interface states are assumed to introduce amphotheric defects into the silicon band-gap. All electronic energy-levels are assumed to be homogeneously distributed due to the amorphous nature of the interfacial layer.

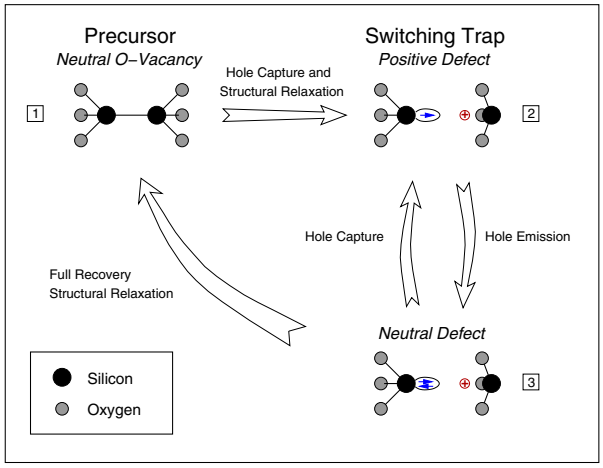


Fig. 7: The HDL model for a switching oxide trap. Initially, a neutral precursor exists (state 1). Upon hole capture, the Si–Si bond breaks and a positively charged E'_γ center is created (state 2). Hole emission (electron capture) neutralizes the E'_γ center (state 3). Being in state 3, two options exist: a hole can be captured again, causing a transition to state 2, or the structure can relax back to its equilibrium configuration (state 1).

NBTI recovery and explains the often observed bias dependence. Furthermore, this can explain the asymmetry between stress and recovery, with the recovery lasting considerably longer than the time required to build up the degradation.

- The fact that the precursor level is below the valence band and the defect level within the silicon bandgap is precisely what is expected of a defect responsible for NBTI. A higher energy level of the precursor, e.g. above the valence band, would cause most precursors to be already initially broken in PMOSFETs. The energy level of the created defect inside the silicon bandgap results in most defects to be positively charged (a donor-like defect) during both stress and recovery. A lower defect level would render the defects electrically neutral and thus not contributing to the threshold voltage shift.
- After NBTI stress an increase in $1/f$ noise has been reported [46, 47]. The prime suspect for $1/f$ noise are the E' centers [29, 39], while P_b centers do not create a suitable $1/f$ spectrum.

Although an extensive amount of literature is available on the E' center and its qualitative behavior is well understood, no rate-equation based model that spans the full cycle of trap creation, recharging and discharging until final annealing seems to be available. As we will demonstrate in the following, however, for NBTI a full model is required in order to account for the asymmetry between stress and relaxation and the correct bias dependence of recovery.

VI. A Two Stage Model

Based on the above observations we formulate a *new model for NBTI*, where defect creation proceeds via a *two stage process*: In stage one, upon application of stress, holes can be trapped into near-interfacial oxygen vacancies via the MPE/MPFAT mechanism. In the second stage, the increased hole concentration considerably enhances the creation of poorly recoverable defects, e.g. P_b-centers in SiO_2 layers and K_N-centers in oxynitrides [48]. We remark that such a two-stage process is the standard assumption for defect creation following irradiation [49]. The total threshold voltage shift is thus given by

$$\Delta V_{\rm th}(t) = -\frac{\Delta Q_{\rm ox}(t) + \Delta Q_{\rm it}(t)}{C_{\rm ox}} \ . \tag{3}$$

In the following, detailed microscopic models for $\Delta Q_{\rm ox}$ and $\Delta Q_{\rm it}$ are developed.

A. Complete E' Switching Trap Model

Our full model for the E' switching trap relies heavily on the HDL switching trap model [37]. We start our derivation with a generalization of the lattice-relaxation multiphonon emission (MPE) theory for phonon-assisted capture of holes and electrons [24, 30, 50] by including a depth dependence and MPFAT field acceleration [32, 33]. We remark that although the final equations are formally equivalent to the familiar Shockley-Read-Hall equations, the underlying microscopic assumptions differ [24].

Consider a trap level $E_{\rm T}$ located at a distance x away from the interface. The hole capture and emission rates are then approximately given by

$$k_p^{\rm c} = p\, v_p^{\rm th}\, \sigma_p\, {\rm e}^{-x/x_{p,0}}\, {\rm e}^{-\beta\Delta E}\, \theta(E_{\rm VT},\ {\rm e}^{-\beta E_{\rm VT}},\ 1)\, {\rm e}^{F^2/F_{\rm c}^2}, \tag{4}$$

$$k_p^{\rm e} = p\, v_p^{\rm th}\, \sigma_p\, {\rm e}^{-x/x_{p,0}}\, {\rm e}^{-\beta\Delta E}\, \theta(E_{\rm VT},\ {\rm e}^{-\beta E_{\rm VF}},\ {\rm e}^{-\beta E_{\rm TF}}), \tag{5}$$

while the corresponding rates for electrons read

$$k_n^{\rm c} = n\, v_n^{\rm th}\, \sigma_n\, {\rm e}^{-x/x_{n,0}}\, {\rm e}^{-\beta\Delta E}\, \theta(E_{\rm TC},\ {\rm e}^{-\beta E_{\rm TC}},\ 1), \tag{6}$$

$$k_n^{\rm e} = n\, v_n^{\rm th}\, \sigma_n\, {\rm e}^{-x/x_{n,0}}\, {\rm e}^{-\beta\Delta E}\, \theta(E_{\rm TC},\ {\rm e}^{-\beta E_{\rm FC}},\ {\rm e}^{\beta E_{\rm TF}}). \tag{7}$$

Here, p and n are the hole and electron concentrations in the channel, $v_p^{\rm th}$ and $v_n^{\rm th}$ their thermal velocities ($\sqrt{8{\rm k_B}T/(\pi m)}$), σ_p and σ_n their capture cross sections ($\sim 3\times10^{14}\,{\rm cm}^2$ [51]), $E_{\rm F}$ the Fermi-level in the channel, $E_{\rm V}$ and $E_{\rm C}$ the valence and conduction bands directly at the interface (classical approximation), ΔE the MPE barrier, and $\beta = 1/{\rm k_B}T$. $F_{\rm c}$ is the reference field for the multiphonon-field-assisted tunneling mechanism which is, due to lack of decisive data, only introduced for hole capture. According to a simplified WKB approximation for large tunneling barriers ϕ [24], $x_0 = \hbar/(2\sqrt{2m\phi})$, $x_{n,0} = 0.8\,$Å for electrons ($\phi_{\rm C} \approx 3.2\,$eV and $m \approx 0.5 m_0$) and $x_{p,0} = 0.5\,$Å for holes ($\phi_{\rm V} \approx 4.65\,$eV and $m \approx 0.8 m_0$). Furthermore, we use the shorthand $E_{\rm AB} = E_{\rm A} - E_{\rm B}$ and the auxiliary function

$$\theta(E_{\rm switch},\ a,\ b) = \begin{cases} a & E_{\rm switch} \geq 0 \\ b & E_{\rm switch} < 0 \end{cases} \tag{8}$$

to account for the fact that thermal activation is required for hole capture into a trap below $E_{\rm V}$ while capture in a trap above $E_{\rm V}$ proceeds without activation (the hole 'bubbles up'), and vice-versa for electrons.

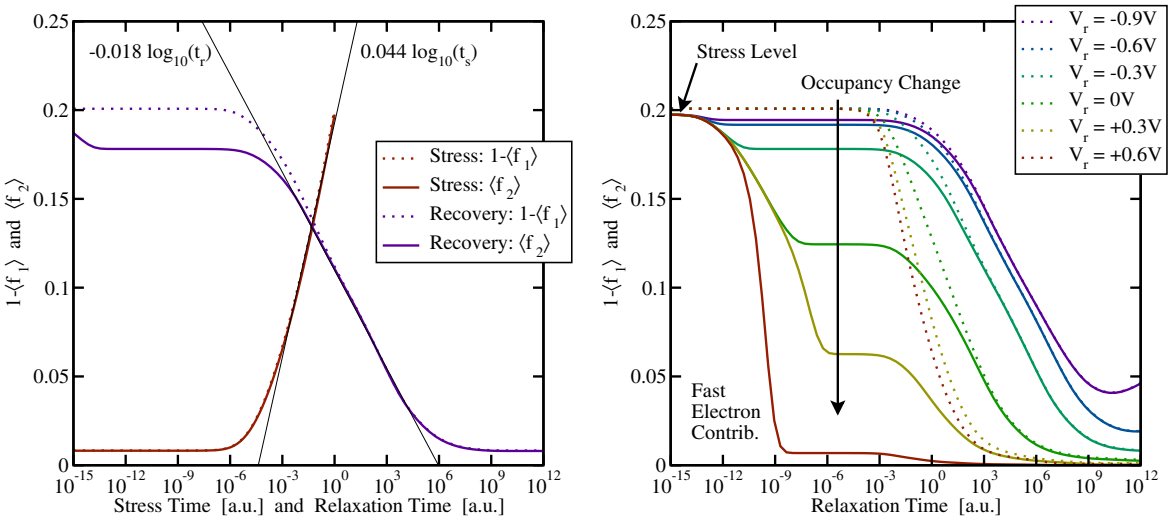


Fig. 8: *Qualitative degradation and recovery behavior* predicted by our E' center model. Slightly artificial parameter values with a narrow distribution for ΔE_B where used to distinctly contrast the change in occupancy (transition from state 2 to state 3) versus true annealing (transition from state 3 to state 1). **Left**: The model can predict the asymmetric behavior during stress and recovery. In the above example, $B_\mathrm{s}/B_\mathrm{r} \approx 2.5$ is obtained, just like in the real data of Fig. 4. **Right**: The model can predict the strong bias sensitivity during recovery. Dotted lines give the number of oxide defects (state 2 and 3), while the solid line corresponds to the positively charged center (state 2 only).

Based on the above rates, the rate equations describing the HDL switching trap are straight-forward to set up. Regarding the rates we take the simplest possible case that can capture the currently available data. We assume that when the defect is in state 1 the trap energy lies at E_T, has a MPE barrier of ΔE_B, and a MPFAT reference field F_c. When in state 2 and 3, the defect level is assumed to be at E'_T with a small charging/discharging MPE barrier ΔE_C. Although the barrier ΔE_C is expected to be considerably lower than the barrier ΔE_B, it is responsible for the E' centers to act as 'slow states' in CP measurements. Nonetheless, for the data investigated here, ΔE_C can be neglected. We also neglect the MPFAT mechanism for charging/discharging since our data taken to sense these characteristics are recorded at relatively low fields. Transition from state 3 to state 1 (full annealing rather than electrical neutralization) proceeds over a barrier ΔE_A.

The rate equations describing the transitions between the three states then read

$$\frac{\partial f_1}{\partial t} = -f_1 k_{12} + f_3 k_{31}, \tag{9}$$

$$\frac{\partial f_2}{\partial t} = +f_1 k_{12} - f_2 k_{23} + f_3 k_{32}, \tag{10}$$

$$\frac{\partial f_3}{\partial t} = +f_2 k_{23} - f_3 k_{32} - f_3 k_{31}. \tag{11}$$

The probability of being in state i is given by f_i while the transition rates from state i to j are given by k_{ij}. Naturally, (9)–(11) are not linearly independent since the defect has to be in one of its three states ($f_1 + f_2 + f_3 = 1$). Being a linear equation system, solution of (9)–(11) is straight-forward. The rates read as follows

$$k_{12} = k_p^\mathrm{c}(E_\mathrm{T}, \Delta E_\mathrm{B}, F_\mathrm{c}) + k_n^\mathrm{e}(E_\mathrm{T}, \Delta E_\mathrm{B}, F_\mathrm{c}), \tag{12}$$

$$k_{23} = k_p^\mathrm{e}(E'_\mathrm{T}, \Delta E_\mathrm{C}, 0) + k_n^\mathrm{c}(E'_\mathrm{T}, \Delta E_\mathrm{C}, 0), \tag{13}$$

$$k_{32} = k_p^\mathrm{c}(E'_\mathrm{T}, \Delta E_\mathrm{C}, 0) + k_n^\mathrm{e}(E'_\mathrm{T}, \Delta E_\mathrm{C}, 0), \tag{14}$$

$$k_{31} = \nu \exp(-\beta \Delta E_\mathrm{A}), \tag{15}$$

where the shorthand k(trap level, MPE barrier, MPFAT reference field) is used and $\nu \sim 10^{13}\,\mathrm{Hz}$ is the typical attempt frequency for thermal transitions over energetic barriers. We remark that under conventional stress and recovery voltages the contribution of electrons is negligible. Only for positive bias which is applied in some of our recovery experiments they provide a significant contribution and impact the recovery dynamics, see Fig. 8.

In order to describe the response of a device to a change in the bias conditions, a certain number of defects N is assumed to exist. Due to the amorphous nature of the Si/SiO$_2$ interface, each defect will be described by a unique configuration of random variables $\mathbf{X} = (x, E_\mathrm{T}, E'_\mathrm{T}, \Delta E_\mathrm{A}, \Delta E_\mathrm{B}, \Delta E_\mathrm{C})$. The joint probability density function is given by $g(\mathbf{X})$ and macroscopically observable quantities will be given by suitable averages

$$\langle m \rangle = N \int \mathrm{d}\mathbf{X}\, m\, g(\mathbf{X}). \tag{16}$$

For instance, the total positive charge contributing to ΔV_th is then given by the statistical average over the E' centers being in state 2, that is, positively charged, and we obtain

$$Q_\mathrm{ox}(t) = \mathrm{q}\langle (1 - x/t_\mathrm{ox}) f_2(t) \rangle, \tag{17}$$

with t_ox the oxide thickness, and $\Delta Q_\mathrm{ox}(t) = Q_\mathrm{ox}(t) - Q_\mathrm{ox}(0)$. The total number of oxide traps is given by all the defects not being in state 1 (the precursor state) and is simply

$$N_\mathrm{ox}(t) = 1 - \langle f_1(t) \rangle. \tag{18}$$

In order to keep the model as simple as possible, we will assume $g(\mathbf{X})$ to be given by independent homogeneous distributions in every variable, unless otherwise noted.

As the recovery in our ultra-thin and ultra-thick oxides looks basically the same, we conclude that the observed dispersion in time constants is primarily a property of the Si/SiO$_2$ interface. Consequently, the depth dispersion will be neglected ($x \approx 0$) to further simplify the model and thus $Q_\mathrm{ox} = \mathrm{q}\langle f_2 \rangle$. Naturally, a variation in x will exist in reality, however, at the moment it does not appear necessary to include it into the model to explain our data.

Two interesting aspects of our E' center model are demonstrated in Fig. 8. First, it can predict asymmetric logarithmic stress and recovery behavior. Second, the model predicts a strong sensitivity to positive bias during recovery: Initially ($t_\mathrm{r} < 10^{-15}$), after the stress bias is

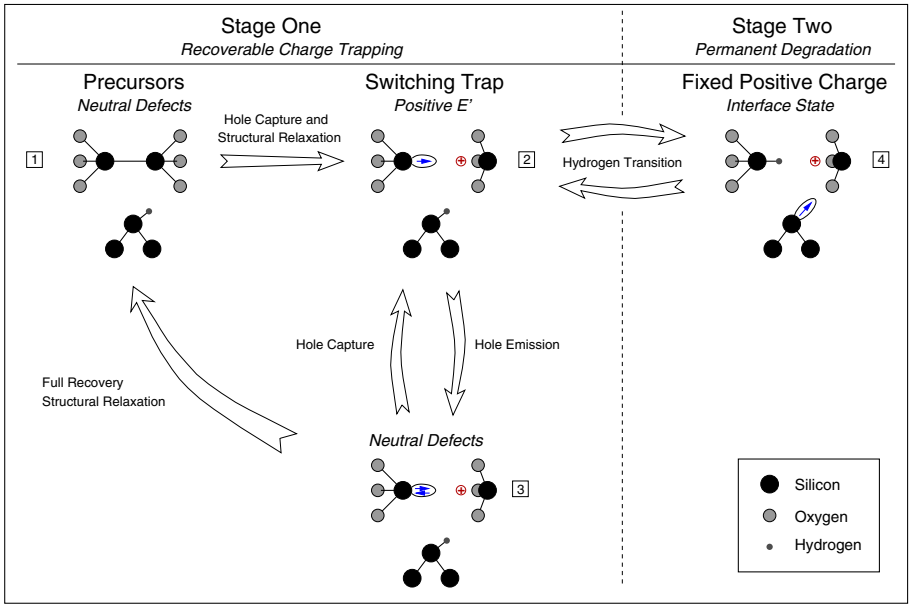
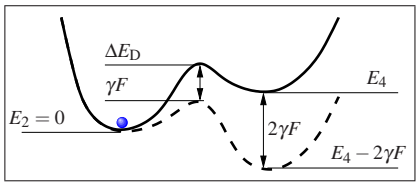


Fig. 9: Left: The HDL model for a switching oxide trap coupled to the creation of a dangling bond at the interface. When the E' center is positively charged (in state 2), the hydrogen passivating a silicon dangling bond at the interface can move to the E' center, thereby effectively locking in the positive charge (state 4). The charge state of the thereby created dangling bond depends on the position of the Fermi-level. **Bottom**: The transition between state 2 and 4 is modeled by assuming a field-dependent thermal transition over a barrier.



switched to the recovery voltage, most defects are positively charged (state 2) and nothing changes. Then, depending on the position of the Fermi-level during recovery, a possibly significant number of defects are electrically neutralized in the pico- and nano-second regime. In particular for positive bias, electrons from the conduction band cause a fast change in occupancy. It is important to realize that although the amount of visible positive charge can be small, the defect is not yet annealed and still there (not in state 1). With the artificially narrow distribution of ΔE_B used for demonstration purposes, *real recovery* only sets in at 10^{-3} with the recovery being fast when most defects are neutral (positive gate voltage) and slow when most defects are positive (negative voltage). With more realistic parameters, these transitions are blurred.

B. Coupled Interface State Generation

Once a hole has been trapped in the oxygen vacancy, a positive E' center is obtained. One half of the E' center is an unpassivated silicon dangling bond (DB). We now follow the arguments of Lenahan [35]: assume that in an unstressed device a certain number of hydrogen passivated Si DBs at the interface exist. During stress, unpassivated Si DBs right next to the interface are created in the form of the E' centers. Using simple thermodynamical arguments it can be shown that the depassivation of interface DBs by the migration of hydrogen to the newly formed DBs at the E' center is energetically favored. Contradicting evidence is available regarding what could happen to the hydrogen now residing at the E' DB. Often, a 1:1 correlation between fixed oxide charge and interface states has been reported, which is obtained in the current model by assuming that the H stays at the E' center. Alternatively, the E' center could just act as a catalyst and the H may migrate further away.

In order to capture this coupling between E' and P_b centers we thus extend the HDL model of the E' center of Fig. 7 by introducing an E'/P_bH complex. The complex is assumed to be in one of 4 states, with states 1-3 being the same as the E' center states and the P_b center being passivated with H (see Fig. 9). However, now once positively charged (state 2), the E' center can attract the H from the P_bH. When this happens, the E'/P_bH complex moves to state 4. This step locks in the positive charge at the E' and creates a DB at the interface, whose charge state quickly follows the Fermi-level in the substrate. The H has a non-zero probability of moving back to the P_b center, thereby resetting the E'/P_b complex to state 2, from which complete annealing is eventually possible. However, for typical stress conditions the full recovery would be outside the measurement window.

The rate equation describing such a E'/P_bH complex follow from a straight-forward extension of (9)–(11)

$$\frac{\partial f_1}{\partial t} = -f_1 k_{12} + f_3 k_{31}, \tag{19}$$

$$\frac{\partial f_2}{\partial t} = +f_1 k_{12} - f_2 k_{23} + f_3 k_{32} - f_2 k_{24} + f_4 k_{42}, \tag{20}$$

$$\frac{\partial f_3}{\partial t} = +f_2 k_{23} - f_3 k_{32} - f_3 k_{31}, \tag{21}$$

$$\frac{\partial f_4}{\partial t} = +f_2 k_{24} - f_4 k_{42}. \tag{22}$$

Again, the probabilities must fulfill $f_1 + f_2 + f_3 + f_4 = 1$ and one of the above equations must be omitted. The transition rates between state 2 and 4 are modeled in the spirit of [1, 22] by thermal activation over a field-dependent barrier (cf. Fig. 9) as

$$k_{24} = \nu e^{-\beta(\Delta E_D - E_2 - \gamma F)}, \tag{23}$$

$$k_{42} = \nu e^{-\beta(\Delta E_D - E_4 + \gamma F)}. \tag{24}$$

Consequently, the probability of being in state 4 corresponds to a fixed positive charge at the E' center *and* a depassivated interface state. However, only interface states above the Fermi level and up to mid-gap are assumed to be positively charged and their occupancy with an electron is given by $f_{it}(t)$, which is determined using the conventional SRH mechanism. Contrary to claims in [13], f_{it} normally reaches its equilibrium value in the nanosecond regime (extending into the millisecond regime only for a switch of the Fermi-level below E_V during stress to mid-gap) and thus f_{it} may be set equal to the Fermi distribution in most cases.

Following [1, 22, 52], the barrier ΔE_D is assumed to be given by a narrow Gaussian distribution ($\sigma \approx 100\,\text{meV}$), while the energy-levels E_{it} of the electrically active states are assumed to be homogeneously distributed over the lower-half of the silicon bandgap. We remark that a strong correlation between these two distributions is to be

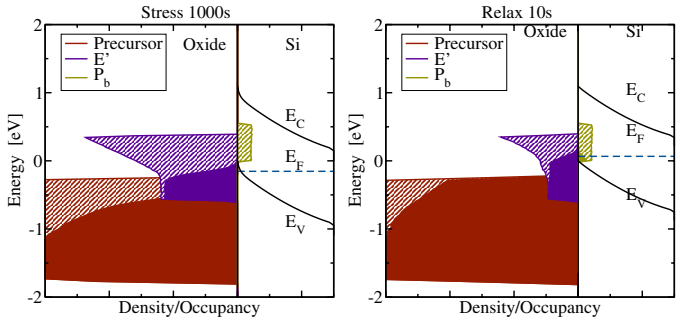


Fig. 10: Slightly schematized simulated evolution of the three densities-of-states for the precursors, E' centers, and the interface states, together with their occupancies. The DOS of the precursor and its occupancy ($\langle f_1 \rangle$) were divided by two while the DOS for the interface states ($\langle f_4 \rangle$) and the occupancy ($\langle f_{\rm it} \rangle$) were multiplied by two in order to highlight the important details. The bandedges in Si are schematically shown. **Left**: During stress, $E_{\rm F}$ is below $E_{\rm V}$ at the interface and precursors are being broken up. The created E' centers with an $E'_{\rm T}$ below $E_{\rm F}$ are neutral and thus not as stable as the centers with $E'_{\rm T} > E_{\rm F}$, which gives rise to the spike in the DOS above $E_{\rm F}$. All interface states are positively charged. **Right**: During recovery, $E_{\rm F}$ moves upwards, neutralizes E' centers and thereby accelerates recovery. Most interface states remain positively charged.

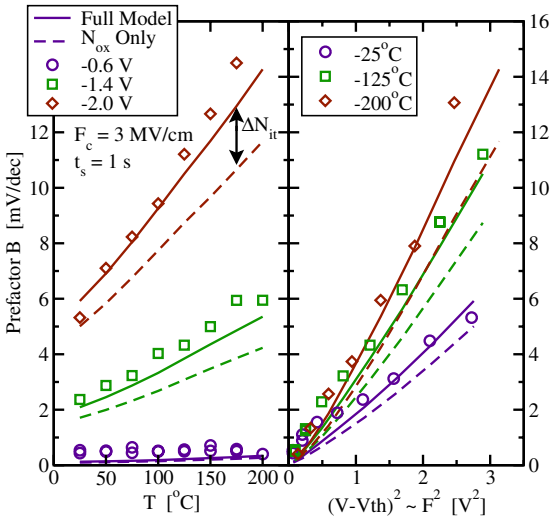


Fig. 11: The measured prefactor $B_{\rm s}(T,V)$ of Fig. 3 compared to the simulated prefactor of the two stage model obtained under the same conditions. Very good agreement for all voltages and temperatures is obtained. Since the two-stage model captures the asymmetry, good agreement during both stress and recovery is possible. We put the remaining deviation down to the mobility error in the OTF data [11, 13].

expected, that is, $E_{\rm it}$ should also be given by two Gaussians in the bandgap [43]. Also note that in contrast to the E' center model, re-passivation of interface states is assumed to be independent of their charge state, which might not be correct. We thus consider the above model a first-order approximation with the minimal number of free parameters. More detailed experimental data would be required to justify a refined version of the model.

The random variables of our $E'/P_{\rm b}$H complex are thus $\mathbf{X} = (x, E_{\rm T}, E'_{\rm T}, \Delta E_{\rm A}, \Delta E_{\rm B}, \Delta E_{\rm C}, \Delta E_{\rm D})$ and their joint probability density function $g(\mathbf{X})$ used in the statistical average (16) is constructed from independent distributions.

The amount of positive charge stored in the E' part of the complexes is now given by the statistical average of complexes being either in state 2 *or* 4, and we have

$$Q_{\rm ox} = {\rm q}\langle (1 - x/t_{\rm ox})(f_2 + f_4) \rangle \tag{25}$$

The amount of positive charge stored in the interface states is given by the average of the probability of having depassivated DBs (f_4) times the probability that the created electrical level is unoccupied ($1 - f_{\rm it}$) and for $E_{\rm it}$ within the lower-half of the silicon band-gap (the donor-like states) given by

$$Q_{\rm it} = {\rm q}\langle f_4 (1 - f_{\rm it}) \rangle, \tag{26}$$

while the total number of available interface states is simply obtained from $N_{\rm it} = \langle f_4 \rangle$. For the data analyzed here, the occupancy of the created interface states is of minor relevance since normally both during stress and recovery the Fermi-level is close to the silicon valence band ($f_{\rm it} \approx 0$) and consequently most interface states will be positively charged and thus $Q_{\rm it} \approx {\rm q}N_{\rm it}$.

In order to explain the presently available data, it is sufficient to use a reduced set of random variables and we use $x \approx 0$, $\Delta E_{\rm A} \approx \Delta E_{\rm B}$, and $\Delta E_{\rm C} \approx 0$ in the following examples. Further simplification, like $E'_{\rm T} \approx E_{\rm T}$, significantly impacts the quality of the model as will be discussed later.

A typical evolution of the densities of the precursors, the created E' centers, and the created interface states is shown in Fig. 10 during stress and after 10 s of recovery. Although uniform distributions are used in the model, the resulting DOS of E' centers is non-uniform. This is because states closer or below the Fermi-level are electrically neutral and thus have a larger annealing rate. Consequently, the higher the trap level $E'_{\rm T}$, the longer it will take the defect to fully anneal, consistent with the interpretation given in [27].

VII. Comparison with Measurements

Simulation results for the SiON devices are compared to the data of Fig. 3 and Fig. 4. The simulated prefactors during stress are in very good agreement with the data, see Fig. 11. Details of the simulation demonstrate also that the model can reproduce the asymmetry between stress and recovery, see Fig. 12. We remark that this is the first time that a model can reproduce both OTF and recovery data in a wide temperature and voltage range, providing a theoretical link between these two measurement techniques.

A particularly challenging data set is given in Fig. 13, where five devices are brought to the same level of degradation. During recovery, various bias switches are used to probe both the occupancy effect (amount of charge visible depending on the Fermi-level) as well as the impact of the occupancy on the recovery dynamics. Again, very good agreement between theory and data is obtained.

Results of similar accuracy have been obtained for the HK device, see Fig. 14. Compared to the SiON-k devices, a considerably larger amount of E' centers is created.

This is in contrast to data taken on thick SiO_2 devices where a considerably smaller amount of E' centers is created, see Fig. 15. Finally, we test the underlying physical assumptions of the model by applying a recently developed poly-heater technology allowing rapid switches of the device temperature [20]. The appealing feature of this technique is that devices can be brought to the same stress level from which recovery can be monitored at different T. As shown in Fig. 16, these insightful experimental data are well reproduced by the model, confirming its validity.

VIII. Alternative Models and Why They Do Not Work

One question that naturally needs to be raised is whether really all features of the E' center are needed in order to explain NBTI. In a nutshell, the answer is simply 'yes' and we will demonstrate in this section why this is the case. We have to keep in mind, however,

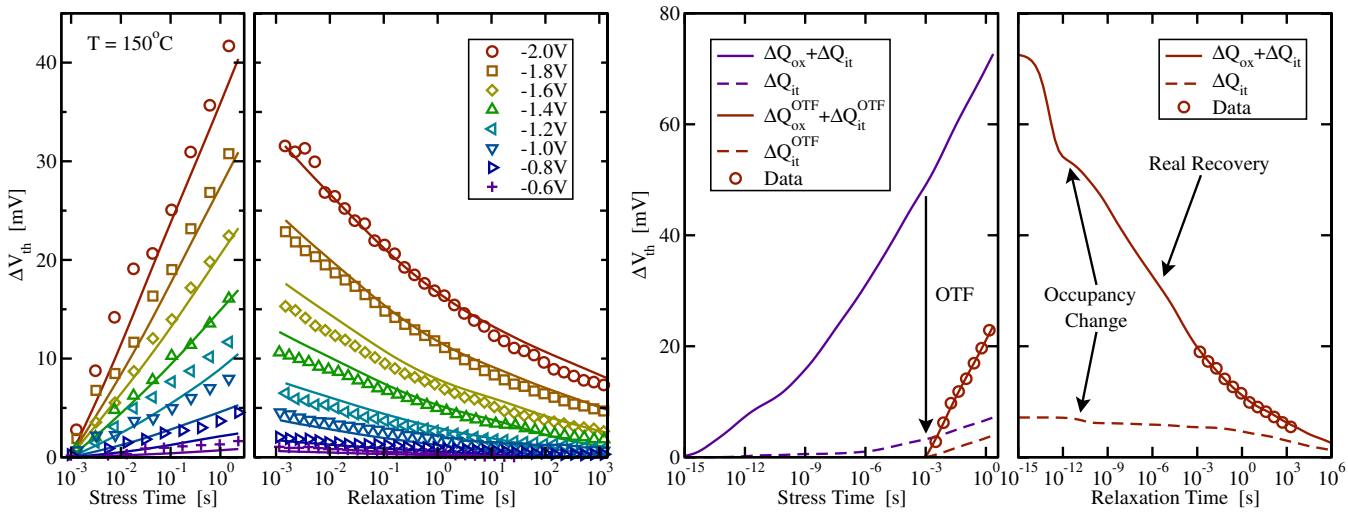


Fig. 12: Left: Comparison of the simulated asymmetry of stress and recovery measured at 150 °C for 8 different stress voltages for the thin SiON devices. The asymmetry is properly reproduced by the model, resulting in a good fit during *both* stress and recovery. **Right**: Detailed simulation results at 50 °C and −2 V. Both measurements miss about 50 mV of the real degradation.

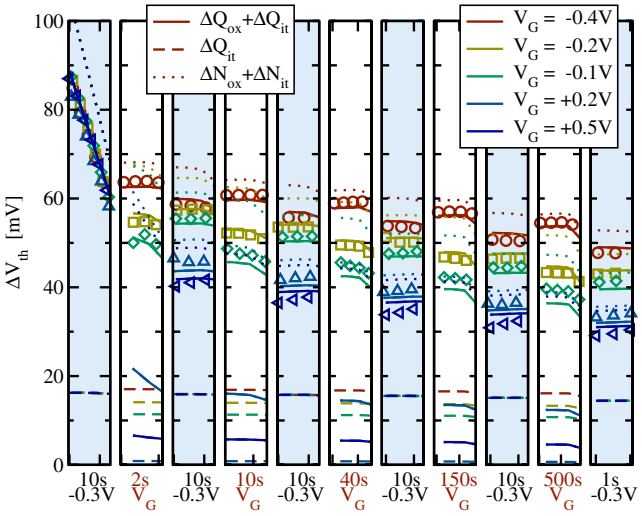


Fig. 13: Five devices were brought to the same level of degradation (t_s = 6000 s, $T = 125\,°C$, $V_{stress} = -2\,V$). First, recovery was monitored for 10 s at $V_G = V_{th} = -0.3\,V$ (panel 1). Then the gate voltage was switched for 2 s to 5 different values, including more negative and more positive values (panel 2). When possible ($V_G > 0\,V$), the change in the drain current was converted to $\Delta V_{th}(V_G)$ which was found to be clearly different from $\Delta V_{th}(V_{th})$ recorded at V_{th}. Next, V_G was switched back to V_{th} (panel 3), were a clear impact of the intermediate bias switch is observed. This procedure was repeated for increasing durations of bias switches (10, 40, 150, 500 s). Simulation results are given by the lines, which show very good agreement with the data, capturing both the occupancy effect (evenly numbered panels) and the acceleration/retardation of recovery as a response to the gate bias.

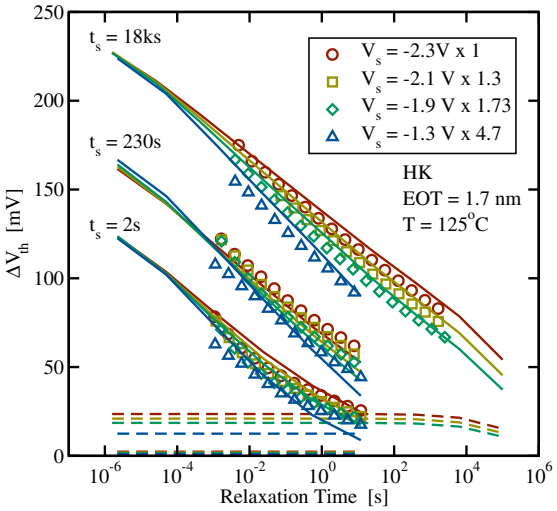


Fig. 14: High-k devices stressed under various conditions were used to extract the model parameters. Shown is the scaling of the recovery data after three selected stress times obtained at different stress voltages. The data was scaled to the first simulated recovery point for the most severe stress condition.

that the following allegedly simpler models are incompatible with the known properties of the E' center. Rather, we wish to show that each aspect of the model has its correspondence in experimental data.

A. *Occupancy Effect*

First, we demonstrate the fundamental impact of the trap occupancy on the simulated stress and recovery characteristics. From an electrical point of view, the occupancy effect (transitions between state 2 and 3) is responsible for a change in the subthreshold slope. It also explains the asymmetry between the degradation and recovery dynamics. A simplified model that neglects state 3 (the electrically neutral state which is not yet fully annealed) by inducing recovery right after hole emission from state 2, inevitably predicts a symmetric stress and recovery behavior, see Fig. 17. Furthermore, since the recovery rate does no longer depend on the occupancy of the defect, only a weak bias dependence is observed.

B. *Level Shift*

The next example effectively demonstrates what happens when the level shift is neglected, that is, when $E'_T = E_T$ is assumed. In order to have a stable precursor configuration, E_T must lie well below the silicon valence band. During stress, holes are trapped at that level. During relaxation the holes are kept in that low level only by the MPE barrier and must go back to the valence band, a process independent of the energetic position of the trap, since the holes just 'bubble up'. Consequently, practically no bias dependence of recovery is observed in such a model, see Fig. 18. Also, the model predicts symmetric stress and recovery behavior.

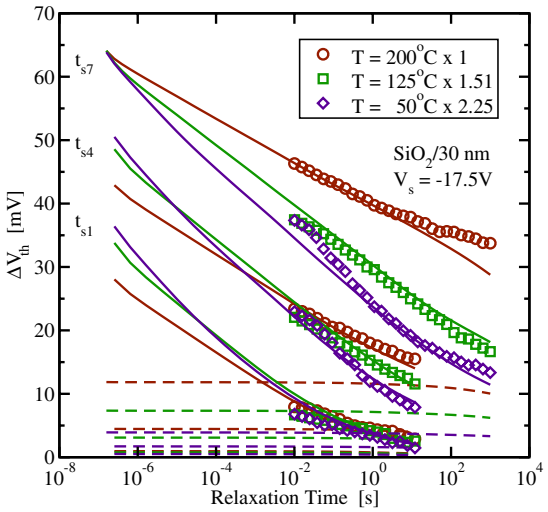


Fig. 15: Similar to Fig. 14, but for a thick SiO_2 with EOT = 30 nm. Shown is the scaling of the recovery data after three selected stress times obtained at different stress temperatures. Note that compared to the SiON and the high-k devices, the contribution of recoverable oxide charges is considerably smaller.

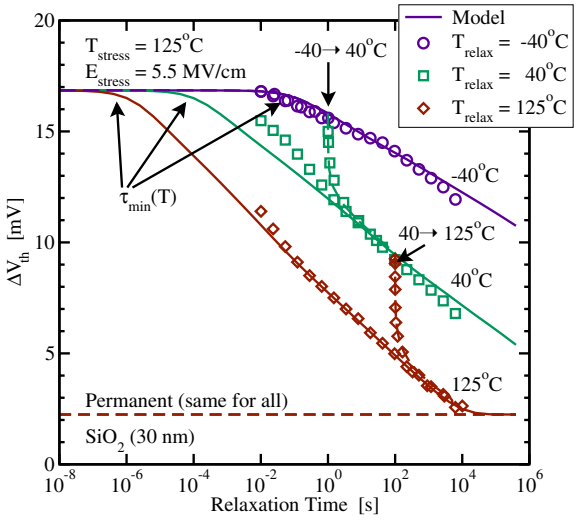


Fig. 16: Application of a recently developed poly-heater technology [20] allows for high-precision switching of the device temperature within 1 s. Identically processed devices were stressed for the same time at 125 °C, while prior recovery the temperature is quickly switched to −40 °C or +40 °C. Finally, in a fourth device the temperature is first switched to -40 °C, then after 1 s recovery to +40 °C, and finally after 100 s to 125 °C. Excellent agreement of the model with data is again obtained.

IX. Conclusions

We suggest a two stage model for the negative bias temperature instability based on established properties of E' centers. Creation of E' centers from their oxygen vacancy precursors is suggested to occur via a *multiphonon-field-assisted* hole trapping mechanism. The created E' centers are then suggested to favor the depassivation of interface states, which results in a coupling of created oxide and interface state component.

This model can explain degradation and recovery over a wide range of bias voltages and stress temperatures, the observed asymmetry between stress and recovery, and the strong sensitivity to bias and temperature during recovery. Excellent agreement with data from three vastly different technologies (thick SiO_2, SiON, and HK) is obtained, supporting the idea that NBTI is determined by the

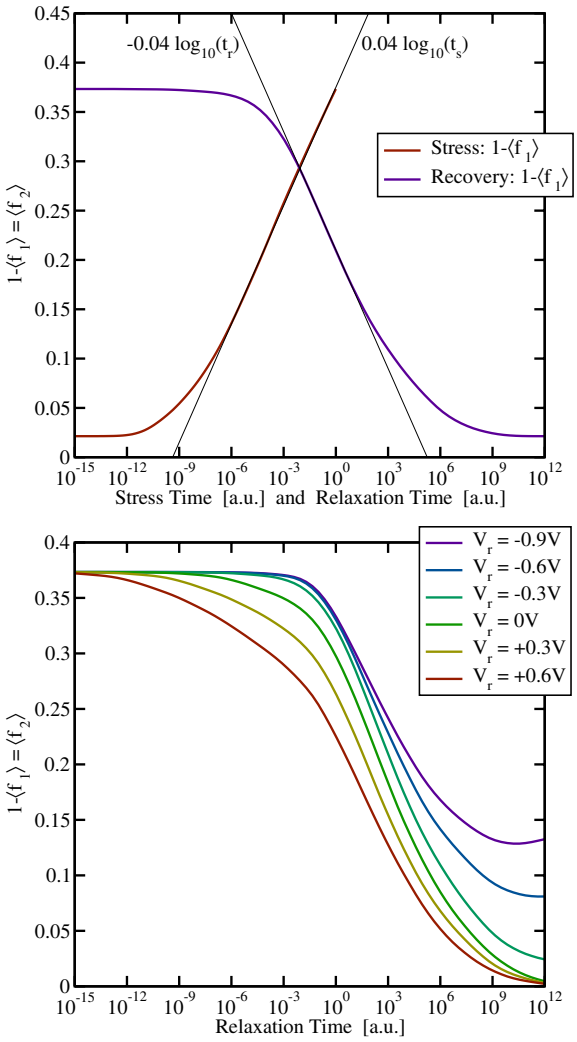


Fig. 17: Characteristic behavior of a model that neglects the occupancy effect (no switching behavior, no state 3 in the HDL model). **Top**: Although it is possible to delay the recovery (shift to larger times), the slopes are always the same during stress and recovery, they are always symmetric. **Bottom**: Devices stressed at the same voltage only show a weak voltage dependence during recovery.

chemistry of the amorphous SiO_2/Si interface region. The model has the minimum number of parameters required to explain the experimentally observed features of NBTI.

X. Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement n°216436 (project ATHENIS). We also gratefully acknowledge stimulating discussions with P. Lenahan and A. Shluger.

References

[1] V. Huard, C. Parthasarathy, N. Rallet, C. Guerin, M. Mammase, D. Barge, and C. Ouvrard, "New Characterization and Modeling Approach for NBTI Degradation from Transistor to Product Level," in *Proc. IEDM*, 2007, pp. 797–800.

[2] C. Shen, M.-F. Li, C. E. Foo, T. Yang, D.M. Huang, A. Yap, G.S. Samudra, and Y.-C. Yeo, "Characterization and Physical Origin of Fast V_{th} Transient in NBTI of pMOSFETs with SiON Dielectric," in *Proc. IEDM*, 2006, pp. 333–336.

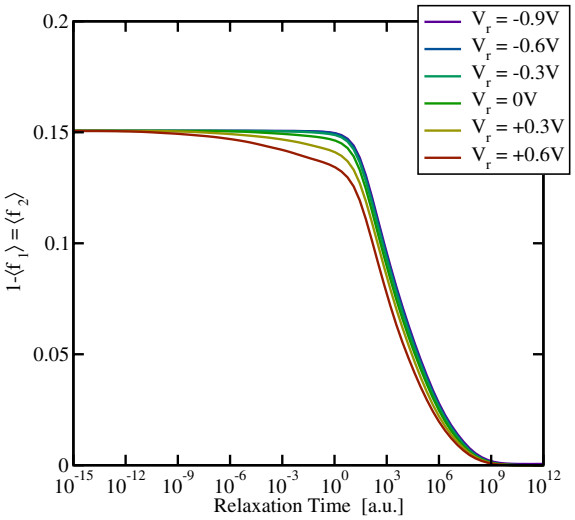


Fig. 18: Characteristic behavior of a model that neglects the occupancy effect and does not consider the level shift (no state 3 and $E'_{\mathrm{T}} = E_{\mathrm{T}}$). Stress and recovery are always symmetric (not shown). Devices stressed at the same voltage show practically no voltage dependence during recovery.

[3] B. Kaczer, T. Grasser, Ph.J. Roussel, J. Martin-Martinez, R. O'Connor, B.J. O'Sullivan, and G. Groeseneken, "Ubiquitous Relaxation in BTI Stressing-New Evaluation and Insights," in *Proc. IRPS*, 2008, pp. 20–27.

[4] T. Grasser, "Negative Bias Temperature Instability: Modeling Challenges and Perspectives," in *Proc. IRPS*, 2008, (Tutorial).

[5] T. Grasser, B. Kaczer, and W. Goes, "An Energy-Level Perspective of Bias Temperature Instability," in *Proc. IRPS*, 2008, pp. 28–38.

[6] T. Grasser, B. Kaczer, Th. Aichinger, W. Goes, and Michael Nelhiebel, "Defect Creation Stimulated by Thermally Activated Hole Trapping as the Driving Force Behind Negative Bias Temperature Instability in SiO_2, SiON, and High-k Gate Stacks," in *IIRW Final Rep.*, 2008.

[7] T. Grasser and B. Kaczer, "Evidence that Two Tightly Coupled Mechanism are Responsible for Negative Bias Temperature Instability in Oxynitride MOSFETs," *IEEE Trans.Electr.Dev.*, 2009, (in print).

[8] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, "Disorder-Controlled-Kinetics Model for Negative Bias Temperature Instability and its Experimental Verification," in *Proc. IRPS*, 2005, pp. 381–387.

[9] A.T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, and S. Krishnan, "NBTI Impact on Transistor and Circuit: Models, Mechanisms and Scaling Effects," in *Proc. IEDM*, 2003, pp. 1–4.

[10] S. Mahapatra, K. Ahmed, D. Varghese, A. E. Islam, G. Gupta, L. Madhav, D. Saha, and M. A. Alam, "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*, 2007, pp. 1–9.

[11] T. Grasser, P.-J. Wagner, Ph. Hehenberger, W. Goes, and B. Kaczer, "A Rigorous Study of Measurement Techniques for Negative Bias Temperature Instability," *IEEE Trans.Dev.Mat.Rel.*, vol. 8, no. 3, pp. 526 – 535, 2008.

[12] H. Reisinger, U. Brunner, W. Heinrigs, W. Gustin, and C. Schlünder, "A Comparison of Fast Methods for Measuring NBTI Degradation," *IEEE Trans.Dev.Mat.Rel.*, vol. 7, no. 4, pp. 531–539, 2007.

[13] A.E. Islam, E. N. Kumar, H. Das, S. Purawat, V. Maheta, H. Aono, E. Murakami, S. Mahapatra, and M.A. Alam, "Theory and Practice of On-the-fly and Ultra-fast V_{T} Measurements for NBTI Degradation: Challenges and Opportunities," in *Proc. IEDM*, 2007, pp. 1–4.

[14] C. Schlünder, R.-P. Vollertsen, W. Gustin, and H. Reisinger, "A Reliable and Accurate Approach to Assess NBTI Behavior of State-Of-The-Art pMOSFETs With Fast-WLR," in *Proc. ESSDERC*, 2007, pp. 131–134.

[15] H. Reisinger, O. Blank, W. Heinrigs, A. Mühlhoff, W. Gustin, and C. Schlünder, "Analysis of NBTI Degradation- and Recovery-Behavior Based on Ultra Fast V_{th}-Measurements," in *Proc. IRPS*, 2006, pp. 448–453.

[16] J.F. Zhang, Z. Ji, M.H. Chang, B. Kaczer, and G. Groeseneken, "Real V_{th} Instability of pMOSFETs Under Practical Operation Conditions," in *Proc. IEDM*, 2007, pp. 817–820.

[17] V. Huard, M. Denais, and C. Parthasarathy, "NBTI Degradation: From Physical Mechanisms to Modelling," *Microelectr.Reliab.*, vol. 46, no. 1, pp. 1–23, 2006.

[18] T. Grasser, W. Goes, V. Sverdlov, and B. Kaczer, "The Universality of NBTI Relaxation and its Implications for Modeling and Characterization," in *Proc. IRPS*, 2007, pp. 268–280.

[19] T. Grasser, B. Kaczer, Ph. Hehenberger, W. Goes, R. O'Connor, H. Reisinger, W. Gustin, and C. Schlünder, "Simultaneous Extraction of Recoverable and Permanent Components Contributing to Bias-Temperature Instability," in *Proc. IEDM*, 2007, pp. 801–804.

[20] Th. Aichinger, M. Nelhiebel, and T. Grasser, "Unambiguous Identification of the NBTI Recovery Mechanism using Ultra-Fast Temperature Changes," in *Proc. IRPS*, 2009, (in print).

[21] M.-F. Li, D. Huang, C. Shen, T. Yang, W.J., W.J. Liu, and Z. Liu, "Understand NBTI Mechanism by Developing Novel Measurement Techniques," *IEEE Trans.Dev.Mat.Rel.*, vol. 8, no. 1, pp. 62–71, March 2008.

[22] A. Haggag, W. McMahon, K. Hess, K. Cheng, J. Lee, and J. Lyding, "High-Performance Chip Reliability from Short-Time-Tests," in *Proc. IRPS*, 2001, pp. 271–279.

[23] Ph. Hehenberger, Th. Aichinger, T. Grasser, W. Goes, O. Triebl, B. Kaczer, and M. Nelhiebel, "Do NBTI-Induced Interface States Show Fast Recovery? A Study Using a Corrected On-The-Fly Charge-Pumping Measurement Technique," in *Proc. IRPS*, 2009, (in print).

[24] T.L. Tewksbury, *Relaxation Effects in MOS Devices due to Tunnel Exchange with Near-Interface Oxide Traps*, Ph.D. Thesis, MIT, 1992.

[25] A.E. Islam, H. Kufluoglu, D. Varghese, S. Mahapatra, and M.A. Alam, "Recent Issues in Negative-Bias Temperature Instability: Initial Degradation, Field Dependence of Interface Trap Generation, Hole Trapping Effects, and Relaxation," *IEEE Trans.Electr.Dev.*, vol. 54, no. 9, pp. 2143–2154, 2007.

[26] J.F. Zhang, C.Z. Zhao, A.H. Chen, G. Groeseneken, and R. Degraeve, "Hole Traps in Silicon Dioxides - Part I: Properties," *IEEE Trans.Electr.Dev.*, vol. 51, no. 8, pp. 1267–1273, 2004.

[27] D.S. Ang, S. Wang, G.A. Du, and Y.Z. Hu, "A Consistent Deep-Level Hole Trapping Model for Negative Bias Temperature Instability," *IEEE Trans.Dev.Mat.Rel.*, vol. 8, no. 1, pp. 22–34, 2008.

[28] M.B. Weissman, "$1/f$ Noise and other Slow, Nonexponential Kinetics in Condensed Matter," *Rev.Mod.Phys*, vol. 60, no. 2, pp. 537–571, 1988.

[29] D.M. Fleetwood, H.D. Xiong, Z.-Y. Lu, C.J. Nicklaw, J.A. Felix, R.D. Schrimpf, and S.T. Pantelides, "Unified Model of Hole Trapping, $1/f$ Noise, and Thermally Stimulated Current in MOS Devices," *IEEE Trans.Electr.Dev.*, vol. 49, no. 6, pp. 2674–2683, 2002.

[30] C.H. Henry and D.V. Lang, "Nonradiative Capture and Recombination by Multiphonon Emission in GaAs and GaP," *Phys.Rev.B*, vol. 15, no. 2, pp. 989–1016, 1977.

[31] M.J. Kirton and M.J. Uren, "Noise in Solid-State Microstructures: A New Perspective on Individual Defects, Interface States and Low-Frequency (1/f) Noise," *Adv.Phys*, vol. 38, no. 4, pp. 367–486, 1989.

[32] S. Makram-Ebeid and M. Lannoo, "Quantum Model for Phonon-Assisted Tunnel Ionization of Deep Levels in a Semiconductor," *Phys.Rev.B*, vol. 25, no. 10, pp. 6406–6424, 1982.

[33] S.D. Ganichev, W. Prettl, and I.N. Yassievich, "Deep Impurity-Center Ionization by Far-Infrared Radiation," *Phys.Solid State*, vol. 39, no. 1, pp. 1703–1726, 1997.

[34] E.H. Poindexter and W.L. Warren, "Paramagnetic Point Defects in Amorphous Thin Films of SiO_2 and Si_3N_4: Updates and Additions," *J.Electrochem.Soc.*, vol. 142, no. 7, pp. 2508–2516, 1995.

[35] P.M. Lenahan, "Atomic Scale Defects Involved in MOS Reliability Problems," *Microelectr.Eng.*, vol. 69, pp. 173–181, 2003.

[36] P.V. Sushko, S. Mukhopadhyay, A.S. Mysovsky, V.B. Sulimov, A. Taga, and A.L. Shluger, "Structure and Properties of Defects in Amorphous Silica: New Insights from Embedded Cluster Calculations," *J.Phys.:Condensed Matter*, vol. 17, pp. 2115–2140, 2005.

[37] A.J. Lelis and T.R. Oldham, "Time Dependence of Switching Oxide Traps," *IEEE Trans.Nucl.Science*, vol. 41, no. 6, pp. 1835–1843, Dec 1994.

[38] J.F. Conley Jr., P.M. Lenahan, A.J. Lelis, and T.R. Oldham, "Electron Spin Resonance Evidence for the Structure of a Switching Oxide Trap: Long Term Structural Change at Silicon Dangling Bond Sites in SiO_2," *Appl.Phys.Lett.*, vol. 67, no. 15, pp. 2179–2181, 1995.

[39] C.J. Nicklaw, Z.-Y. Lu, D.M. Fleetwood, R.D. Schrimpf, and S.T. Pantelides, "The Structure, Properties, and Dynamics of Oxygen Vacancies

in Amorphous SiO_2," *IEEE Trans.Nucl.Science*, vol. 49, no. 6, pp. 2667–2673, Dec 2002.

[40] A. V. Kimmel, P.V. Sushko, A.L. Shluger, and G. Bersuker, "Positive and Negative Oxygen Vacancies in Amorphous Silica," in *Silicon Nitride, Silicon Dioxide, and Emerging Dielectrics 10*, R. Sah, J. Zhang, Y. Kamakura, M. Deen, and J. Yota, Eds. ECS Transactions, 2009, (in print).

[41] V. Reddy, A.T. Krishnan, A. Marshall, J. Rodriguez, S. Natarajan, T. Rost, and S. Krishnan, "Impact of Negative Bias Temperature Instability on Digital Circuit Reliability," in *Proc. IRPS*, 2002, pp. 248–254.

[42] A.T. Krishnan, C. Chancellor, S. Chakravarthi, P.E. Nicollian, V. Reddy, A. Varghese, R.B. Khamankar, and S. Krishnan, "Material Dependence of Hydrogen Diffusion: Implications for NBTI Degradation," in *Proc. IEDM*, 2005, pp. 688–691.

[43] C.H. Helms and E.H. Poindexter, "The Silicon-Silicon-Dioxide System: Its Microstructure and Imperfections," *Rep.Prog.Phys.*, vol. 57, pp. 791–852, 1994.

[44] C. Schlunder, M. Hoffmann, R.-P. Vollertsen, G. Schindler, W. Heinrigs, W. Gustin, and H. Reisinger, "A Novel Multi-Point NBTI Characterization Methodology Using Smart Intermediate Stress (SIS)," in *Proc. IRPS*, May 2008, pp. 79–86.

[45] D. Brisbin and P. Chaparala, "The Effect of the Subthreshold Slope Degradation on NBTI Device Characterization," in *IIRW Final Rep.*, 2008, pp. 96–99.

[46] G. Kapila, N. Goyal, V.D. Maheta, C. Olsen, K. Ahmed, and S. Mahapatra, "A Comprehensive Study of Flicker Noise in Plasma Nitrided SiON p-MOSFETs: Process Dependence of Pre-Existing and NBTI Stress Generated Trap Distribution Profiles," in *Proc. IEDM*, 2008, pp. 103–106.

[47] B. Kaczer, T. Grasser, J. Martin-Martinez, E. Simoen, M. Aoulaiche, Ph.J. Roussel, and G. Groeseneken, "NBTI from the Perspective of Defect States with Widely Distributed Time Scales," in *Proc. IRPS*, 2009.

[48] J.P. Campbell, P.M. Lenahan, C.J. Cochrane, A.T. Krishnan, and S. Krishnan, "Atomic-Scale Defects Involved in the Negative-Bias Temperature Instability," *IEEE Trans.Dev.Mat.Rel.*, vol. 7, no. 4, pp. 540–557, 2007.

[49] D.B. Brown and N.S. Saks, "Time Dependence of Radiation-Induced Trap Formation in Metal-Oxide-Semiconductor Devices as a Function of Oxide Thickness and Applied Field," *J.Appl.Phys.*, vol. 70, no. 7, pp. 3734–3747, 1991.

[50] D.V. Lang and C.H. Henry, "Nonradiative Recombination at Deep Levels in GaAs and GaP by Lattice-Relaxation Multiphonon Emission," *Physical Review Letters*, vol. 35, no. 22, pp. 1525–1528, 1975.

[51] J.F. Conley Jr., P.M. Lenahan, and W.F. McArthur, "Preliminary Investigation of the Kinetics of Postoxidation Rapid Thermal Anneal Induced Hole-Trap-Precursor Formation in Microelectronic SiO_2Films," *Appl.Phys.Lett.*, vol. 73, no. 15, pp. 2188–2190, 1998.

[52] A. Stesmans, "Dissociation Kinetics of Hydrogen-Passivated P_b Defects at the (111)Si/SiO_2 Interface," *Phys.Rev.B*, vol. 61, no. 12, pp. 8393–8403, 2000.