

AN ENERGY-LEVEL PERSPECTIVE OF BIAS TEMPERATURE INSTABILITY

Tibor Grasser*, Ben Kaczer^o, and Wolfgang Goes*

* Christian Doppler Laboratory for TCAD at the Institute for Microelectronics, TU Wien, Gusshausstr. 27–29, A-1040 Wien, Austria

^o IMEC, Kapeldreef 75, B-3001 Leuven, Belgium

ABSTRACT

Many recent publications discussing the stress and recovery behavior of bias temperature instability (BTI) have suggested the existence of two components contributing to the phenomenon. One of these components was found to be quickly relaxing while the other was only slowly relaxing or even permanent. Curiously, although the most likely suggested mechanisms are the generation of interface states and the capture of holes into pre-existing traps, there is no agreement on which mechanism corresponds to which component and both possibilities have been suggested. Alternatively, other groups have suggested evidence that BTI is dominated by a single mechanism, and used the reaction-diffusion (RD) model to describe the degradation. However, RD theory cannot explain the recovery and related intricacies of the phenomenon.

We present a new modeling framework based on the various possible energetic configurations of the system and tentatively assign these levels to the hydrogen binding/transport levels in an amorphous oxide. We investigate the possibility that the often observed recoverable and permanent components are in fact two facets of a single degradation mechanism proceeding as a series of steps. We finally subject the model to various experimental data (DC, AC, duty-factor, negative and positive stress, mixed stresses) which are all well reproduced by the model.

I. INTRODUCTION

After four decades of research, bias temperature instability (BTI) is still a highly puzzling phenomenon which has so far eluded our complete understanding [1–4]. Particularly intriguing are the complex degradation/recovery patterns which are observed when the gate bias is modulated [5–8]. Understanding this behavior is mandatory for any estimation of device degradation in a circuit setting.

Quite contrary to these experimental observations, most modeling approaches published so far have focused on constant gate bias stress and are remarkably oblivious to any recovery of the degradation which sets in as soon as the stress is removed. This dilemma is easily illustrated using the reaction-diffusion (RD) model [9, 10] as the most prominent example: the RD model describes the stress phase as a delicate interplay between the reaction-limited forward rate and the diffusion-limited backward rate [11]. The forward rate alone would predict a degradation directly proportional to the stress time t_s , while only in combination with the backward rate the ominous $\sim t_s^{1/6}$ time-dependence is obtained. During recovery, the backward-rate can be studied on its own, predicting recovery to occur basically within four decades of time, in stark contrast to the experimentally observed log-like recovery over twelve decades [8, 12]. Recent attempts to rectify this problem [13–16] have so far been unsatisfactory [4].

Triggered by the interest in the origin of the recovery [5, 17], the

last couple of years have seen a lively debate on whether NBTI is caused by interface states alone [3] or whether the observed degradation is a superposition of interface state generation and hole capture into pre-existing traps [1, 18–21]. In particular, charge-pumping (CP) experiments have delivered results which are quite different from degradation observed during electrical measurements. For instance, it has been observed that the CP signals show a remarkably different recovery behavior compared to conventional data. This and similar facts have been used to separate the overall degradation into interface state and hole trapping contributions.

In our recent studies [8, 12] we have demonstrated how the wealth of information present in the recovery data can be used to deepen our understanding of the phenomenon. Utilizing the universality of the recovery [12, 22] we have extracted a fast and universally recovering component which is observed on top of a permanent or very slowly recovering component.

The existence of two components appears to be consistent with a growing number of recent publications [1, 18–20]. The fast universally recovering component is often identified as hole trapping while the permanent or slowly recovering component may be caused by interface states. Curiously, however, contradictory explanations have also been given speculating that deep hole traps form the permanent component [18].

From a physical point of view these interpretations would mean that two independent degradation mechanisms act in parallel with only the overall degradation being visible. While an NBTI model may eventually be constructed using such two individual components [1, 12], one may ask if these two components are in fact *not parallel but act in series and are therefore two facets of a single degradation mechanism proceeding in steps*. Such a view is supported by the fact that the degradation of the drain current (actually, $I_D(t_s) - I_D(1\text{ms})$) in a simple on-the-fly (OTF) measurement obtained at different temperatures and voltages can be made to overlap using simple scaling (see Fig. 1 and Fig. 2). If two independent mechanisms were to contribute to NBTI, it would be highly unlikely for these mechanisms to have a similar voltage and temperature acceleration, because only then would a superposition of these two mechanisms result in such a scalability. In fact, quite different activation energies have been extracted for such two possible mechanisms [8] which appears to be inconsistent with the apparent scalability of the data. Also, PBTI and mixed negative/positive stresses (NBTI/PBTI), as shown in the following do not seem to be consistent with the hole-trapping model suggested by Tewksbury [23] and suggested for NBTI by Huard *et al.* [1], since independently of the stressing voltage, *positive* charge is created in the device (at least for the samples used in this study). However, creation of positively charged defects is consistent with donor-like (or amphoteric) states [24, 25] in the silicon

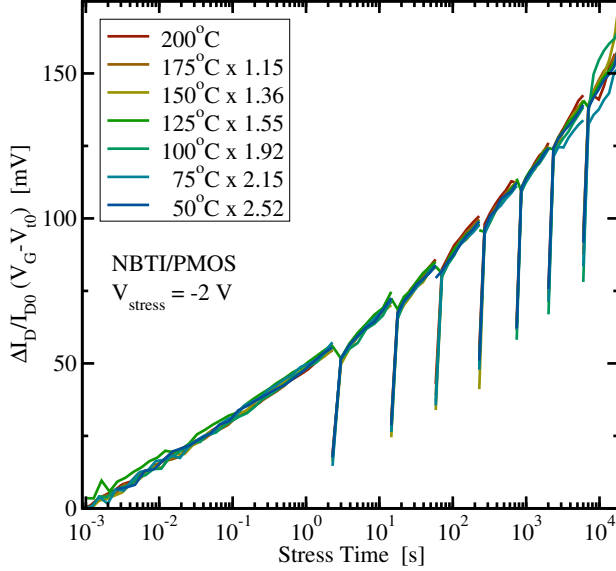


Fig. 1: The change in the drain current recorded in an on-the-fly (OTF) measurement with an initial delay of 1ms at different temperatures for an oxynitride PMOS with EOT = 1.4nm. The OTF measurement is interrupted periodically to record 12s long relaxation sequences. Data obtained at the same stress voltage (-2V) can be made to overlap by multiplication with a suitably chosen (temperature-dependent) factor, given in the legend.

bandgap [26] which are able to rapidly exchange charges with the silicon substrate and thus appear as positive charges when measured at $V_G = V_{th}$.

As previously suggested, we first direct our efforts at understanding the peculiarities of recovery [12]. A particularly noteworthy property of the recovery is its nearly logarithm-like character lasting at least twelve decades in time [8, 27]. This kind of recovery can be empirically described by various analytic expressions, including the stretched-exponential form [12], which is characteristic to an abundance of diverse physical phenomena (dielectric relaxation, voltage recovery in Leyden bottles, recovery of mechanical strain in wires, decay of luminescence in wires, etc. [28]) and commonly attributed to some kind of disorder.

One modeling approach considers such systems as having two closely coupled energetic minima, the equilibrium and a slightly higher second minimum separated by an energy barrier [29, 30]. This barrier may either be surmounted by thermal excitation or by quantum-mechanical tunneling [30, 31]. During the application of stress, the second minimum may be lowered below the equilibrium minimum, making it the preferred energetic state. Depending on the transition mechanism and the statistical distribution of these barrier energies, a time-dependent transition between the two wells is observed. Once the stress is removed, the equilibrium well will be occupied again, which commonly involves long recovery transients.

Based on these principles we will develop a new model for bias temperature instability in the following which provides an explanation for the issues raised in our previous work [4, 8, 12]. In particular, we do not just consider the stress phase for the formulation of our theory, but put equal emphasis on a correct description of the recovery

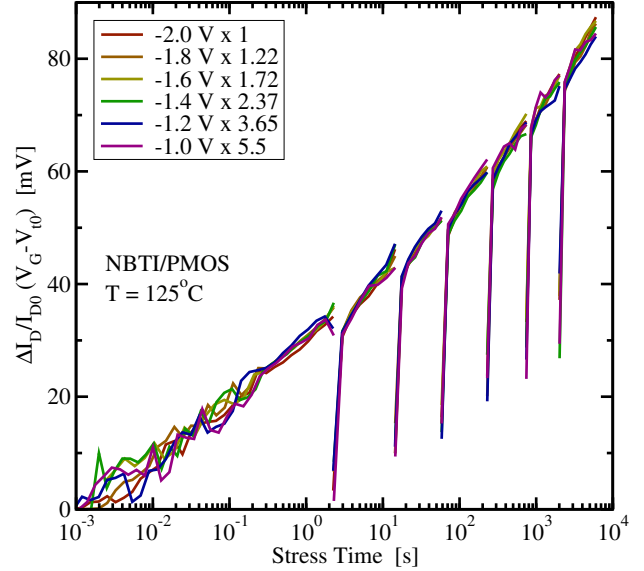


Fig. 2: Same as Fig. 1, but now at constant temperature (125°C) with varying stress voltages. Again, the data can be made to overlap. Consequently, the quantity $I_D(t_s) - I_D(1\text{ms})$ is proportional to a temperature and voltage dependent prefactor. This is not necessarily true for the 'real' $\Delta V_{th}(t_s)$. Nevertheless, this could possibly indicate the dominance of a single (complex) mechanism.

phase [5]. Consequently, the model is evaluated for DC NBTI/PBTI stress and recovery, duty factor dependence, and mixed NBTI/PBTI stresses.

II. AN ENERGY-LEVEL PERSPECTIVE

Our new model can be understood in terms of previously suggested mechanisms for the dissociation of silicon-hydrogen bonds, which creates an electrically active defect, for instance a P_b center at the SiO_2/Si interface or a K_N center inside SiON [26]. In equilibrium, that is, after the passivation step used during fabrication and before application of stress, most hydrogen atoms are bound to silicon dangling bonds, which represents the ground state. Such a scenario is schematically sketched in Fig. 3. However, this is not the case for all configurations as wells are assumed to exist where a site next to the dangling bond is energetically favored. This is a reasonable assumption because during passivation it is possible to passivate a large number of interface states (e.g. 10^{12}cm^{-2}) but not all (e.g. 10^9cm^{-2} remain).

Upon application of an electric field and at elevated temperatures the hydrogen atom may overcome the barrier to the neighboring transport state via thermal emission [30]. Alternatively, particularly at low temperatures, the hydrogen may tunnel from an excited level to the transport state. Such a mechanism has been suggested for dielectric relaxation in order to explain the linear temperature dependence of the specific heat at very low temperatures ($T < 1\text{K}$) [32]. However, as we are interested in temperatures above room temperature, we will not further consider this possibility. Furthermore, the theory we present in the following is purely classical based on the ground-state only. In a possible interpretation of the energy levels, the first transport state may be the anti-bonding (AB) site [33], which has been

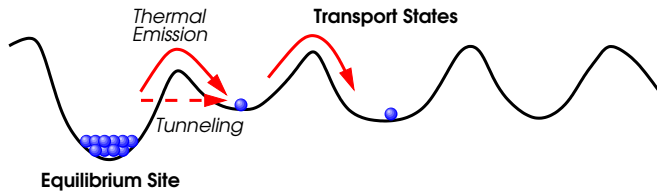


Fig. 3: The energy-levels involved during creation of a dangling bond in an amorphous network, after [34]. Hydrogen is released from a Si–H bond over a barrier into transport states (tetrahedral interstitial sites, bond center sites, etc.). Emission may occur via thermal emission or, at very low temperatures, via tunneling.

associated with electrically active trap levels in the silicon bandgap. The anti-bonding site is located behind the Si atom on the other side of the dangling bond. In order to move the hydrogen atom from the dangling bond (DB) to the AB site, lattice relaxation is required, thereby coupling the dissociation process to the silicon-silicon vibrational modes [33].

Upon capture of a carrier, the binding energy is further reduced, thereby favoring complete dissociation of the hydrogen from the silicon dangling bond. In the following we will assume this capture of carriers to happen instantaneously, that is, capture of carriers via quantum-mechanical tunneling [23] will not be assumed to be rate-limiting. Alternatively, instead of complete dissociation, when the stress is removed, the particle will go back to the ground state, with the time constant determined by the barrier height separating the first transport state from the equilibrium position. Consequently, the energy levels used in the final model may be considered 'effective energy levels', comprising more than a single effect.

Since the gate insulators we are dealing with are amorphous, the barriers separating the ground state from the transport state will be different for each silicon dangling bond. For example, with an effective maximum density of interface/oxide states of $10^{12} \dots 10^{13} \text{ cm}^{-2}$, we obtain for a device with a channel length of 100nm and a width of $1 \mu\text{m}$ a total number of possible bonds equal to $N = 10^3 \dots 10^4$, each of which with somewhat different barriers. Experimental support for such a scenario has been given by Stesmans *et al.* [35, 36] for both the transition from the ground state to the transport state and back. Also, the fact that interface states at the SiO_2/Si interface show a rather broad Gaussian distribution of energy levels in the silicon bandgap [37] may support the concept of similarly distributed hydrogen binding energies.

For a pure SiO_2 sample, it has been shown that the creation of dangling bonds (P_b centers) directly at the Si/SiO_2 interface is primarily responsible for degradation [38]. In SiON , on the other hand, it has been suggested based on spin-dependent tunneling and recombination measurements that defects are created inside the oxide but still close to the interface, so called K_N centers [26]. These defects have been suggested to also be silicon dangling bonds, but back-bonded to three nitrogen atoms rather than three silicon atoms as in the case of the P_b centers. This observation is compatible with the similar degradation kinetics observed in SiO_2 and SiON samples, by placing the common Si–H bond at the heart of the problem. Due to the amorphous nature of SiON , the possible reactions paths may occur in various orientations as shown in Fig. 4, a possibility we will make use of to explain PBTI together with mixed NBT/PBT stresses.

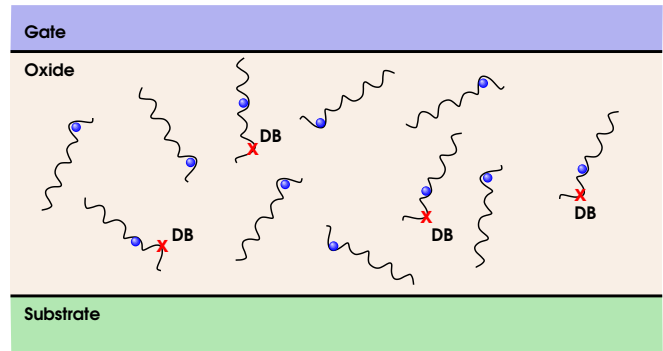


Fig. 4: Schematic distribution of hydrogen bonding sites, possibly K_N centers, in an amorphous nitrided oxide. The orientation of the reaction paths is arbitrary and in most well configurations the DB site is the energetically preferred site.

It is interesting to note that this approach is very similar to a model successfully used for the description of defect formation and annealing in hydrogenated amorphous silicon [39]. In fact, many of the experimental observations typical for this process seem to have their direct correspondence in bias temperature instabilities (stretched-exponential-like degradation and recovery, considerably better efficiency of holes compared to electrons, etc.)

In order to model the behavior of the system depicted in Fig. 4 we should include the ground state, the transition to the first transport state, the transition to the other transport states, and eventually the transition to another ground state after a certain number of 'hops'. However, to study the characteristic features of our model we try to limit the number of energy levels to the absolutely necessary minimum in the following, similarly to the approach taken in [39].

It is also important to realize that previously published NBTI models are in principle fully compatible with this physical picture. However, in the final model formulation of existing models some crucial simplifying assumptions are introduced which may not be justified.

1. The reaction-diffusion model [9–11] assumes that the transition rates between the ground state and the first transport state are in equilibrium and that the degradation dynamics are controlled by the redistribution of hydrogen between *identical* transport states (diffusive process).
2. Reaction-dispersive-diffusion models [5, 40–42] also assume that the transition rates between the ground state and the first transport state are in equilibrium. By contrast, the degradation dynamics are here controlled by the redistribution of hydrogen between energetically *different* transport states (dispersive-diffusion process). However, the coupling of the distributed transport states to a supposedly fixed ground state and fixed barrier at the interface appears to result in some inconsistencies [42].
3. The rate-limited model used by Huard *et al.* [1] assumes that the barrier from the first transport state back to the ground state is so large that practically no re-passivation of the dangling bonds occurs. Such an assumption seems to be a contradiction with the dangling bond being the equilibrium configuration. (The recovery is attributed solely to hole trapping and detrapping into pre-existing traps.)

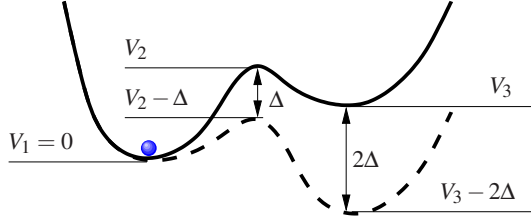


Fig. 5: A single double-well configuration and the impact of the electric field. In equilibrium, well 1 (V_1) is the energetically preferred configuration. Application of the electric field tilts the wells, in analogy to [30], favoring the second well (V_3). The system is assumed to be positively charged when in the second minimum and the total charge Q is given by the sum over all possible configurations of $(1 - f_1)$. Transitions between the wells are assumed to follow first-order kinetics via thermal activation.

III. A DOUBLE-WELL MODEL

Based on the previous considerations we first investigate a double-well model for BTS, where only the transitions from the ground state to the first transport state (possibly the anti-bonding site) are considered. This is schematically depicted in Fig. 5. Application of an electric field is assumed to lower the barrier by an amount of Δ and the second minimum by 2Δ .

Assuming that a particular well is characterized by given barriers V_1 , V_2 , and V_3 , the rate equations describing the depassivation process are then

$$\frac{\partial f_1(t)}{\partial t} = -f_1 k_{13} + f_3 k_{31}, \quad (1)$$

$$\frac{\partial f_3(t)}{\partial t} = f_1 k_{13} - f_3 k_{31} \quad (2)$$

with the rates given by [30]

$$k_{13}(V_1, V_2) = \nu \exp(-\beta(V_2 - V_1 - \Delta)), \quad (3)$$

$$k_{31}(V_2, V_3) = \nu \exp(-\beta(V_2 - V_3 + \Delta)). \quad (4)$$

The probability of the hydrogen atom being in well one is given by f_1 , the probability of it being in well 2 f_3 , while having a defect is the probability of the particle not being in well 1, thus $1 - f_1$. Furthermore, we use $1/\beta = k_B T$ and an attempt frequency $\nu \approx 10^{13} \text{ s}^{-1}$. Obviously, as only energy differences enter the rates, V_1 may be set to zero and eliminated altogether. However, since from a physical point of view each energy may follow a different distribution, the full equation system is retained. We furthermore assume that the individual energy wells are independent and that there is a single hydrogen atom in each well, that is, $f_1 + f_3 = 1$. This assumption may not be justified for unhydrogenated interfaces/oxides or devices with a large background concentration of hydrogen.

Altogether, the threshold voltage shift due to the created defects (assuming instantaneous capture and release of carriers once the defect is created or annealed) is given by the sum of all possible reaction paths and reads

$$\Delta V_{\text{th}}(t) = -\frac{q}{C_{\text{ox}}} \sum_{i=1}^N \left(1 - \Delta f_{i,1}(t)\right) \left(1 - \frac{x_i}{t_{\text{ox}}}\right), \quad (5)$$

with $\Delta f_{i,1} = f_{i,1} - f_{i,1}^{\text{eq}}$, $f_{i,1}$ the occupancy of the first well in the i -th reaction path, $f_{i,1}^{\text{eq}}$ its equilibrium value, and x_i the distance from the Si/SiO₂ interface. Assuming that most defects are still close to this interface, we will assume $x_i \approx 0$ in the following (we remark that this approximation can be easily relaxed). Each of the N double-wells is

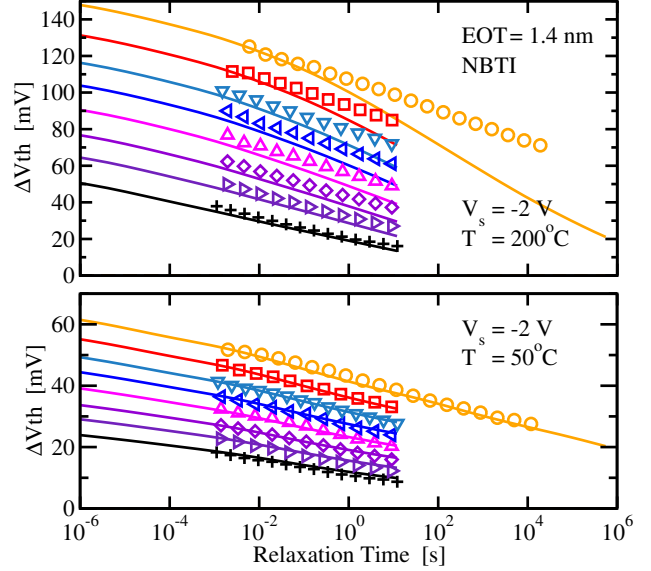


Fig. 6: Evaluation of the double-well model (lines) against pMOS recovery data (syms) at two different stress temperatures. While the accuracy at 50°C is very good, at 200°C the 'lock-in' is not reproduced, resulting in too fast relaxation.

then characterized by a unique combination of V_1 , V_2 , and V_3 , which are assumed to be given by independent Gaussian distributions [35, 36]. For computational reasons, however, we will approximate the Gaussian distribution functions using the Fermi-derivative function [1, 43]

$$g_i(V_i, V_{i,m}, \sigma_i) = \frac{1}{\sigma_i} \frac{\exp\left(\frac{V_{i,m} - V_i}{\sigma_i}\right)}{\left(1 + \exp\left(\frac{V_{i,m} - V_i}{\sigma_i}\right)\right)^2}. \quad (6)$$

Under these approximations it is easy to show that the maximum degradation obtainable in the system is given by

$$\Delta V_{\text{th}} \approx -\frac{qN}{C_{\text{ox}}} \left(\frac{1}{1 + \exp\left(\frac{V_{3,m} - (V_{1,m} + 2\Delta)}{\sigma_3}\right)} - \frac{1}{1 + \exp\left(\frac{V_{3,m} - V_{1,m}}{\sigma_3}\right)} \right). \quad (7)$$

but a complete analytical solution will be published elsewhere.

A. Double-Well Model: Relaxation Behavior

The double-well model was calibrated to relaxation data obtained on two different devices at the temperatures 50°C and 200°C, see the appendix for experimental details. The calibration to both data sets was performed simultaneously in order to extract physically feasible parameters, see Table I for the extracted values. During calibration of the double-well model it was observed that only data recorded at lower temperatures could be well described. Consequently, the weights between the two data sets during calibration were set slightly in favor for the 50°C data, for which excellent results could be obtained. However, the long logarithmic tails at higher temperatures and larger stress times could not be well reproduced, see Fig. 6. This is put down to the emerging 'permanent/slow' component

Parameter	Calibrated Value
σ_2	0.34 eV
σ_3	0.46 eV
$V_{2,m}$	2.53 eV
$V_{3,m}$	0.09 eV
$\Delta(50^\circ\text{C})$	0.60 eV
$\Delta(200^\circ\text{C})$	0.44 eV

TABLE I: The parameters extracted from a simultaneous calibration of the double-well model to relaxation data obtained at 50°C and 200°C. Since the voltage applied during relaxation was different at these temperatures (0.34V @ 50°C vs. 0.23V @ 200°C), Δ was allowed to be temperature-dependent to accommodate this difference. Interestingly, the value for the first barrier height is very close to the value given in literature for the dissociation energy of the Si-H bond (2.4 – 2.6eV [44]). The same parameter values are of course used for both the stress and relaxation phases in Fig. 6 and Fig. 7.

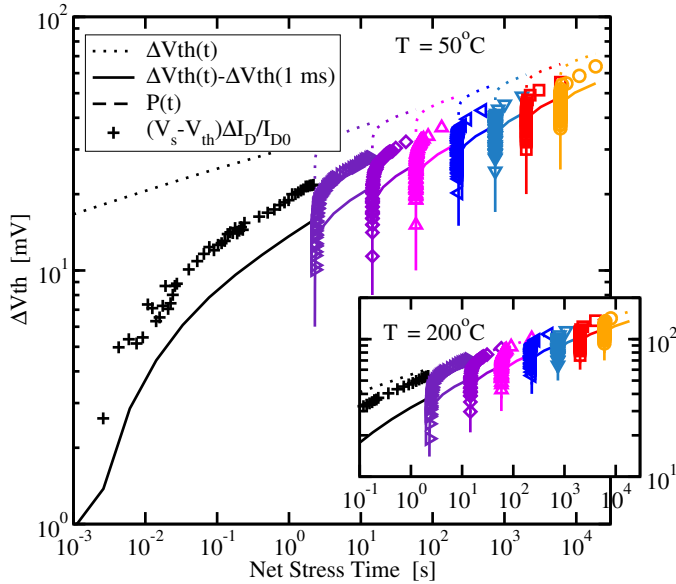


Fig. 7: Data from the same measurement as in Fig. 6 but taken in between the relaxation phases using the OTF approach. The model was calibrated against the relaxation data only. While the basic features are captured, it was not possible to calibrate for the correct amplitude (see the solid lines, which emulate the initial measurement delay of 1 ms and are too low.)

which begins to slow down the relaxation. We were not able to accommodate such a *permanent/slow component*, or, the oft-cited 'lock-in' phenomenon [17], with a double-well model.

B. Double-Well Model: Stress Behavior

A comparison of the calibrated double-well model with the data obtained during the intermediate stress phases is shown Fig. 7. Although the basic features of the stress data can be captured, the magnitude was found to be too low, which can be traced back to the fact that the double-well model is not able to maintain the logarithmic decay for enough decades and results in a too early roll-off at short relaxation times.

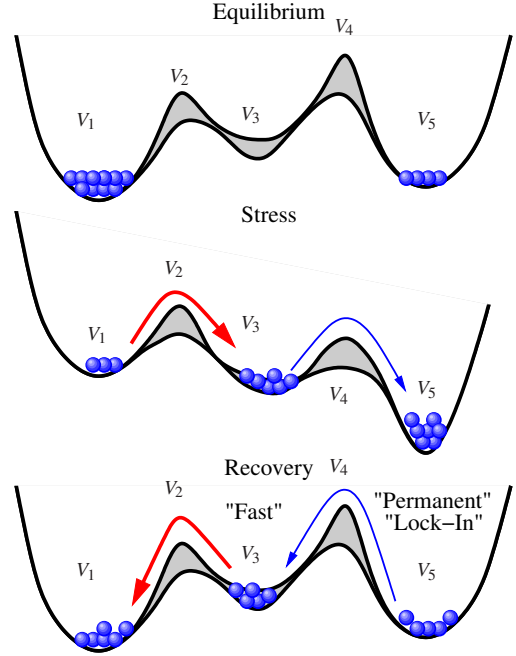


Fig. 8: Addition of the second transport state results in a triple-well model. The second well is energetically higher than the first and the third well and forms a transitional saddle point. Transitions from the second well back to the first well are fast, while the third well represents the permanent component/lock-in.

IV. A TRIPLE-WELL MODEL

As the double-well model cannot accommodate the whole range of experimentally observed features, we include the next transport level (a third well) which accounts for a *more permanent deviation* of the system from its equilibrium configuration. The third well would be separated by a different barrier and could be interpreted as the hydrogen diffusing farther away from the defect (first extension towards a multi-hop limit as suggested in [39]). Such a triple-well model is depicted in Fig. 8. Each well is now characterized by a particular combination of $V_1, V_2, V_3, V_4,$ and V_5 and we have to solve the linear equation system

$$\frac{\partial f_1}{\partial t} = -f_1 k_{13} + f_3 k_{31}, \quad (8)$$

$$\frac{\partial f_3}{\partial t} = +f_1 k_{13} - f_3 k_{31} - f_3 k_{35} + f_5 k_{53}, \quad (9)$$

$$\frac{\partial f_5}{\partial t} = +f_3 k_{35} - f_5 k_{53}. \quad (10)$$

By assuming again that wells do not interact and that there is exactly one hydrogen atom in each well we make use of

$$f_1 + f_3 + f_5 = 1. \quad (11)$$

The latter assumption seems to be justified as wells without a single hydrogen atom are dangling bonds already, contribute to the initial threshold voltage and will, at least initially, not react to changes in the electric field. By contrast, wells with more than one hydrogen atom will have larger difficulties to react to changes in the field since possible states are already occupied.

The transition rates are modeled in analogy to the double-well model, however, we allow for a different impact of the stress-field

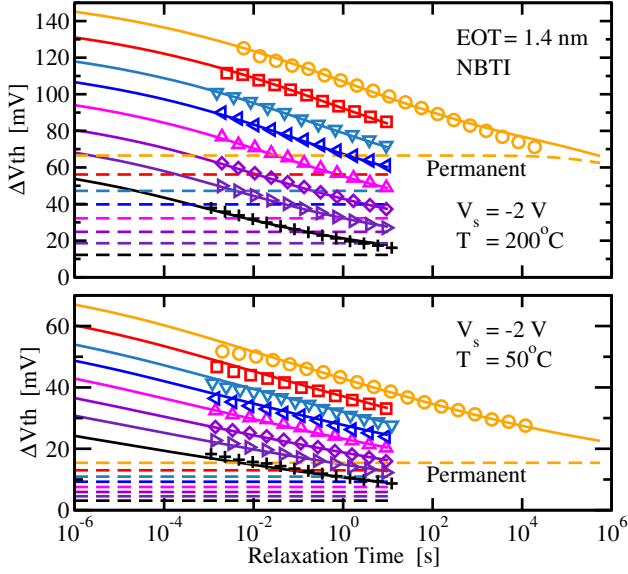


Fig. 9: Evaluation of the triple-well model (solid lines: total, dashed lines: ‘permanent’ component) against the same data as in Fig. 6. For the model evaluation the measurement sequence was simulated using fixed parameters in a single simulation run, just like the real measurement. Excellent accuracy is obtained for both temperatures during the relaxation phase which is not sensitive to mobility changes [46].

on the two wells. The physical picture behind this is that the two barriers are likely to either have a different dipole moment [30, 45], a different tunneling distance, or some other, currently unspecified, mechanisms which accelerates bond-breaking. Thus, quite generally, we write

$$k_{13} = v \exp(-\beta(V_2 - V_1 - \Delta_2)) , \quad (12)$$

$$k_{31} = v \exp(-\beta(V_2 - V_3 + \Delta_2)) , \quad (13)$$

$$k_{35} = v \exp(-\beta(V_4 - V_3 - \Delta_4)) , \quad (14)$$

$$k_{53} = v \exp(-\beta(V_4 - V_5 + \Delta_4)) . \quad (15)$$

In the following, however, we assume that the dispersion of well 1 and well 3 can be neglected, $\sigma_1 = \sigma_5 = 0$, as well as a similar response of the wells to the electric field, $\Delta_2 = \Delta_4$.

Parameter	Calibrated Value
σ_2	0.42 eV
σ_3	0.45 eV
σ_4	0.41 eV
$V_{2,m}$	2.56 eV
$V_{3,m}$	0.80 eV
$V_{4,m}$	3.00 eV
$V_{5,m}$	0.19 eV
$\Delta(50^\circ\text{C})$	0.47 eV
$\Delta(200^\circ\text{C})$	0.38 eV

TABLE II: The parameters extracted from a simultaneous calibration of the triple-well model to relaxation data obtained at 50°C and 200°C. Note that we assumed $\sigma_1 = \sigma_5 = 0$ and $\Delta_2 = \Delta_4$. The calibrated values for σ_i were very similar, $\sigma_2 \approx \sigma_3 \approx \sigma_4$, and may be substituted by a single parameter. Again, as for the double-well model, the first barrier energy $V_{2,m}$ is in excellent agreement with literature data for the dissociation energy of the Si–H bond.

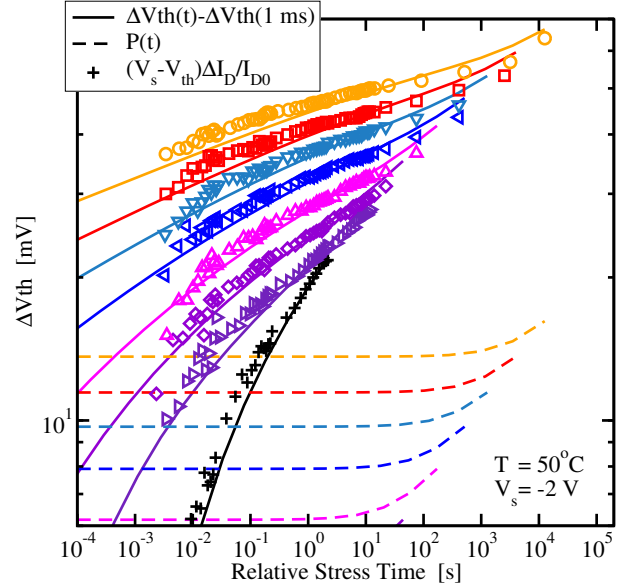
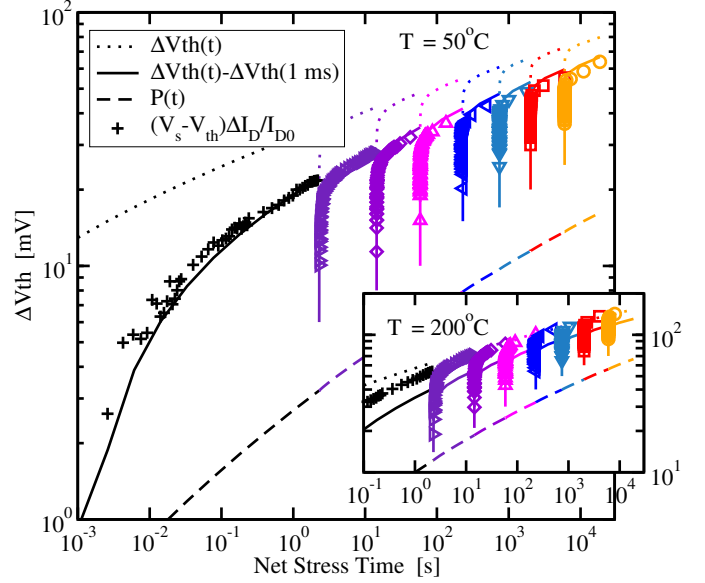


Fig. 10: Top: The threshold-voltage shift predicted by the triple-well model is also in excellent agreement during each stress phase, at least at 50°C. Also, the initial delay in the first measurement was taken into account by subtracting $\Delta V_{th}(1 \text{ ms})$ from the simulation result (shifts the dotted line to the solid line). At higher stresses (temperature or time), the agreement is not that perfect, see the inlay for 200°C. We put this down to the error in the simple OTF prediction related to mobility degradation and the compact modeling error [27, 46]. **Bottom:** The excellent agreement is confirmed by plotting the individual stress phases on a log scale with the time measured with respect to the beginning of each stress phase.

A. Triple-Well Model: Relaxation Behavior

The relaxation behavior predicted by the triple-well model is evaluated against the same data used for the double-well model (cf. Fig. 6) in Fig. 9. In contrast to the double-well model, the *triple-well model* also reproduces the *permanent component/lock-in* very well which becomes visible at larger stresses.

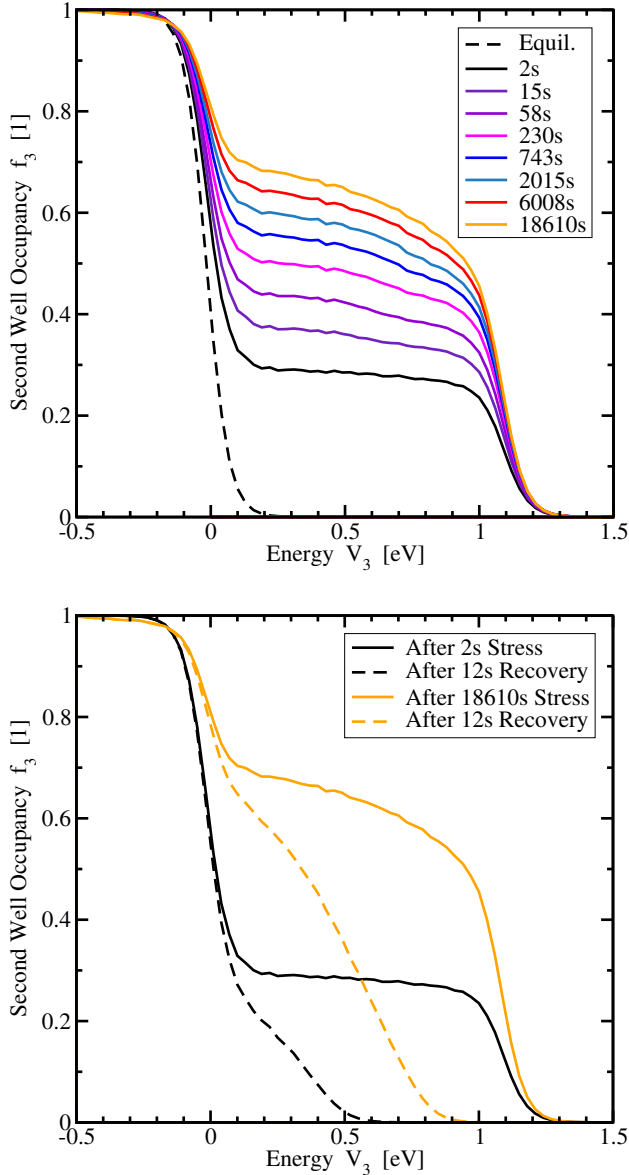


Fig. 11: Top: Occupancy of the individual states in the second well as a function of the energy after the eighth stress phase. **Bottom:** Impact of a 12s relaxation phase on the occupancy. Shown is the impact after the first and the last stress phase.

The extracted parameter values are given in Table II. Although the model appears to require a large number of parameters, it is important to realize that this is due to the complicated nature of BTI and that for example the latest refined version of the reaction-diffusion model (explicit conversion of H to H_2 , diffusion of both species, dynamic occupancy of the interface states [16]) requires an even larger number of parameters without giving comparable accuracy [4].

B. Triple-Well Model: Stress Behavior

The prediction of the triple-well model during stress is evaluated in Fig. 10 against the OTF data obtained between the relaxation sequences. Note that all data was recorded in a single measurement/stress/measurement sequence. During stress the recorded I_{Dlin} is converted to a threshold voltage shift using the simple on-the-

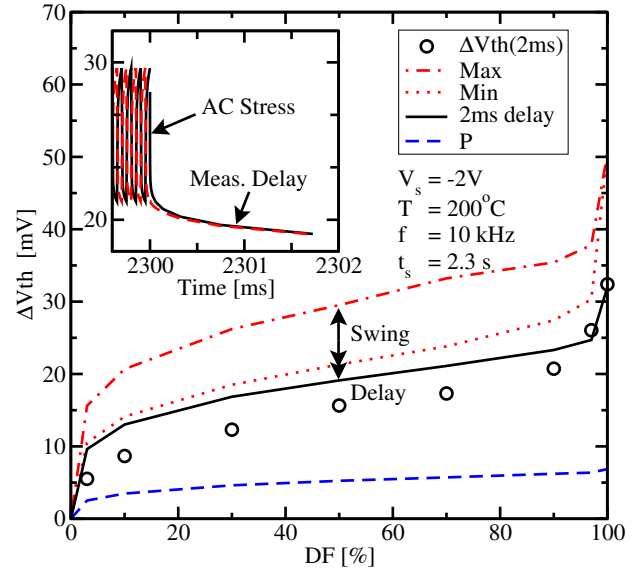


Fig. 12: The influence of the duty factor on the overall degradation is well reproduced by the triple-well model. The measured degradation is strongly influenced by the measurement delay, 2ms in our case. Since the period of the oscillation for 10kHz DF stress is much smaller than the measurement delay, the measured degradation does not depend on whether the stress is interrupted for the measurement during the on or off phase (see inlay).

fly expression, $(V_G - V_{th})\Delta I_D / I_{D0}$. Recall that this expression is only approximately accurate and prone to overestimation of the real threshold-voltage shift [46].

In particular for moderate stresses, the triple-well model gives excellent agreement during the stress phase, when the impact of the error introduced by the time-zero measurement of I_D is emulated by subtracting $\Delta V_{th}(1ms)$ from the simulation data. Quite convincingly, all characteristic features of the data are fully reproduced. For larger stresses, however, a multiplicative offset is observed (a parallel shift in a log-log plot), while the main features of the data are reproduced. We put this down to the mobility variations in the OTF data [27, 46], which have been suggested to produce such an offset for larger stress [16]. Note that ΔV_{th} during relaxation, which is also calculated from I_D using the single point I_D method [47], appears to be insensitive to these errors [46].

The occupancy of the second well as a function of the stress time is shown in Fig. 11. With increasing stress, the bottom of well 2 is filled, while particles with higher energy are more likely to be transferred to well 3 (the 'permanent' component). Also shown is the impact of a recovery phase which predominantly empties well 2, beginning with the higher-energetic particles.

C. Triple-Well Model: Duty Factor Dependence

As a next step in the evaluation of the triple-well model, its prediction of the duty-factor (DF) dependency is considered. Available measurement data [8, 21, 48] show a plateau around a DF of about 50% in addition to a strong sensitivity for duty-factors close to 0% and 100%. This issue has been notoriously difficult to model, with the first successful attempt given only recently in [21]. The conventional reaction-diffusion model, for instance, predicts a DF dependence of

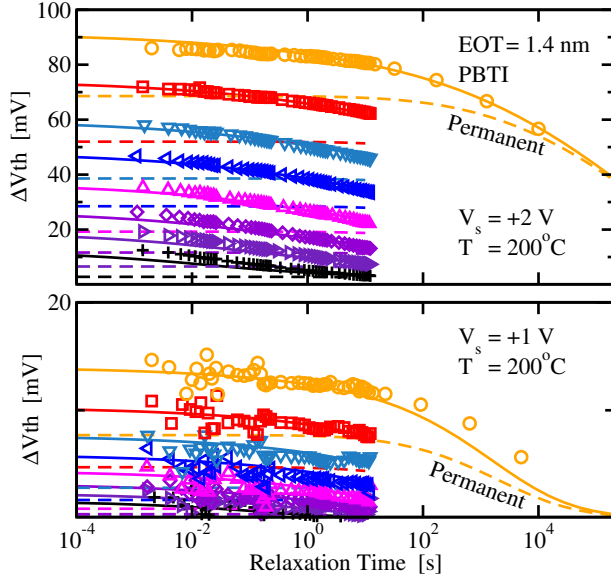


Fig. 13: The triple-well model can also be used to describe PBTI stress. However, NBTI and PBTI are apparently asymmetric in our samples, resulting in different parameters of the triple-well model in our current formulation. The two stress bias conditions can be well reproduced using a single parameter set. *Note that PBTI stress also creates positive charge and thus a negative ΔV_{th} , just like NBTI.*

the form $DF^{1/3}$ [4] and thus no sensitivity for a DF approaching 100%. In particular, RD theory predicts an AC/DC ratio of 80%, see [49], in contrast to the nearly universally observed ratio about 50% [8, 21, 48, 50, 51]. In Fig. 12 we show how the triple-well model can be used to characterize a *duty-factor dependent stress*. While the agreement is not as perfect as for the relaxation and stress data, the triple-well model correctly predicts the plateau, and the other features of the data.

D. Triple-Well Model: PBTI Stress

In addition to NBTI, the triple-well model can also be used to describe *PBTI*, see Fig. 13, which, remarkably, also generates *positive charge*. However, the degradation and relaxation is not completely symmetric to NBTI [8] in our 1.4 nm SiON samples. In a first attempt, this can be described using different energy barriers in the triple-well model. Whether this is an artifact of the abstract formulation of the barriers where the microscopic origin is left unidentified (the substrate being in inversion or accumulation does so far not enter the model) or a feature of the interfacial layer remains to be investigated.

E. Triple-Well Model: Mixed NBTI/PBTI Stress

Our current interpretation of the asymmetry relies on the assumption that during PBT stress the same kind of defect is created. For this we assume that the defects responsible for degradation exist with reaction coordinates in two orientations, one parallel and one anti-parallel to the applied electric field, see also Fig. 4. Then, upon application of a positive field to the triple-wells of Fig. 8, the NBTI-wells would not dramatically change their occupancy of well 1. On the other hand, the wells oriented the other way, the PBTI-wells, would start to degrade.

Such an assumption is confirmed when looking at the behavior of a sample first subjected to NBT stress followed by an analogous PBT

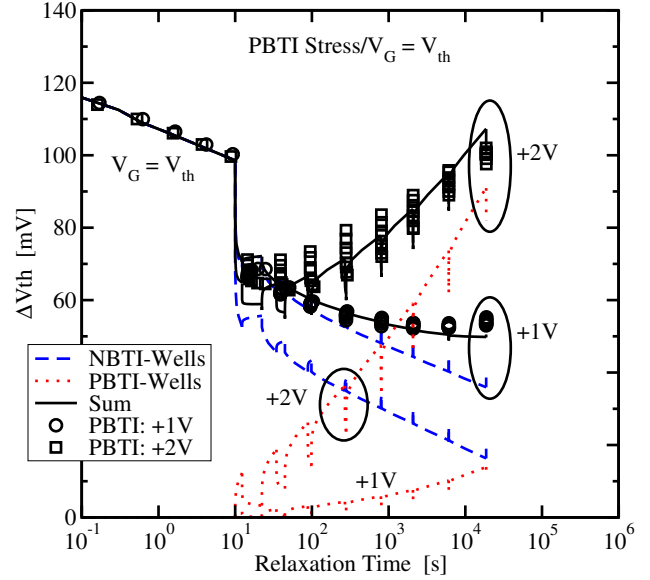


Fig. 14: Application of a PBTI stress sequence during the relaxation phase after NBTI stress leads to hitherto unexplainable relaxation patterns (data: syms). Using the NBTI-wells calibrated for NBTI stress (Fig. 10) and the PBTI-wells calibrated for PBTI stress (Fig. 13), these patterns can be straightforwardly explained using superposition (solid lines). In particular, the model explains why large positive biases result in stronger degradation rather than complete recovery as expected from an electron injection model.

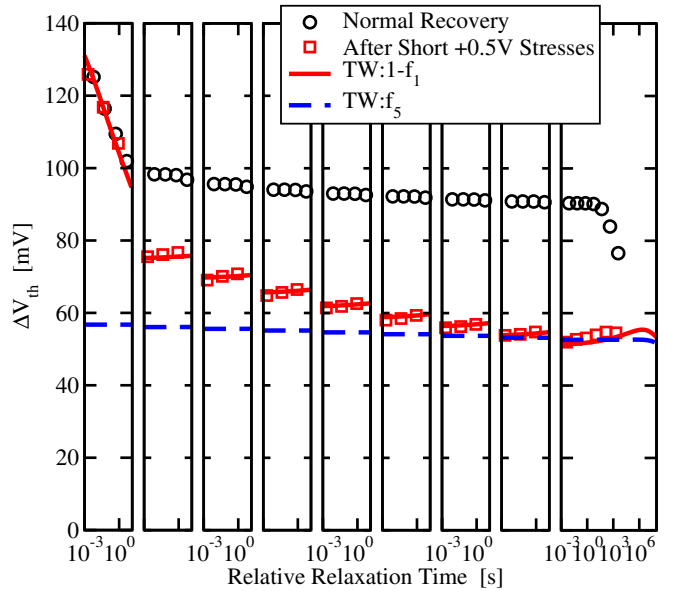


Fig. 15: Detailed view of the recovery after application of small positive bias pulses in between the relaxation sequences. The difference to the normal recovery at $V_G \approx V_{th}$ is obvious. While the positive bias is too small to trigger a significant response of the PBTI-wells, the occupancy of well 2 of the NBTI-wells is lowered below its equilibrium value, resulting in *degradation* during the 'recovery' phase. The permanent component (that is, the occupancy of well 3, f_5), is given by the dashed line and is insensitive to the small positive bias. Consequently, during CP measurements only f_5 would be visible.

