Characterization and Physical Modeling of the Temporal Evolution of Near-Interfacial States Resulting from NBTI/PBTI Stress in nMOS/pMOS Transistors

T. Grasser,* B. Stampfer,* M. Waltl,* G. Rzepa,* K. Rupp,* F. Schanovsky,‡ G. Pobegen,† K. Puschkarsky,• H. Reisinger,• B. O'Sullivan,° and B. Kaczer°

*Institute for Microelectronics, TU Wien, Austria ‡ Global TCAD Solutions, Austria
† KAI, Villach, Austria • Infineon, Munich, Germany ° imec, Leuven, Belgium

***Abstract*— The last decade of BTI research has seen a frantic search for ultra-fast measurement methods to correctly understand the impact of fast as-grown traps which impact the initial phase (1 ks) of the degradation and recovery. These methods focus mostly on determining the time-dependence of the threshold voltage shift. Other experimental methods able to resolve the energetic distribution of traps in the bandgap have recently not been as frequently employed as they introduce a large delay and also require switching the device into accumulation, which considerably accelerates recovery. Here we use detailed CV measurements to study NBTI/PBTI in SiO_2 nMOS/pMOS capacitors. We extract a unique defect band inside the SiO_2 insulator which can describe the build-up of near-interfacial states over time in all four combinations using our recently suggested gate-sided hydrogen release model. Our results suggest that depending on the transistor type and stress bias conditions, the generated slowly-recovering near-interfacial states are due to a combination of slower oxide traps and faster P_b centers.**

I. Introduction

Due to the rapid onset of recovery well below the microseconds regime and the fact that the accumulated degradation can be heavily perturbed by abrupt changes in the gate-bias, experimental challenges have been an important aspect of BTI research [1–6]. Conventionally used $\Delta V_{\mathrm{th}}(t)$ measurements sense the charge stored in the oxide and the interfacial layer for the surface potential close to the silicon valence band (in pMOS transistors) or close to the conduction band (in nMOS transistors). As such, these measurements are insensitive to the details of the charge distribution inside the bandgap and – as a consequence – a considerable amount of physically relevant information is missed. This includes for instance the flat-band voltage shift $\Delta V_{\mathrm{FB}}(t)$ and the details of the time evolution of the energetic profile of near-interfacial traps in the bandgap, $D_{\mathrm{it}}(E,t)$, as for instance provided by CV measurements.

During the early days of bias temperature instability (BTI) research, CV measurements had been an integral part of the characterization procedure [7, 8]. It appears that together with the realization that BTI recovery – and in particular the accelerated recovery in accumulation – can heavily perturb measured threshold voltage shifts, the wealth of information contained in CV data [9] has recently received a limited amount of attention in favor of ultra-fast characterization methodologies [1, 10, 11]. However, ΔV_{th} measurements only provide information on the threshold voltage and are virtually completely oblivious to the energetic profiles of the responsible defects. Compared to CP and DCIV measurements, CV measurements appear to provide the best energetic resolution with a minimum amount of interference: While CP measurements require rapid changes between inversion and accumulation with incompletely understood consequences on the accumulated degradation [12], DCIV measurements [13, 14] are most sensitive to defects in the middle of the bandgap – which are the most efficient recombination centers – and the sensitivity of the method decreases exponentially with increasing distance from midgap.

Thus, in order to better understand the temporal change in the distributions of interfacial and oxide traps, we take a closer look at what can be learned from CV data and correlate the results with what has been learned during the last decade. This comparison is insofar of fundamental importance, as published models have focused on the evolution of $\Delta V_{\mathrm{th}}(t)$ and only qualitatively considered the evolution of the average of $D_{\mathrm{it}}(t) = \int D_{\mathrm{it}}(E,t)\Delta E$, while ignoring its energetic distribution. A word of caution is required here to clearly define the termini *interface* and *oxide states* as used in this work:

- **Experimental:** CV measurements are sensitive to all defects fast enough to respond to the AC frequency of the test signal, f_{AC}. These defects produce a true capacitive addition to the oxide capacitance, C_{it}. Also, CV measurements are sensitive to defects/charges in the oxide and at the interface, ΔV, which result in a stretch-out of the CV curve. Experimentally, however, it appears impossible to clearly separate ΔV and C_{it}. For instance, as will be elaborated below, the temporal change in the capacitance defined as $\Delta C(t) = C(t) - C(0)$ will always contain a *mixture* of both, ΔV and C_{it}.
- **Physical:** From a physical perspective, both interface defects (like P_b centers [15]) and defects deeper in the oxide (such as hydroxyl E'centers [16–18]) can contribute to both ΔV and C_{it}, depending on the AC and the sweep frequency. Again, it is very difficult to clearly separate these defects using electrical measurements only. As will be shown, the conventional assumption that C_{it} in SiO_2/Si devices – such as those used in the present study – is solely due to P_b centers is inconsistent with the presented data and the known electrical responses of P_b centers [19]. While P_b centers are definitely created during BTI stress [20], it has been repeatedly suggested that they are neither the only nor the dominant defects responsible for C_{it} and that a hydrogen-containing oxide defect is likely responsible [16, 21–25]. These arguments are consistent with our previous work [26–28] on the gate-sided hydrogen release mechanism (GSHR) [29].

- **Modeling:** In our gate-sided hydrogen release (GSHR) model, we will model the temporal evolution of interface defects, D_{I}, as well as the temporal evolution of oxide defects, D_{O}. We deliberately refer to these components with the unusual names D_{I} and D_{O} rather than using the conventional terms D_{it} and D_{ot}, since their experimental determination is very difficult and a multitude of definitions has been used. Compared to the experiment, the theoretical assignment is clear and both can contribute to ΔV and C_{it}, depending on the experimental conditions.

In the following we will put forward experimental and theoretical evidence supporting the hydrogen release mechanism. One of the strengths of the GSHR model is that it provides a natural and fully consistent explanation both for the creation of oxide and interfacial states, which contribute to both C_{it} as well as ΔV. For NBTI these defects are created near the channel by the release of neutralized protons from the gate side. For PBTI/pMOS, protons are released at the channel-side and also lead to the creation interface states as well as of defects at the gate side, which, however, are too far away from the channel to contribute to C_{it} and only impact ΔV. Finally, due to their energetic alignment, no protons are neutralized during PBTI/nMOS, which leads to negligible degradation. It is particularly this asymmetry that leads us to argue that studying *all four combinations*, N/PBTI in n/pMOS, is required to reveal the true physics of BTI degradation [30, 31], even though from an industrial perspective only the most relevant combination NBTI/pMOS is typically considered.

II. Experimental

Recent results have established that BTI essentially follows the same dynamics in production-quality SiO_2, SiON, and high-K devices [32–34], all of which seem to be dominated by the Si/SiO_2 interfacial region. Since our goal is the understanding of the fundamentals of BTI rather than any particular latest technology, we use large-area ($50\,\mu\mathrm{m} \times 50\,\mu\mathrm{m}$) pMOS and nMOS capacitors on $\langle 100 \rangle$ Si with p- and n-poly gates, respectively. To avoid excessive leakage currents, which would interfere with our CV measurements, we chose relatively thick SiO_2 capacitors with $t_{\mathrm{ox}} = 5\,\mathrm{nm}$. To assure comparability between the different measurement frequencies, the frequency was varied from 1 kHz to 1 MHz in the *inner loop* to avoid recovery induced by bias changes [35]. In the outer loop, the voltage was swept from $-2\,\mathrm{V}$ to $2\,\mathrm{V}$. Otherwise, if for instance the low-frequency CV curves had been measured first followed by the higher frequencies, the highest frequency CV curve would have seen a considerably amount of recovery which could be mistaken for a frequency dependence.

III. CV Considerations

Before we begin to analyze our CV data, we investigate the expected response using theoretical models. Regrettably, the extraction of interface state profiles, $D_{\mathrm{it}}(E)$, from CV data is notoriously difficult and various methods have been suggested [36], none of which is free of artifacts: for example, a high-frequency reference curve measured for instance at 1 MHz still contains the response of interface states while

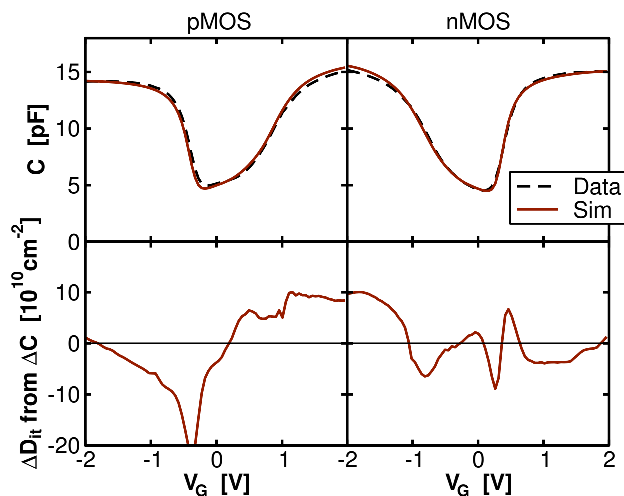


Fig. 1. **Top**: Comparison of measured (dashed) and simulated CV (solid line) for pMOS (left) and nMOS (right). **Bottom**: Despite the use of a sophisticated TCAD model, the difference ΔC between measurement and simulation is on the order of $10^{11}\,\mathrm{cm}^{-2}$ when converted to an effective interface density of states, which is on the same order of magnitude as the degradation built up in our experiments.

it is extremely challenging to create precise theoretical CV reference curves for devices with non-trivial doping profiles, non-ideal interfaces and oxides. This problem is illustrated in Fig. 1, in which a TCAD simulator was calibrated to measured CV curves of the MOSCAPs used in this study. Data were recorded in the temperature range 125 – 175°C and the 150°C fits together with the difference between data and simulation, ΔC, converted to an effective density of interface states ($\Delta D_{\mathrm{it}} = \Delta C / q_0 A$, with A being the device area and q_0 the elementary charge), are shown. Despite the use of sophisticated TCAD models (Fermi-Dirac statistics, quantum correction models [37, 38], non-uniform doping profiles, poly-depletion, temperature-dependent band-edges, effective masses, density-of-states and thermal velocities), the difference ΔC is on the order of $10^{11}\,\mathrm{cm}^{-2}$ when converted to an effective interface density of states. While the fits of the CV curves are apparently very good, this effective D_{it} is of a similar magnitude as the recorded drifts in this work and would thus severely compromise the accuracy of the extracted D_{it} profiles.

Due to the reasons stated above, we do not attempt to extract $D_{\mathrm{it}}(t)$ but rather restrict out analysis to the change of D_{it} over time, $\Delta D_{\mathrm{it}}(t) = D_{\mathrm{it}}(t) - D_{\mathrm{it}}(0)$. In the simplest case, $\Delta D_{\mathrm{it}}(t)$ is obtained from $\Delta C = C(t) - C(0)$, as has been done previously [9]. However, it is important to realize that the shift ΔC is not necessarily due to the true capacitive contribution caused by interface states but can contain a large spurious component caused by the bias-dependent stretch-out $\Delta V(V_{\mathrm{G}})$.

This problem is elaborated using the calibrated TCAD model for the nMOS and pMOS capacitors. We start with Fig. 2, which shows ΔC resulting from a rigid shift of the CV curve by $\Delta V = -30\,\mathrm{mV}$, caused e.g. by fixed positive charges. Thus, in this simplest case $\Delta C(V_{\mathrm{G}}) = C(V_{\mathrm{G}} - 30\,\mathrm{mV}) - C(V_{\mathrm{G}})$. This value of ΔV was chosen to be similar to the ΔV observed in our experiments. As can be seen, a spurious $D_{\mathrm{it}} \approx 2 \times 10^{11}\,\mathrm{cm}^{-2}$ is obtained, which has of course nothing to do

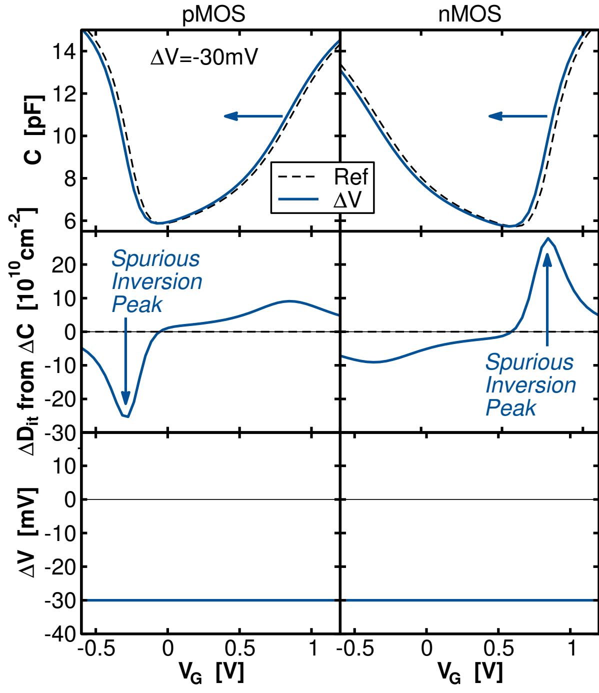


Fig. 2. **Top**: Simulated change in $C(V_G)$ resulting from a rigid shift by $\Delta V = -30\,\mathrm{mV}$. **Middle**: The D_{it} extracted from $\Delta C(V_G) = C(V_G + \Delta V) - C(V_G)$ is completely spurious and a direct consequence of the shift ΔV. Particularly when dC/dV_G is large, e.g. in inversion, very large peaks are obtained. **Bottom**: The trivial ΔV.

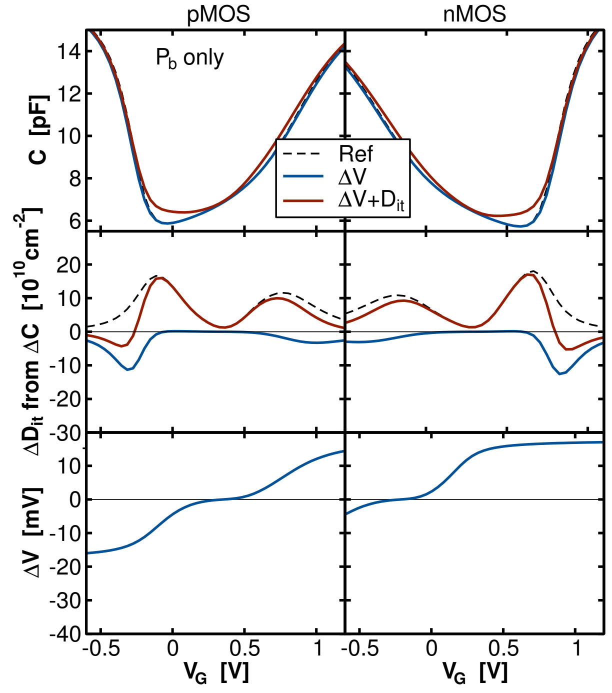


Fig. 3. **Top**: Simulated change in C for a P_b center density of about $15 \times 10^{10}\,\mathrm{cm}^{-2}$ at $f_{AC} = 1\,\mathrm{MHz}$ and $S = 1\,\mathrm{V/s}$. In addition to the stretch-out caused by ΔV (solid blue line), the capacitance shows a true increase by ΔC_{it} (solid red) compared to the reference curve (dashed). **Middle**: The D_{it} extracted from ΔC (solid red) is only little affected by the change in D_{it} caused solely by the stretch-out ΔV (solid blue), e.g. at very high f_{AC}. The dashed black line is the true ΔC_{it} put into the simulator. **Bottom**: P_b centers are positively charged below midgap and negatively charged above, the integral of which determines ΔV.

with the creation of interface states. On the other hand, Fig. 3 shows the resulting contribution of P_b centers to D_{it} assuming a density of $15 \times 10^{10}\,\mathrm{cm}^{-2}$, also in the range of what will be measured later. For this value of D_{it}, a maximum $\Delta V = -35\,\mathrm{mV}$ is obtained in our 5 nm oxides, provided all defects can follow the CV sweep rate S. The parameters of the characteristic double humps of the P_b centers were taken to match the data of [19], with the normally distributed +/0 transition levels at $0.235/0.125\,\mathrm{eV}$ (mean/variance) and the 0/- transition at $0.8/0.1\,\mathrm{eV}$. These values were obtained on 2.9 nm thick SiO_2 grown on $\langle 100 \rangle$ wafers, which should be quite similar to our devices.

In this simulation study it is possible to separate the impact of the true capacitive contribution C_{it} and the spurious impact of the stretch-out ΔV. For example, for a reasonably slow sweep rate S (around $1\,\mathrm{V/s}$) and a very high AC frequency f_{AC} outside feasible experimental ranges, only ΔV will be relevant and C_{it} will be zero. As can be seen in the middle panel of Fig. 3, the shape of the characteristic humps is hardly affected by the stretch-out ΔV. In a number of publications it has been observed that the number of created fast states, D_{it}, appear to be in a 1:1 relation with fixed oxide charges [7, 8]. A similar conclusion was drawn by Huard *et al.* [39], who observed a strong shift in V_{th} while no shift in V_{FB} was seen. Such a scenario is theoretically investigated in Fig. 4. However, the impact of the additional Q_{ox} on ΔC appears to be small as the energetic position of the created D_{it} is rather close to the band-edges.

Finally, typical CV/GV signals produced by oxide defects modeled via a two-state non-radiative multiphonon (NMP)

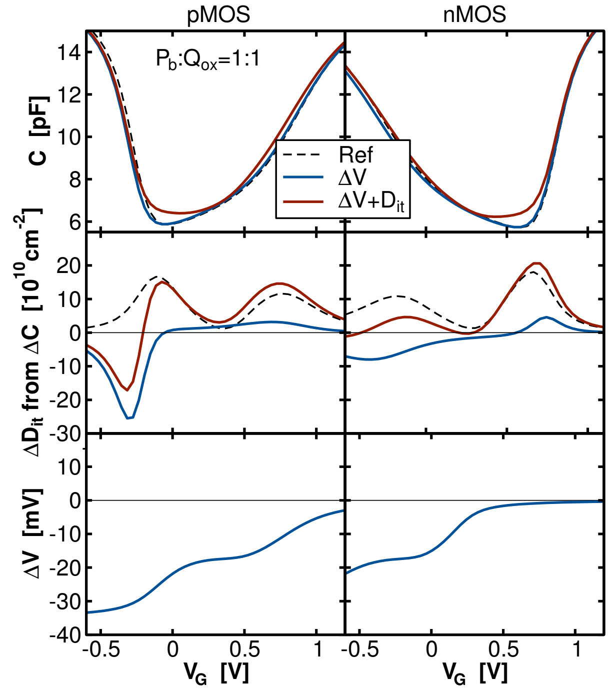


Fig. 4. **Top**: Like Fig. 3 but assuming that P_b centers and fixed positive charges are created in a 1:1 ratio. **Middle**: Due to the larger amount of charges below midgap, the D_{it} extracted from ΔC is more strongly affected by the stretch-out ΔV in the lower half of the bandgap. **Bottom**: For each P_b center, a positive charge is added, resulting in zero net charge for large positive V_G, thereby rigidly shifting the ΔV curve of Fig. 3.

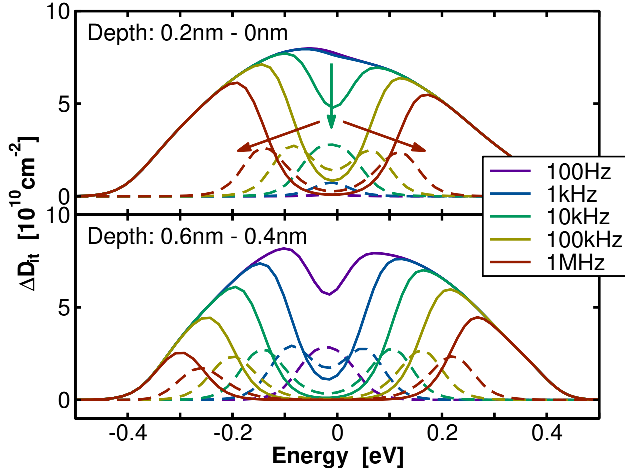


Fig. 5. Theoretical ΔD_{it} from a two-state NMP defect normally distributed in energy around midgap with $\sigma = 0.2\,\mathrm{eV}$ and over a certain depth into the oxide of a pMOS: right at the interface (top) and 0.5 nm into the oxide (bottom). Solid lines are CV and dashed lines GV. Note the similarity to the symmetric NBTI/nMOS case of Fig. 8.

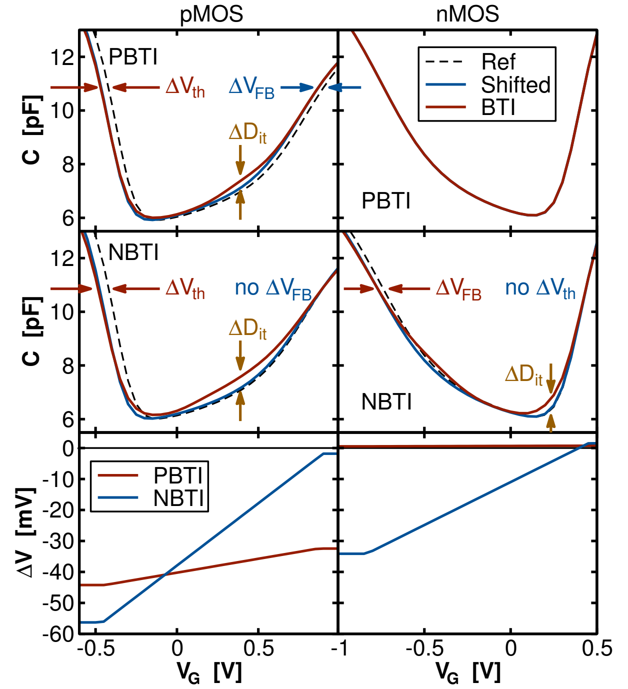


Fig. 6. The reference CV (dashed line, measured at 100 kHz) corrected by the stretch-out (solid blue), approximated by ΔV_{th} and ΔV_{FB} and linearly interpolated in-between. The difference is converted to ΔD_{it}. NBTI ($-5.5\,\mathrm{V}/125°\mathrm{C}/7.5\,\mathrm{ks}$, middle row) results in similar changes in both n/pMOS while PBTI ($+5.5\,\mathrm{V}/125°\mathrm{C}/7.5\,\mathrm{ks}$, top row) leads to positive charging (pMOS) but has no impact on the nMOS. The approximate ΔV is shown in the bottom row.

model are shown in Fig. 5 for a broad normally-distributed defect band directly at and slightly displaced from the channel, which already shows some resemblance with the present data shown in Section IV. Notably, defects around midgap have the largest time constants and can thus only follow lower CV frequencies. In particular, this slow response of midgap states explains the characteristic "dip" in the data at typically used higher frequencies, which makes this broad distribution split into two peaks, quite reminiscent of the well-known peaks of P_{b} centers shown in Figs. 3 and 4.

IV. Results

Based on these theoretical considerations we can investigate the impact of NBTI and PBTI stress on the CV characteristics of both our p- and nMOSCAPs. As can be seen in Fig. 6, BTI stress has a marked impact on the CV curves in three out of the four cases considered:

- PBTI on pMOS capacitors leads mostly to the creation of positive charges visible as similar ΔV_{th} and ΔV_{FB} shifts plus a small contribution in D_{it}.
- Quite to the contrary, PBTI on nMOS capacitors does neither lead to an appreciable amount of positive charges nor a change in D_{it} as both ΔV_{th} and ΔV_{FB} are very small.
- NBTI, on the other hand, leads to comparable shifts in ΔV_{th} and ΔV_{FB} for both n- and pMOS capacitors when one considers that they are given by $\Delta V_{\mathrm{v}} = \Delta V(E_{\mathrm{v}})$ and $\Delta V_{\mathrm{c}} = \Delta V(E_{\mathrm{c}})$, which are invariant to the transistor type and correspond to ΔV_{th} and ΔV_{FB} in a pMOS and vice versa for an nMOS.

In order to approximately separate ΔC caused by the stretch-out ΔV and the "true" contribution due to fast states, we follow a pragmatic approach and compare a CV measured after stress, $C(V_{\mathrm{G}})$, to a reference curve taken before stress, $C_0(V_{\mathrm{G}} + \Delta V)$. Irrespective of the precautions taken, extraction of ΔD_{it} and ΔV will always be prone to errors and should be used with care. In the following, we will merely use this separation for visualizing the changes in CV data, while for matching the data with the model the full ΔC curves are taken. This is done since particularly for technologically relevant drifts the changes would otherwise be difficult to visualize, see Fig. 6, which already uses relatively high stress voltages ($\pm 5.5\,\mathrm{V}$). In order to visualize ΔD_{it}, we will use a pragmatic ansatz for ΔV by first approximately extracting ΔV_{v} and ΔV_{c}. Then,

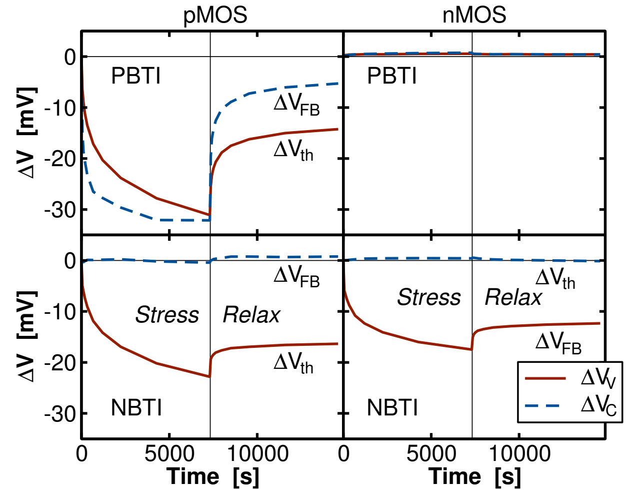


Fig. 7. Temporal evolution of ΔV_{c} and ΔV_{v} for NBTI/PBTI (stress at $\pm 4.5\,\mathrm{V}/125°\mathrm{C}/7.5\,\mathrm{ks}$ and recovery at $0\,\mathrm{V}/125°\mathrm{C}/7.5\,\mathrm{ks}$). The corresponding conventional notation ΔV_{th} and ΔV_{FB} is also given.

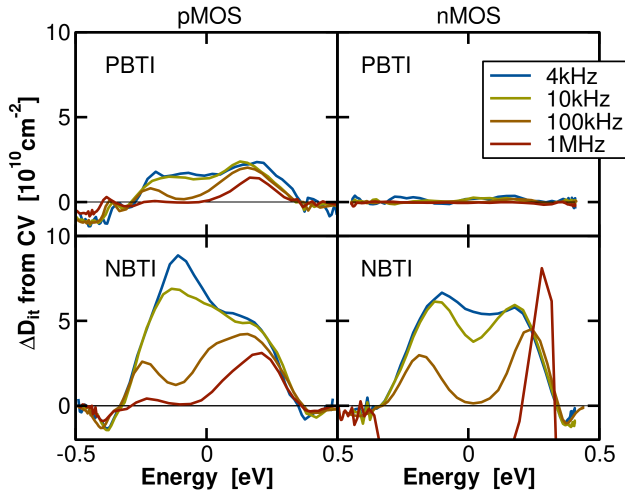


Fig. 8. Extracted ΔD_{it} profiles for the four cases using the simple correction scheme explained in Fig. 6. NBTI results in similar ΔD_{it} for nMOS and pMOS, while PBTI in pMOS produces mostly positive charge. Conversely, PBTI on nMOS has little impact on ΔD_{it}.

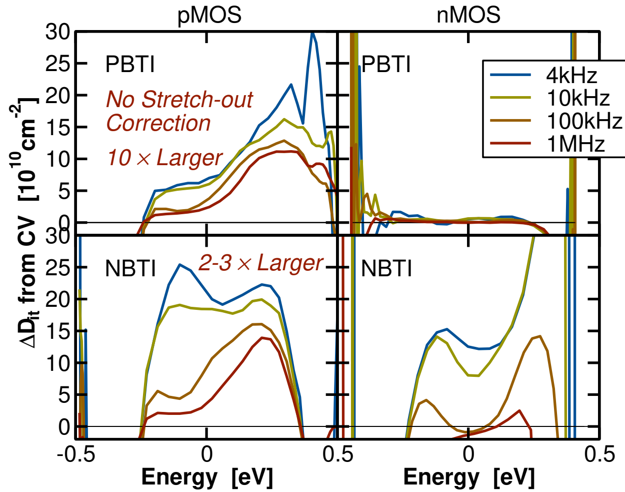


Fig. 9. Spurious ΔD_{it} resulting from simple ΔC extraction without correction for the stretch-out. While for NBTI ΔD_{it} can be 2-3×larger, a 10×larger ΔD_{it} is obtained for PBTI/pMOS.

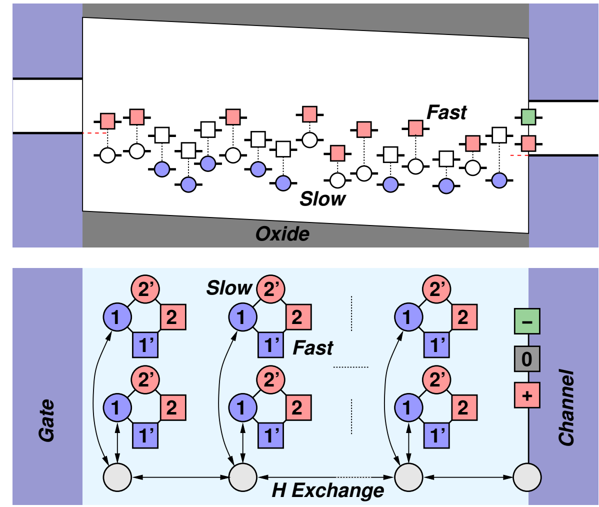


Fig. 10. A schematic view of the H-release model using 4-state NMP defects (H-complexes). In the primary configuration (◯) the trap level is lower than in the distorted secondary configuration (☐). From state 1, the H can be released, leaving an inactive precursor site behind. The released H can either bind to an empty precursor to create a new active defect or depassivate a P_{b} center at the interface, which can then be either positive, neutral, or negative, depending on the Fermi level position.

a corrected reference curve is constructed by mapping the original CV curve onto these points and stretching it linearly in-between, see Fig. 6 (bottom). This corresponds to assuming a constant defect density between E_{c} and E_{v} (as a function of V_{G} and thus only approximately constant as a function of the surface potential) and any deviation in the real D_{it} will result in a spurious ΔD_{it}.

Extracted ΔD_{it} profiles for the considered four cases are shown in Fig. 8. NBTI creates the largest ΔD_{it} in both nMOS and pMOS, while PBTI creates positive charges in pMOS but hardly affects the nMOS. The impact of the stretch-out is shown in Fig. 9. If neglected, a simple ΔC will show spurious peaks (typically negative in the lower half of the bandgap and positive in the upper-half), which can be easily confused with ΔD_{it}. While the NBTI cases result in a 2-3 times overestimation, the massive amount of positive charge for PBTI/pMOS gives nearly a 10 times large spurious ΔD_{it} if the stretch-out is not considered.

Typically it is *implicitly* assumed that ΔD_{it} is due to the creation of P_{b} centers. In contrast, ΔD_{it} peaks have been occasionally observed to move between the lower and upper half of the bandgap [40], an observation sometimes explained by suggesting the involvement of different defect types evolving with different kinetics. When comparing the measured CV peaks to what is expected from P_{b} centers, it is quite obvious that P_{b} centers alone cannot explain the data: As shown in Figs. 3 and 4, P_{b} centers form peaks in the lower and upper half of the bandgap which grow in unison, the energetic location of which show little frequency dependence [41]. It has been previously suggested that – although ESR-active P_{b} centers are also created – the *dominant* defect is ESR-inactive and possibly hydrogen related [42].

V. Theoretical Modeling

We extend our previously suggested gate-sided hydrogen-release model [26–28] for NBTI to describe the present data. In the first formulation of this mechanism we only modeled ΔV_{th} and the precise location of the defects inside the bandgap was not considered. For this purpose we used a simple two-state NMP model for the hydrogen-related defects, most likely hydroxyl E' centers, which were allowed to release H from their neutral state. For modeling CV data, the more complicated nature of the defects has to be taken into account: TDDS studies have clearly shown that oxide defects have at least four active states and their transitions are described by two defect levels [17, 43]: the slow transitions between states 1 and 2' typically have an energy level close to E_{v}(Si), while the faster transitions between states 2 and 1' (the "switching trap level") are inside the Si bandgap and dominate the CV response.

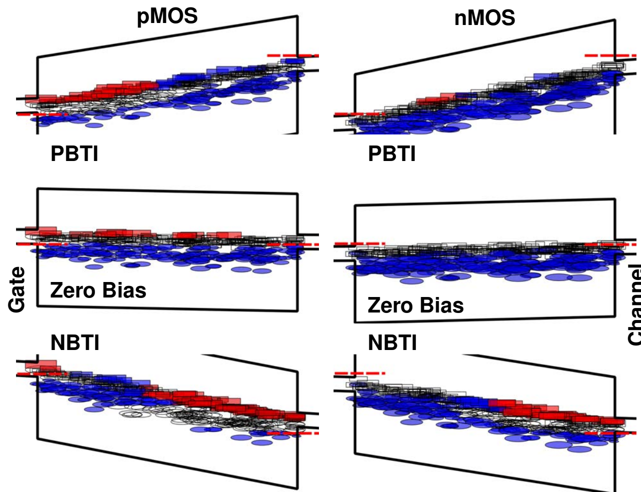


Fig. 11. Creation of charge and defects for the four cases. In the middle row the initial zero bias case is shown. Neutralized defects (blue) can release hydrogen, which preferentially moves to regions where it can create a positive defect (red). The circles correspond to the primary configuration (states 1 and 2') while the squares represent the secondary configuration (states 2 and 1'). Empty symbols are precursor defect sites without a trapped H.

These four-state defects are now coupled by a hydrogen exchange mechanism via the neutral state 1, see Fig. 10. Application of negative/positive bias for the nMOS and pMOS results in charge transitions at those defect sites, see Fig. 11, as well as H release from neutralized sites. H then quickly migrates to newly available sites which support a positive – and thus preferred – defect site, see Fig. 12. In addition, the released H can easily depassivate SiH bonds to create P_b centers [42].

In a first attempt to extend the GSRH model from two-state to four-state defects, we expressed the problem as a complete state machine with hydrogen now hopping between the four defect states of each defect and the neutral interstitial position. However, this formulation results in a very large equation system and severe numerical difficulties related to the vastly different time scales (TS) implied by the individual processes:

- TS1 Hydrogen diffusion is extremely fast, leading to extremely fast H redistribution. For instance, using established parameters for H diffusion [44], at 125°C interstitial H would hop to a next neighbor separated by 5 Å in a nanosecond.
- TS2 Charge trapping and detrapping in the secondary configuration of the four-state defects (2 $\leftrightarrow$ 1') occurs over a wide range of time scales. This begins with very fast transitions being able to contribute to D_{it} for defects very close to the interface (with time constants in the milliseconds to microseconds range) and ends at slower defects deeper in the oxide or those with larger relaxation energies (seconds).
- TS3 Activation of defects via a transition from the primary to the secondary configuration (1 $\rightarrow$ 2) is relatively slow, with time constants in the microseconds to kiloseconds regime.
- TS4 Finally, release of H from the defects proceeds over a barrier in range of 1–1.5 eV, leading to time constants well above the kiloseconds regime.

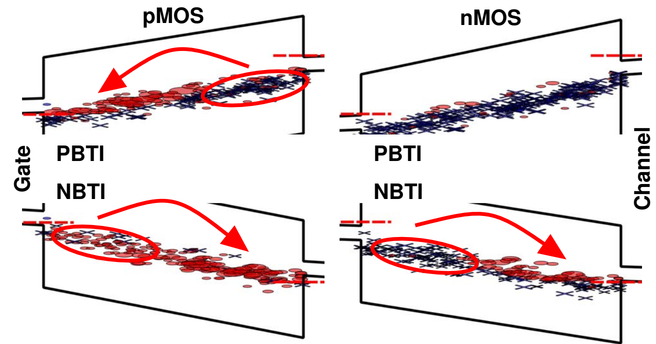


Fig. 12. Depending on the bias scenario and the transistor type, H relocates from different regions. For NBTI/pMOS, H comes predominantly directly from the gate and moves to a relatively wide region near the channel, explaining the asymmetric frequency dependence. For NBTI/nMOS, the H comes from a relatively broad region on the gate-sided half of the oxide and moves closer to the channel interface. For PBTI/pMOS mostly positively charged defects are created while for PBTI/nMOS no H relocates. The crosses mark the locations where H is lost while the red circles mark the newly activated defect sites relative to the initial equilibrium solution.

As a consequence of the various time scales, the resulting equation system was found to be extremely stiff and as such ill-conditioned and virtually unsolvable even with sophisticated quad-precision solvers such as PETSc [45]. In order to obtain a tractable model, the following simplifications were introduced:

- Hopping of interstitial H is so fast that no spatial distribution of H needs to be considered. As such, a simple H reservoir can be used which couples with the individual defects. This approximation is equivalent to assuming quasi-equilibrium at time scale TS1.
- While not perfectly accurate, the time scales TS2/TS3 and TS4 were separated by assuming the transitions 0 $\leftrightarrow$ 1 to be the slowest.

Using the above two simplifications, a relatively simple model is obtained which is nearly trivial to solve (compared to the original problem), as it consists only of a large number of independent 4×4 linear equation systems (for which a closed form solution exists), which feed into a single non-linear equation with one unknown, the amount of interstitial H^0: Based on a given density of oxide traps (or precursors, depending on the availability of H), oxide traps are activated. Their (virtual) occupancy is dynamically calculated using an implicit Euler algorithm as done previously for four-state oxide traps [46], leading to the occupancies in the states $f_1, f_{2'}, f_2$, and $f_{1'}$.

Contrary to our previous version of the model, which considered the number of H atoms in each state, we now introduce the total number of trapped hydrogens at each "oxide site", $H_{\text{T},i}$. An "oxide site" could for instance be one of the N-member rings in the SiO_2 matrix but are here introduced for numerical convenience so that not every H trapping site needs to be considered individually. Each "oxide site" can trap $H_{\text{T,max}}$ hydrogens by exchanging H with the H^0 reservoir from state 1. Using the previously calculated occupancies for

the current time-step, the number of positively charged defects in state 2 can then be obtained as $f_2 \times H_{\mathrm{T},i}$.

The dynamics of each $H_{\mathrm{T},i}$ are then given by the following ordinary differential equation

$$\frac{\mathrm{d}H_{\mathrm{T},i}}{\mathrm{d}t} = -k_{10,i}\, f_{1,i}\, H_{\mathrm{T},i} + k_{01,i}\, \mathrm{H}^0 \left(H_{\mathrm{T,max}} - H_{\mathrm{T},i}\right). \quad (1)$$

Note that each $H_{\mathrm{T},i}$ at each time-step can be calculated independently of all the others. Using an implicit Euler scheme, the above can be easily solved for a time-step Δt to give

$$H_{\mathrm{T},i} = \frac{aH_{\mathrm{T},i}^{\mathrm{old}} + k_{01,i}\, \mathrm{H}^0\, H_{\mathrm{T,max}}}{a + k_{10,i}\, f_{1,i} + k_{01,i}\, \mathrm{H}^0} \quad (2)$$

with $a = 1/\Delta t$. Since the total number of hydrogens in the system must stay constant (H_{tot}) at every time-step, we obtain the following equation to express this hydrogen conservation

$$\mathrm{H}^0 + \sum_i H_{\mathrm{T},i} = H_{\mathrm{tot}} \quad (3)$$

which via substitution of eq. (2) results in

$$\mathrm{H}^0 + \sum_i \frac{aH_{\mathrm{T},i}^{\mathrm{old}} + k_{01,i}\, \mathrm{H}^0\, H_{\mathrm{T,max}}}{a + k_{10,i}\, f_{1,i} + k_{01,i}\, \mathrm{H}^0} = H_{\mathrm{tot}}. \quad (4)$$

At every time-step eq. (4) is a simple non-linear equation with one unknown (H^0, the number of interstitial hydrogens) which was found to be very stable and converge quickly within a few iterations using a Newton algorithm. With H^0 available, the individual $H_{\mathrm{T},i}$ can be calculated using eq. (2). With $H_{\mathrm{T},i}$ available, the contribution of each trapping site to the stretch-out ΔV can be calculated by considering its positive charge in the states 2 and 2'. Thus, one obtains

$$\Delta V = \frac{q_0 t_{\mathrm{ox}}}{A\varepsilon} \sum_i H_{\mathrm{T},i} \left(f_2 + f_{2'}\right) \left(1 - \frac{x}{t_{\mathrm{ox}}}\right), \quad (5)$$

with x being the distance of the trap from the interface and ε the permittivity of SiO_2. Analogously, the individual contributions to the total capacitance as well as the conductance at a particular bias can be obtained as

$$\Delta C_{\mathrm{it},i} = \frac{\Delta C_{\mathrm{it},i}^0}{1 + \left(\frac{\omega}{k_i}\right)^2}, \quad (6)$$

$$\frac{\Delta G_{\mathrm{it},i}}{\omega} = \Delta C_{\mathrm{it},i}\, \frac{\omega}{k_i}, \quad (7)$$

with the auxiliary quantities

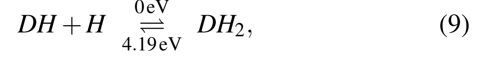
$$\Delta C_{\mathrm{it},i}^0 = q_0\, H_{\mathrm{T},i}\, \frac{k'_{21',i} f_{2,i} - k'_{1'2,i} f_{1',i}}{k_i},$$
$$k_i = k_{1'2,i} + k_{21',i},$$
$$k'_{lm,i} = \frac{\mathrm{d}k_{lm,i}}{\mathrm{d}V_{\mathrm{G}}},$$

and the angular frequency $\omega = 2\pi f$. Hereby it was assumed that only the fast transitions in the secondary configuration ($2 \leftrightarrow 1'$) contribute to the capacitance.

Before proceeding to the calculation of the interface states, we remark on an important approximation employed above: in order to keep the model tractable, it was assumed that a defect can either exist as a precursor, D_0, or as an active defect, DH. Activation of the defect requires the capture of a hydrogen via the reversible reaction

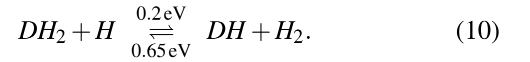
$$D_0 + H \underset{k_{10}}{\overset{k_{01}}{\rightleftharpoons}} DH. \quad (8)$$

As outlined previously using DFT calculations [47], however, an active defect can capture another H and become inactive via



$$DH + H \underset{4.19\,\mathrm{eV}}{\overset{0\,\mathrm{eV}}{\rightleftharpoons}} DH_2, \quad (9)$$

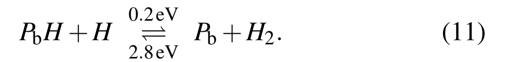
where the average barrier is given above/below the arrows for the case of the hydroxyl E' center. At a first glance this reaction implies that provided sufficient H is available, all precursors will be passivated and inactive, DH_2. However, when DH_2 encounters a third hydrogen, the defect is re-activated via



$$DH_2 + H \underset{0.65\,\mathrm{eV}}{\overset{0.2\,\mathrm{eV}}{\rightleftharpoons}} DH + H_2. \quad (10)$$

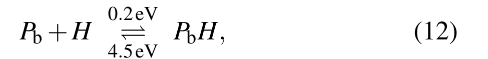
As a consequence, the actual concentration of active defects, DH will depend on the complicated interplay of H, H_2, D, $H_{\mathrm{T,max}}$, and the solubility of H_2 in the surrounding material layers. While the barriers for the various reactions have been estimated using DFT calculations in [47], the prefactors remain unknown, making the correct consideration of the above reactions challenging. To keep the model complexity within bounds, all H_2 related reactions have therefore been neglected in the derivation of the model.

In addition to the rearrangement of H inside the oxide, which leads to the creation of oxide defects, D_{O}, the highly reactive H can also depassivate P_{b} centers: Once the H is released inside the oxide, the creation of interface states can be calculated via



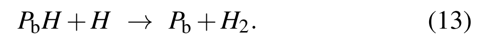
$$P_{\mathrm{b}}H + H \underset{2.8\,\mathrm{eV}}{\overset{0.2\,\mathrm{eV}}{\rightleftharpoons}} P_{\mathrm{b}} + H_2. \quad (11)$$

Conversely, depassivated dangling bonds will also be easily passivated again via the released H via



$$P_{\mathrm{b}} + H \underset{4.5\,\mathrm{eV}}{\overset{0.2\,\mathrm{eV}}{\rightleftharpoons}} P_{\mathrm{b}}H, \quad (12)$$

making the whole process extremely complicated, in particular since both P_{b} and D compete for the released H. At the present stage of understanding, we keep the model tractable by only considering the forward reaction of eq. (11)



$$P_{\mathrm{b}}H + H \;\rightarrow\; P_{\mathrm{b}} + H_2. \quad (13)$$

Furthermore, we assume that the density of P_{b} centers is relatively small and thus the competition for H is negligible. Then, the concentration of P_{b} centers can be calculated independently of D, resulting basically in the reaction-limited model proposed by Huard *et al.* [31] with a forward rate only.

Finally, the model is calibrated to the experimental data in Fig. 13. Shown are all four cases of NBTI/PBTI in pMOS and nMOS using the same universal defects bands. As discussed above, the model was calibrated to ΔC to include both the contribution of the stretch-out as well as the creation of fast states. Excellent agreement with experimental data is obtained for the build-up of ΔD_{it} near the interface, as well as the

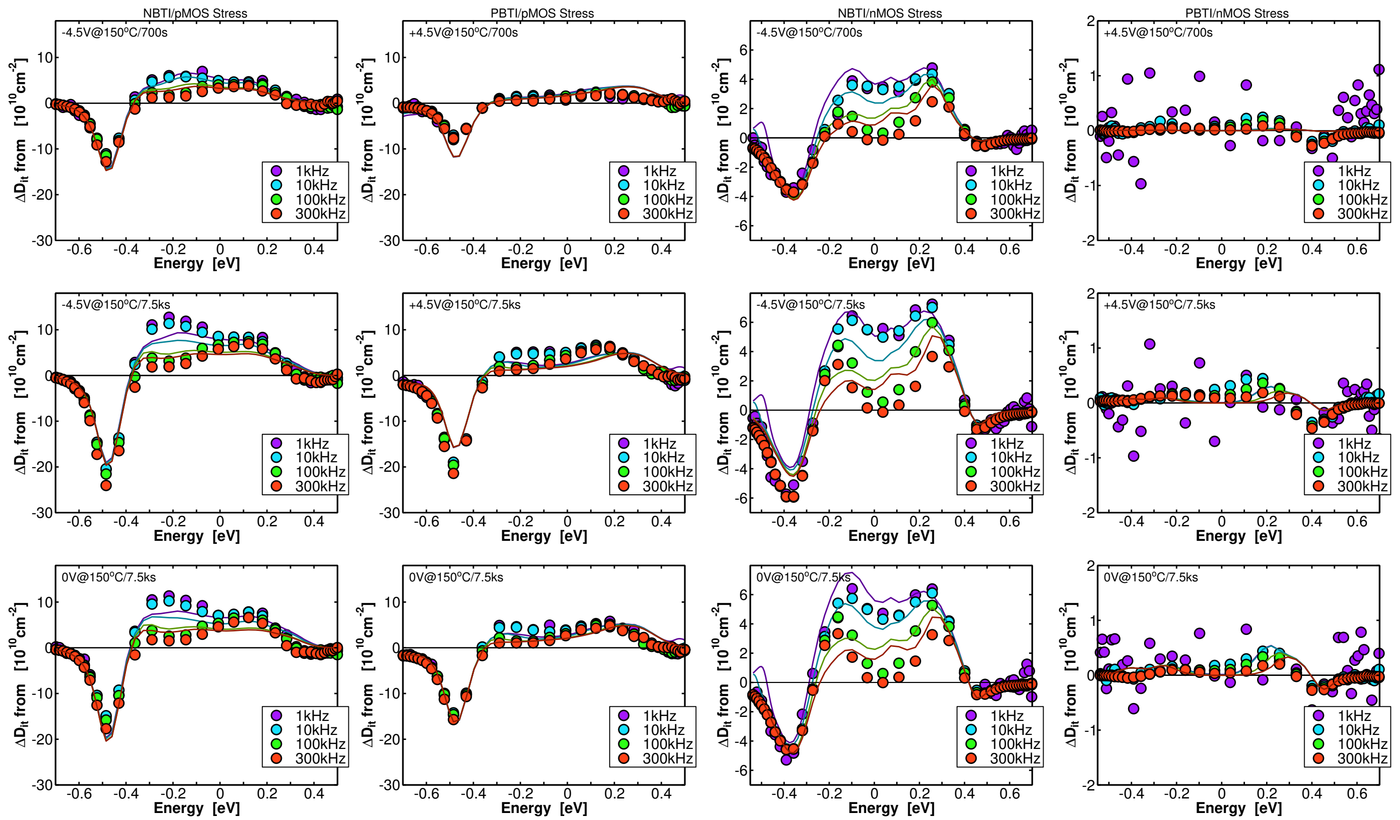


Fig. 13. Comparison of measured and simulated ΔD_{it} profiles from ΔC measurements for the four combinations of NBTI/PBTI in pMOS/nMOS during both two stress times (top, middle) and after a 7.5 ks recovery (bottom). Note the zoomed y-axis scales for the two nMOS cases to better bring out the details. The gate-sided hydrogen-release model captures both the build-up of ΔD_{it} over time as well as the energetic location and the asymmetric frequency dependence. The noise visible particularly for the NBTI/nMOS case is due to the stochastic solution method employed.

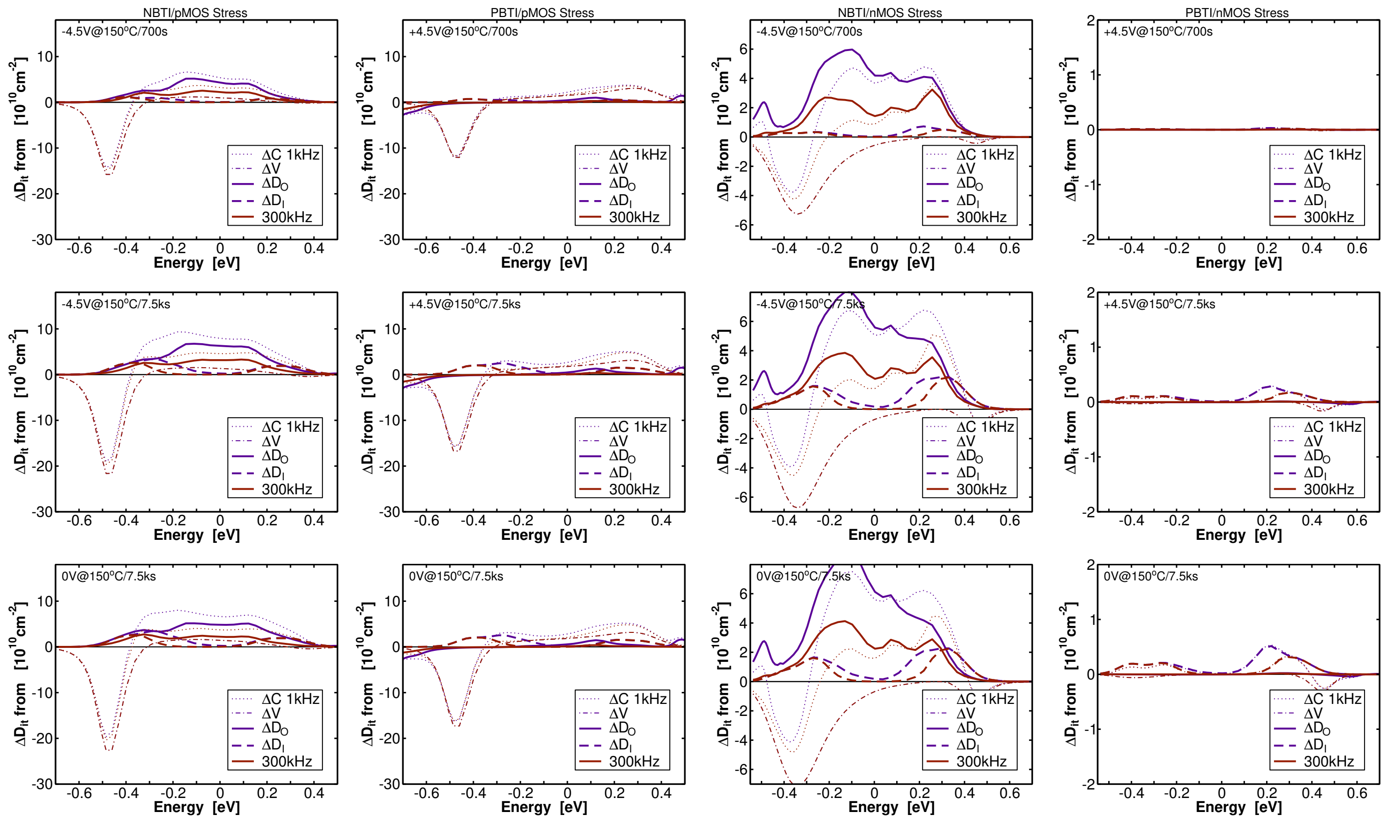


Fig. 14. The details of the simulation of Fig. 13 showing how the three components, ΔV, ΔC_{it} and ΔC_{ot} contribute to the overall observed ΔC and thus ΔD_{it}. Again note the zoomed y-axis scales for the two nMOS cases to better bring out the details.

peculiar frequency dependence: While NBTI/nMOS results in a narrow depth distribution close to the channel interface, which is symmetric in its frequency dependence in the lower and upper half of the bandgap, NBTI/pMOS produces a deeper distribution that results in an asymmetric frequency dependence.

As can be clearly seen by the large negative peaks in ΔD_{it}, ΔC contains a significant *spurious* ΔD_{it} contribution which is not due to fast states but merely due to the stretch-out ΔV. This stretch-out is due to D_O, that is, defects which are too slow to contribute to ΔD_{it} and thus primarily act as positive charge. The individual contributions are shown in Fig. 14. The contributions to the "real" D_{it} is strongly frequency dependent and ΔC_{it} only dominant at higher frequencies above 100 kHz. Only for the PBTI/pMOS case, D_{it} is mostly due to P_b-centers.

VI. Conclusions

We have studied slow NBTI/PBTI degradation in nMOS/pMOS capacitors using detailed frequency-dependent CV measurements. Using our previously suggested gate-sided hydrogen release model we explain a wide range of features observed in the data. Consistent with previous results, our data confirm that while hints of the well-known energetic distribution of P_b centers are visible in our data, P_b centers can neither account for the majority of ΔD_{it} nor the even larger stretch-out, particularly at lower CV frequencies. In addition, we argue that there is a large contribution of oxide defects that dominates the permanent degradation after BTI. We demonstrate for the first time that the unique defect band of these *spatially distributed hydrogen-related defects* can fully explain the degradation for all four stress combinations. It is also this spatial distribution which is responsible for the observed strong frequency dependence. Contrary to NBTI in both pMOS and nMOS, where hydrogen is released at the gate-side of the SiO_2, during PBTI/pMOS the neutralization of positive defects at the channel side leads to the release of hydrogen there. Finally, for the PBTI/nMOS case, only a small number of hydrogen-related defects are neutralized and little hydrogen is released, leading to a very small overall degradation. In addition, the highly reactive hydrogen released during the neutralization of oxide defects can consistently explain the creation of P_b centers, which has so far remained a mystery.

Acknowledgments

Stimulating discussions with Prof. Alex. Shluger, Prof. Valeri Afanas'ev, Prof. Andre Stesmans, Prof. Jian Zhang, Prof. Luca Larcher, Dr. Jason Campbell, and Dr. Andrea Padovani are gratefully acknowledged. The research leading to these results has received funding from the Austrian Science Fund (FWF) projects n°26382-N30 and n°29119.

References

[1] A. Islam, E. N. Kumar, H. Das, S. Purawat, V. Maheta, H. Aono, E. Murakami, S. Mahapatra, and M. Alam, "Theory and Practice of On-the-fly and Ultra-fast V_T Measurements for NBTI Degradation: Challenges and Opportunities," in *Proc. Intl.Electron Devices Meeting (IEDM)*, 2007, pp. 1–4.

[2] 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.

[3] J. Zhang, Z. Ji, M. Chang, B. Kaczer, and G. Groeseneken, "Real V_{th} Instability of pMOSFETs Under Practical Operation Conditions," in *Proc. Intl.Electron Devices Meeting (IEDM)*, 2007, pp. 817–820.

[4] T. Grasser, P.-J. Wagner, P. 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.

[5] R. Gao, Z. Ji, S. M. Hatta, J. F. Zhang, J. Franco, B. Kaczer, W. Zhang, M. Duan, S. D. Gendt, D. Linten, G. Groeseneken, J. Bi, and M. Liu, "Predictive As-grown-Generation (A-G) Model for BTI-induced Device/Circuit Level Variations in Nanoscale Technology Nodes," in *2016 IEEE International Electron Devices Meeting (IEDM)*, Dec 2016, pp. 31.4.1–31.4.4.

[6] R. Gao, Z. Ji, A. B. Manut, J. F. Zhang, J. Franco, S. W. M. Hatta, W. D. Zhang, B. Kaczer, D. Linten, and G. Groeseneken, "NBTI-Generated Defects in Nanoscaled Devices: Fast Characterization Methodology and Modeling," *IEEE Transactions on Electron Devices*, vol. 64, no. 10, pp. 4011–4017, Oct 2017.

[7] B. Deal, M. Sklar, A. S. Grove, and E. H. Snow, "Characteristics of the Surface-State Charge (Qss) of Thermally Oxidized Silicon," *J.Electrochem.Soc.*, vol. 114, no. 3, p. 266, 1967.

[8] K. Jeppson and C. Svensson, "Negative Bias Stress of MOS Devices at High Electric Fields and Degradation of MNOS Devices," *J.Appl.Phys.*, vol. 48, no. 5, pp. 2004–2014, 1977.

[9] A. Krishnan, C. Chancellor, S. Chakravarthi, P. Nicollian, V. Reddy, A. Varghese, R. Khamankar, and S. Krishnan, "Material Dependence of Hydrogen Diffusion: Implications for NBTI Degradation," in *Proc. Intl.Electron Devices Meeting (IEDM)*, 2005, pp. 688–691.

[10] 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. Intl.Rel.Phys.Symp. (IRPS)*, 2006, pp. 448–453.

[11] B. Kaczer, T. Grasser, P. Roussel, J. Martin-Martinez, R. O'Connor, B. O'Sullivan, and G. Groeseneken, "Ubiquitous Relaxation in BTI Stressing – New Evaluation and Insights," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, 2008, pp. 20–27.

[12] T. Grasser, T. Aichinger, G. Pobegen, H. Reisinger, P.-J. Wagner, J. Franco, M. Nelhiebel, and B. Kaczer, "The 'Permanent' Component of NBTI: Composition and Annealing," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, Apr. 2011, pp. 605–613.

[13] A. Neugroschel, C.-T. Sah, K. Han, M. Carroll, T. Nishida, J. Kavalieros, and Y. Lu, "Direct-Current Measurements of Oxide and Interface Traps on Oxidized Silicon," *IEEE Trans.Electron Devices*, vol. 42, no. 9, pp. 1657–1662, 1995.

[14] A. Neugroschel, G. Bersuker, R. Choi, C. Cochrane, P. Lenahan, D. Heh, C. Young, C. Kang, B. Lee, and R. Jammy, "An Accurate Lifetime Analysis Methodology Incorporating Governing NBTI Mechanisms in High-k/SiO2 Gate Stacks," in *Proc. Intl.Electron Devices Meeting (IEDM)*, 2006, pp. 1–4.

[15] P. Lenahan and J. Conley Jr., "What Can Electron Paramagnetic Resonance Tell Us about the Si/SiO_2 System?" *J.Vac.Sci.Technol.B*, vol. 16, no. 4, pp. 2134–2153, 1998.

[16] J. de Nijs, K. Druijf, V. Afanas'ev, E. van der Drift, and P. Balk, "Hydrogen Induced Donor-Type Si/SiO_2 Interface States," *Appl.Phys.Lett.*, vol. 65, no. 19, pp. 2428–2430, 1994.

[17] T. Grasser, W. Goes, Y. Wimmer, F. Schanovsky, G. Rzepa, M. Waltl, K. Rott, H. Reisinger, V. Afanas'ev, A. Stesmans, A. El-Sayed, and A. Shluger, "On the Microscopic Structure of Hole Traps in pMOSFETs," in *Proc. Intl.Electron Devices Meeting (IEDM)*, Dec. 2014.

[18] W. Goes, Y. Wimmer, A.-M. El-Sayed, G. Rzepa, M. Jech, A. Shluger, and T. Grasser, "Identification of Oxide Defects in Semiconductor Devices: A Systematic Approach Linking DFT to Rate Equations and Experimental Evidence," *Microelectronics Reliability*, vol. 82, 2018.

[19] L. Ragnarsson and P. Lundgren, "Electrical Characterization of P_b Centers in (100)Si/SiO_2 Structures: The Influence of Surface Potential on Passivation During Post Metallization Anneal," *J.Appl.Phys.*, vol. 88, no. 2, pp. 938–942, 2000.

[20] J. Campbell, P. Lenahan, C. Cochrane, A. 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.

[21] E. Cartier, J. Stathis, and D. Buchanan, "Passivation and Depassivation of Silicon Dangling Bonds at the Si(111)/SiO_2 Interface by Atomic Hydrogen," *Appl.Phys.Lett.*, vol. 63, no. 11, pp. 1510–1512, 1993.

[22] E. Cartier and J. Stathis, "Atomic Hydrogen-Induced Degradation of the Si/SiO_2 Structure," *Microelectronic Engineering*, vol. 28, no. 1-4, pp. 3–10, 1995.

[23] V. Afanas'ev and A. Stesmans, "Hydrogen-Induced Valence Alternation State at SiO_2 Interfaces," *Physical Review Letters*, vol. 80, pp. 5176–5179, 6 1998.

[24] J. Zhang, C. Zhao, H. Sii, G. Groeseneken, R. Degraeve, J. Ellis, and C. Beech, "Relation between Hole Traps and Hydrogenous Species in Silicon Dioxides," *Solid-State Electron.*, vol. 46, pp. 1839–1847, 2002.

[25] J. Stathis and S. Zafar, "The Negative Bias Temperature Instability in MOS Devices: A Review," *Microelectronics Reliability*, vol. 46, no. 2-4, pp. 270–286, 2006.

[26] T. Grasser, M. Waltl, Y. Wimmer, W. Goes, R. Kosik, G. Rzepa, H. Reisinger, G. Pobegen, A. El-Sayed, A. Shluger, and B. Kaczer, "Gate-Sided Hydrogen Release as the Origin of "Permanent" NBTI Degradation: From Single Defects to Lifetimes," in *Proc. Intl.Electron Devices Meeting (IEDM)*, Dec. 2015.

[27] T. Grasser, M. Waltl, G. Rzepa, W. Goes, Y. Wimmer, A.-M. El-Sayed, A. Shluger, H. Reisinger, and B. Kaczer, "The "Permanent" Component of NBTI Revisited: Saturation, Degradation-Reversal, and Annealing," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, Apr. 2016, pp. 5A.2.1–5A.2.8.

[28] T. Grasser, M. Waltl, K. Puschkarsky, B. Stampfer, G. Rzepa, G. Pobegen, H. Reisinger, H. Arimura, and B. Kaczer, "Implications of Gate-Sided Hydrogen Release for Post-Stress Degradation Build-Up after BTI Stress," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, Apr. 2017, pp. 6A.2.1–6A.2.6.

[29] E. Cartier, "Characterization of the Hot-Electron-Induced Degradation in Thin SiO_2 Gate Oxides," *Microelectronics Reliability*, vol. 38, no. 2, pp. 201–211, 1998.

[30] M. Denais, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, N. Revil, and A. Bravaix, "Interface Trap Generation and Hole Trapping under NBTI and PBTI in Advanced CMOS Technology with a 2-nm Gate Oxide," *IEEE Trans.Dev.Mat.Rel.*, vol. 4, no. 4, pp. 715–722, 2004.

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

[32] H. Reisinger, R. Vollertsen, P. Wagner, T. Huttner, A. Martin, S. Aresu, W. Gustin, T. Grasser, and C. Schlünder, "The effect of recovery on NBTI characterization of thick non-nitrided oxides," in *Proc. Intl.Integrated Reliability Workshop*, 2008, pp. 1–6.

[33] S. Mahapatra, V. Huard, A. Kerber, V. Reddy, S. Kalpat, and A. Haggag, "Universality of NBTI - From Devices to Circuits and Products," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, 2014, pp. 3B.1.1–3B.1.8.

[34] S. Novak, C. Parker, D. Becher, M. Liu, M. Agostinelli, M. Chahal, P. Packan, P. Nayak, S. Ramey, and S. Natarajan, "Transistor Aging and Reliability in 14nm Tri-Gate Technology," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, 2015, pp. 2F.2.1–2F.2.5.

[35] T. Grasser, B. Kaczer, and W. Goes, "An Energy-Level Perspective of Bias Temperature Instability," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, 2008, pp. 28–38.

[36] E. Nicollian and J. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology*. New York: Wiley, 1982.

[37] C. Jungemann, C. Nguyen, B. Neinhüs, S. Decker, and B. Meinerzhagen, "Improved Modified Local Density Approximation for Modeling of Size Quantization in NMOSFETs," in *Proc. 4th Intl. Conf. on Modeling and Simulation of Microsystems*, Hilton Head Island, South Carolina, USA, Mar. 2001, pp. 458–461.

[38] C. Nguyen, C. Jungemann, B. Neinhus, B. Meinerzhagen, J. Sedlmeir, and W. Molzer, "Compatible Hole Channel Mobility and Hole Quantum Correction Models for the TCAD Optimization of Nanometer Scale PMOSFETs," in *Proc. 3st IEEE Conference on Nanotechnology*, 2003.

[39] V. Huard, "Two Independent Components Modeling for Negative Bias Temperature Instability," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, May 2010, pp. 33–42.

[40] T. Ma, "Interface Trap Transformation in Radiation or Hot-Electron Damaged MOS Structures," *Sem.Sci.Tech.*, vol. 4, pp. 1061–1079, 1989.

[41] P. Hurley, B. O'Sullivan, V. Afanas'ev, and A. Stesmans, "Interface States and P_b Defects at the Si(100)/HfO_2 Interface," *Electrochemical and Solid-State Letters*, vol. 8, no. 2, pp. G44 – G46, 2005.

[42] E. Cartier and J. Stathis, "Hot-Electron Induced Passivation of Silicon Dangling Bonds at the Si(111)/SiO_2 Interface," *Appl.Phys.Lett.*, vol. 69, no. 1, pp. 103–105, 1996.

[43] T. Grasser, "Stochastic Charge Trapping in Oxides: From Random Telegraph Noise to Bias Temperature Instabilities," *Microelectronics Reliability*, vol. 52, pp. 39–70, 2012.

[44] D. Griscom, "Diffusion of Radiolytic Molecular Hydrogen as a Mechanism for the Post-Irradiation Buildup of Interface States in SiO_2-on-SiStructures," *J.Appl.Phys.*, vol. 58, no. 7, pp. 2524–2533, 1985.

[45] S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, V. Eijkhout, D. Kaushik, M. G. Knepley, D. A. May, L. C. McInnes, W. D. Gropp, K. Rupp, P. Sanan, B. F. Smith, S. Zampini, H. Zhang, and H. Zhang, "PETSc Users Manual," Argonne National Laboratory, Tech. Rep. ANL-95/11 - Revision 3.8, 2017.

[46] T. Grasser, H. Reisinger, P.-J. Wagner, W. Goes, F. Schanovsky, and B. Kaczer, "The Time Dependent Defect Spectroscopy (TDDS) for the Characterization of the Bias Temperature Instability," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, May 2010, pp. 16–25.

[47] A. El-Sayed, Y. Wimmer, W. Goes, T. Grasser, V. Afanas'ev, and A. Shluger, "Theoretical Models of Hydrogen-Induced Defects in Amorphous Silicon Dioxide," *Physical Review B*, vol. 92, no. 11, p. 014107, 2015.