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 SiO₂ nMOS/pMOS capacitors. We extract a unique defect band inside the SiO₂ 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 Pb centers.

I. INTRODUCTION

Due to the rapid onset of recovery well below the microsecond 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 $\Delta V_{th}(t)$ 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 insinuous 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 $\Delta V_{FB}(t)$ and the details of the time evolution of the energetic profile of near-interfacial traps in the bandgap, $D_a(E,t)$, 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, $\Delta V_{th}$ 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 $\Delta V_{th}(t)$ and only qualitatively considered the evolution of the average of $D_a(E,t)\Delta E$, 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, $C_0$. Also, CV measurements are sensitive to defects/charges in the oxide and at the interface, $\Delta V$, which result in a stretch-out of the CV curve. Experimentally, however, it appears impossible to clearly separate $\Delta V$ and $C_0$. 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, $\Delta V$ and $C_0$.

- **Physical**: From a physical perspective, both interface defects (like Pb centers [15]) and defects deeper in the oxide (such as hydroxyl E’ centers [16–18]) can contribute to both $\Delta V$ and $C_0$, 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 $C_0$ in SiO₂/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 $C_0$ 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, \(D_1\), as well as the temporal evolution of oxide defects, \(D_O\). We deliberately refer to these components with the unusual names \(D_1\) 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 \(\Delta V\) 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 \(\Delta V\). 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 \(\Delta V\). 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 \(\mu\)m \(\times\) 50 \(\mu\)m) pMOS and nMOS capacitors on ⟨100⟩ 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 \(t_{ox} = 5\) 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 V to 2 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, \(D_{it}(E)\), 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 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, \(\Delta C\), converted to an effective density of interface states \(\Delta D_{it} = \Delta C/q_0 A\), 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 \(\Delta C\) is on the order of \(10^{11}\) 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 \(D_{it}(t)\) but rather restrict out analysis to the change of \(D_{it}\) over time, \(\Delta D_{it}(t) = D_{it}(t) - D_{it}(0)\). In the simplest case, \(\Delta D_{it}(t)\) is obtained from \(\Delta C = C(t) - C(0)\), as has been done previously [9]. However, it is important to realize that the shift \(\Delta C\) 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 \(\Delta V(V_G)\).

This problem is elaborated using the calibrated TCAD model for the nMOS and pMOS capacitors. We start with Fig. 2, which shows \(\Delta C\) resulting from a rigid shift of the CV curve by \(\Delta V = -30\) mV, caused e.g. by fixed positive charges. Thus, in this simplest case \(\Delta C(V_G) = C(V_0 - 30\) mV\) − \(C(V_G)\). This value of \(\Delta V\) was chosen to be similar to the \(\Delta V\) observed in our experiments. As can be seen, a spurious \(D_{it} \approx 2 \times 10^{11}\) cm\(^{-2}\) is obtained, which has of course nothing to do...
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 \( 15 \times 10^{10} \text{cm}^{-2} \), also in the range of what will be measured later. For this value of \( D_{it} \), a maximum \( \Delta V = -35 \text{mV} \) is obtained in our 5nm 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 0.235/0.125eV (mean/variance) and the 0/- transition at 0.8/0.1eV. These values were obtained on 2.9nm thick SiO\(_2\) grown on (100) 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 \( \Delta V \). For example, for a reasonably slow sweep rate \( S \) (around 1 V/s) and a very high AC frequency \( f_{AC} \) outside feasible experimental ranges, only \( \Delta V \) 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 \( \Delta V \). 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 \( V_{fb} \) 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 \( \Delta C \) 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)
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 frequencies. In particular, this slow response of midgap states explains the characteristic “dip” in the data at typically used frequencies. In particular, this slow response of midgap states explains the characteristic “dip” in the data at typically used frequencies.

In order to visualize \( \Delta V \) for NBTI/PBTI (stress at \( \pm 5.5 \text{V}/125^\circ \text{C}/7.5 \text{ks} \) and recovery at 0V/125^\circ C/7.5ks). The corresponding conventional notation \( \Delta V_{th} \) and \( \Delta V_{FB} \) is also given.

In order to approximately separate \( \Delta C \) caused by the stretch-out \( \Delta V \) and the “true” contribution due to fast states, we follow a pragmatic approach and compare a CV measured after stress, \( C(V_G) \), to a reference curve taken before stress, \( C_0(V_G + \Delta V) \). Irrespective of the precautions taken, extraction of \( \Delta D_h \) and \( \Delta V \) 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 \( \Delta C \) 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 (\( \pm 5.5 \text{V} \)). In order to visualize \( \Delta D_h \), we will use a pragmatic ansatz for \( \Delta V \) by first approximately extracting \( \Delta V_{th} \) and \( \Delta V_c \). Then,

\[
\Delta V = \Delta V_{th} + \Delta V_c.
\]

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

IV. Results

Fig. 5. Theoretical \( \Delta D_h \) from a two-state NMP defect normally distributed in energy around midgap with \( \sigma = 0.2 \text{eV} \) and over a certain depth into the oxide of a pMOS; right at the interface (top) and 0.5nm 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 100kHz) corrected by the stretch-out (solid blue), approximated by \( \Delta V_{th} \) and \( \Delta V_{FB} \) and linearly interpolated in-between. The difference is converted to \( \Delta D_h \). NBTI (\( -5.5 \text{V}/125^\circ \text{C}/7.5 \text{ks} \), middle row) results in similar changes in both n/pMOS while PBTI (\( +5.5 \text{V}/125^\circ \text{C}/7.5 \text{ks} \), top row) leads to positive charging (pMOS) but has no impact on the nMOS. The approximate \( \Delta V \) is shown in the bottom row.

Fig. 7. Temporal evolution of \( \Delta V_{th} \) and \( \Delta V_c \) for NBTI/PBTI (stress at \( \pm 4.5 \text{V}/125^\circ \text{C}/7.5 \text{ks} \) and recovery at 0V/125^\circ C/7.5ks). The corresponding conventional notation \( \Delta V_{th} \) and \( \Delta V_{FB} \) is also given.
confused with bandgap and positive in the upper-half), which can be easily spurious peaks (typically negative in the lower half of the is shown in Fig. 9. If neglected, a simple but hardly affects the nMOS. The impact of the stretch-out and pMOS, while PBTI creates positive charges in pMOS Δ in a spurious shown in Fig. 8. NBTI creates the largest ΔPBTI on nMOS has little impact on Δ. Conversely, times overestimation, the massive amount of positive charge a constant defect density between in-between, see Fig. 6 (bottom). This corresponds to assuming original CV curve onto these points and stretching it linearly for the stretch-out. While for NBTI ΔD it can be 2-3 times larger, a 10× larger ΔD it is obtained for PBTI/pMOS.

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

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 c(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. 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,
Hydrogen diffusion is extremely fast, leading to ex-

Finally, release of H from the defects proceeds over a

Charge trapping and detrapping in the secondary config-

– and thus preferred – defect site, see Fig. 12. In addition,

migrates to newly available sites which support a positive

results in charge transitions at those defect sites, see Fig. 11,

exchange mechanism via the neutral state 1, see Fig. 10.

different time scales (TS) implied by the individual processes:

system and severe numerical difficulties related to the vastly

defect states of each defect and the neutral interstitial position.

state machine with hydrogen now hopping between the four

to four-state defects, we expressed the problem as a complete

centers [42].

Empty symbols are precursor defect sites without a trapped H.

2') while the squares represent the secondary configuration (states 2 and 1').

hydrogen, which preferentially moves to regions where it can create a positive

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 distri-

ution 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 ↔ 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

H0; 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

trap 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 ex-

remely fast H redistribution. For instance, using estab-
lished parameters for H diffusion [44], at 125°C intersti-
tial H would hop to a next neighbor separated by 5 Å in

a nanosecond.

TS2 Charge trapping and detrapping in the secondary config-

uration of the four-state defects (2 ↔ 1') occurs over a

wide range of time scales. This begins with very fast

transitions being able to contribute to $D_h$ 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 → 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.

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 $P_i$

centers [42].

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').

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”, $H_{T,i}$. An “oxide site” could for instance be one of the

N-member rings in the SiO2 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 $H_{T,max}$ hydrogens by exchanging H with the H0

reservoir from state 1. Using the previously calculated occupancies for

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

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

$$\frac{dH_{T,i}}{dt} = -k_{10,i} f_{i1} H_{T,i} + k_{01,i} H^0 (H_{T,max} - H_{T,i}).$$  \hspace{1cm} (1)

Note that each $H_{T,i}$ 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 $\Delta t$ to give

$$H_{T,i} = \frac{aH_{T,i}^{old} + k_{01,i} H^0 H_{T,max}}{a + k_{10,i} f_{i1} + k_{01,i} H^0}$$  \hspace{1cm} (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

$$H^0 + \sum_i H_{T,i} = H_{tot}$$  \hspace{1cm} (3)

which via substitution of eq. (2) results in

$$H^0 + \sum_i \frac{aH_{T,i}^{old} + k_{01,i} H^0 H_{T,max}}{a + k_{10,i} f_{i1} + k_{01,i} H^0} = H_{tot}.$$  \hspace{1cm} (4)

At every time-step eq. (4) is a simple non-linear equation with one unknown ($H_{tot}$), 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 $H_{T,i}$ can be calculated using eq. (2). With $H_{T,i}$ available, the contribution of each trapping site to the stretch-out $\Delta V$ can be calculated by considering its positive charge in the states 2 and 2'. Thus, one obtains

$$\Delta V = \frac{q_0 \alpha_{ox}}{AE} \sum_i H_{T,i} (f_2 + f_2') \left( 1 - \frac{x}{t_{ox}} \right),$$  \hspace{1cm} (5)

with $x$ being the distance of the trap from the interface and $\epsilon$ 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

$$\Delta C_{ni,j} = \frac{\Delta C^{0}_{ni,j}}{1 + \left( \frac{\omega}{k_i} \right)^2},$$  \hspace{1cm} (6)

$$\Delta G_{ni,j} = \frac{\Delta C_{ni,j} \omega}{k_i},$$  \hspace{1cm} (7)

with the auxiliary quantities

$$\Delta C^{0}_{ni,j} = q_0 H_{T,i} \frac{k'_2 f_{i2,j} - k_2 f_{i2,j}}{k_i},$$

$$k_i = k_{1/2,i} + k_{2'},$$

$$k'_i = \frac{d \epsilon_{ni,j}}{d V_G},$$

and the angular frequency $\omega = 2 \pi f$. Hereby it was assumed that only the fast transitions in the secondary configuration (2 ↔ 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

$$D_0 + H \overset{k_{01}}{\underset{k_{01}}{\leftrightarrow}} DH.$$  \hspace{1cm} (8)

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

$$DH + H \overset{0.2eV}{\underset{0.65eV}{\leftrightarrow}} DH_2,$$  \hspace{1cm} (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 reactivated via

$$DH_2 + H \overset{0.2eV}{\underset{0.65eV}{\leftrightarrow}} DH + H_2.$$  \hspace{1cm} (10)

As a consequence, the actual concentration of active defects, $DH$, will depend on the complicated interplay of $H$, $H_2$, $D$, $H_{T,max}$, 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, $D_0$, the highly reactive H can also depassivate $P_h$ centers: Once the H is released inside the oxide, the creation of interface states can be calculated via

$$P_h H + \overset{0.2eV}{\underset{2.0eV}{\leftrightarrow}} P_h + H_2.$$  \hspace{1cm} (11)

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

$$P_h + \overset{0.2eV}{\underset{4.3eV}{\leftrightarrow}} P_h H,$$  \hspace{1cm} (12)

making the whole process extremely complicated, in particular since both $P_h$ 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)

$$P_h H + \overset{0.2eV}{\underset{4.3eV}{\leftrightarrow}} P_h + H_2.$$  \hspace{1cm} (13)

Furthermore, we assume that the density of $P_h$ centers is relatively small and thus the competition for $H$ is negligible. Then, the concentration of $P_h$ 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 $\Delta C$ 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 $\Delta D_d$ near the interface, as well as the
Fig. 13. Comparison of measured and simulated $\Delta D_\text{it}$ profiles from $\Delta C$ measurements for the four combinations of NBTI/PBTI in pMOS/nMOS during both two stress times (top, middle) and after a 7.5ks 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 $\Delta D_\text{it}$ 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, $\Delta V'$, $\Delta C_{\text{at}}$ and $\Delta C_{\text{it}}$ contribute to the overall observed $\Delta C$ and thus $\Delta D_\text{it}$. 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 $\Delta D_{it}$, $\Delta C$ contains a significant spurious $\Delta D_{it}$ contribution which is not due to fast states but merely due to the stretch-out $\Delta V'$. This stretch-out is due to $D_0$, that is, defects which are too slow to contribute to $\Delta D_{it}$ 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 $\Delta C_{it}$ only dominant at higher frequencies above 100kHz. 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 $\Delta D_{it}$ 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


