The 'Permanent' Component of NBTI: Composition and Annealing

T. Grasser, Th. Aichinger†,‡, G. Pobegen†, H. Reisinger•,*
P.-J. Wagner, J. Franco°, M. Nelhiebel□, and B. Kaczer°*

* Christian Doppler Laboratory for TCAD at the Institute for Microelectronics, TU Wien, Austria
† KAI, Villach, Austria ‡ Now at Penn State University, USA • Infineon, Munich, Germany
° imec, Leuven, Belgium □ Infineon, Villach, Austria

***Abstract*— A number of recent publications explain NBTI to consist of a recoverable and a more permanent component. While a lot of information has been gathered on the recoverable component, the permanent component has been somewhat elusive. We demonstrate that oxide defects commonly linked to the recoverable component also form an important contribution to the permanent component of NBTI. As such, they can contribute to both the threshold voltage shift as well as to the charge pumping current. Under favorable conditions, particularly when subjected to continuous charge-pumping measurements, the permanent component can show recovery rates comparable to that of the recoverable component. We argue that this enhanced recovery is due to a recombination enhanced defect reaction mechanism. We introduce a simple extension to our switching trap model to also capture the impact of charge pumping measurements on the transition rates between the defect states.**

I. Introduction

Recent research indicates that two components dominantly contribute to the negative bias temperature instability (NBTI) [1–6]: while one component dominates the recovery (R) the other one has been suspected to be more or less permanent (P). It has been recently shown that the complete NBTI induced degradation can be annealed at higher temperatures [6–8], implying that P is recoverable as well, albeit at larger timescales compared to R. The most important aspect regarding P is that it might dominate device degradation at long times and could thus be the crucial degradation mechanism eventually determining the lifetime [6]. Unfortunately, the extraction of P is challenging as within conventional measurement windows ($1\,\mu\mathrm{s}$ – $100\,\mathrm{ks}$) it is normally overshadowed by R. As such, our understanding of P is somewhat vague, also regarding its constituents, be it interface and/or oxide defects [3, 6], or fixed positive charges [6]. We show that considerable precautions have to be taken for accurate extraction of P, as it suffers from similar issues than those typically related to the extraction of R, such as measurement *delay*, measurement *duration*, as well as stress/recovery *artifacts introduced by the measurement procedure* itself. Contrary to the work of Huard [6], who links P to interface states and an equal amount of fixed positive charge, *our analysis demonstrates that a significant fraction of P is due to switching oxide traps*, which contribute to both the threshold voltage shift ΔV_{th} and to the frequency dependent fraction of the charge pumping current.

II. Erroneous Extraction of P

The most straight-forward approach for the extraction of P would be to wait until the recovery of ΔV_{th} has leveled at a plateau, thus directly exposing P. However, the fundamental problem here is the large timescales involved in the recovery of R, as even a short stress of $t_{\mathrm{s}} = 1\,\mu\mathrm{s}$ can lead to recovery transients of up to $1\,\mathrm{ks}$, not to mention the recovery of P itself. On the other hand, P is created at a slower rate than R, making it difficult to locate plateaus within reasonable measurement times (< 1 week). As a consequence, plateaus in the recovery are rarely reported in literature [6]. (The plateau reported in [9] was later found to be not reproducible.)

Using different test technologies, from thick $\mathrm{SiO_2}$ to SiON and high-κ gate stacks, we investigated a number of possible extraction methods. Most unfortunately, the extraction of P turned out to be much more complicated than expected. In particular, despite the fact that some attempts are incorrect altogether, it seemed almost as if P was trying to evade our characterization attempts. Fig. 1 summarizes some potential mistakes related to the extraction of P:

M1: The recovery of ΔV_{th} has to be plotted on a *relative logarithmic* scale following the end of stress, otherwise a spurious plateau appears. Such 'plateaus' are commonly found in literature but are completely irrelevant and simply a consequence of the inadequate presentation of the data.

M2: Switching to a lower temperature temporarily freezes recovery, resulting in a spurious plateau [10]. While the example temperature switch from 80 °C to 40 °C given in Fig. 1 may appear pathological, a typical real-world case appears to be given in [6]: The recovery of the devices was monitored at a high temperature on a probe-station for a day. Then, the devices were taken off the probe-station and stored at room temperature to be re-measured after some time. The plateaus obtained from this method are completely arbitrary.

M3: Application of a short positive bias partially removes oxide charges, temporarily accelerating recovery. This is because the emission time constant of switching traps depends strongly on the gate bias [13, 15]. Back at the original recovery voltage, these defects have already been annealed, resulting in a spurious plateau until the original recovery continues.

M4: In order to minimize the recovery, short stress times and low stress voltages can be chosen. This leads to relatively weak stresses and relatively short recovery times. However, particularly in thin oxides, the difference between stress and recovery voltage can be small, leading to notable degradation at the recovery voltage, interfering with the actual recovery. As a consequence, spurious plateaus can appear.

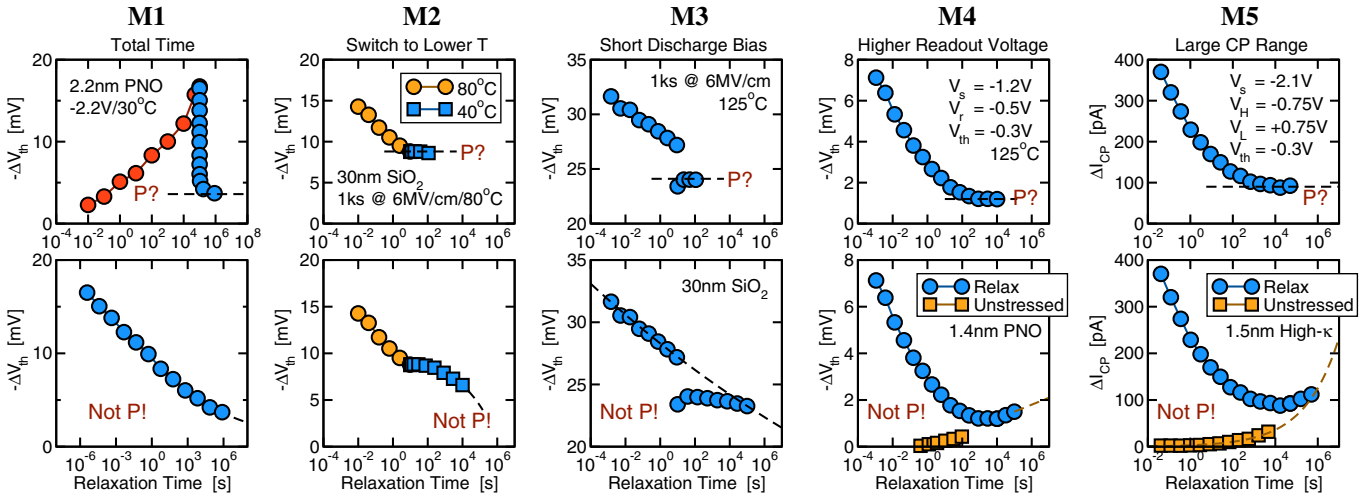

Fig. 1. Potential mistakes encountered while trying to locate plateaus in $\Delta V_{\rm th}$ recovery traces. The top figures show apparent plateaus, which have nothing to do with permanent degradation. The reasons for the occurrence of these plateaus are illustrated in the bottom figures. From left to right: **(M1)** The recovery of $\Delta V_{\rm th}$ is plotted as a function of the total time, rather than the recovery time $t_{\rm r}$. Even if the recovery perfectly follows $\log(t_{\rm r})$, a spurious plateau will appear if the data is plotted this way, which has nothing whatsoever to do with P. Similar considerations relate to plotting the data on a linear scale, where the spurious plateau depends solely on the measurement time. **(M2)** Due to the large recovery time required, the device is only kept at stress temperature for a short amount of time [6] and, to ease measurement, recovery is continued at a lower temperature. This, however, is pointless, as a switch to a lower temperature freezes the recovery [10], which results in a spurious plateau. **(M3)** In order to remove R, which is due to trapped holes in the oxide [6, 11–13], a positive bias could be applied [14]. However, since the trap sites are switching traps, this has basically the same effect as a temperature switch, because such a bias switch only removes a few decades from the recovery trace, which continues after that. **(M4)** Relaxation gate voltages only slightly larger than the threshold voltage can already lead to degradation, in this example $V_{\rm relax} = -0.5\,{\rm V}$, with $V_{\rm th} = -0.3\,{\rm V}$. As a result, degradation overlaps with the 'normal' recovery, resulting in a spurious plateau for a certain amount of time. The signature of this plateau is that it disappears when either stress or relaxation voltages are changed. **(M5)** If the charge pumping amplitude is chosen too large, for this $1.5\,{\rm nm}$ high-κ device for example from $\pm 0.75\,{\rm V}$, degradation is observed during the CP measurement, again resulting in a spurious plateau as in (M4). Note the strong relaxation of $\Delta I_{\rm CP}$, which is *anything but constant*.

M5: Similarly to M4, charge pumping (CP) measurements can lead to degradation of $\Delta I_{\rm CP}$ when the charge pumping amplitude is chosen too large. Balancing the recovery of $\Delta I_{\rm CP}$, this can lead to spurious plateaus as well, just like M4. M5 already highlights an important issue [16]: $\Delta I_{\rm CP}$ is not constant, even within conventional measurement windows, contradicting claims that $\Delta I_{\rm CP}$ is nearly constant and equal to P [1, 6]. In particular, the resemblance between the recovery of $\Delta V_{\rm th}$ in M4 and $\Delta I_{\rm CP}$ in M5 is indeed striking.

III. Attempts at Extracting P

We proceed by analyzing $\Delta V_{\rm th}$ recovery traces recorded after carefully selected stress/recovery voltages, stress/recovery times, and temperature. A typical plateau at the end of the recovery is shown in Fig. 2. According to Huard [6], this plateau is due to semi-permanent interface states $\Delta N_{\rm it}$ and fixed oxide charges. Interface states are fast and can quickly follow changes in the bias ($< 1\,{\rm ms}$). Thus, a change of the interfacial Fermi-level would result in a rapid change of the charge stored in these interface states, $\Delta Q_{\rm it}(E_{\rm F})$, according to their density-of-states. In particular, after a temporary bias change, the same $\Delta V_{\rm th}$ would be expected back at the original bias. This is clearly not the case. In fact, $\Delta V_{\rm th}$ only slowly goes back to its original value, an apparent degradation during the recovery phase [14]. We call this phenomenon *reverse recovery*, which thus indicates that a significant part of P is due to slow oxide defects, $\Delta N_{\rm ot}$, such as those observed previously [11, 13, 17]. The explanation of the reverse recovery effect is as follows: during stress, defects are created inside the oxide. These defects have an energy level in the silicon

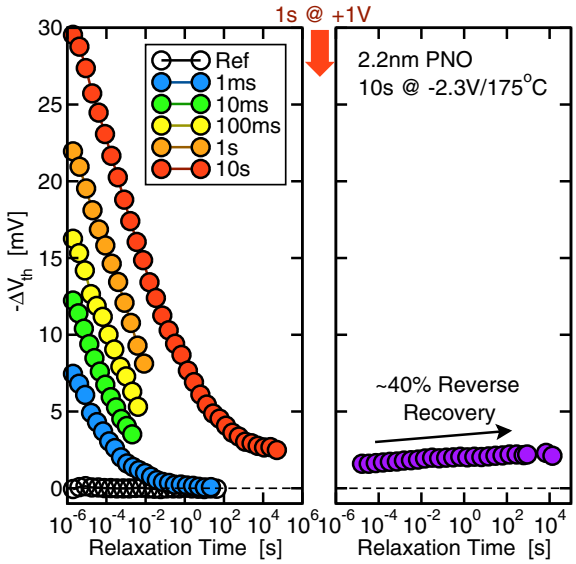

Fig. 2. Typical plateau observed under medium stress conditions. After the plateau has been reached, a positive bias was applied for a short time. For $P \sim \Delta N_{\rm it}$, one would expect $\Delta V_{\rm th}$ to rapidly follow bias changes (within a $1\,{\rm ms}$). In fact, a pronounced reverse recovery is observed with time constants as large as $10\,{\rm ks}$, indicating that $\Delta N_{\rm ot}$ contributes to P.

bandgap and their occupancy depends on the position of the Fermi-level. During application of a positive bias, the defects are discharged. This does not mean that the defects are annealed, the discharging step just makes them electrically neutral and thus invisible in $\Delta V_{\rm th}$. Once in this metastable neutral state, the defects can either completely anneal or they can be charged again when the Fermi-level is moved back to the threshold voltage. However, as the time constants responsible for charging and discharging can be considerably

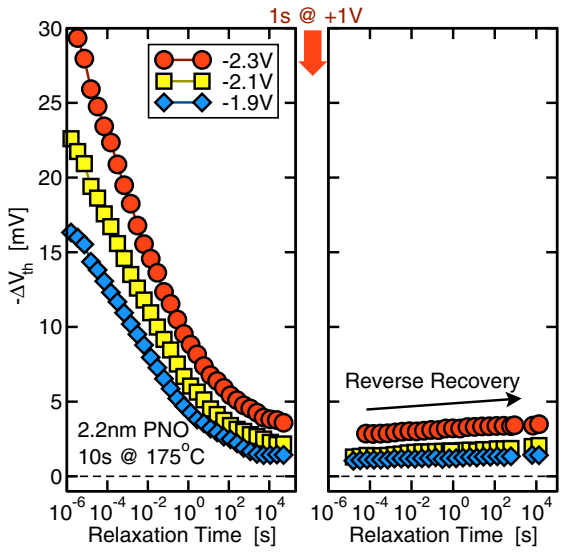

Fig. 3. **Left**: Plateaus in the recovery of $\Delta V_{\rm th}$ for three different stress voltages. **Right**: After short application of a positive bias pulse, again considerable reverse recovery is observed, which indicates a significant contribution of slow switching oxide traps to the plateau.

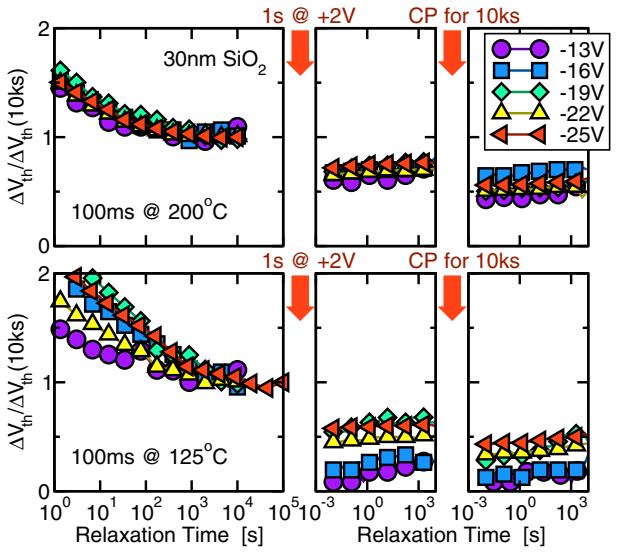

Fig. 5. The same effect as in Fig. 2 is observed on thick SiO_2 devices. The reverse recovery time constants are either somewhat larger or the application of positive bias anneals a fraction of the oxide defects, this being more pronounced at lower T. Continuous CP for 10 ks removes a further fraction, cf. Fig. 9.

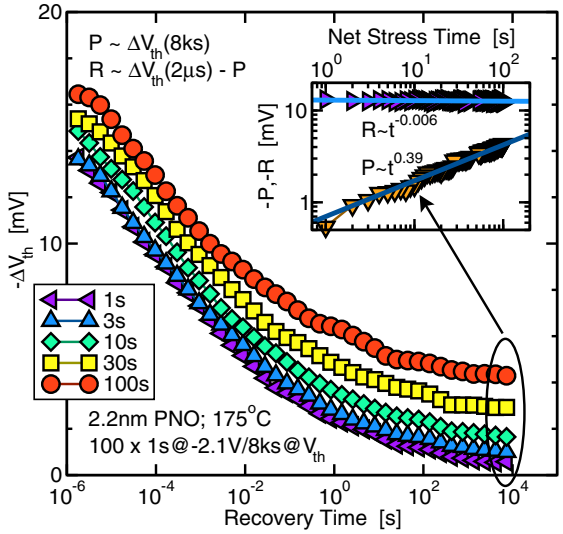

Fig. 4. Low duty factor experiment $(1/8000)$ with 1 s stress followed by 8 ks recovery repeated 100 times. $\Delta V_{\rm th}(8\,{\rm ks})$ recorded at the end of every trace increases as shown in the inset, indicating the build-up of slowly recoverable damage. At the same time, R is slightly reduced, albeit by a much smaller amount than the creation of P (inset).

larger than those typically associated with interface states, their charging is visible as a reverse recovery transient [3]. Note that this is markedly different from the conventional picture of hole trapping, where holes are simply trapped in the oxide. Such trapped holes would not react that sensitively to changes of the Fermi-level and when discharged, they are already fully annealed [3]. Moving back to the threshold voltage would not result in a reverse recovery as these defects can only be charged again by application of a large negative stress pulse.

Fig. 3 shows the bias dependence of these plateaus, demonstrating that the recovery settles at a higher level when larger stress voltages are employed. Fig. 3 also demonstrates the fundamental dilemma regarding the characterization of the plateaus, namely that even after a stress time of only $10\,{\rm s}$, the plateaus may only become gradually visible after a recovery of $10^5\,{\rm s}$ (about a day), which is already close to the maximum experimentally feasible recovery time.

One possibility to stimulate the build-up of P without excessive creation of R is by repeated stress and recovery experiments with a very low duty factor. Such an experiment results in a slow additive component P to the otherwise unchanged recoverable component R. Unfortunately, due to the intricate dynamic nature of the experiment, the analysis of the data is also much more involved and not possible without assumptions taken from a sensible model. An example of such an experiment with a duty factor of $1/8000$ is shown in Fig. 4. While P increases with a relatively large power-law exponent of about 0.4, the reduction of R is only weak. Still, this reduction in R might be indicative of a coupling between R and P [4, 6].

Fig. 5 documents our search for plateaus on thick SiO_2 devices. In order to make the plateaus clearly visible so that their bias and temperature dependence can be studied, we kept the stress short $(10\,{\rm ms})$. At $200\,^\circ{\rm C}$, where both degradation as well as recovery are strongest, the recovery levels to a plateau after about $10^4\,{\rm s}$. Again, the large disparity between stress and recovery times, which differ by six orders in magnitude, is hard to miss. After the plateau had been reached, the devices were driven into accumulation for $1\,{\rm s}$. Back at the original read-out voltage ($V_{\rm G} = V_{\rm th}$), $\Delta V_{\rm th}$ was found to be reduced by 40%, the same percentage as observed for the thin SiON devices. Again, reverse recovery was observed, resulting in $\Delta V_{\rm th}$ to slowly *increase* following this bias switch. However, even after $10^4\,{\rm s}$ the original degradation of the plateaus level was not reached, contrary to the thin SiON devices. Also, the effect appears to be about the same for all stress voltages used. Following this reverse recovery phase, a continuous

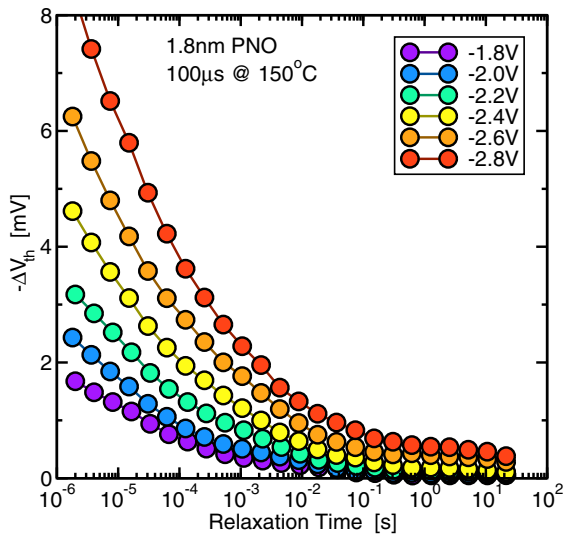

Fig. 6. Plateaus are occasionally also observed after ultra-short stress times, shown for a 1.8 nm PNO device. The plateaus are not permanent and strongly depend on the stress bias. Particularly after such weak stresses, it is important that the degradation at $V_G = V_{th}$ is negligible in order to avoid mistake M4 of Fig. 1.

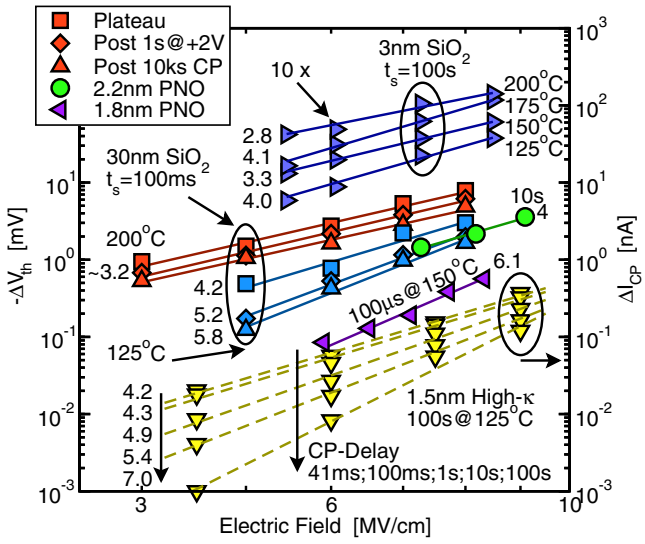

Fig. 7. The field dependence of the 'permanent' part. For the 1.8 nm PNO device from Fig. 6 we take $P \sim \Delta V_{th}(t_r = 1\,s)$, for the ones of Figs. 3 and 5 $P \sim \Delta V_{th}(t_r = 10\,ks)$ was chosen. In addition, for the thick SiO_2 devices, the last value at the end of the second and third relaxation cycle are shown (10 ks after the bias switch and 10 ks after the 10 ks CP measurement cycles). Also shown is the field dependence of the CP current, which appears to behave in a similar manner. In any case, P can be fitted by a power-law $\sim E_{ox}^{\gamma}$.

CP measurement lasting 10^4 s was performed. At the end of this CP measurement, the degradation level was reduced by about 60% relative to the original plateau value. Again, even after such a long CP measurement, reverse recovery was visible. The same experiment was repeated at 125 °C where the plateaus were only reached after a considerably longer time. Now, the response of the devices to the short accumulation pulse and the long charge pumping measurement depended on the stress voltage used.

IV. Bias Dependence of P

For lifetime back-extrapolation, the bias dependence of P is crucial. This is mostly due to the fact that at higher stress voltages and temperatures, the contribution of P relative to R apparently increases, which has to be corrected for when extrapolating back to operating conditions. Huard [6] observed $P \sim E_{ox}^{\gamma}$ with a technology-independent $\gamma = 4$, without giving details of the extraction scheme for P.

Another example showing experimentally observable plateaus which appear already after a $t_s = 100\,\mu s$ stress is given in Fig. 6. The E_{ox} dependence of these plateaus is shown in Fig. 7, together with the plateaus of Figs. 3–5 and related experiments. Contrary to the universal exponent of 4 given by Huard, a wider range is observed, with values smaller and larger than 4. Also, the bias dependence of the plateaus in the thick SiO_2 devices (Fig. 5) shows exponents around 3.2 at 200 °C and in the range 4.2–5.6 at 125 °C. The latter is insofar interesting as the initial plateau has $\gamma = 4.2$, which increases to 5.2 after application of +2 V for 1 s, and even to 5.6 after continuous CP measurements for 10 ks. This again demonstrates that P, whatever it is and by whatever means it is extracted, is not really permanent.

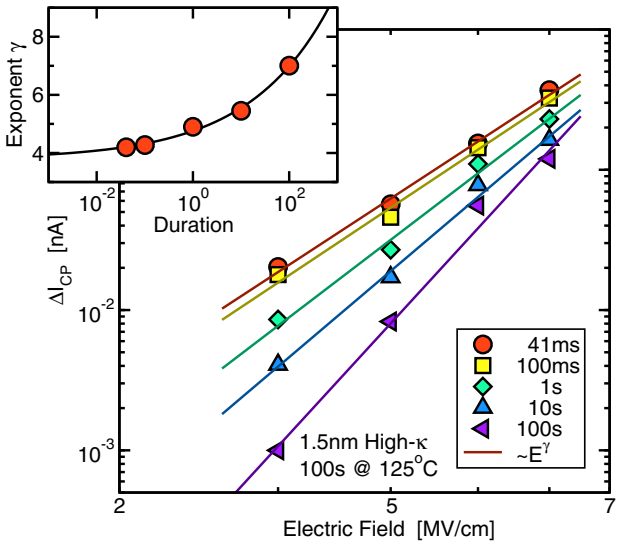

Fig. 8. The permanent component has been suggested to correlate with the change in the charge pumping current, $P \sim \Delta I_{CP}$. Just like in ΔV_{th} measurements, the CP current is very sensitive to the measurement time in all our investigated technologies. This has a profound impact on the extracted bias dependence of P. Only for short CP measurements an E_{ox}^4 dependence as in [6] is obtained.

V. Correlation with Charge-Pumping Data

It has been occasionally suggested [6, 19] that P is correlated to the CP current, ΔI_{CP}. In that context, ΔI_{CP} has been interpreted as being proportional to the number of interface states. Particularly at lower frequencies it has been observed, however, that I_{CP} also contains considerable contributions from oxide traps [20]. This issue has been commonly neglected in the context of NBTI [21].

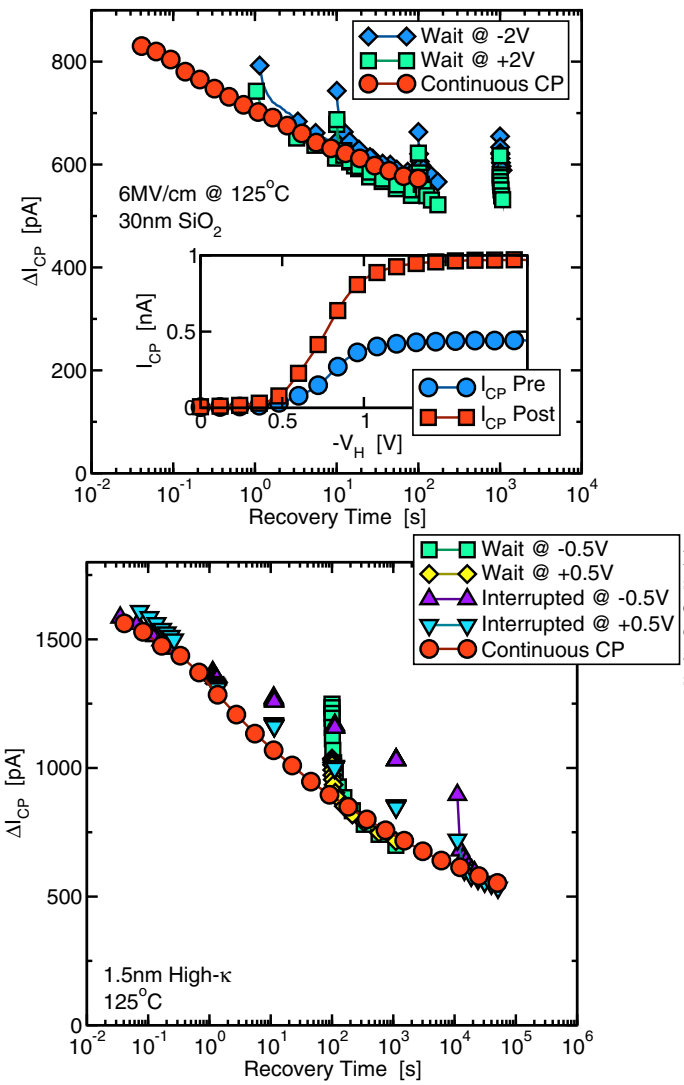

Fig. 9. Recovery of the CP current at a frequency of 1 MHz. Contrary to other observations [6], the CP current is never constant in any of our technologies, consistent with [16]. Most importantly, continuous CP accelerates recovery. When the beginning of the CP measurement is delayed for variable amounts of time, with V_G being either the CP high- or low-level, rapid recovery is observed at the beginning of the delayed continuous CP measurement. Application of the CP high- and low-levels during the wait phase proves that it is not the voltage but rather the charge pumping itself that causes the rapid recovery. **Top**: For a thick 30 nm SiO_2 device. The inset shows a constant-base-level CP sweep before and after stress. **Bottom**: For a 1.5 nm high-κ gate stack device. The same effect is observed in both dramatically different technologies.

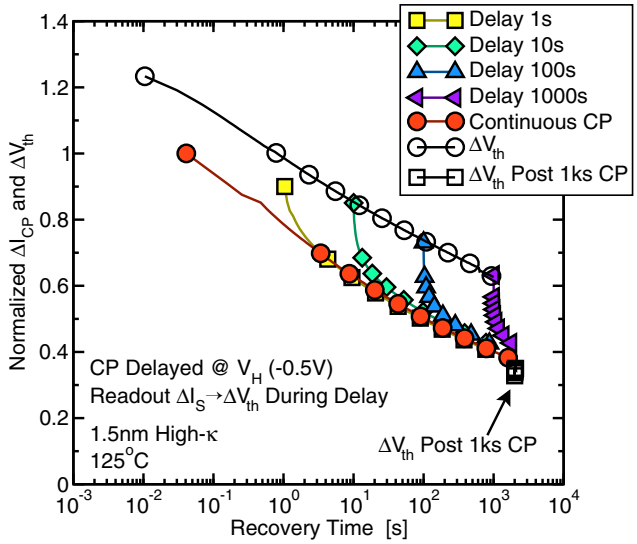

Fig. 10. Extension of the experiment in Fig. 9: the change in the linear source current ΔI_S at the CP high-level V_H is read during the wait phase and converted to ΔV_{th} following [18]. This clearly demonstrates that the onset of the accelerated CP recovery follows the recovery of ΔV_{th}. CP induced accelerated recovery in ΔI_{CP} is paralleled by a recovery in ΔV_{th} by the same fraction.

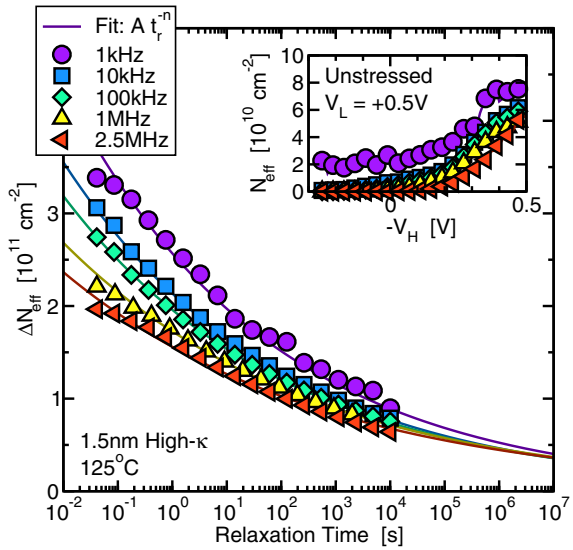

Fig. 11. Frequency dependence of the CP recovery. With lower frequency, a larger amount of oxide defects contributes to ΔN_{eff}. With increasing CP time, however, the oxide defects are 'pumped-away', giving a frequency independent ΔN_{eff} as expected from a pure ΔN_{it} contribution.

The bias dependence of ΔI_{CP} is compared in Fig. 7 to the bias dependence of the ΔV_{th} plateaus, which particularly for fast CP measurements seems to agree well, indicating a correlation between the defects visible in these two experiments. Such a correlation has led previous studies to conclude that ΔV_{th} as well as ΔI_{CP} are dominated by interface states [22].

Interestingly, the bias dependence of ΔI_{CP} is very sensitive to the measurement duration as shown in Fig. 8. While for fast CP experiments (41 ms) we obtain $\gamma = 4$, we observe a strong dependence of the extracted exponent on the duration of the CP measurement. This is reminiscent to ΔV_{th} measurements [23], where the exponent also increases with the measurement duration. The reason for this behavior is that ΔI_{CP} shows similar recovery rates as ΔV_{th} which is contrary to Huard's work, but consistent with the observation of Rangan *et al.* [16].

Remarkably, recovery is accelerated by the CP measurement, see Fig. 9. The effect is quite similar for thick SiO_2 and thin high-κ devices. At first glance one might relate this to the bias-dependence of the defect annealing rates, since the transistor is continuously pulsed between accumulation and inversion. As it has already been shown for ΔV_{th} recovery, recovery can be accelerated when the transistor is driven towards accumulation [3, 23]. This effect, so undoubtedly present, does not provide the full answer, though. This can be seen in Fig. 9 which also shows reference CP measurements which were interrupted by constant bias phases at the low-

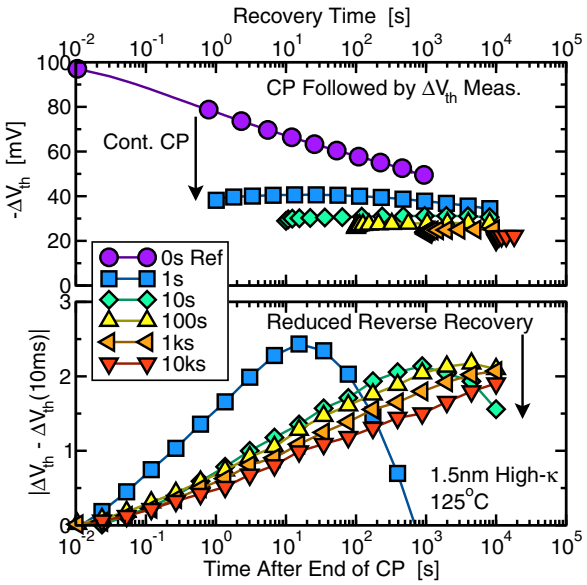

Fig. 12. Impact of continuous CP measurements with variable duration on $\Delta V_{\rm th}$. A considerable recovery in $\Delta V_{\rm th}$ is observed. Also, with increasing CP duration, the reverse recovery is reduced, indicating that CP anneals slow oxide defects, $\Delta N_{\rm ot}$.

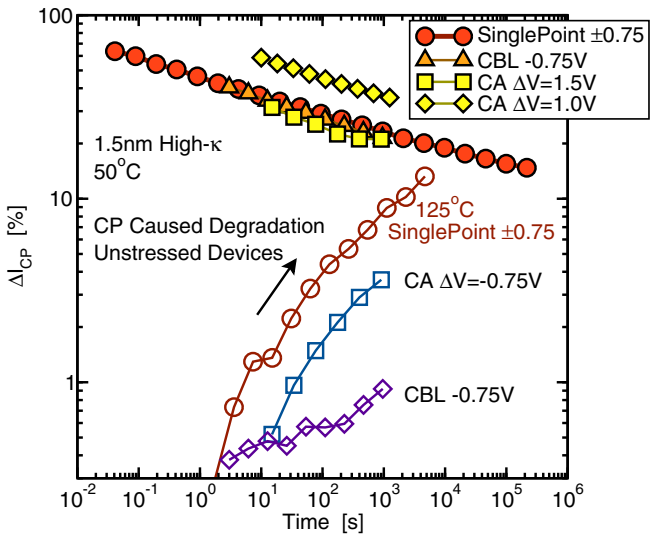

Fig. 13. All CP methods see the same recovery rate ($\Delta I_{\rm CP}$/decade), plotted relative to $I_{\rm CP,0}$. The single-point CP method is the fastest and gives the same $\Delta I_{\rm CP}$ as the full constant-base-level sweep (CBL). Comparable full constant-amplitude (CA) sweeps with proper ΔV also see the same $\Delta I_{\rm CP}$. Particularly at higher temperatures, CP measurements can lead to degradation.

and high-level of the charge pumping pulse, which according to the previous argument should contain the most effective recovery case. However, the $\Delta I_{\rm CP}$ recovery in the continuous CP measurement is even larger, implying that it is *accelerated by the pulsing event itself.*

Fig. 10 shows that $\Delta I_{\rm CP}$ at the beginning of delayed CP measurements follows the recovery of $\Delta V_{\rm th}$. Also, the amount of recovery induced by the CP measurement is mirrored in the recovery of $\Delta V_{\rm th}$. This data strongly suggests that both $\Delta V_{\rm th}$ and $\Delta I_{\rm CP}$ are at least partially related to the same microscopic defect, namely switching oxide traps [13, 24–26]. This conclusion is also confirmed by the frequency dependence of the $\Delta I_{\rm CP}$ recovery shown in Fig. 11 which gradually becomes smaller after long CP times, indicating that it is oxide defects which can be 'pumped-away'.

Further confirmation that defects visible in $\Delta V_{\rm th}$ react to CP measurements is given in Fig. 12: following a CP cycle, $\Delta V_{\rm th}$ shows reverse recovery due to slow oxide defects reaching their equilibrium occupancy after long times. Since this reverse recovery becomes smaller and smaller with increasing recovery time, it must be concluded that the defects responsible for reverse recovery can recover as well.

In [27] it was argued that the recovery of $I_{\rm CP}$ was not due to the actual recovery of interface states, but rather due to a reduction of the swing of the surface potential. Although we fail to see why the 'width of the CP hat' should be related to the density of interface states, we performed full constant amplitude and constant-base-level CP measurements in addition to the single-point CP experiments shown in the previous figures. The result is shown in Fig. 13 and quite reassuringly, the same level of degradation and recovery is observed in all measurements, except for the 1 V constant amplitude CP measurements, which use an amplitude too small to cover the whole bandgap. Fig. 13 also shows the detrimental impact of CP measurements which can lead to degradation and thus to artificial plateaus, cf. **(M5)**, particularly at higher temperatures.

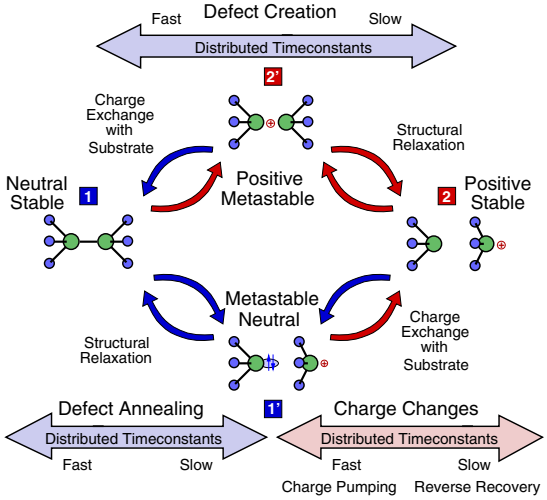

Fig. 14. The switching oxide trap model developed recently [13] is consistent with the experimentally observed behavior. Created defects can be switching traps and exchange charge with the substrate ($1' \leftrightarrow 2$), some of them faster, contributing to $\Delta I_{\rm CP}$ in a frequency-dependent manner, some of them very slow, thus causing reverse recovery or transient RTN [13], thereby contributing to $\Delta V_{\rm th}$ only (if in state 2).

VI. The Defect Model

Except for the newly discovered effect of CP-induced recovery, which will be discussed separately below, all features observed so far are consistent with the detailed defect properties identified using our recent time-dependent defect spectroscopy (TDDS) measurements [13, 15]. The microscopic model we use for the description of the defects is an extension of the switching trap model proposed by Lelis *et al.* [24] and shown in Fig. 14:

- The defects are switching traps, that is, have an energy-level in the Si bandgap. Prior to stress, the defect is in the neutral state 1, while stress transfers it into the positive state 2. Depending on the defect properties, the defect

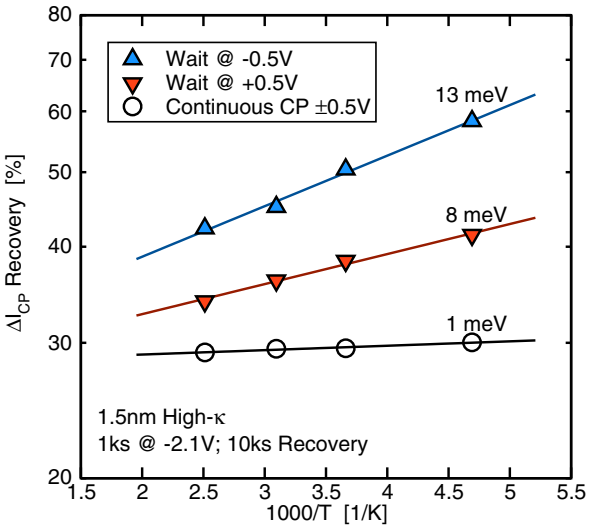

Fig. 15. Temperature dependence of the CP recovery: Continuous CP recovers ΔI_{CP} down to 30% of the post-stress ΔI_{CP}, nearly independently of temperature. Remaining at $-0.5\,\mathrm{V}$ maintains 60% at $-60\,^{\circ}\mathrm{C}$. Mere application of $+0.5\,\mathrm{V}$, the worst-case in a bias-driven interpretation, reduces ΔI_{CP} only down to 40% at $-60\,^{\circ}\mathrm{C}$.

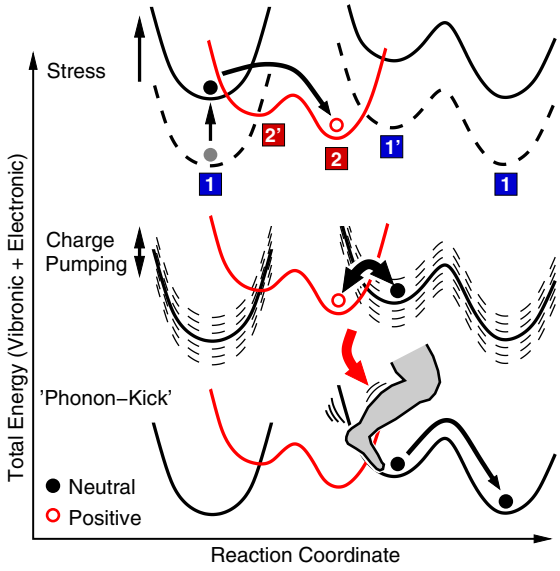

Fig. 16. Schematic illustration of the REDR effect: **Top**: During stress, defects become positively charged. **Middle**: During CP, the neutral level is rapidly moved up and down, causing frequent transitions. **Bottom**: The excess energy of the recombination events is deposited into the accepting mode, which leads to a reduction of the activation energy, known as the 'phonon-kick' or REDR effect [30].

may have a neutral metastable state $1'$, which provides the aforementioned energy-level in the bandgap. Transitions from 2 to $1'$ are particularly likely during switches toward accumulation or during CP measurements. While in $1'$, the defect is uncharged and thus not visible in ΔV_{th}.

- The defects may contribute to the CP signal in two ways. First, on an unstressed device, transitions between 1 and $2'$, provided they are sufficiently fast, can create recombination events. Contrary to interface states, whose contribution is temperature-independent, these switching traps will provide a larger contribution at higher temperatures. In particular, they will form the temperature-dependent tail of constant-base-level CP measurements. In addition, as shown in Fig. 13, CP can lead to degradation, which corresponds to a transition to state 2. Once in state 2, transitions between 2 and $1'$ can also contribute to I_{CP}, resulting in a temperature- and bias-dependent hysteresis of the CP curve [21].
- During NBTI stress, the defects move from the neutral state 1 to stable state 2. There, again, transitions between 2 and $1'$ can also contribute to I_{CP}, resulting in a temperature- and frequency-dependent contribution. This appears to be a significant contribution to ΔI_{CP} following NBTI stress.
- Since the time-constants are widely distributed, only the faster transitions between 2 and $1'$ can contributed to I_{CP}. As seen in Figs. 2, 3, 5, and 12, the slower states constitute the reverse recovery effect.
- The reason why these switching traps can contribute to both ΔV_{th} and ΔI_{CP} is simply because once created, these defects can be either positive (state 2) or neutral (state $1'$), depending on the Fermi-level, the former contributing to ΔV_{th}. This Fermi-level dependent defect occupancy also causes the change in the sub-threshold slope reported after NBTI stress [28, 29].

VII. Understanding CP-Induced Recovery

In order to understand the CP-induced recovery, we first studied its temperature dependence which is shown in Fig. 15. At all temperatures, wait phases at $-0.5\,\mathrm{V}$ show relatively weak recovery but a relatively strong temperature dependence. Compared to wait phases at $-0.5\,\mathrm{V}$, wait phases at $+0.5\,\mathrm{V}$ result in a stronger recovery but have a weaker temperature dependence. Finally, continuous CP measurements without an intermediate wait phases accelerate recovery down to 30% of the stress level after $10\,\mathrm{ks}$, nearly independent of temperature.

A possible explanation for this behavior is as follows: the CP measurement at $I_{\mathrm{CP}}^{\mathrm{max}}$ is designed to maximize the number of recombination events. Each event releases an energy of the order of the silicon bandgap. With 10^6 cycles per second, this accumulates to an *enormous amount of energy* which has to be *dissipated via phonons*. In due course, reactions near the defect site can be dramatically enhanced, a phenomenon known as the 'phonon-kick', or more recently as *recombination enhanced defect reaction (REDR)* [30, 31]. This is schematically shown in Fig. 16, where the REDR accelerates the transition from the neutral metastable state to the neutral equilibrium state. Following the arguments of Weeks *et al.* [31], the thermal transition rate from state 1 to state $1'$ of our switching trap model [13],

$$k_{1'1} = \nu\, \mathrm{e}^{-\beta \varepsilon_{1'1}} \tag{1}$$

with $\varepsilon_{1'1}$ as the thermal barrier separating the states $1'$ and $1'$, is replaced by $k_{1'1} + k^*_{1'1}$ with the enhanced rate

$$k^*_{1'1} = \nu^*\, \mathrm{e}^{-\beta(\varepsilon_{1'1} - \varepsilon^*)}, \tag{2}$$

with $\beta^{-1} = \mathrm{k_B}T$. From our experimental data we extract $\varepsilon^* \approx 60\,\mathrm{meV}$ and $\nu^* = 2.5 \times 10^{14}\,\mathrm{s}^{-1}$, with the convincing calibration result shown in Fig. 17. We remark that, as noted

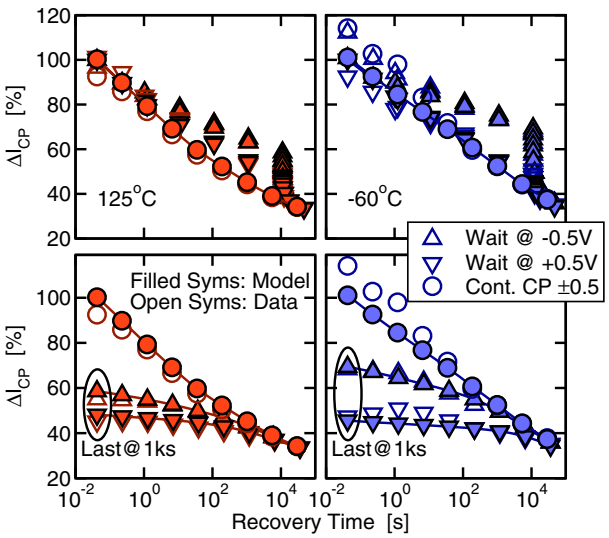

Fig. 17. Consideration of REDR in our switching trap model [3, 13] allows to reproduce the experimental data very well. **Top**: Data from Fig. 15 at 125 °C and −60 °C. **Bottom**: The last long recovery trace during continuous CP at 125 °C and −60 °C. The scatter particularly in the initial data (up to 20%) makes the model calibration challenging.

previously for $\Delta V_{\rm th}$ recovery [11], temperature-activated microscopic defect time constants again result in an apparent temperature-independent macroscopic behavior.

VIII. Special Case: Hydrogen-Rich Wafer

The log-like recovery of $\Delta I_{\rm CP}$ as for instance shown in Figs. 9 and 13 is clearly incompatible with the recovery predicted by the reaction-diffusion (RD) model [32], $(1+\sqrt{t_{\rm s}/t_{\rm r}})^{-1}$, which does not depend on bias or temperature [33]. A peculiar exception has been observed on a hydrogen-rich 30 nm SiO_2 split-wafer. Measuring $\Delta I_{\rm CP}$ only once per decade results in $\Delta I_{\rm CP} \sim$ const. By contrast, a continuous CP measurement produces recovery traces which bear a striking resemblance to the RD prediction, particularly for $t_{\rm s} = 10$ ks, see Fig. 18. After studying different stress times, however, we found the measured recovery to be practically independent of the stress-time, not scaling universally over $t_{\rm s}/t_{\rm r}$ as expected from RD theory [33]. Still, under continuous CP conditions, recovery could be a diffusion-limited process in this particular wafer. An intriguing feature is that after longer stress times the devices continue to degrade after the end of stress. This is consistent with the idea that hydrogen is released during stress which then depassivates interface states and creates oxide defects [34–37]. Otherwise, degradation after termination of the stress would not be possible. We remark that this is the standard model of irradiation damage [38, 39].

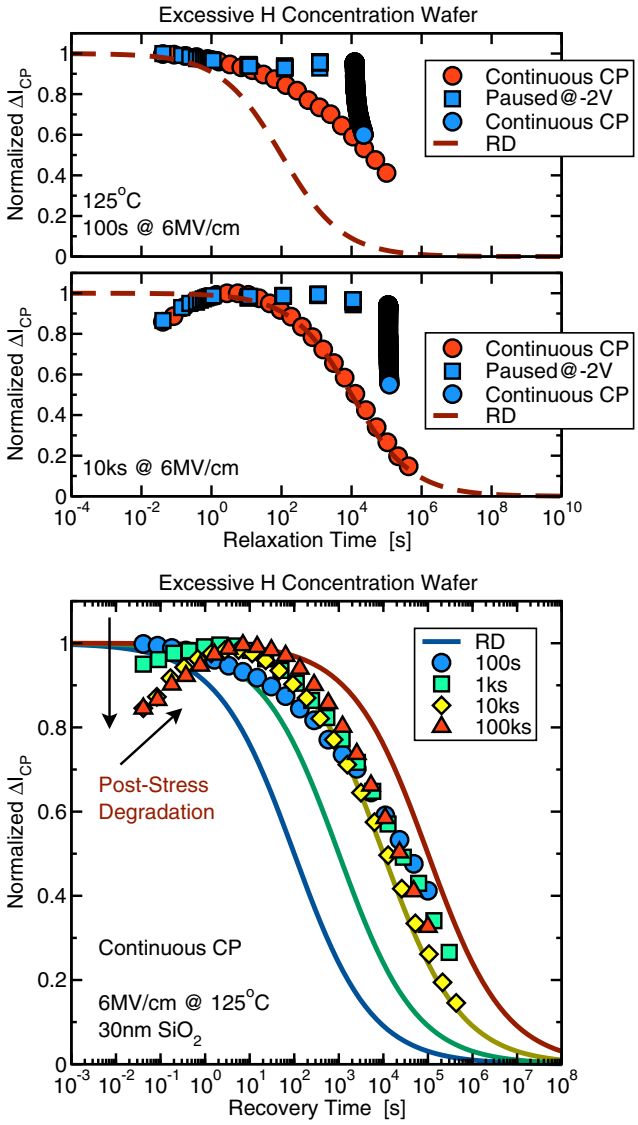

Fig. 18. **Top**: In an extremely H-rich wafer, recovery in $\Delta I_{\rm CP}$ is basically absent provided only occasional CP measurements are made. A diffusion-limited recovery behavior seems to dominate for continuous CP measurements. **Bottom**: The recovery rate appears to be roughly independent of the stress time, meaning that the hydrogen profile is only weakly disturbed during stress. With increasing stress, though, the initial degradation during the recovery phase can last for up to 10 s and amount to 20%. This is a degradation of $\Delta I_{\rm CP}$ which was found in this hydrogen-rich wafer only and must not be confused with the reverse recovery visible in $\Delta V_{\rm th}$ only, cf. Fig. 12.

IX. Conclusions

We have demonstrated that the plateaus occasionally observed in carefully tuned stress/recovery experiments consist of contributions from interface states as well as slower donor-like switching oxide traps. These plateaus are not permanent and normally not too well developed, making a precise definition and extraction difficult. In particular, the plateaus can be annealed by applying short positive bias pulses or, more effectively, by continuous CP measurements. Particularly the latter provides an efficient means for annealing NBTI degradation, likely due to a recombination enhanced defect reaction mechanism. Under normal recovery conditions, the recovery of $\Delta V_{\rm th}$ determines the starting level of $\Delta I_{\rm CP}$, which starts recovering quickly once CP measurements are performed. The latter demonstrates that oxide defects contribute to both $\Delta V_{\rm th}$ and $\Delta I_{\rm CP}$. Overall, considering P as permanent will lead to serious errors, even within conventional measurement windows. Finally, we have suggested a simple extension of our switching trap model to also account for recombination enhanced defect reaction (REDR) effects.

ACKNOWLEDGMENT

This work has received funding from the EC's FP7 grant agreement n°216436 (ATHENIS) and from the ENIAC MODERN project n°820379.

REFERENCES

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

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

[3] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel, "A Two-Stage Model for Negative Bias Temperature Instability," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, 2009, pp. 33–44.

[4] T. Grasser and B. Kaczer, "Evidence that Two Tightly Coupled Mechanism are Responsible for Negative Bias Temperature Instability in Oxynitride MOSFETs," *IEEE Trans.Electron Devices*, vol. 56, no. 5, pp. 1056–1062, 2009.

[5] T. Aichinger, M. Nelhiebel, and T. Grasser, "A Combined Study of p- and n-Channel MOS Devices to Investigate the Energetic Distribution of Oxide Traps after NBTI," *IEEE Trans.Electron Devices*, vol. 56, no. 12, pp. 3018–3026, 2009.

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

[7] A. Katsetos, "Negative Bias Temperature Instability (NBTI) Recovery with Bake," *Microelectronics Reliability*, vol. 48, no. 10, pp. 1655–1659, 2008.

[8] C. Benard, G. Math, P. Fornara, J. Ogier, and D. Goguenheim, "Influence of Various Process Steps on the Reliability of PMOSFETs Submitted to Negative Bias Temperature Instabilities," *Microelectronics Reliability*, vol. 49, pp. 1008–1012, 2009.

[9] H. Reisinger, O. Blank, W. Heinrigs, W. Gustin, and C. Schlünder, "A Comparison of Very Fast to Very Slow Components in Degradation and Recovery Due to NBTI and Bulk Hole Trapping to Existing Physical Models," *IEEE Trans.Dev.Mat.Rel.*, vol. 7, no. 1, pp. 119–129, 2007.

[10] T. Aichinger, M. Nelhiebel, and T. Grasser, "Unambiguous Identification of the NBTI Recovery Mechanism using Ultra-Fast Temperature Changes," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, 2009, pp. 2–7.

[11] H. Reisinger, T. Grasser, W. Gustin, and C. Schlünder, "The Statistical Analysis of Individual Defects Constituting NBTI and its Implications for Modeling DC- and AC-Stress," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, 2010, pp. 7–15.

[12] B. Kaczer, T. Grasser, P. Roussel, J. Franco, R. Degraeve, L. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, "Origin of NBTI Variability in Deeply Scaled PFETs," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, 2010, pp. 26–32.

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

[14] 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.

[15] T. Grasser, H. Reisinger, P.-J. Wagner, and B. Kaczer, "The Time Dependent Defect Spectroscopy for the Characterization of Border Traps in Metal-Oxide-Semiconductor Transistors," *Physical Review B*, vol. 82, no. 24, p. 245318, 2010.

[16] S. Rangan, N. Mielke, and E. Yeh, "Universal Recovery Behavior of Negative Bias Temperature Instability," in *Proc. Intl.Electron Devices Meeting (IEDM)*, 2003, pp. 341–344.

[17] T. Grasser, H. Reisinger, W. Goes, T. Aichinger, P. Hehenberger, P. Wagner, M. Nelhiebel, J. Franco, and B. Kaczer, "Switching Oxide Traps as the Missing Link between Negative Bias Temperature Instability and Random Telegraph Noise," in *Proc. Intl.Electron Devices Meeting (IEDM)*, 2009, pp. 729–732.

[18] 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.

[19] T. Aichinger, S. Puchner, M. Nelhiebel, T. Grasser, and H. Hutter, "Impact of Hydrogen on Recoverable and Permanent Damage following Negative Bias Temperature Stress," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, 2010, pp. 1063–1068.

[20] R. Paulsen and M. White, "Theory and Application of Charge-Pumping for the Characterization of Si-SiO_2 Interface and Near-Interface Oxide Traps," *IEEE Trans.Electron Devices*, vol. 41, no. 7, pp. 1213–1216, 1994.

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

[22] S. Mahapatra, K. Ahmed, D. Varghese, A. E. Islam, G. Gupta, L. Madhav, D. Saha, and M. A. Alam, "On the Physical Mechanism of NBTI in Silicon Oxynitride p-MOSFETs: Can Differences in Insulator Processing Conditions Resolve the Interface Trap Generation versus Hole Trapping Controversy?" in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, 2007, pp. 1–9.

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

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

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

[26] J. Ryan, P. Lenahan, T. Grasser, and H. Enichlmair, "Recovery-Free Electron Spin Resonance Observations of NBTI Degradation," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, 2010, pp. 43–49.

[27] M. Denais, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, D. Roy, and A. Bravaix, "Perspectives on NBTI in Advanced Technologies: Modelling & Characterization," in *Proc. ESSDERC*, 2005, pp. 399–402.

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

[29] D. Brisbin and P. Chaparala, "The Effect of the Subthreshold Slope Degradation on NBTI Device Characterization," in *Proc. Intl.Integrated Reliability Workshop*, 2008, pp. 96–99.

[30] H. Sumi, "Dynamic Defect Reactions Induced by Multiphonon Nonradiative Recombination of Injected Carriers at Deep Levels in Semiconductors," *Physical Review B*, vol. 29, no. 8, pp. 4616–4630, 1984.

[31] J. Weeks, J. Tully, and L. Kimerling, "Theory of Recombination-Enhanced Defect Reactions in Semiconductors," *Physical Review B*, vol. 12, no. 8, pp. 3286–3292, 1975.

[32] M. Alam, "A Critical Examination of the Mechanics of Dynamic NBTI for pMOSFETs," in *Proc. Intl.Electron Devices Meeting (IEDM)*, 2003, pp. 345–348.

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

[34] L. Tsetseris, X. Zhou, D. Fleetwood, R. Schrimpf, and S. Pantelides, "Physical Mechanisms of Negative-Bias Temperature Instability," *Appl.Phys.Lett.*, vol. 86, no. 14, pp. 1–3, 2005.

[35] S. Volkos, E. Efthymiou, S. Bernardini, I. Hawkins, A. Peaker, and G. Petkos, "The Impact of Negative-Bias-Temperature-Instability on the Carrier Generation Lifetime of Metal-Oxynitride-Silicon Capacitors," *J.Appl.Phys.*, vol. 100, no. 12, pp. 124 103–1–124 103–9, 2006.

[36] M. Houssa, V. Afanas'ev, A. Stesmans, M. Aoulaiche, G. Groeseneken, and M. Heyns, "Insights on the Physical Mechanism behind Negative Bias Temperature Instabilities," *Appl.Phys.Lett.*, vol. 90, no. 4, p. 043505, 2007.

[37] D. Danković, I. Manić, V. Davidović, S. Djorić-Veljković, S. Golubović, and N. Stojadinović, "Negative Bias Temperature Instability in n-Channel Power VDMOSFETs," *Microelectronics Reliability*, vol. 48, pp. 1313–1317, 2008.

[38] F. McLean, "A Framework for Understanding Radiation-Induced Interface States in SiO_2 Structures," *IEEE Trans.Nucl.Sci.*, vol. 27, no. 6, pp. 1651–1657, Dec 1980.

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