Refined NBTI characterization of arbitrarily stressed PMOS devices at ultra-low and unique temperatures

Thomas Aichinger\textsuperscript{a,}\textsuperscript{*}, Michael Nelhiebel\textsuperscript{b}, Tibor Grasserc

\textsuperscript{a}KAI, Europastraße 8, 9524 Villach, Austria
\textsuperscript{b}KAI, Europastraße 8, 9524 Villach, Austria
\textsuperscript{c}The Technical University of Vienna, Institute for Microelectronics, Gußhausstraße 27-29, 1040 Wien, Austria

\textbf{A B S T R A C T}

We reexamine degradation and recovery dynamics in the negative bias temperature instability (NBTI) of p-channel metal oxide semiconductor field effect transistors (PMOSFETs) by making use of the recently developed in situ polyheater technique. The capability of switching the device temperature extremely fast and almost arbitrarily allows for measuring differently stressed devices directly after the termination of stress at a unique and much lower characterization temperature (e.g. $-60^\circ$C). This procedure (‘degradation quenching’) is a powerful extension of the conventional measure–stress–measure (MSM) technique and provides a cleaner way for comparing threshold voltage shifts and charge pumping (CP) currents of arbitrarily stressed devices. We find that increasing the stress bias predominantly activates a larger number of defects with similar (short) recovery time constants causing steeper threshold voltage recovery transients after the termination of stress. Increasing the stress temperature has a very similar effect on the threshold voltage shift as increasing the stress time. In both cases, defects with larger recovery time constants are activated while the number of defects with short recovery time constants remains essentially the same. A comparison of $V_{TH}$ shift and CP data suggests that the total threshold voltage shift is due to at least two fundamentally different types of defects, one being readily recoverable and uncorrelated to the CP current while the other is ‘quasi-permanent’ and proportional to the CP current. By converting CP currents into corresponding threshold voltage shifts, we find that only about 50% of the ‘quasi-permanent’ $V_{TH}$ damage is due to slowly-recoverable interface states. The remaining fraction is due to another, yet undefined, positively charged defect generated at virtually the same rate.

\textsuperscript{*}Corresponding author. Present address: Infineon Technologies AG, Siemensstrasse 2, 9500 Villach, Austria. Tel.: +43 (0)5 1777 18602; fax: +43 (0)5 1777 6887.

\textsuperscript{E-mail addresses: thomas.aichinger@infineon.com (T. Aichinger), michael.nelhiebel@infineon.com (M. Nelhiebel), grasser@iue.tuwien.ac.at (T. Grasserc).
exponentially harder to perform, CP data is typically presented with much larger time delays compared to \(V_{nn}\) shift data. In any case, a bias change and initialization of a subsequent (recovery) measurement involves an unavoidable stress-measure time delay, which results in unknown recovery prone to distort already the first measurement point after the termination of the stress. Because of the thermal activation of relaxation, the amount of \(V_{nn}\) recovery within this time delay and the following recovery transients depend on the characterization temperature [13]. Similar conclusions hold for ‘recovery-free’ on-the-fly measurements, where the initial degradation in the first reference measurement point depends on the temperature as well [14].

In this work we characterize NBTI degradation and recovery of PMOS devices stressed at different temperatures and fields. As opposed to previous attempts, which were limited to the constraint that the stress temperature must equal the recovery temperature, we decouple the stress and the recovery phase in the MSM experiment by making use of the in situ polyheater technique [15,16]. In-situ polyheaters are deposited poly-Si wires which embed the active device. Our key approach is to provide identical characterization conditions within a stress-measure time delay of only 10 ms for devices stressed at almost arbitrary fields and temperatures. Using this clean characterization technique, we reexamine NBTI dynamics of threshold voltage shift and CP current and evaluate the relative contributions of different defect types involved in threshold voltage degradation and recovery.

The study is performed on isolated lateral p-channel MOSFETs equipped with 30 nm pure SiO\(_2\) gate oxides with n++ doped poly-Si on top. Beside the economical and scientific relevance of high voltage devices [17], thick oxide transistors provide considerable benefits for NBTI characterization. For example, due to the absence of gate tunneling currents at conventional stress biases \((j_c < 10^{-7} A/cm^2)\), we do not have to correct our CP data for leakage currents or worry about non-NBTI related degradation components due to e.g. fixed oxide charges caused by impact ionization. The use of pure SiO\(_2\) gate oxides further guarantees that our general conclusions are not distorted by the strongly process dependent impact of nitridation [18]. Our main conclusions should hold for thin dielectric transistors as well since it has recently been shown by several independent studies that the basic mechanisms of NBTI are essentially the same in thin and thick SiO\(_2\) and SiON technologies [19–22].

2. The role of temperature in NBTI characterization

All NBTI models agree that the temperature plays a significant role in both defect creation and relaxation. This and other arguments rule out defect neutralization by elastic tunneling which is almost temperature independent. On the other hand, diffusion or inelastic tunneling processes are considerably accelerated at elevated device temperatures. Assuming an Arrhenius-like recovery process [23,24], we can define the annealing time of a positively charged defect at a recovery temperature \(T\) by

\[
\tau = \tau_0 \exp \left( \frac{\Delta E_a}{kT} \right)
\]  

(1)

In (1), \(\tau\) is the annealing time of an arbitrary trap at a temperature \(T\), \(\tau_0\) is the minimum time constant at barrier height zero, \(\Delta E_a\) is an effective relaxation barrier, \(k\) is the Boltzmann constant, and \(T\) is the device temperature during recovery. To determine the degradation level at the very beginning of the recovery phase and, consequently, the exact value of \(\tau_0\), sophisticated test methods have been developed in the past that accomplish \(\Delta V_{nn}\) measurements only a couple of microseconds after the removal of the stress bias [25]. In most reported cases the extracted recovery curves are straight lines in a semi-logarithmic time plot, suggesting that the activation energies are more or less uniformly distributed [26,27]. On small devices, discrete steps can be observed which are due to only a handful of defects with a few characteristic annealing time constants [28,29]. The fact that in general recovery starts almost immediately after the termination of stress [25,30–32] attaches considerable importance on the transition event between the stress and the relaxation phase. Eq. (1) and the discussion above suggest that it would be most expedient to either measure as fast as possible or at the lowest temperature possible (or both) in order to capture a maximum of the actual damage at the end of stress. Note that if the recovery temperature (i.e. the characterization temperature) is much lower than the stress temperature, recovery would be decelerated significantly. The exponential decrease of the recovery time constants with temperature can be interpreted as a stretching on the time axis [23,33]. For example, let’s assume traps with an average barrier height \(\Delta E_a\) of 0.3 eV. If on such a device the \(\Delta V_{nn}\) shift is measured with a MSM time delay of 10 ms at a temperature as low as \(-60\) °C, a larger fraction of the degradation can be captured compared to a measurement of \(\Delta V_{nn}\) with a time delay of only 1 \(\mu s\) at a much higher (stress) temperature of 125 °C [16,33].

In summary, a clean comparison of devices stressed at different temperatures is challenging due to the strong temperature dependence of the recovery [23]. Data recorded with the same time delay but at different temperatures is afflicted with different recovery [16,23,33]. Additional theoretical difficulties with different stress and characterization temperatures arise when attempting to correlate the measured \(\Delta V_{nn}\) shift to CP data. The CP technique, traditionally used to characterize fast interface states [34–36], is inefficient at conventional stress temperatures since it covers just a narrow range of the silicon band gap when performed at high device temperatures. To convert the CP current to a corresponding \(V_{nn}\) shift at such high temperatures, the standard practice is to assume a flat density-of-states profile for the interface traps. However, this approach is known to be a very crude approximation [37–39] which likely introduces a considerable error into CP data measured at different temperatures. Similar difficulties also arise for \(V_{nn}\) shift data recorded at the same operating point but at different characterization temperatures [16,33].

To overcome all these experimental and theoretical difficulties, devices stressed at different stress temperatures must be characterized directly post stress at the same characterization/recovery temperature. To accomplish this, one needs a tool/technique that allows for ultra-fast temperature switches. Such a tool is the recently developed polyheater measurement technique [15,16,33]. Using in situ heated devices on test structures (i.e. devices embedded between electrically-isolated poly-Si wires), we can generate different stress temperatures by applying different heater powers. The ambient (characterization) temperature is defined by the ground temperature of the thermo chuck. During stress, the chuck temperature is constant and typically set to a much lower value (e.g. \(-60\) °C) than the stress temperature. At the end of stress, the heater is turned off and the device cools down within 1–10 s [15]. After 0.1 s, the target temperature is reached within 3% accuracy, after 1 s within 1%, and after 10 s, the device temperature is less than 0.1% away from the target temperature (i.e. the chuck temperature of \(-60\) °C). The time required for cooling the device from typical stress temperatures down to \(-60\) °C within 0.1% accuracy is always about 10 s. During temperature quenching, the stress field (bias) is maintained. Since recovery is dominantly triggered by the switch from the stress bias to the read-out bias (i.e. the threshold voltage), maintaining the stress bias conserves the degradation level during the short cooling phase [16,23,33]. This procedure is called ‘degradation quenching’. Note that since the stress bias remains applied during the cooling phase, there is
The stress biases correspond to equivalent oxide electric fields of $/C_0/$. Devices were stressed at three different gate biases ($V_{GS}$) and temperatures ($T_{Chuck}$). During stress, the polyheater provides the stress temperature. 10 s before the recovery measurement is initiated ($t_{CP}$), the heater is turned off and the device quickly cools down to a constant recovery temperature of –60 °C. Only then the gate bias is switched from the stress level to the read-out level, and this is the moment where recovery starts. Our first measurement is typically afflicted with a 10 ms recovery time delay. We subjected each device to four stress/recovery runs with increasing stress ($t_S$) and recovery ($t_R$) durations (1/10/100/1000 s). The CP cycles $t_{CP}$ and $t_{RP}$ and the recovery phase $t_R$ each lasted for 0.5 s independently of $t_S$ and $t_R$. The only difference between experiment (1) and experiment (2) is that the initial CP cycle right after stress ($t_{CP}$) was omitted in (1) and included in (2) (for details see text).

no recovery during cool-down, although some very small additional low-temperature degradation cannot be ruled out. This very small stress contribution, however, can be safely ignored compared to the much larger degradation during the actual stress phase performed at a much higher temperature [16,23,33]. This is demonstrated in Fig. 1. Different PMOS devices were stressed for 1000 s at 125 °C and 6.0 MV/cm. After the stress time had elapsed, the heater was turned off allowing the devices to cool down to the chuck temperature of –60 °C (with stress bias applied). The delay between turning off the heater and switching the gate bias from the stress level to $V_{TH}$ is referred as time $t_D$. The true recovery time delay is the time between switching the gate bias from the stress level to $V_{TH}$ and performing the $\Delta V_{TH}$ measurement. In Fig. 1, this time delay is labeled as ‘time since end of stress’. Once the device can reach the chuck temperature (–60 °C) with sufficient accuracy (0.1%) before the measurement is initialized (>10 s after switching off the heater), the $\Delta V_{TH}$ measurements can reasonably be assessed at the same temperature and the obtained shifts and recovery traces are seen to be independent of $t_D$. This demonstrates the conservation of the degradation level during $t_D$: even with a 1000 s delay, neither more nor less shift is seen compared to a 1 s delay, which means that $V_{TH}$-relevant defects neither appeared nor disappeared during $t_D$. More detailed information on calibration and application of in situ polyheaters is given in [15,16,40].

3. Experimental setup

We designed the two MSM sequences illustrated in Fig. 2 to measure degradation and recovery dynamics of threshold voltage shift and CP current. We applied each sequence to a separate set of PMOS devices. In order to achieve a good correlation of the degradation levels after stress, devices from the same wafer were carefully selected by their virgin CP characteristics. This was done because preliminary investigations have indicated a good correlation between the degradation levels observed in equally fabricated devices provided they had identical CP characteristics before stress [16,41]. Each device set consisted of nine test structures, which were stressed at three different gate biases (–18.0/–16.0/–14.0 V) and three different stress temperatures (125/100/75 °C). The stress biases correspond to equivalent oxide electric fields of 5.6/5.0/4.3 MV/cm. During LIBTS, the stress temperature was generated by applying a certain (pre-calibrated) power to the polyheater. We adjusted a low and unique characterization (i.e. recovery) temperature for all samples by setting the thermo chuck to 60 °C. This is demonstrated in Fig. 1. Different PMOS devices were stressed for 1000 s at 125 °C and 6.0 MV/cm. After the stress time had elapsed, the heater was turned off allowing the devices to cool down to the chuck temperature of –60 °C (with stress bias applied). The delay between turning off the heater and switching the gate bias from the stress level to $V_{TH}$ is referred as time $t_D$. The true recovery time delay is the time between switching the gate bias from the stress level to $V_{TH}$ and performing the $\Delta V_{TH}$ measurement. In Fig. 1, this time delay is labeled as ‘time since end of stress’. Once the device can reach the chuck temperature (–60 °C) with sufficient accuracy (0.1%) before the measurement is initialized (>10 s after switching off the heater), the $\Delta V_{TH}$ measurements can reasonably be assessed at the same temperature and the obtained shifts and recovery traces are seen to be independent of $t_D$. This demonstrates the conservation of the degradation level during $t_D$: even with a 1000 s delay, neither more nor less shift is seen compared to a 1 s delay, which means that $V_{TH}$-relevant defects neither appeared nor disappeared during $t_D$. More detailed information on calibration and application of in situ polyheaters is given in [15,16,40].

### 4. Results and general observations

The $V_{TH}$ shifts recorded during $t_R$ in experiment (1) are illustrated in Fig. 3. For clarity, we have grouped different stress biases and stress temperatures in nine separate graphs on the left-hand side of Fig. 3. There are four separate measurement curves in each graph corresponding to the four subsequent stress/recovery runs (1/10/100/1000 s). The $V_{TH}$ shift increases with stress time, stress bias and stress temperature. Within the observed experimental window, all recovery traces show a perfectly linear decrease on the semi-logarithmic time plot. Two features define the shape of the recovery traces: (i) the slope (mV/decade) and (ii) the offset at an arbitrary time. Considering that every recoverable defect...
particular stress temperature but different stress fields. The field-scaling factors for different stress temperatures are very similar. By appropriate scaling factors (factors are displayed in the lower right corner of each figure), perfect scalability can be obtained for devices which were stressed at one particular stress temperature but different stress fields. The field-scaling factors for different stress temperatures are very similar. $V_{TH}$ shifts recorded for different stress times and different stress temperatures are shifted parallel on the $y$-scale.

Fig. 3. $\Delta V_{TH}$ recovery traces measured during $t_e$ at $-60$°C in experiment (1). Nine different devices were stressed for 1/10/100/1000 s using three different stress temperatures (125/100/75 °C) and three different oxide fields (5.6/5.0/4.3 MV/cm). The results for different stress fields are illustrated from left to right, while different stress temperatures are illustrated as diamonds (125 °C), circles (100 °C) and triangles (75 °C) from the top to the bottom. On the right hand side, we show the raw data multiplied by appropriate scaling factors (factors are displayed in the lower right corner of each figure). Perfect scalability can be obtained for devices which were stressed at one particular stress temperature but different stress fields. The field-scaling factors for different stress temperatures are very similar. $V_{TH}$ shifts recorded for different stress times and different stress temperatures are shifted parallel on the $y$-scale.

Fig. 4. Recovery rates (slopes of the recovery curves per decade) extracted from the data of Fig. 3. Different stress temperatures are illustrated as diamonds (125 °C), circles (100 °C) and triangles (75 °C). The stress field has a significant influence on the recovery rate. Stress time and stress temperature leave the recovery rates almost unaffected. However, in the classical MSM approach the scaling factors did not coincide for different stress temperatures. In Fig. 3, devices stressed at different temperatures at the same stress bias show an offset in their $V_{TH}$ recovery traces. This is different from results obtained by conventional MSM measurements which showed scalability for different stress biases and different stress temperatures [6,7]. The discrepancy is most likely due to the different characterization temperatures in the classical MSM approach which are known to affect the recovery rates.

Experiment (2) was performed to check whether recovery of fast interface defects (defects which contribute to 500 kHz CP) plays a significant role in the $V_{TH}$ recovery illustrated in Fig. 3. Therefore, we compared CP currents recorded during $t_{CP}$ in experiment (1) to CP currents recorded during $t_{CP}$ in experiment (2). Between $t_{CP}$ and $t_{CP}$, the $V_{TH}$ recovery was measured at a constant bias ($t_e$). Changes in the CP current ($\Delta I_{CP}$) after different stress times are illustrated in Fig. 5a and b. The results for a stress temperature of 125 °C and a stress field of 5.6 MV/cm are representative for all other stress conditions which show similar characteristics and lead to the same conclusions. Remarkably, both experiments show very similar $\Delta I_{CP}$ values, although the CP measurement of experiment (1) is afflicted with a much larger stress-measure time delay (i.e. $t_e$). This indicates that the type of defect which contributes to CP is ‘quasi-permanent’ during $t_e$ at $-60$ °C. Thus, this defect type cannot be responsible for the obtained $V_{TH}$ recovery illustrated in Fig. 3. In the following, we denote defects which do not anneal within the scope of the performed recovery experiment as ‘quasi-permanent’. We remark that this does not imply that these defects are permanent forever (or at all temperatures) [5,42,43]; however, such defects would appear permanent in comparison to other defects which recover on a much shorter time scale [5,42].

The finding that the defects which contribute to CP are ‘quasi-permanent’ during constant bias recovery is in contradiction to results obtained by on-the-fly charge-pumping (OTFT).
measurements reported in [44]. The disagreement could be a consequence of our low characterization temperature (−60 °C). Other possible reasons are given in [45].

5. Correlation between fast interface states and the total \( V_{TH} \) Shift

Having measured time, bias, and temperature dependent \( V_{TH} \) shifts as well as CP current degradation and recovery at identical characterization conditions, we proceed to check whether there is any correlation between the number of defects contributing to the CP signal and the total \( V_{TH} \) shift. To compare these two quantities, one has to convert the increase in the CP current (\( \Delta I_{CP} \)) into an equivalent threshold voltage shift due to fast interface states (\( \Delta V_{TH} \)). Similar attempts have been made by others, however, previous studies were always bound to the constraint that the analyzing temperature equals the stress temperature, the implications of which being quite significant, as will be shown in the following.

To accomplish the conversion from \( \Delta I_{CP} \) into \( \Delta V_{TH} \), it is crucial to consider that the \( V_{TH} \) shift measurement (performed at DC gate bias) and the CP measurement (performed with pulsed gate bias) are in general probe different ranges of the silicon band gap (\( \Delta V_{TH} \) and \( \Delta E_{CP} \)). Furthermore, one has to establish the amphoteric nature of the fast interface traps contributing to CP [46,47]. In Si/SiO\(_2\) systems these traps are most likely \( P_0 \) centers [26]. \( P_0 \) centers have their amphoteric transition level around mid-gap, i.e. the intrinsic Fermi level (\( E_F \)) [48,49]. Traps above \( E_F \) are considered as acceptor-like while traps below \( E_F \) are treated as donor-like. When measuring the \( V_{TH} \) shift at a particular read-out gate bias (\( V_{GR} \)), the Fermi level position (\( E_F \)) at the interface governs the charge state of the traps. At \( V_{GR} = −1.1 \text{V} \) and at −60 °C, \( E_F \) was simulated numerically (process simulator TSUPREM4; device simulator MEDICI) to be about \( E_F + 120 \text{meV} \) [50]. Due to the amphoteric nature of \( P_0 \) centers, the states in the lower half of the silicon band gap (between \( E_F \) and \( E_F \)) are positively charged; traps above mid-gap are neutral − \( \Delta E_{TH} = E_F − E_F \).

Following [35], we can also calculate the probed energy range during CP (\( \Delta E_{CP} \)). At −60 °C and using our pulse setup, we obtain \( \Delta E_{CP} = 800 \text{meV} \), where we have assumed an energetically homogeneous capture cross section of \( \sigma = 10^{-13} \text{cm}^2 \) [51,52]. The lower boundary of \( \Delta E_{CP} \) (i.e. \( E_F + 150 \text{meV} \)) almost perfectly coincides with the Fermi level position at \( V_{GR} \) (i.e. \( E_F + 120 \text{meV} \)). That is a particular benefit of our low characterization temperature (−60 °C). At typical stress temperatures (100 °C −75 °C) and using the same pulse setup, \( \Delta E_{CP} \) probes a considerably narrower fraction of the silicon band gap (500 meV −300 meV). This leads to a significant mismatch between \( \Delta E_{CP} \) and \( \Delta E_{TH} \). To account for the mismatch in the conversion, one has to assume a particular density-of-states profile. Using a characterization temperature of −60 °C, such an assumption is not required because \( \Delta E_{CP} \) probes nearly the whole range of \( \Delta E_{TH} \). Since we used a symmetrical pulse shape, \( \Delta E_{CP} \) also probes about the same energy range in the upper half of the silicon band gap [35]. To exclude the neutral defect states above mid-gap in the conversion, we divide the CP current by a factor 2. This is reasonable since the density-of-states profile of fast interface traps is to a very good approximation symmetric around mid-gap [48,49,53]. The threshold voltage shift due to fast interface states (\( \Delta V_{TH} \)) is then given by

\[
\Delta V_{TH}^{\text{RT}} \approx \Delta Q_g \frac{C_{ox}}{2f\text{AC}_{ox}} \Delta I_{CP}
\]

In (2), \( \Delta Q_g \) is the increase in positive interface charge, \( C_{ox} \) is the specific oxide capacitance, \( A \) is the device area, \( \Delta I_{CP} \) is the increase of the maximum CP current and \( f \) is the gate pulsing frequency. Using (2), we can calculate \( \Delta V_{TH}^{\text{RT}} \) from \( \Delta I_{CP} \) for every stress bias, stress temperature and stress time.

Fig. 6 illustrates an attempt to scale \( \Delta V_{TH}^{\text{RT}} \) and \( \Delta I_{CP} \). We compare \( \Delta V_{TH}^{\text{RT}} \) recorded 10 ms post stress in experiment (1) to \( \Delta I_{CP} \) recorded (via CP) 10 ms post stress in experiment (2). In (a), \( \Delta V_{TH}^{\text{RT}} \) and \( \Delta I_{CP} \) are illustrated as a function of the stress time for all tested stress fields and stress temperatures. Both shifts show a power-law increase with stress time (\( t_S \)): \( \Delta V_{TH}^{\text{RT}} \sim t_S^n \) and \( \Delta I_{CP} \sim t_S^m \), with \( n_H \) and \( n_{ac} \) being the characteristic power law exponents. Two different sets of scaling factors had to be used to scale \( \Delta V_{TH}^{\text{RT}} \) and \( \Delta I_{CP} \) individually for different stress conditions, c.f. (b) and (c). By multiplying the scaled \( \Delta V_{TH}^{\text{RT}} \) shifts with a factor of eight, scalability can be obtained for stress time of 1000 s but fails for all shorter stress times. The fact that \( \Delta V_{TH}^{\text{RT}} \) and \( \Delta I_{CP} \) are not simply proportional to each other indicates that they are most likely not coupled in a direct way.

In Fig. 7, the extracted power law exponents of \( \Delta V_{TH}^{\text{RT}} \) \( (n_{ac}) \) and \( \Delta I_{CP} \) \( (n_H) \) are illustrated for different stress temperatures and stress fields. Apparently, \( n_{ac} \) and \( n_H \) differ by almost a factor of two, the first being around 0.22 while the latter is around 0.12. The different degradation dynamics emerge because in DC \( V_{TH} \) measurements (Fig. 3), different types of traps contribute to the power law exponent \( (n_{ac}) \) whereas in CP measurements performed at high frequencies (500 kHz) and with moderate pulse amplitudes, the majority of contributing traps are fast interface states \( (n_H) \). This finding is consistent with the results of Teo et al. [54] on p-channel MOSFETs with 2.8 nm SiO\(_2\) gate dielectrics. Using CP with ultra-fast 100 ns pulses, they found a power law exponent of roughly 0.15 (c.f. 0.12). Using a conventional CP MSM sequence, they reported a considerably reduced CP current and a different
amplitudes, e.g. when the pulse levels are re-used as stress bias. The same conclusions were also drawn by Hehenberger et al. in [45]. Since we perform CP at −60 °C applying 500 kHz gate pulses with only 3.0 V pulse amplitudes to 30 nm dielectrics, our CP measurements can be considered “conventional” and virtually unaffected by near interfacial oxide traps. Most of these traps recover either within the switching time delay between the end of DC stress and the start of AC gate pulsing during which the gate is floating or during the CP measurement itself.

Within the range of our experiment (75–125 °C), the exponents \(n_\text{it} \) and \(n_\text{tot} \) appear almost independent of the stress temperature and the stress field. While independence of the electric field was already observed by others, independence of the stress temperature is in contradiction to most previous studies [43,55] that suggested a \(kT\)-like increase of \(n_\text{it} \) with the stress temperature. Teo et al. [54] attributed the apparent variations in the power-law exponents with temperature to an artifact due to different activation energies of oxide and interface trap generation. Another reason could be the temperature dependence of the scanned energy range during CP [35]. In our experiments, we measure CP only 10 ms post stress always at the same low characterization temperature while others compare CP currents measured with larger time delay at different and much higher temperatures where different regions of the silicon band gap (Δ\(E_G\)) are probed. It has been shown [41,42] that long continuous gate pulsing at elevated temperatures leads to enhanced recovery of the CP current, a phenomenon possibly linked to a process known as recombination enhanced defect reaction [56,57].

To check whether the reported larger values of \(n_\text{it} \) and its strong temperature \(V_{TH} \) dependence are measurement artifacts due to different characterization temperatures, we have repeated experiment (2) on our devices using the conventional MSM approach, where the stress temperature equals the recovery temperature (stress field 5.6 MV/cm). The results are compared to our original data in Fig. 8. Consistent with our hypothesis, the conventional MSM technique gives completely different power law exponents for \(n_\text{it} \) which also increase with temperature, just as reported in literature [43,45,58,59].

All evidence collected in these experiments suggests that the total \(V_{TH} \) shift must be due to at least two or more different types of defects which have different bias and temperature acceleration as well as different recovery behavior. One of them recovers readily as a function of bias, temperature and time while the other appears ‘quasi-permanent’ within the scope of our experiment. Consistent with previous studies [5,43,62], we believe that the ‘quasi-permanent’ fraction in the total shift is related to the defects detected in

power law exponent of 0.25−0.27 (c.f. 0.22). The discrepancy was attributed to the contribution of fast interfacial oxide traps which can constitute a significant contribution to the CP current at high temperatures, low gate pulsing frequencies, and large gate pulsing
CP, while the fast recoverable fraction is of somewhat different origin. To retrieve the ‘quasi-permanent’ component ($\Delta V_{\text{TH}}^{\text{perm}}$) from the total $V_{\text{TH}}$ shift ($\Delta V_{\text{TH}}^{\text{tot}}$), we have the option to either monitor constant bias recovery at $V_{\text{TH}}$ until the curves flatten, which may take more than $10^{10}$ seconds [42,45,60], or try to accelerate the recovery by providing electrons at the interface [6,32,61–62], i.e. biasing the device for a short time in accumulation. Practicability dictates the use of the latter method. During CP, the gate bias is switched periodically from inversion to accumulation. After CP, we obtain a considerably smaller and virtually constant $\Delta V_{\text{TH}}$ degradation level (not shown). This suggests that most recoverable components (with short time constants) have been annealed. It should be noted, however, that recovery will eventually continue after a long enough time interval [23,33].

In Fig. 9 we compare the ‘quasi-permanent’ $V_{\text{TH}}$ shift ($\Delta V_{\text{TH}}^{\text{perm}}$), recorded during $t_e$ in experiment (1), to $\Delta V_{\text{TH}}^{\text{tot}}$ calculated from the CP current in experiment (2). In (a), $V_{\text{TH}}^{\text{perm}}$ and $V_{\text{TH}}^{\text{tot}}$ are illustrated as a function of the stress time for all analyzed stress biases and stress temperatures. We obtain almost exactly the same bias and temperature scaling factors for $\Delta V_{\text{TH}}^{\text{perm}}$ and $\Delta V_{\text{TH}}^{\text{tot}}$. (c) In (c), the scaled $V_{\text{TH}}$ shifts are illustrated as a function of the stress time. The evolutions of $\Delta V_{\text{TH}}^{\text{perm}}$ and $\Delta V_{\text{TH}}^{\text{tot}}$ are almost parallel, indicating similar power law factors and a direct correlation between the two components. However, apparently the magnitude of $\Delta V_{\text{TH}}^{\text{perm}}$ and $\Delta V_{\text{TH}}^{\text{tot}}$ is not the same. In fact, $\Delta V_{\text{TH}}^{\text{tot}}$ is universally about a factor 2.5 smaller than $\Delta V_{\text{TH}}^{\text{perm}}$. If $\Delta V_{\text{TH}}^{\text{perm}}$ would be exclusively due to defects which also contribute to the CP current, $\Delta V_{\text{TH}}^{\text{perm}}$ and $\Delta V_{\text{TH}}^{\text{tot}}$ would be identical which would give a scaling factor of one. The actually obtained scaling factor of 2.5 indicates that the defects detected in CP only account for about 50% of the total ‘quasi-permanent’ damage in the $V_{\text{TH}}$ shift. This suggests that $\Delta V_{\text{TH}}^{\text{perm}}$ most likely includes another type of positively charged defect which is tightly related to fast interface states but does not show up in the CP measurement [63–65].

Alternatively, one may argue that $\Delta V_{\text{TH}}^{\text{tot}}$ could have been underestimated in (2) by making the assumption of an amphoteric nature of fast interface traps. In fact, $\Delta V_{\text{TH}}^{\text{perm}}$ would be about a factor of two larger when assuming pure donor-like interface defects which cover the whole silicon band gap [18]. However, such an assumption would be in contradiction with most of the defect literature. For example, it has been shown that fast interface states charge up negatively in NMOS devices, where the Fermi level is pinned close to the conduction band edge at the threshold voltage, causing a net smaller or even positive threshold voltage shift after NBTS [2,44].

6. Discussion on the defect structure

Our results indicate three different types of NBTS induced defects: (i) a defect which recovers readily as a function of bias and temperature, (ii) another positive defect which cannot be annealed easily by biasing the device in accumulation and which does not contribute to CP at $-60 \, ^\circ\text{C}$ and 500 kHz; this defect is considered ‘quasi-permanent’ since it has recovery time constants beyond the range of our experiments and (iii) fast interface states that contribute to CP; these defects were found ‘quasi permanent’ too, at least as long as the device is biased at the negative threshold voltage at $-60 \, ^\circ\text{C}$. We could further establish a universal correlation between defect type (ii) and defect type (iii). The defects are created at virtually the same rate and in a ratio of approximately 50:50 [5].

While electrical measurements can study densities, energy levels and time constants of traps, they do not provide information on the microscopic nature of point defects. Electron paramagnetic resonance (EPR) [66] and electrically detected magnetic resonance (EDMR) [67] have the analytical power to study the atomistic structure of defects. EPR and EDMR studies of irradiated Si/SiO$_2$ systems have revealed $E'$ and $P_5$ centers [68,69]. The $E'$ center is an oxide defect which consists of a dangling bond on a silicon atom and an adjacent positive defect in a puckered configuration (the $E'_1$ center) [70,71]. The $P_5$ center is a dangling bond defect on a silicon, located at the semiconductor/oxide interface. $P_6$ centers are conventionally assumed to be created through hydrogen release from passivated Si–H bonds at the interface. Recent EPR investigations [72] and EDMR studies via spin dependent recombination [73,74] have shown that similar defect types also emerge after NBTS in significant portion. In the following, we attempt to relate the three different defect types to defects identified with EPR and EDMR. For better readability, the different defect types, their recovery behavior, their appearance in the electrical measurement as well as their possibly related point defect are summarized in Table 1. Defect (i): It is known from random telegraph noise and EPR experiments that near interfacial oxide defects like $E'$ centers can act as a switching traps which can be either positively charged or neutral, depending on the Fermi level position at the interface [75,76]. A microscopic degradation/recovery model developed by

---

**Table 1**

<table>
<thead>
<tr>
<th>Scaling factors</th>
<th>$\Delta V_{\text{TH}}^{\text{perm}}$</th>
<th>$\Delta V_{\text{TH}}^{\text{tot}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.6 MV/cm</td>
<td>$\times 1.00$</td>
<td>$\times 1.00$</td>
</tr>
<tr>
<td>5.0 MV/cm</td>
<td>$\times 1.65$</td>
<td>$\times 1.65$</td>
</tr>
<tr>
<td>4.3 MV/cm</td>
<td>$\times 2.85$</td>
<td>$\times 2.85$</td>
</tr>
<tr>
<td>125°C</td>
<td>$\times 1.00$</td>
<td>$\times 1.00$</td>
</tr>
<tr>
<td>100°C</td>
<td>$\times 1.50$</td>
<td>$\times 1.50$</td>
</tr>
<tr>
<td>75°C</td>
<td>$\times 2.20$</td>
<td>$\times 2.20$</td>
</tr>
</tbody>
</table>

---

**Fig. 9.** A scaling attempt of $\Delta V_{\text{TH}}^{\text{perm}}$ (full symbols) and $\Delta V_{\text{TH}}^{\text{tot}}$ (open symbols). The unscaled data, obtained 10 ms post stress, is depicted in (a) as a function of the stress time. Different stress temperatures are illustrated by diamonds (125 °C), circles (100 °C) and triangles (75 °C). Remarkably, one single set of scaling factors is sufficient to make the individual $V_{\text{TH}}$ shifts overlap for all stress temperatures, stress fields and stress times, c.f. (b) and (c). Also, universal scalability (factor 2.5) between $\Delta V_{\text{TH}}^{\text{perm}}$ and $\Delta V_{\text{TH}}^{\text{tot}}$ can be obtained. This is strong indication for a tight coupling of the two quantities.
Lelis et al. [77] and extended by Grasser et al. [6,78,79] suggests oxygen vacancies close to the interface as precursors for $E'$ centers. In the positive charge state, the $E'$ center is stable. In the neutral charge state, after capturing an electron from the substrate, the $E'$ center may undergo a structural relaxation which annihilates the defect permanently by reforming the oxygen vacancy precursor. This behavior is very similar to our recoverable defect (i). Thus, it is likely that defect (i) is the $E'$ center.

Defect (iii): These defects are traps which can respond to 500 kHz gate pulsing at $-60$ °C (i.e. CP). From EPR and EMDR studies it is well known that $P_0$ centers exist at the Si/SiO$_2$ interface and that they follow fast Shockley–Read–Hall like trapping/detrapping. Thus, defect (iii) is most likely the $P_0$ center. It should be noted that $E'$ centers close to the semiconductor/oxide interface can under certain circumstances also contribute to the CP current [42,45,54]. However, as already pointed out previously, the relative contribution of $E'$ centers to the CP current is only relevant at low pulsing frequencies, large pulse amplitudes and high temperatures, c.f. for instance Hehenberger et al. in Ref. [45].

Defect (ii): These defects are positively charged traps which contribute to the $V_{TH}$ shift but do not respond to 500 kHz CP at $-60$ °C. The microscopic origin of defect (ii) is, at the present state of knowledge, not fully clear. One might argue that defect (ii) is simply a more stable variant of defect (i) with considerably longer response/recovery times (e.g. an $E'$ center deeper in the oxide or with a different energy level). However, our measurement data does not support this idea. The lacking correlation between defect (i) and defect (ii) suggest a different microscopic structure of defect (ii). EPR studies [80,81] and theoretical studies using density functional theory calculations [82–85] suggest $E''$ center–hydrogen complexes and/or $P_0$ center–hydrogen complexes [86] as possible alternatives. If such hydrogen centers exist, the obtained correlation between defect (iii) and defect (ii) could be due to a mechanism involving hydrogen transition between Si–H bonds at the interface and $E'$ centers in the oxide [6,87,88]. The release of hydrogen from the Si–H bond would create a $P_0$ center; the capture of hydrogen at the $E'$ center's dangling bond site would block its recovery and lock its positive defect site in a puckered configuration. A correlation between NBTI induced amphoteric interface states and fixed positive oxide charges was also suggested by Ushio et al. [89] for Si/SiO$_2$ and Si/SiO$_2$/N$_x$ interfaces based on density functional calculations. They explained the simultaneous creation of fixed positive oxide charges and interface states through Si–H bond breakage (which leaves behind an amphoteric $P_0$ center) and subsequent H capture at a Si–O–Si group within the gate-oxide, near the interface. The H transition is promoted by hole capture thereby forming a positive oxide charge. Consistent with our results, Denais et al. [90] also reported three different types of traps, namely interface trapped charges, fixed charges and oxide trapped holes for advanced node technologies containing nitrogen and Aoualiche et al. [91] for high-k metal gates.

7. Conclusions

We have measured $V_{TH}$ shift and CP current dynamics of identically processed PMOS devices, stressed at different electric fields and temperatures. By making use of degradation quenching and the polyheater technique, devices stressed at various temperatures and fields could be characterized right after stress at a much lower and unique characterization temperature. Our investigations reveal that the degradation and recovery dynamics of $V_{TH}$ shifts and CP currents are different and widely uncorrelated. While the CP current seems to be almost permanent at $-60$ °C, the $V_{TH}$ shift recovers readily after the removal of the stress bias. We attributed the different behavior and the lack in universal scalability to a recoverable oxide defect which shows up in the $V_{TH}$ shift, but does not contribute to the CP current. We have shown that the number of recoverable defects with similar (short) time constants is mainly determined by the stress field, while defects with larger time constants are more likely generated at higher stress temperatures and longer stress times. Once the majority of the recoverable component has been removed from the total $V_{TH}$ shift (e.g. by gate pulsing), universal scalability between the CP current and the remaining ‘quasi-permanent’ $V_{TH}$ shift is obtained. However, only about 50% of the ‘quasi-permanent’ fraction of the $V_{TH}$ shift can be explained by fast interface states. This suggests a third type of defect which is ‘quasi-permanent’ and strongly coupled to the formation of fast interface states. By comparing our observations to characteristics of various point defects reported in literature, we found good agreement between the fast recoverable defect and the $E'$ center as well as between the CP defect and the $P_0$ center. The structure of the third defect which is created at virtually the same rate as the $P_0$ center remains uncertain at the present date. The defect may be created through hydrogen exchange between passivated $P_0$ centers (Si–H bonds) and recoverable $E'$ centers or alternatively through hydrogen exchange between passivated $P_0$ centers and Si–O–Si groups. Such reactions or similar hydrogen transitions between Si–H bonds at the interface and intrinsic interface/oxide defects could explain the universal coupling of the ‘quasi-permanent’ $V_{TH}$ shift and the CP current.

Acknowledgments

This work was jointly funded by the Federal Ministry of Economics and Labour of the Republic of Austria (Contract 98.362/0112-C1/10/2005) and the Carinthian Economic Promotion Fund (KWF) (Contract 98.362/0112-C1/10/2005).

References

Application market forecast tool. iSuppli Corporation. [accessed April 2010].


