Observing two stage recovery of gate oxide damage created under negative bias temperature stress

Thomas Aichinger,[1,a)] Michael Nelhiebel,[2] Sascha Einspieler,[1] and Tibor Grasser[3]
[1]*KAI (Kompetenzzentrum Automobil-und Industrieelektronik), Europastrasse 8, 9524 Villach, Austria*
[2]*Infineon Technologies Austria, Siemensstrasse 2, 9500 Villach, Austria*
[3]*Christian Doppler Laboratory for TCAD, Institute for Microelectronics, Technische Universität Wien, Gusshausstrasse 27-29, 1040 Wien, Austria*

(Received 30 July 2009; accepted 23 November 2009; published online 26 January 2010)

We study the hysteresis in threshold voltage shift during alternating gate bias ramps (drain current vs gate voltage (IdVg) sweeps) after negative bias temperature stress and compare the results with carefully recorded charge pumping measurements. This allows us to clearly identify three different types of defects. All defect types have in common that their charge state depends on the position of the Fermi level and that they introduce a broad density of states (DOS) in the vicinity or within of the silicon band gap. Defect I is fully recoverable, defect II is similar to defect I in terms of DOS but does not recover, while defect III can be attributed to the conventional interface states. With a precise microstructural model in mind, and by using specific test chips, which allow us to vary stress bias and temperature quasiarbitrarily, we come to the conclusion that the carrier trapping and detrapping characteristics of stress induced defects can be controlled by temperature and electric field in a similar way, but that irrevocable structural relaxation is mainly influenced by temperature. Based on these ideas, we present a measurement method which can be used to energetically profile the relaxation of stress induced oxide defects. © *2010 American Institute of Physics*.
[doi:10.1063/1.3276178]

I. INTRODUCTION

When a metal-oxide semiconductor (MOS) transistor is exposed to high electric fields and temperatures, various energetically widely spread defects have been reported to emerge via bond breakage or hydrogen release from passivated dangling bonds or strained Si–Si bonds within the gate oxide and at the gate oxide-silicon substrate interface.[1–6] The damage is believed to be partly permanent, partly recoverable. During the past decades, many attempts have been made to understand the phenomenon. Unfortunately, it has turned out to be more complicated than expected, mostly because of the fast recovery of the degradation once the stress bias is removed. In addition, the increasingly complicated device structures and layer sequences that have been introduced in recent technologies make a comparison and understanding of published literature data a difficult task. In particular, a generally accepted and universally valid model[7] is still missing.

In order to avoid additional complications introduced by complicated processing issues, we use simple isolated p-type MOS (PMOS) structures with 30 nm pure SiO_2 gate oxides in our study. This is done in order to prevent direct tunneling currents through the gate oxide during stress and in order not to distort our general conclusions by the strongly process dependent impact of nitridation.[8]

Our aim is to unambiguously clarify some basic characteristics of the traps created under negative bias temperature stress (NBTS). In order not to mix fundamentally different experimental parameters such as voltage and temperature, we make use of a recently established polyheater technique,[9–11] which allows us to control these parameters independently. For instance, this technique allows us to stress devices at an elevated stress temperature (125 °C) but characterize them at a much lower relaxation temperature (−60 °C). As such, this procedure allows complete decoupling of stress and recovery parameters,[10,11] and permits modification of the recovery conditions in subsequent experiments on separate PMOS devices which all have been brought to strictly the same degradation level. In particular, it allows us to perform low temperature charge pumping (CP) measurements right after stressing the device at a much higher stress temperature. This is very beneficial since a low characterization temperature expands the accessible energy range within the silicon band gap during CP, which improves the measurement resolution and hence the density of states (DOS) approximation.

II. DEGRADATION-RECOVERY PROCESS

Our experimental setup is particularly adapted to probe an established damage/recovery hypothesis based on oxygen vacancies[12,13] that are believed to be generated as a result of NBTS and then can act as switching traps[14–16] during recovery. This so-called Harry Diamond Laboratories (HDL) model was introduced by Lelis *et al.*[17] following irradiation studies, later confirmed by ESR studies[18] and finally revisited and extended by Grasser *et al.*[19,20] in the context of the negative bias temperature instability (NBTI). The model-predicted transitions linked to our experimental setup are discussed in Fig. 1.

In our standard experiment, we expose identically processed PMOS devices to NBTS by subjecting them to a

[a)]Electronic mail: thomas.aichinger@k-ai.-at.

 © 2010 American Institute of Physics

Downloaded 10 Jan 2011 to 128.130.68.53. Redistribution subject to AIP license or copyright; see http://jap.aip.org/about/rights_and_permissions

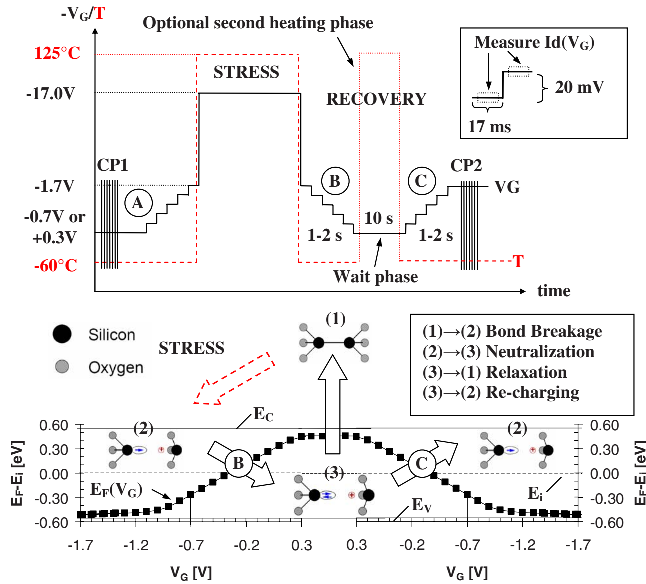

FIG. 1. (Color online) A schematic description of the trap characterization experiment. CP1: CP measurement before stress; (a) ramp up before stress; (b) ramp down poststress; (c) ramp up poststress; CP2: CP measurement poststress+recovery. The gate voltage is given by a solid line, the temperature generated by the heater as a dashed (dotted) line. Timing parameters and step width during the staircase sweeps are depicted in the inset. At the bottom, a numerical simulation of the Fermi level position at the substrate gate oxide interface at −60 °C is illustrated (b) when going from inversion to accumulation and (c) when going back from accumulation to inversion during the recovery cycle. Superimposed, we have sketched the expected model transitions involved.

stress field of −5.5 MV/cm at a temperature of 125 °C for a defined time. This "stress" labeled phase in Fig. 1 follows an initial characterization procedure. Without loss of generality, we assume that during stress, interface and oxide traps can be created and exchange carriers with the silicon substrate. Since during stress, the device is in strong inversion, the Fermi level is close to the valence band edge and most interface states are positively charged. Since there are almost no electrons present at the interface as long as the gate bias is at stress level, hole-capture dominates the carrier exchange process between NBTS induced traps and the silicon substrate, resulting in predominantly positively charged defects at the end of stress.[8,21,22] Bond breakage and charging during stress is symbolized by the transition $1\rightarrow2$ from a trap precursor state (1) (e.g., an oxygen vacancy) to a broken bond with a captured hole (e.g., an E′ center), state (2) in Fig. 1.

After stress, we quench degradation.[10] In general, this means that we turn off the heater first but do not remove the stress bias until the device has cooled down to −60 °C. This cooling event takes just a few seconds and does not change much in the Fermi level position and in the dominance of holes at the interface. However, the generation of additional traps should be suppressed assuming that the temperature is a driving factor in bond breakage. Thus, cooling down to −60 °C freezes stress, while recovery only sets in when the gate bias is changed.

The starting signal for recovery is given by the switch of the gate voltage from stress level (−17.0 V) to a lower recovery value (−1.7 V). As long as this gate bias value still corresponds to strong inversion, the Fermi level remains pinned close to the valence band edge (cf. Fig. 1). Consequently, the occupancy of the positively charged oxide defects will not change dramatically. This is why we switch the gate bias from stress level to a value well above threshold here to −1.7 V for a device with V_{TH}=−1.0 V. From this point a "down sweep" is started, which slowly turns the device off. During this sweep, the drain current is monitored and converted to a threshold voltage shift using an initial sweep on an unstressed device.[23] A schematic illustration of the step timing and the voltage parameters during a standard ramp is given in the inset of Fig. 1. The Fermi level during ramp down (B), during the 10 s "wait" phase at constant gate bias, and during ramp up (C) has been simulated numerically (see the bottom of Fig. 1). The proposed structural transitions involved are also illustrated.

As soon as the gate bias approaches and then passes the threshold voltage during ramp down (B), the Fermi level quickly moves from the valence band edge toward the conduction band edge, which influences the interface carrier concentrations considerably. As a result, a new equilibrium condition for carrier exchange between the silicon substrate and the oxide defects develops, which favors electrical neutralization of the positively charged switching traps (transition $2\rightarrow3$). Basically, this $2\rightarrow3$ transition (which we propose to call electrical neutralization, since electrically visible damage disappears upon neutralization of positively charged defects) is reversible, that is, a subsequent "up sweep" leads to recapture of a hole from the substrate and restores the electrically visible damage in state (2) (transition $3\rightarrow2$). However, provided the defect remains electrically neutral for a sufficiently long time, the defect can undergo a structural relaxation (transition $3\rightarrow1$), which lowers the defect level in energy and brings it back to the initial state (1). This recovery is irrevocable (as opposed to the reversible electrical *recovery* mentioned before) in the sense that a large stress voltage is now again required to trigger the transition $1\rightarrow2$. Since this recovery can only occur when the trap state is neutral [that is, in state (3)] it will happen predominantly during the wait phase (10 s) of our experiment which lasts much longer than the actual ramping durations (B) and (C) (1–2 s).

Optionally, we may increase the temperature during the wait phase, which will enhance the relaxation process $3\rightarrow1$, provided it is temperature activated. Note that increasing the temperature at a constant gate bias does not modify the Fermi level position considerably.

Our analysis is based on the observed difference in V_{TH} shift between the two gate bias sweeps (B) and (C) which provides insight into the above mentioned transitions: Pure electrical neutralization ($2\rightarrow3$) is reversible, meaning that the positive charge can be restored in the "up sweep." On the other hand, the real chemical relaxation ($3\rightarrow1$) is irreversible under the applied bias conditions of the sweep which results in observable differences in the up sweep compared to the down sweep (cf. Fig. 1).

We remark that in general we can never return exactly to the initial prestress characteristics of the device not even after the final up sweep (C). This indicates that we have to consider superimposed *permanent* damage additionally to

Downloaded 10 Jan 2011 to 128.130.68.53. Redistribution subject to AIP license or copyright; see http://jap.aip.org/about/rights_and_permissions

the *recoverable* damage described in Fig. 1. Following Grasser *et al.*,[19,20] the formation of this permanent damage may be explained by hydrogen released during stress from stretched bonds at the interface. This released hydrogen is suspected to create an interface state and also block the $2 \rightarrow 3$ and the $3 \rightarrow 1$ transitions. For our experiment described here, this effectively reduces the number of available state (2) defects and accounts for a permanent offset in the extracted V_{TH} shifts.

Having explained the concept of our experiment, tailored to the physical mechanisms which we expect to observe, we will briefly summarize the detailed experimental conditions and present the results. Possible alternative explanations are discussed in the Appendix using additional experiments supporting our interpretation.

III. EXPERIMENT DESCRIPTION AND RESULTS

The measurement begins with a CP (CP1) sequence and a full gate bias ramp (A) at a chuck temperature of −60 °C. This was done in order to record the virgin conditions of the devices with regard to interface state and transfer characteristics. The initial CP characterization is also used to select identical devices for the various experiments performed. In Ref. 10 it was found that identically processed devices on the same wafer degrade in a comparable way, provided they have an initially similar CP current. In the next step, we apply a defined power to the previously calibrated polyheater in order to heat up the device to the stress temperature (125 °C). After a few seconds, the temperature has stabilized and we subject the device to electrical stress (−5.5 MV/cm). At the end of stress, we switch off the heater while keeping the stress bias applied and wait for several seconds until the device has cooled down to the chuck temperature (−60 °C). Not before the device temperature is stable at −60 °C, we switch the gate bias from stress level to its first recovery value (−1.7 V), which is well above the threshold voltage of the device (−1.0 V). The above described sequence is equivalent for all devices under test discussed in the following. After this procedure, every device is at the same defined degradation level confirmed by nearly identical initial recovery traces. From that point on, the experiment was carried out in four different ways.

A. Separate experiment discriptions

(i) Right after stress, we ramp down (B) the gate voltage from deep inversion (−1.7 V) to depletion (−0.7 V) and simultaneously record the linear drain current at −100 mV source-drain bias. The sweep takes about one second and consists of 50 equidistant voltage steps of 20 mV. After remaining for 10 s at −0.7 V, we ramp up (C) the gate junction back from −0.7 to −1.7 V and record again the linear drain current at every gate voltage step. Note that identical sweeps are performed on the unstressed device to allow $I_d \rightarrow V_{TH}$ conversion. The sequence is completed by a final CP measurement (CP2).

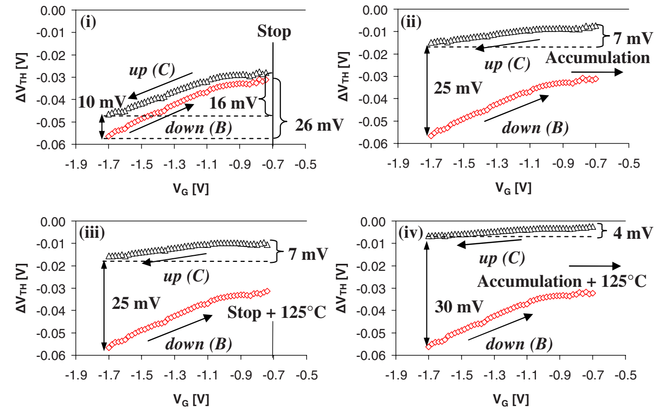

FIG. 2. (Color online) Development of the V_{TH} shift in experiment (i), (ii), (iii), and (iv) during (b) ramp down (open diamonds) and (c) ramp up (open triangles). Considerable recovery of ΔV_{TH} can be observed, as we ramp down the gate voltage toward depletion ($2 \rightarrow 3$). During ramp up, some of the previously neutralized defects can be charged positively again ($3 \rightarrow 2$), the others are assumed to recover permanently by structural relaxation ($3 \rightarrow 1$) during the wait phase.

(ii) Same measurement as in (i), however, ramp down (B) and ramp up (C) were recorded from deep inversion (−1.7 V) till accumulation (+0.3 V) including a 10 s wait period at +0.3 V between the ramps.

(iii) Same measurement as in (i), however, during the wait phase at −0.7 V, the polyheater was turned on to heat the device to 125 °C for 10 s.

(iv) Mixture of experiments (ii) and (iii). During the 10 s wait period at +0.3 V between ramp down (B) and ramp up (C), a 125 °C heating pulse has been introduced.

B. Discussion of the results

In the analysis of experiments (i)–(iv), we evaluate a gate voltage dependent hysteresis effect that emerges, when the ramp down (B) and the ramp up (C) are compared to the virgin transfer curve (A). These particular stress/recovery induced shifts, extracted from the experiments (i)–(iv), are separately depicted as gate voltage dependent threshold voltage variations in Fig. 2. Note that as the gate bias exceeds −0.7 V and approaches accumulation (+0.3 V) in experiments (ii) and (iv), the V_{TH} shift cannot be monitored anymore. This is a general experimental limitation originating from the lack of an inversion channel as the Fermi level approaches and finally exceeds the intrinsic level E_i (more electrons than holes at the interface).

We remark that the V_{TH} shift in ramp down (B) is identical for all experiments discussed in Fig. 2. In all experiments we observe a considerable reduction in ΔV_{TH} by 26 mV as we drive the gate voltage from deep inversion (−1.7 V) toward depletion (−0.7 V). This reduction is the same for all devices since the measurement sequences are identical until the gate bias exceeds the depletion voltage of −0.7 V. The good correlation is evidence of a comparable degradation level in all devices at the end of stress. Keeping our model of energetically widely distributed oxide defects in mind, we propose that this significant 26 mV reduction is

Downloaded 10 Jan 2011 to 128.130.68.53. Redistribution subject to AIP license or copyright; see http://jap.aip.org/about/rights_and_permissions

mainly caused by electrical oxide trap neutralization as the Fermi level travels from the valence band edge toward midgap (transition $2 \rightarrow 3$). Possible alternative interpretations and objections are discussed in detail by means of separate experiments in the Appendix.

As opposed to sweep down (B) the results of sweep up (C) depend significantly on the bias range of the sweep and the temperature during the 10 s wait phase. According to our model, we suggest that the *difference* in the ΔV_{TH} shift between sweep down (B) and sweep up (C) is mainly due to irreversible relaxation of neutralized switching oxide traps ($3 \rightarrow 1$). Repassivation of interface states and time delay effects are assumed to be negligible. For a justification of these assumptions, we refer to the Appendix.

In the following we discuss the particular experimental procedures separately and highlight the agreement with the predicted structural transitions involved:

1. Experiment (i)

After the 10 s wait time at −60 °C and a gate voltage of −0.7 V, we ramp the gate bias back toward inversion (−1.7 V). As a result of sweep up (C), we observe an increase in the V_{TH} shift of about 16 mV which is, however, not as large as the amount of ΔV_{TH} lost during ramp down (B). According to the model of Grasser *et al.*,[19,20] the final difference in V_{TH} shift of 10 mV between ramp down (B) and ramp up (C) would be attributed to permanently recovered oxide defects (transition $3 \rightarrow 1$). On the other hand, the remaining increase of 16 mV in sweep up (C) indicates that electrical neutralization of oxide traps does not necessarily result in a structural relaxation which would erase them irrecoverably. Some of the defects can obviously be charged positively again as the gate bias reapproaches inversion in ramp up (C) (transition $3 \rightarrow 2$).

2. Experiment (ii)

We study the influence of a Fermi level shift toward the conduction band on the permanent recovery of oxide defects by driving the device from inversion till accumulation (+0.3 V) during sweep down (B) and keep the bias there for another 10 s. As we ramp up the gate voltage back to inversion (C), an even larger difference of about 25 mV emerges in the ΔV_{TH} shift, indicating that during accumulation irrevocable oxide trap recovery ($3 \rightarrow 1$) was much more effective than during depletion. This is likely due to the fact that in accumulation, almost all state (2) traps become electrically neutralized in state (3). Consequently, the more traps are in state (3), the higher is the possibility of a $3 \rightarrow 1$ transition. Also, the remaining increase in the ΔV_{TH} shift during ramp up (C) is considerably reduced to about 7 mV, which is consistent with the suggestion that most oxide traps visible in ramp down (B) have already been erased irrevocably during the accumulation phase.

3. Experiment (iii)

We study the influence of temperature on the permanent recovery of oxide defects. As opposed to experiment (i), we heat the device to 125 °C during the 10 s wait phase in depletion between ramp down (B) and ramp up (C). Remarkably, as a consequence of the 10 s temperature pulse, we observe again a 25 mV reduction in the V_{TH} shift similar as in experiment (ii). The subsequent increase in ramp up (C) also looks comparable to experiment (ii). This indicates that device heating in depletion (−0.7 V) has a similar impact on permanent oxide trap recovery as biasing the device in accumulation (+0.3 V) without heating for the same time.

The origin of this equivalence can be explained as follows: During the low temperature accumulation phase in experiment (ii) most oxide traps are immediately electrically neutralized (transition $2 \rightarrow 3$). This means that at the beginning of the 10 s accumulation phase nearly all traps are in state (3) and therefore available for the $3 \rightarrow 1$ transition. Although the probability of structural relaxation is low at −60 °C, some of the state (3) traps will still permanently recover within the 10 s wait phase.

On the other hand, if we bias the device in depletion during the wait phase, a smaller fraction of the previously generated state (2) traps are in state (3) at the beginning of the 10 s wait phase. As a consequence, we would expect less $3 \rightarrow 1$ transitions, cf. experiment (i). However, if we start to heat up the device [cf. experiment (iii)], the temperature activated $3 \rightarrow 1$ transition is accelerated considerably. Furthermore, additional $2 \rightarrow 3 \rightarrow 1$ transitions are likely to occur as a consequence of inelastic tunneling ($2 \rightarrow 3$) and subsequent structural relaxation ($3 \rightarrow 1$).

4. Experiment (iv)

A combination of experiments (ii) and (iii), accumulation phase plus heating pulse, leads to the highest degree of relaxation. The difference in the V_{TH} shift between ramp down (B) and ramp up (C) exceeds 30 mV at −1.7 V, which is almost full recovery except for a small apparently permanent offset. It seems that nearly the entire oxide charge contribution, visible in ramp down (B), has been neutralized ($2 \rightarrow 3$) and subsequently recovered permanently ($3 \rightarrow 1$) by heating the device in accumulation. This is consistent with the Fermi level dependence of the $2 \rightarrow 3$ neutralization and the temperature acceleration of $3 \rightarrow 1$ relaxation. The small remaining ΔV_{TH} increase of about 4 mV in ramp up (C) and the permanent offset of 3 mV visible at −0.7 V can be attributed to interface state charging (cf. Appendix) and permanent locked-in oxide defects. Consistent with the extended model of Grasser *et al.*,[19,20] this permanent damage gives approximately a 1:1 correlation of interface states and permanent oxide traps which emerge as a result of hydrogen exchange between passivated interface traps and state (2) oxide traps.

We remark that interface states (Pb centers) alone can neither be made responsible for the offset between sweep down (B) and sweep up (C) nor for the remaining degradation at $V_G = -0.7$ V. This is simply due to the fact that P_b centers are amphoteric[24,25] and therefore commonly assumed to be neutral, when the Fermi level is close to midgap. The small fraction of interface states, which are positively charged at −0.7 V, is much too small to explain the large remaining V_{TH} shift of 30 mV visible at the end of sweep down (B) in Fig. 2.

Downloaded 10 Jan 2011 to 128.130.68.53. Redistribution subject to AIP license or copyright; see http://jap.aip.org/about/rights_and_permissions

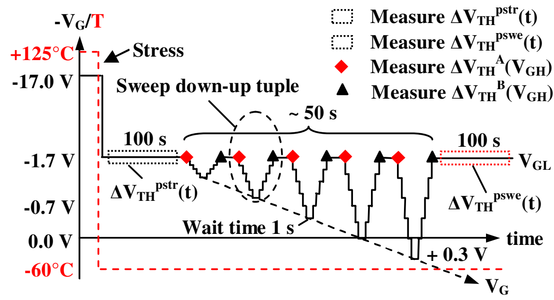

FIG. 3. (Color online) Description of the incremental sweep experiment. Right after stress and after the ~50 s sweep, we apply a constant gate bias of $V_{GL}=-1.7$ V for 100 s and record the time dependent threshold voltage recovery $\Delta V_{TH}^{pstr}(t)$ and $\Delta V_{TH}^{pswe}(t)$, respectively. During the intermediate sweep procedure, we ramp the gate bias down to a variable high level V_{GH}, stay there for 1 s and afterward ramp it back to -1.7 V (constant V_{GL} level). In the following, we repeat this down-up cycle while increasing V_{GH} continuously by 0.2 V from -1.7 V till +0.3 V. Before and after the double staircase sweep $V_{GL}\rightarrow V_{GH}\rightarrow V_{GL}$, we measure a ΔV_{TH} tuple and reference it to the intermediate bias V_{GH}.

IV. ENERGETIC TRAP PROFILING

The results and conclusions drawn by the above described experiments identified the Fermi level position and the temperature as the main parameters influencing our model transitions. In order to correlate the Fermi level position to the amount of V_{TH} recovery, we have performed an additional extended experiment which is schematically depicted in Fig. 3.

As opposed to starting with a full down sweep immediately after stress, we include a 100 s constant gate bias phase at $V_G=-1.7$ V prior to the sweep measurement. This was done in order to differentiate between time dependent and bias dependent recovery effects. Within this initial 100 s, the ΔV_{TH} shift was determined as a function of time. After the constant bias phase, we start with an incremental down sweep beginning from $V_{GL}=-1.7$ V till $V_{GH}=-1.5$ V, followed by a constant bias wait phase of 1 s at V_{GH}. Subsequently, we ramp up the gate bias back toward $V_{GL}=-1.7$ V and stay there for another second. In the following, we repeat this $V_{GL}\rightarrow V_{GH}\rightarrow V_{GL}$ cycle keeping V_{GL} constantly at -1.7 V while increasing V_{GH} incrementally by 0.2 V till a final value of +0.3 V. Right before and immediately after such a double staircase sweep, we record tuples of ΔV_{TH} values and reference them to the intermediate V_{GH} levels (cf. Fig, 3). After the last down-up cycle, we monitor again the ΔV_{TH} recovery for 100 s at a constant gate bias of -1.7 V.

The results of the 100 s time dependent threshold voltage recovery recorded right after stress $[\Delta V_{TH}^{pstr}(t)]$ and at the end of the sweeping procedure $[\Delta V_{TH}^{pswe}(t)]$ are illustrated in Fig. 4(a). Note that the sweep routine follows right after the initial 100 s constant bias phase at -1.7 V. Within this time interval, we observe a logarithmic recovery with a slope of +2.8 mV/dec. Consequently, any kind of time dependent recovery associated with the large voltage drop from stress bias (-17.0 V) to V_{GL} (-1.7 V) is negligible during the attached 50 s sweeping phase due to the large delay following the actual stress phase. What remains is mostly Fermi level and temperature controlled carrier exchange. This is

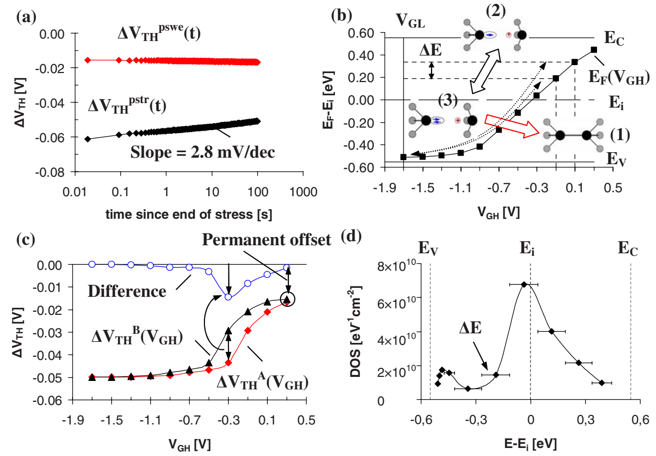

FIG. 4. (Color online) The results of the incremental sweep experiment. The time dependent recovery of the threshold voltage shift, recorded for 100 s at V_{GL} poststress $[\Delta V_{TH}^{pstr}(t)]$ and after the sweep procedure $[\Delta V_{TH}^{pswe}(t)]$, is illustrated in (a). A numerical simulation of the Fermi level $E_F(V_{GH})$ is depicted in (b). By increasing V_{GH} incrementally, the Fermi level increasingly moves into the upper regions of the silicon band gap, which results in the neutralization of positively charged traps ($2\rightarrow3$). Within the 1 s wait time at V_{GH}, neutralized traps energetically located below $E_F(V_{GH})$ tend to relax permanently ($3\rightarrow1$). In (c), the ΔV_{TH} tuples $\Delta V_{TH}^{A}(V_{GH})$ and $\Delta V_{TH}^{B}(V_{GH})$, measured before and after the down-up ramp at V_{GL}, are illustrated as a function of V_{GH}. The difference between $\Delta V_{TH}^{A}(V_{GH})$ and $\Delta V_{TH}^{B}(V_{GH})$ is representative for the number of $3\rightarrow1$ transitions during the V_{GH} phase. (d) From that difference, we evaluate a density of state profile and reference the energy to the position of the Fermi level within the silicon band gap. A significant peak of permanently recovered donorlike traps emerges around midgap.

consistent with the observation that the degradation level is frozen at the end of the sweep routine [cf. $\Delta V_{TH}^{pswe}(t)$ in Fig. 4(a)].

As we incrementally move the V_{GH} level from inversion toward accumulation in our sweep experiment, the Fermi level $E_F(V_{GH})$ increasingly approaches the conduction band edge [cf. Fig. 4(b)]. In fact, when going from one V_{GH} level (V_{GH}') to the other (V_{GH}''), the accessible energy range for the $2\rightarrow3$ transition within the silicon band gap extends by $\Delta E=E_F(V_{GH}'')-E_F(V_{GH}')$. Traps within this energy range may be neutralized during the following down sweep ($V_{GL}\rightarrow V_{GH}''$). Once neutralized, some of these additional state (3) traps within ΔE may manage to relax permanently during the 1 s wait phase at V_{GH} ($3\rightarrow1$). The ones who made the $3\rightarrow1$ transition are finally missing in the subsequent up sweep and therefore account for a difference between $\Delta V_{TH}^{A}(V_{GH}'')$ and $\Delta V_{TH}^{B}(V_{GH}'')$ [cf. Fig. 4(c)].

By referencing the differences $\Delta V_{TH}^{A}(V_{GH})-\Delta V_{TH}^{B}(V_{GH})$ to the current Fermi level position $E_F(V_{GH})$ and the corresponding energy levels ΔE, we can determine an average DOS within ΔE of primarily neutralized and finally permanently recovered state (2) traps ($2\rightarrow3\rightarrow1$) [cf. Fig. 4(d)]. Note that the stated energy levels in Fig. 4(d) need not necessarily correspond exactly to the actual energy levels of the defects located within the oxide, especially when assuming that the carrier exchange mechanism between the silicon substrate and the state (2) trap is inelastic. The DOS of permanently recovered oxide traps is calculated under the assumption that all traps are located close to the oxide-silicon interface.

Downloaded 10 Jan 2011 to 128.130.68.53. Redistribution subject to AIP license or copyright; see http://jap.aip.org/about/rights_and_permissions

$$\overline{D_{OX}^{3\to 1}}\left(E \pm \frac{\Delta E}{2}\right) = \frac{C_{OX}\cdot\{\Delta V_{TH}^{A}[E(V_{GH})] - \Delta V_{TH}^{B}[E(V_{GH})]\}}{q\cdot\Delta E}. \quad (1)$$

In Eq. (1), C_{OX} is the area related gate oxide capacitance and q is the elementary charge. In the DOS profile illustrated in Fig. 4(d) we obtain a significant peak around midgap.

We remark that the occurrence of this peak can in general have two different reasons: (a) There is a large density of state (2) traps located around midgap. After being neutralized by Fermi level variations ($2\to 3$), many $3\to 1$ transitions may occur during the 1 s wait phase at V_{GH} causing a significant reduction in positive charge at the end of the subsequent up sweep. (b) In general, oxide trap neutralization ($2\to 3$) and subsequent relaxation ($3\to 1$) of state (2) traps might become efficient not before the free electron concentration exceeds the free hole concentration at the oxide-substrate interface.

Based on our previous experiments, we have to favor explanation (a): we have shown in Fig. 2 that an increase in temperature during depletion has a similar impact on permanent oxide trap recovery than biasing the device in accumulation for the same time [cf. Fig. 2 experiments (ii) and (iii)]. Since elevating the temperature does not change the ratio of free interface carriers significantly but considerably enhances the number of $2\to 3$ transitions by inelastic tunneling and simultaneously accelerates the thermally activated $3\to 1$ transition, a dominance of electrons does not seem to be a mandatory requirement for relaxation.

Note that the final degradation level of −16 mV after V_{GH}=+0.3 V in Fig. 4(c) is permanent and correlates perfectly with ramp up (C) in experiments (ii) and (iii) in Fig. 2.

V. CONCLUSION

In our experiments we have identified and examined the dominant type of switching oxide trap created during NBTS, which is positively charged during stress and can change its charge state during recovery. According to the HDL model, we differentiate strictly between electrical neutralization and chemical bond reconstruction. The neutralization process was found to be influenced considerably by temperature and the position of the Fermi level. Furthermore, at −60 °C, the observed carrier exchange mechanisms are slow enough to produce time dependent recovery transients as we switch the gate bias from stress level to −1.7 V. Due to the temperature influence[11] and the relatively slow recovery transients at −60 °C, we have to conclude that oxide trap neutralization by carrier exchange ($2\to 3$) is due to an inelastic rather than an elastic tunneling process.

In the neutral charge state (3) the switching oxide trap can either undergo a structural relaxation which anneals the trap permanently ($3\to 1$) or can re-emit the previously captured carrier and charge up positively again ($3\to 2$). Although electrical neutralization was identified to be a basic requirement for chemical relaxation, the intermediate neutral trap state requires thermal activation and time in order to lower in energy and relax irrevocably.

By incrementally moving the Fermi level position into the silicon band gap, we could stimulate oxide trap neutralization and subsequent permanent relaxation in a controlled manner. From the results we conclude that oxide trap neutralization is accelerated when the Fermi level moves beyond the intrinsic energy level within the silicon band gap. By referencing the observed threshold voltage shift reduction to the position of the Fermi level, we obtain an oxide trap DOS profile which has a significant peak around midgap.

The remaining degradation, after having annealed the switching traps, seems to consist of a mixture of interface states and another type of fixed oxide trap.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of Stefan Decker (Infineon Technologies Munich) in numerical Fermi level simulation. This work was jointly funded by the Federal Ministry of Economics and Labour of the Republic of Austria (Contract No. 98.362/0112-C1/10/2005) and the Carinthian Economic Promotion Fund (KWF) (Contract No. 98.362/0112-C1/10/2005).

APPENDIX: DISCUSSION OF ALTERNATIVE INTERPRETATIONS

The interpretation of our measurement results given above may meet opposition if alternative explanations cannot be ruled out. Without the loss of generality, the observed recovery effects depicted in Fig. 2 could possibly lead to alternative conclusions competing with the ones described in our model approach. In order to validate our hypothesis, we have performed additional experiments supporting our conclusions rather than confirming the below listed concerns.

(i) *Time dependent recovery:* In sweep down (B) there is a time delay of approximately 1 s involved between the first measured point recorded at V_G=−1.7 V (15 ms poststress) and the last one recorded at V_G= −0.7 V. This time interval is usually needed to ramp the gate voltage and measure the drain current in parallel at every voltage step (cf. inset Fig. 1). From this point of view, it may be natural to assume that time dependent recovery effects are likely to interfere with our simple gate bias-related model predictions.

(ii) *Interface states recovery:* Once generated during stress, interface states have been reported to recover moderately by hydrogen repassivation. In our interpretations we neglect this chemical interface state recovery and claim that dangling interface bonds remain mainly permanent at a recovery temperature of −60 °C. Indeed, they may change their charge state by Fermi level variation, however, they should not relax significantly by hydrogen capture.

(iii) *Interface states neutralization:* The considerable ΔV_{TH} reduction of 26 mV observed during sweep down (B) in Fig. 2 might be at least to some extend due to interface trap charging. As the Fermi level moves from the valence band edge toward midgap, we expect a continuous neutralization of positively

Downloaded 10 Jan 2011 to 128.130.68.53. Redistribution subject to AIP license or copyright; see http://jap.aip.org/about/rights_and_permissions

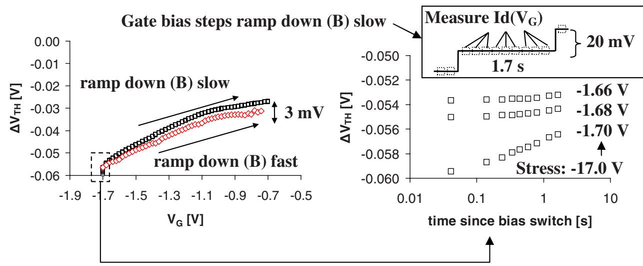

FIG. 5. (Color online) A comparison of the ΔV_{TH} shift during fast ramp down (B) and slow ramp down (B) is depicted in the left figure. As we increase the step duration of the voltage ramp by a factor of 100 from 17 ms (diamonds) to 1.7 s, the total sweeping time of ramp down (B) extends from 1 to 100 s (squares). As a consequence, an additional 3 mV ΔV_{TH} recovery can be observed at −0.7 V. At each gate bias level of the slow ramp down (b) we have measured nine $\Delta V_{TH}(t)$ points within the 1.7 s step duration (cf. inset right figure). The recovery transients of the first three step levels after the end of stress (−17 V→−1.7/−1.68/−1.66 V) are illustrated in the right figure. A significant time dependent transient $\Delta V_{TH}(t)$ emerges only as we switch the gate bias from stress level to −1.7 V.

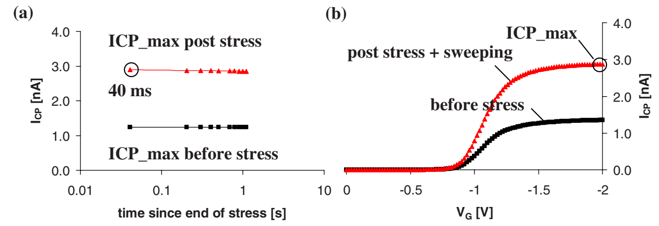

FIG. 6. (Color online) The recovery of the maximum CP current (ICP_max) recorded directly poststress (a) and a full CP curve recorded at the end of experiment (ii) (b). For comparison, the unstressed CP currents of the two reference devices are also included in (a) and (b). In (a), the first CP current recorded 40 ms poststress is nearly identical to the maximum CP current recorded at the end of experiment (ii) (b). This indicates that interface state recovery is negligible or at least of minor significance during the sweep.

charged interface states in the lower half of the silicon band gap.[28,29]

The first point—*time dependent recovery*—can be checked by a reference experiment where we extend the step time in ramp down (B) from 17 ms to 1.7 s. Consequently, the total ramp time extends from 1 to 100 s. In order to visualize possible recovery transients, we sample several points at each gate voltage step. A comparison of the extracted V_{TH} shift in slow and fast [experiment (i)] ramp is illustrated in Fig. 4. Time dependent recovery transients obviously emerge only within the first couple of voltage steps right after stress. The reason for that may be the large bias step when going from stress bias (−17.0 V) to the first recovery value (−1.7 V). The transients corresponding to the first three bias switches are illustrated as a function of time in the inset of Fig. 5. However, as the gate voltage moves further toward depletion, the variation in bias seems to have a much larger impact on recovery than the elapsing of time. As a consequence, the time-transients become negligible. If we compare the V_{TH} shift at $V_G = -0.7$ V of the slow (100 s) and fast (1 s) measured ramp, we observe just a small discrepancy of 3 mV. This indicates that time dependent recovery does account just for a small fraction of the total observed 26 mV V_{TH} shift reduction in experiment (i).

The second point—*repassivation of positively charged interface states*—can be ruled out by comparing the maximum CP currents recorded right after stress and at the end of the sweeping procedure [cf. experiment (ii) and CP2 in Fig. 1]. In order to get an idea of the interface degradation right after stress, we have skipped the sweeps in Fig. 1 on a reference device and measured the maximum CP current immediately (40 ms) after removing the stress bias. The CP results of both measurement sequences [maximum CP current right after stress and full CP curve at the end of experiment (ii)] are illustrated in Fig. 6. If there were significant interface state recovery involved in the sweeping procedure of experiment (ii), the first measured CP point (40 ms poststress) in Fig. 6(a) should be much higher than the maximum CP current at the end of experiment (ii) [cf. Fig. 6(b)]. However, as can be seen, there is nearly no difference. We consequently conclude that chemical interface state recovery can be neglected during our "sweep down-up" experiment. This argument holds as long as we assume that the first measured CP current recorded 40 ms poststress at −60 °C is representative for the actual total interface degradation active at the very beginning of our sweep measurement (cf. Fig. 2).

We remark that other groups report that gate pulsing from inversion to accumulation and a short time delay between the end of stress and the first CP measurement may already introduce considerable interface state recovery.[30–32] However, we point out that our stress-measurement delay (40 ms) is much shorter than the one achieved by other groups. Furthermore, we quench degradation to a much lower recovery temperature (−60 °C), which should considerably reduce temperature driven interface state repassivation. Besides that, recent investigations[33] indicate that the fast interface state recovery component[30–32] may be afflicted with an error. This error is a consequence of the different voltage ranges used during the stress and recovery measurements, which leads to a different amount of defects switching between states (2) and (3) and contributing to the CP current.

The third point—*neutralization of charged interface states*—can be assessed by evaluating the maximum conceivable threshold voltage shift caused by positively charged interface states located in the lower half of the silicon band gap. This can be done by converting the stress induced increase in the CP current into an equivalent V_{TH} shift. For the evaluation, we use the maximum CP current recorded 40 ms poststress [cf. Fig. 6(a)]. For our CP measurements, we use a pulse frequency of 1 MHz and rising/falling slopes of 32 V/μs. Following Refs. 26 and 27, the scanned energy interval ΔE_{CP} in the maximum CP region can be calculated analytically. By inserting our pulse setup and recovery temperature (−60 °C), we obtain $\Delta E_{CP}=0.83$ eV. This is approximately 75% of the silicon band gap. Following Fig. 1, the energy interval ΔE_{VG} profiled during ramp down (B) equals approximately a quarter of the silicon band gap $\Delta E_{VG} \approx Eg/4$. If we consider the mismatch of the two energy regions by a weight factor and assume that the average DOS within ΔE_{CP} and ΔE_{VG} are similar, we can estimate a maximum threshold voltage shift due to the charging of interface states.

Downloaded 10 Jan 2011 to 128.130.68.53. Redistribution subject to AIP license or copyright; see http://jap.aip.org/about/rights_and_permissions

$$\Delta V_{TH}^{Dit} = \frac{\Delta E_{VG}}{\Delta E_{CP}} \times \frac{\Delta I_{CP}^{max}}{A_G \times f \times C_{OX}}. \quad (A1)$$

In Eq. (A1), ΔI_{CP}^{max} is the increase in the maximum CP current as a result of NBTI stress, A_G is the gate area of the device, f is the gate pulse frequency, and C_{OX} is the area related gate oxide capacitance. By inserting all parameters into Eq. (A1), we obtain a maximum value of 1.9 mV for ΔV_{TH}^{Dit}, which is definitely just a small fraction of the total V_{TH} reduction of 26 mV observed during ramp down (B) irrespective of the simple treatment above. From this result, we conclude that the main contribution to V_{TH} shift in ramp down (B) actually originates from the neutralization of positive oxide traps.

[1]D. K. Schroder, Microelectron. Reliab. **47**, 841 (2007).
[2]V. Huard, M. Denais, and C. Parthasarathy, Microelectron. Reliab. **46**, 1 (2006).
[3]M. A. Alam and S. Mahapatra, Microelectron. Reliab. **69**, 71 (1991).
[4]T. L. Tewksbury and H.-S. Lee, IEEE J. Solid-State Circuits **29**, 239 (1994).
[5]D. M. Fleetwood, M. R. Shaneyfelt, W. L. Warren, J. R. Schwank, T. L. Meisenheimer, and P. S. Winokur, Microelectron. Reliab. **35**, 403 (1995).
[6]K. O. Jeppson and C. M. Svensson, J. Appl. Phys. **48**, 2004 (1977).
[7]T. Grasser, W. Goes, V. Sverdlov, and B. Kaczer, Proceedings of the IEEE International Reliability Physics Symposium, 2007 (unpublished), Vol. 268.
[8]S. Mahapatra, K. Ahmed, D. Varghese, A. E. Islam, G. Gupta, L. Madhav, D. Saha, and M. A. Alam, Proceedings of the IEEE International Reliability Physics Symposium, 2007 (unpublished), Vol. 1.
[9]W. Muth and W. Walter, Microelectron. Reliab. **44**, 1251 (2004).
[10]T. Aichinger, M. Nelhiebel, and T. Grasser, Microelectron. Reliab. **48**, 1178 (2008).
[11]T. Aichinger, M. Nelhiebel, and T. Grasser, Proceedings of the IEEE International Reliability Physics Symposium, 2009 (unpublished), Vol. 1.
[12]E. H. Poindexter and W. L. Warren, J. Electrochem. Soc. **142**, 2508 (1995).
[13]P. M. Lenahan, Microelectron. Eng. **69**, 173 (2003).
[14]C. J. Nicklaw, Z.-Y. Lu, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, IEEE Trans. Nucl. Sci. **49**, 2667 (2002).
[15]A. V. Kimmel, P. V. Sushko, A. L. Shluger, and G. Bersuker, ECS Trans. **19**(2), 3 (2009).
[16]T. L. Tewksbury, Ph.D. thesis, Massachusetts Institutes of Technology, 1992.
[17]A. J. Lelis and T. R. Oldham, IEEE Trans. Nucl. Sci. **41**, 1835 (1994).
[18]J. F. Conley, Jr., P. M. Lenahan, A. J. Lelis, and T. R. Oldham, Appl. Phys. Lett. **67**, 2179 (1995).
[19]T. Grasser, B. Kaczer, W. Goes, T. Aichinger, Ph. Hehenberger, and M. Nelhiebel, Proceedings of the IEEE International Reliability Physics Symposium, 2009 (unpublished), Vol. 33.
[20]T. Grasser, B. Kaczer, W. Goes, T. Aichinger, Ph. Hehenberger, and M. Nelhiebel, Microelectron. Eng. **86**, 1876 (2009).
[21]A. T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, and S. Krishnan, Proceedings of the IEEE International Electron Devices Meeting, 2003 (unpublished), Vol. 1.
[22]V. Reddy, A. T. Krishnan, A. Marshall, J. Rodriguez, S. Natarajan, T. Rost, and S. Krishnan, Proceedings of the IEEE International Reliability Physics Symposium, 2002 (unpublished), Vol. 248.
[23]B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, Proceedings of the IEEE International Reliability Physics Symposium, 2005 (unpublished), Vol. 381.
[24]J. Bardeen, Phys. Rev. **71**, 717–727 (1947).
[25]Z. Chen, B. B. Jieb, and C.-T. Sahc, Proceedings of the International Conference on Solid-State and Integrated-Circuit Technology, 2006 (unpublished), Vol. 1227.
[26]G. Van den bosch, G. V. Groeseneken, P. Heremans, and H. E. Maes, IEEE Trans. Electron Devices **38**, 1820 (1991).
[27]G. Groeseneken, H. E. Maes, N. Beltran, and R. F. De Keersmaecker, IEEE Trans. Electron Devices **31**, 42 (1984).
[28]M.-F. Li, D. Huang, C. Shen, T. Yang, W. J. Liu, and Z. Liu, IEEE Trans. Device Mater. Reliab. **8**, 62 (2008).
[29]J. F. Zhang, Z. Ji, M. H. Chang, B. Kaczer, and G. Groeseneken, Proceedings of the IEEE International Electron Devices Meeting, 2007 (unpublished), Vol. 817.
[30]T. Yang, C. Shen, M. F. Li, C. H. Ang, C. X. Zhu, Y.-C. Yeo, G. Samudra, and D.-L. Kwong, IEEE Electron Device Lett. **26**, 758 (2005).
[31]D. S. Ang, S. Wang, G. A. Du, and Y. Z. Hu, IEEE Trans. Device Mater. Reliab. **8**, 22 (2008).
[32]S. Mahapatra, D. Saha, D. Varghese, and P. B. Kumar, IEEE Trans. Electron Devices **53**, 1583 (2006).
[33]Ph. Hehenberger, T. Aichinger, T. Grasser, W. Goes, O. Triebl, B. Kaczer, and M. Nelhiebel, Proceedings of the IEEE International Reliability Physics Symposium, 2009 (unpublished), Vol. 1033.

Downloaded 10 Jan 2011 to 128.130.68.53. Redistribution subject to AIP license or copyright; see http://jap.aip.org/about/rights_and_permissions