Understanding negative bias temperature instability in the context of hole trapping

T. Grasser a,*, B. Kaczer c, W. Goes a, Th. Aichinger d, Ph. Hehenberger b, M. Nelhiebel e

a Christian Doppler Laboratory for TCAD in Microelectronics, Institute for Microelectronics, TU Wien, Gußhausstraße 27–29, A-1040 Wien, Austria
b Institute for Microelectronics, TU Wien, Gußhausstraße 27–29, A-1040 Wien, Austria
c IMEC, Leuven, Belgium
d KAI, Villach, Austria
e Infineon Technologies, Villach, Austria

ARTICLE INFO

Article history:
Received 3 March 2009
Received in revised form 15 March 2009
Accepted 15 March 2009
Available online 26 March 2009

Keywords:
Negative bias temperature instability
Reliability
Modeling hole trapping
Oxide charges
Interface states

ABSTRACT

Hole trapping is often considered a parasitic component clouding the real degradation mechanism that is responsible for the negative bias temperature instability (NBTI). As such, it is often dealt with in a rather sketchy way that lacks physical rigor. We review hole trapping mechanisms that go beyond the conventional elastic tunneling mechanism by including structural relaxation and field effects. Contrary to some previous studies, it is shown that the rich spectrum of experimentally observed features of the most commonly observed defect in amorphous oxides, the E centered in oxynitrides to create an unpassivated silicon dangling bond. We formulate a complete model and evaluate it against experimental data.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

The current understanding of the negative bias temperature instability (NBTI) sees the creation of interface states and oxide charges as the main reason for device degradation following bias temperature stress [1–3]. However, nearly every aspect is controversial, ranging from the creation mechanism of interface states and their recovery behavior to the importance and mechanisms behind hole trapping. Existing models fail in explaining the temperature and voltage dependence of the recovery behavior and the experimentally observed correlation between interface and oxide charges [3], to name just a few. In particular, the popular reaction-diffusion theory [4], which assumes predominant creation of interface states controlled by the diffusion of molecular hydrogen in the oxide/polysilicon layers also fails for instance in reproducing fast recovery, recovery lasting considerably longer than the preceding stress, the bias and temperature dependence of the recovery, and the strong sensitivity to the stress duty-factor [3].

Before suggesting a new model, we briefly summarize some key experimental observations obtained by extended measure/stress/measure (eMSM) experiments [2], which provide important insights regarding the dynamics of NBTI. In particular, they can be used to effectively rule out a number of alternative models.

1.1. Scalability

The data recorded using the eMSM scheme can often be made to overlap by multiplication with a suitably chosen scaling factor. While the initial behavior up to about 1 s is well approximated by a logarithmic time dependence and the long-term data approximately follows a power-law, the same scaling factor can be used in both regimes [5]. A similar scalability is observed during recovery [5], which, to first-order, follows a logarithmic time dependence.

1.2. Bias and temperature dependence of stress

Degradation of devices subjected to short stresses of about 1 s follows \( B_s(F,T) \log t_0 / t \), whereas recovery can be fit by \( B_r(F,T) \log [1 + t_0 / t] + P(t) \), \( P(t) \) is a roughly permanent contribution. For what is typically considered NBTI stress in this oxide thickness range (\( V_{\text{stress}} \approx -1 \) V), the pre-factor \( B_s \) can be approximated as \( B_s(F,T) \approx B_{1,F} T^{1/2} \), with \( \Theta \approx 2 \), see Fig. 1.

1.3. Asymmetry between stress and relaxation

It has been long understood that recovery takes substantially longer than the time used to build up the degradation. Using the extracted pre-factors \( B_s \) (fit to a log in the range 1 ms...1 s) and \( B_r \) (fit to a log in the range 1 ms...100 ms) we observe that the ratio \( B_s / B_r \) is about 2.5, independent of temperature and voltage.

* Corresponding author.
E-mail address: Grassel@iue.tuwien.ac.at (T. Grasser).
can be well approximated by during recovery\[6,7,1,8\]. In particular, application of positive bias
1.4. Bias dependence of recovery
it turns out to be a considerable challenge for any modeling
1. Hole trapping process
Hole trapping is often modeled using elastic tunneling into pre-
existing traps located at various distances away from the interface.
Elastic hole trapping is to first-order temperature independent and
linearly dependent on the stress field [9]. Furthermore, the model
predicts symmetric degradation and recovery behavior. As such,
this is incompatible with our data and a different explanation
has to be sought.
Of particular interest are hole trapping models that have been
applied since the 1970s in attempts to understand 1/f-noise and
thermally stimulated currents at semiconductor surfaces [10,11].
These models are based on a dispersion of activation energies
which for a homogeneous distribution results in a 1/f behavior.
In these models it is assumed that holes can be captured via a
(thermally activated) multiphonon emission (MPE) process into
deep near-interfacial states/border traps [12,13,9,11], probably
into oxygen vacancies (\(E^0\) centers) [11]. The MPE process differs
from the conventionally invoked elastic tunneling process, notably
due to its temperature activation and the larger time constants
resulting therefrom [12].

In order to accommodate the field dependence, we have to recall
that the MPE mechanism is derived under the assumption of neg-
ligible electric fields. This assumption is definitely violated in the
case of NBTI. An extension of MPE to the large electric field case
has been developed for the emission of particles from deep traps
and known as multiphonon-field-assisted tunneling (MPFAT)
[14,15]. The signatures of this mechanism are its \(\exp(\frac{\theta}{C})\) field
dependence (note that only the logarithm of the enhancement
factor enters \(B\) and a considerable temperature activation (from
the MPE process). MPFAT is process is schematically illustrated in
Fig. 2: a hole can either be in the valence band or in a trapped
state. These two states are represented by the two solid parabola
which give the total energy of the system. The vibrational modes
can be approximated using a simple oscillator model and at the
intersection point a transition can occur. The intersection point
determining the barrier \(\Delta E_p\) is dramatically lowered by the appli-
cation of an electric field, resulting in an enhancement of
\(\exp(\frac{\theta}{C})\).

3. Properties of the \(E^0\) center

In order to develop an accurate microscopic model for hole
trapping in the context of NBTI, we summarize the most important
features collected in a long line of studies on oxide defects which
provides a solid basis for our NBTI model.

- The most likely microscopic candidate for the ‘trapped hole’ are
defects from the \(E^0\) center family, most notably the \(E^0\) center
[16,17]. An \(E^0\) center is thought to be created when a hole is
trapped in the precursor structure, which is commonly assumed
to be a neutral oxygen vacancy.
- The energy level for hole trapping is roughly located at about
1 eV below the Si valence band [16], see Fig. 3.
- Once the silicon bond is broken, the distance between the two
silicon atoms increases into a new equilibrium position, which
requires a large-range structural relaxation of the surrounding
lattice (10 Å [18]), and an \(E^0\) center is obtained. The \(E^0\) is visible
in ESR (electron spin resonance) when positively charged, that
is, right after hole capture.
- An important peculiarity of the \(E^0\) center is that it can be repeat-
edly charged and discharged. The corresponding energy-levels
lie within the silicon bandgap [16]. The idea behind this cyclabil-
ity is that once the hole is emitted (that is, an electron is cap-
tured), the bond between the two silicon atoms does not fully
refurn but the defect remains in a dipole state which can easily
lose an electron again. The fact that the \(E^0\) center can act as a
switching trap was suggested by Lelis et al. [19] and led to the
formulation of the Harry-Diamond-Laboratories (HDL) model
of the \(E^0\) center. The model was later confirmed by ESR studies
[20] and theoretical calculations [21], see Fig. 4 for a schematic
representation.

Although this asymmetry may look quite innocent at a first glance,
it turns out to be a considerable challenge for any modeling
attempt.

1.4. Bias dependence of recovery

Recovery has been shown to depend on the bias voltage applied
during recovery [6,7,1,8]. In particular, application of positive bias
accelerates recovery. Just as the asymmetry, this is also challenging
to reproduce correctly by a quantitative model.

2. Hole trapping process
Hole trapping is often modeled using elastic tunneling into pre-
existing traps located at various distances away from the interface.
Elastic hole trapping is to first-order temperature independent and
linearly dependent on the stress field [9]. Furthermore, the model
predicts symmetric degradation and recovery behavior. As such,
Only after having been in its electrically neutral state for a while, the structure relaxes again to the initial dimer configuration and the defect is completely healed.

The $E_0^c$ is often considered a donor-like defect [16], that is, it is either neutral (ESR inactive) or positively charged (ESR active). The oxygen vacancy can also act as an electron trap, with a trap level close to the silicon conduction band [21,22]. Due to the amorphous nature of the interfacial layer, a considerable distribution of the energy-levels is to be expected.

The above summary bears some important and interesting consequences regarding our understanding of NBTI:

- Only after having been in its electrically neutral state for a while, the structure relaxes again to the initial dimer configuration and the defect is completely healed.
- The $E_0^c$ is often considered a donor-like defect [16], that is, it is either neutral (ESR inactive) or positively charged (ESR active).
- The oxygen vacancy can also act as an electron trap, with a trap level close to the silicon conduction band [21,22].
- Due to the amorphous nature of the interfacial layer, a considerable distribution of the energy-levels is to be expected.

The above summary bears some important and interesting consequences regarding our understanding of NBTI:

- So far, hole trapping has been mainly considered as being into pre-existing traps which rapidly fill but are not related to the actual NBTI mechanism, a somewhat parasitic component which has to be removed to get to the heart of NBTI [23].
- Using CV measurements, NBTI has been shown to be due to donor-like defects [24,25,23]. In contrast to interface states, which are amphotheric, that is, donor-like only in the lower half of the silicon band-gap, the $E_0^c$ center is considered a donor-like defect [16].
- NBTI recovery is strongly bias-dependent, in particular when the gate voltage is moved from inversion into accumulation [6]. This is intuitively consistent with carrier trapping [26].
- Describing hole trapping via the known properties of the $E_0^c$ center promotes the positive oxide charge component from a purely parasitic component to the central contributor to NBTI. Indeed, some key experimental features of NBTI which are incompatible with the $P_b$ center and simple hole trapping models follow directly from the properties of the $E_0^c$ center.
- The amount of positive charge visible and thus contributing to $\Delta V_{th}$ will depend on the position of the Fermi-level, that is, the gate voltage at which the degradation is monitored. For example, during OTF experiments, the Fermi-level is below the valence band edge and most defects will be positively charged (state 2 in Fig. 4) and thus visible. During recovery, the Fermi-level is moved towards mid-gap and a smaller frac-
tion of the defects will be positive and visible. This occupancy effect will manifest itself as a change in the subthreshold slope, and is consistent with recent experimental results [27,26].

Since full annealing of oxide defects is only possible when the defect is neutral (state 3), defects having an electrically higher trapping level will show a slower recovery rate. Indeed, this is fully compatible with what is observed during NBTI recovery and explains the often observed bias dependence. Furthermore, this can explain the asymmetry between stress and recovery, with the recovery lasting considerably longer than the time required to build up the degradation.

The fact that the precursor level is below the valence band and the defect level within the silicon bandgap is precisely what is expected of a defect responsible for NBTI. A higher energy level of the precursor, e.g. above the valence band, would cause most precursors to be already initially broken in PMOSFETs. The energy level of the created defect inside the silicon bandgap results in most defects to be positively charged (a donor-like defect) during both stress and recovery. A lower defect level would render the defects electrically neutral and thus not contributing to the threshold voltage shift.

Although an extensive amount of literature is available on the $E_0$ center and its qualitative behavior is well understood, no rate-equation based model that spans the full cycle of trap creation, recharging and discharging until final annealing seems to be available. However, for NBTI a full model is required in order to account for the asymmetry between stress and relaxation and the correct bias dependence of recovery [30].

---

**Fig. 4.** Left: The HDL model for a switching oxide trap coupled to the creation of a dangling bond at the interface. When the $E$ center is positively charged (in state 2), the hydrogen passivating a silicon dangling bond at the interface can move to the $E$ center, thereby effectively locking in the $E$ center (state 4). The charge state of the thereby created dangling bond depends on the position of the Fermi-level. Bottom: The transition between states 2 and 4 is modeled by assuming a field-dependent thermal transition over a barrier.

**Fig. 5.** The measured prefactor $B(F,T)$ of Fig. 1 compared to the simulated prefactor of the two stage model obtained under the same conditions. Very good agreement for all voltages and temperatures is obtained. Since the two stage model captures the asymmetry, good agreement during both stress and recovery is possible. We put the remaining deviation down to the mobility error in the OTF data [34].
4. A two stage model

Based on the above observations we formulate a new model for NBTI, where defect creation proceeds via a two stage process: In stage one, upon application of stress, holes can be trapped into near-interfacial oxygen vacancies via the MPE/MPFAT mechanism. In the second stage, the increased hole concentration considerably enhances the creation of poorly recoverable defects, e.g. P$_b$-centers in SiO$_2$ layers and K$_o$-centers in oxynitrides [31].

Our full model for the E$^\circ$/P$_b$ H complex relies heavily on the HDL switching trap model [19]. In order to capture this coupling between E$^\circ$ and P$_b$ centers we extend the HDL model by introducing an E$^\circ$/P$_b$H complex. The complex is assumed to be in one of 4 states, with states 1–3 describing the dynamics of the E$^\circ$ center and the P$_b$ center being passivated with H (see Fig. 4). Once positively charged (state 2), the E$^\circ$ center can attract the H from the P$_b$H. When this happens, the E$^\circ$/P$_b$H complex moves to state 4. This step locks in K$_o$-centers in oxynitrides [31].

Alternatively, it might be possible that the hole capture event into the E$^\circ$ center releases atomic hydrogen which then depasses a passivated interface state by forming H$_2$ [32]. As the microscopic details of the coupling mechanism are not evident from the data available to us at the moment, we anticipate some refinement of our model. However, we expect the resulting equation systems to be of a similar form as the one suggested below.

The rate-equation describing such a E$^\circ$/P$_b$H complex follow in a straight-forward manner from Fig. 4

$$\frac{df_1}{dt} = -f_1 k_{12} + f_1 k_{31},$$

$$\frac{df_2}{dt} = f_2 k_{12} - f_2 k_{23} + f_2 k_{32} - f_2 k_{24} + f_2 k_{42},$$

$$\frac{df_3}{dt} = f_3 k_{23} - f_3 k_{32} - f_3 k_{31},$$

$$\frac{df_4}{dt} = f_4 k_{24} - f_4 k_{42}.$$  (4)

Naturally, (1)–(4) are not linearly independent since the defect has to be in one of its four states ($f_1 + f_2 + f_3 + f_4 = 1$). The probability of being in state $i$ is given by $f_i$ while the transition rates from state $i$ to $j$ are given by $k_{ij}$. The rates $k_{ij}$ are formulated using a generalization

---

**Fig. 6.** Top: Comparison of the simulated asymmetry of stress and recovery measured at 150 °C for 8 different stress voltages for the thin SiON devices. The asymmetry is properly reproduced by the model, resulting in a good fit during both stress and recovery. Bottom: Detailed simulation results at 50 °C and −2 V. Both measurements are taken at 1 ms and miss about 50 mV of the real degradation.
of the lattice-relaxation multiphonon emission (MPE) theory for phonon-assisted capture of holes and electrons [12,9] by including a depth dependence and MPFAT field acceleration [14,15].

Consider a trap level $E_T$ located at a distance $x$ away from the interface. The hole capture and emission rates are then approximately given by

$$k_p^c = \alpha_p^p \sigma_p e^{-x/x_0} e^{-\beta E_F} \left( E_F, e^{-\beta E_F}, 1 \right) e^{E_F/\beta},$$

$$k_n^c = \alpha_n^p \sigma_n e^{-x/x_0} e^{-\beta E_F} \left( E_F, e^{-\beta E_F}, 1 \right),$$

while the corresponding rates for electrons read

$$k_p^e = \alpha_p^p \sigma_p e^{-x/x_0} e^{-\beta E_F} \left( E_F, e^{-\beta E_F}, 1 \right),$$

$$k_n^e = \alpha_n^p \sigma_n e^{-x/x_0} e^{-\beta E_F} \left( E_F, e^{-\beta E_F}, 1 \right).$$

Here, $p$ and $n$ are the hole and electron concentrations in the channel, $v_p^\text{th}$ and $v_n^\text{th}$ their thermal velocities ($\sqrt{8k_B T/(\pi\hbar)}$), $\sigma_p$ and $\sigma_n$ their capture cross sections ($\sim 3 \times 10^{14} \text{cm}^2$ [33]), $E_F$ the Fermi-level in the channel, $E_k$ and $E_c$ the valence and conduction bands directly at the interface (classical approximation), $\Delta E$ the MPE barrier, and $\beta = 1/k_BT$. $F_k$ is the reference field for the multiphonon-field-assisted tunneling mechanism which is, due to lack of decisive data, only introduced for hole capture. According to a simplified WKB approximation for large tunneling barriers $\phi$ [9], $x_0 = \hbar/(2\sqrt{2m\phi})$, $x_0,0 = 0.8\text{Å}$ for electrons ($\phi_k \approx 3.2eV$ and $m \approx 0.5m_0$) and $x_p,0 = 0.5\text{Å}$ for holes ($\phi_k \approx 4.65eV$ and $m \approx 0.8m_0$). Furthermore, we use the shorthand $E_{\text{Ar}}$ $E_{\text{B}}$ and the auxiliary function

$$\theta\left(E_{\text{switch}}, a, b\right) = \begin{cases} a & E_{\text{switch}} > 0 \\ b & E_{\text{switch}} < 0 \end{cases}$$

to account for the fact that thermal activation is required for hole capture into a trap below $E_v$ while capture in a trap above $E_v$ proceeds without activation (the hole ‘bubbles up’), and vice-versa for electrons.

We take the simplest possible case that can capture the currently available data. We assume that when the defect is in state 1 the trap energy lies at $E_T$, has a MPE barrier of $\Delta E_b$, and a MPFAT reference field $F_c$. When in states 2 and 3, the defect level is assumed to be at $E_T$ with a small charging/discharging MPE barrier $\Delta E_c$. Although the barrier $\Delta E_c$ is expected to be considerably lower than the barrier $\Delta E_b$, it is responsible for the ‘slow states’ in CP measurements. Nonetheless, for the data investigated here, $\Delta E_c$ can be neglected. We also neglect the MPFAT mechanism for charging/discharging since our data taken to sense these characteristics are recorded at relatively low fields. Transition from state 3 to state 1 (full annealing rather than electrical neutralization) proceeds over a barrier $\Delta E_n$. Thus, the rates read as follows

$$k_{12} = k_p^c\left(E_T, \Delta E_b, F_c\right) + k_n^c\left(E_T, \Delta E_b, F_c\right),$$

$$k_{23} = k_p^c\left(E_T, \Delta E_c, 0\right) + k_n^c\left(E_T, \Delta E_c, 0\right),$$

$$k_{32} = k_p^c\left(E_T, \Delta E_c, 0\right) + k_n^c\left(E_T, \Delta E_c, 0\right),$$

$$k_{31} = \exp\left(-\beta \Delta E_n\right).$$

where the shorthand $k\text{trap level, MPE barrier, MPFAT reference field}$ is used and $\nu = 10^{13} \text{Hz}$ is the typical attempt frequency for thermal transitions over energetic barriers.

The transition rates between states 2 and 4 are modeled in the spirit of [1] by thermal activation over a field-dependent barrier (cf. Fig. 4) as

$$k_{34} = \nu e^{-\beta \left( \phi_0 - E_v - \gamma F \right)},$$

$$k_{42} = \nu e^{-\beta \left( \phi_0 - E_v + \gamma F \right)}.$$ 

Consequently, the probability of being in state 4 corresponds to a positive charge at the $E$ center and a depassivated interface state.

In order to describe the response of a device to a change in the bias conditions, a certain number of defects $N$ is assumed to exist. Due to the amorphous nature of the Si/SiO$_2$ interface, each defect will be described by a unique configuration of random energy-levels and the overall created interface and oxide charges are obtained from suitably defined averages.

In order to explain the presently available data, a reduced set of random energy-levels is sufficient and we use $x \approx 0$, $\Delta E_A = \Delta E_b$, and $\Delta E_c \approx 0$ in the following examples. Further simplification, like $E_r \approx E_t$, significantly deteriorates the quality of the model [30].

5. Comparison with measurements

Simulation results for the SION devices are compared to the data of Fig. 1. The simulated prefactors during stress are in very good agreement with the data, see Fig. 5. Details of the simulation demonstrate also that the model can reproduce the asymmetry between stress and recovery [30]. Details of the simulation demonstrate also that the model can reproduce the asymmetry between stress and recovery, see Fig. 6. We remark that this is the first time that a model can reproduce both OTF and recovery data in a wide temperature and voltage range, providing a theoretical link between these two measurement techniques. Also, good agreement between experiment and model is obtained for high-$k$ and ultrathick SiO$_2$ devices [30].

A particularly challenging data set is given in Fig. 7, where five devices are brought to the same level of degradation. During recovery, various bias switches are used to probe both the occupancy effect (amount of charge visible depending on the Fermi-level) as well as the impact of the occupancy on the recovery dynamics. Again, very good agreement between theory and data is obtained.

6. Conclusions

We suggest a two stage model for the negative bias temperature instability based on established properties of $E$ centers. Creation of
$E^0$ centers from their oxygen vacancy precursors is suggested to occur via a multiphonon-field-assisted hole trapping mechanism. The created $E^0$ centers are then suggested to favor the depassivation of interface states, which results in a coupling of created oxide and interface state component.

This model can explain degradation and recovery over a wide range of bias voltages and stress temperatures, the observed asymmetry between stress and recovery, and the strong sensitivity to bias and temperature during recovery. Excellent agreement with data from three vastly different technologies (thick SiO$_2$, SiON, and HK) is obtained, supporting the idea that NBTI is determined by the chemistry of the amorphous SiO$_2$/Si interface region. The model has the minimum number of parameters required to explain the experimentally observed features of NBTI.

Acknowledgements

The research leading to these results has received funding from the European Community’s Seventh Framework Programme under Grant agreement No. 216436 (project ATHENIS). We also gratefully acknowledge stimulating discussions with P. Lenahan and A. Shluger.

References