Switching Oxide Traps as the Missing Link Between Negative Bias Temperature Instability and Random Telegraph Noise

T. Grasser*, H. Reisinger$^\bullet$, W. Goes*, Th. Aichinger†, Ph. Hehenberger$^\diamond$,
P.-J. Wagner*, M. Nelhiebel‡, J. Franco$^\circ$, and B. Kaczer$^\circ$

* Christian Doppler Laboratory for TCAD at the $^\diamond$Institute for Microelectronics, TU Wien, A-1040 Wien, Austria
$^\bullet$ Infineon, München, Germany † KAI, Villach, Austria ‡ Infineon, Villach, Austria $^\circ$ IMEC, B-3001 Leuven, Belgium

Abstract

Due to the ongoing reduction in device geometries, the statistical properties of a few defects can significantly alter and degrade the electrical behavior of nano-scale devices. These statistical alterations have commonly been studied in the form of random telegraph noise (RTN). Here we show that a switching trap model previously suggested for the recoverable component of the negative bias temperature instability (NBTI) can more accurately describe the bias and temperature dependence of RTN than established models. We demonstrate both theoretically and experimentally, that the recovery following bias temperature stress can be considered the non-equilibrium incarnation of RTN, caused by similar defects. We furthermore demonstrate that the recoverable component is solely constituted by individual and uncorrelated discharging of defects and that no diffusive component exists. Finally it is highlighted that the capture and emission times of these defects are uncorrelated.

Introduction

In future nano-scale MOSFETs only a handful of defects will be present in the oxide above the channel region which can have a significant stochastic impact on their operation [1]. In order to understand circuits using such devices, as well as being able to estimate their reliability, one has to study the dynamic behavior of these defects [2]. We have recently reported experimental data which show that both $1/f$ noise and negative bias temperature instability (NBTI) in pMOSFETs are due to defects with very similar properties [3] and we have already successfully described their voltage and temperature dependence [4]. Consequently, the defects responsible for random telegraph noise, which have been suspected to also be the fundamental building blocks of $1/f$ noise [5, 6], could play a similar role in NBTI.

It has been demonstrated that the reduction in the random telegraph noise power brought about by bias switching can be described by a charge trapping model because the defects have to adjust to the new bias condition [7]. Here we will show that the same is also valid under the heavy stress conditions typical for NBTI by demonstrating that both the *quasi-equilibrium* (RTN) and the *non-equilibrium* (i.e., NBTI stress and recovery) behavior can be successfully described by our switching trap model. The main difference is that *defects with larger time constants are activated in NBTI*, resulting in the characteristic long relaxation curves with time constants below $1\,\mu$s and longer than 11 days [8]. The theoretical predictions obtained from our model will be confirmed by carefully designed experiments recorded on nano-scale pMOSFETs ($W/L = 150\,\mathrm{nm}/100\,\mathrm{nm}$). A particularly intriguing observation further supporting the idea that individual switching traps constitute the overall degradation is that the capture and emission times are essentially uncorrelated. Based on this evidence it also has to be concluded that the characteristic switching behavior *is not due to a diffusive process*, such as assumed in the popular reaction-diffusion (RD) theory [9, 10].

Previous Modeling Approaches for RTN

Modeling of RTN and $1/f$ noise dates back to the work of McWhorter [11]. Irrespective of the fact that variants of this elastic tunneling model are still frequently used, it has been repeatedly shown that they can neither explain the temperature nor the bias dependence of the time constants [6, 12–14]. Kirton and Uren have used a lattice-relaxation multiphonon emission (LRME) process [6], see Fig. 1, but also observed that the required capture cross sections showed a bias dependence stronger than the expected $1/p$ behavior (p being the surface hole concentration), see Fig. 2. This mismatch has repeatedly been demonstrated and prompted various authors to introduce empirical corrections to the capture cross sections [15].

For thicker oxides such a strong bias dependence has often been successfully explained by the Coulomb blockade [13, 14], which is introduced by having to move the mirror charges against the external bias source. However, this explanation fails for thin oxides, as there for a defect inside the oxide the Coulomb barrier is dominated by the mirror charge on the gate, resulting even in a turn-around of the bias dependence [16], see Fig. 3. We consequently have to conclude that the experimentally observed bias dependence cannot be properly explained by existing theories [15].

The Switching Trap Model for RTN and NBTI

We will thus model RTN using the Harry Diamonds Lab (HDL) switching trap model [17] previously suggested for the description of recoverable charge trapping in NBTI [4], see Fig. 4. It assumes that charge trapping *creates a defect*, possibly an E' center, which can be repeatedly charged and discharged and anneals only when in the neutral state. For most purposes, charging of the already created defect can be considered fast, allowing us to derive simpler expressions for the effective capture and emission times (the slow defect creation and annealing) using an 'effective Poissonian model' (middle of Fig. 4). In comparison to previously published models, the switching trap model can predict a much stronger bias dependence of the time constants: for the capture time this is a consequence of the pre-cursor defect level lying below (rather than above) the silicon valence band, which requires the hole to be thermally activated rather than just captured. On the other hand, the emission time becomes strongly bias dependent due to the fact that only neutral defects can anneal. Quite intriguingly, the switching trap model, previously developed for NBTI, can also best explain the bias dependence of the RTN capture and emissions times in thin oxides, see Fig. 4. The striking fact that standard charge trapping models fail to completely reproduce the bias dependence of RTN and the recoverable component of NBTI [4], while the switching trap model captures both phenomena, is strong evidence for its correctness.

The Link between RTN and NBTI: A qualitative demonstration of how such defects can be responsible for both RTN and NBTI is given in Fig. 5. For this the switching trap model is written as a stochastic differential equation. Prior to stress, the defects are in quasi-equilibrium and depending on their capture and emission times some are mostly neutral, some mostly positively charged, while the remaining defects randomly capture and release charge, thereby creating detectable RTN. When the bias is changed to the stress voltage, this equilibrium is disrupted by the strong bias dependence of the time constants. Depending on these time constants, a previously neutral defect can start producing RTN or even become predominantly positively charged. When the bias is switched back to the initial value, each defect responds following its own time constants, thereby restoring the previous equilibrium, visible in experimental data obtained on nano-scale devices as discrete recovery steps [3]. However, as such the steps observed in the experimental data are itself no proof that the underlying mechanism is given by such a charge trapping mechanism.

The Stochastic Behavior of the Switching Trap Model: In order to properly interpret these steps, we have to consider their stochastic nature [18]. Stochastic in this case means that the mean capture and emission times only give the probability of the occurrence of a step. This implies that it is not sufficient to study a single recovery trace which is just one possible realization of the stochastic process. Rather, a sufficiently large number of full recovery traces has to be averaged to reveal the characteristic Poissonian behavior, $A_i \exp(-t_\mathrm{r}/\tau_i)$, with τ_i being the emission time of defect i, A_i the bias and trap-location dependent amplitude [1], and t_r the recovery time. In a nano-scale device only a small number of defects contribute to the recovery, and a $\sum_i A_i \exp(-t_\mathrm{r}/\tau_i)$ recovery behavior would be expected (steps of height A_i at $t_\mathrm{r} = \tau_i$), see Fig. 6. Consequently, since each device has its own collection of defects, each with its own time constants, averaging recovery traces of different nano-scale devices would produce a distinctly different pattern of time constants τ_i for each device. Only in the limit of a large-area device with many defects, these individual steps are washed out to give the familiar $\log(t_\mathrm{r})$ behavior [8].

The Stochastic Behavior of the RD Model: The situation is quite different for the case of the RD model which assumes that hole capture and emission are directly linked to the creation and annealing of interface states, controlled by a diffusive process. Although RD theory has as of yet only been studied in its macroscopic form, a nano-scale representation of RD theory can be obtained by expressing the electrochemical reaction at the interface and the subsequent diffusion using stochastic differential equations. The recovery behavior of the stochastic RD model is shown in Fig. 7, which, as expected, also proceeds in steps. However, when a number of such traces is averaged, the behavior of the macroscopic RD model with a single transition lasting about 4 decades is obtained. Furthermore, since the RD model assumes no dispersion in either the interface reaction nor in the diffusion [19], all devices behave identically. We finally remark that in charge trapping models the steps always occur at the same time, while in a diffusive mechanism the single-big-step has its inflection point when the recovery time equals the stress time [20] ($t_\mathrm{s} = t_\mathrm{r}$), see Fig. 8. These theoretical predications can be used to experimentally differentiate between a charge trapping or a diffusive mechanism.

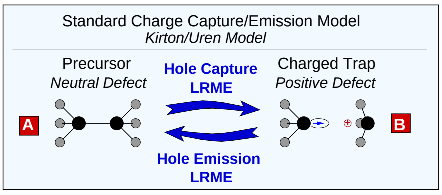
Experimental Validation

Experimental evidence was gathered on narrow SiON devices with $t_\mathrm{ox} = 2.2\,\mathrm{nm}$ and $1.8\,\mathrm{nm}$, see [21] for details. Ensuring complete recovery, devices were repeatedly stressed under the same conditions and the data were averaged. Depending on the stress condition a permanent degradation was also observed [4, 22], probably due to interface states, which, however, does not contribute to the slow switching behavior.

A typical example showing the contribution of a single defect is given in Fig. 9. Longer stresses activate more defects as shown in Fig. 10 and the capture and emission times are clearly uncorrelated. Fig. 11 demonstrates the temperature dependence of both the capture and the emission time of a single defect, consistent with the switching trap model, while Fig. 12 shows that the averaged traces do not depend on the stress time. These data convincingly confirm that, just like with RTN, the physical mechanism behind the recoverable component of NBTI is random trapping and detrapping of charge as described by the switching trap model because *(i)* the characteristic emission times of each trap are fixed in time (if we were dealing with a diffusion-controlled mechanism, at least a hint of 'moving traces' should be detectable while not a single one was found), *(ii)* each averaged switching event covers about 1.3 decades in time as predicted by the model (classical diffusion covers about 3.8 decades), *(iii)* the capture and emission times are uncorrelated (this is definitely not possible with a diffusion controlled mechanism), and *(iv)* both NBTI and RTN can be explained by a single model.

Conclusions

We have demonstrated both theoretically and experimentally that RTN and the recoverable component of NBTI are due to charge trapping in switching oxide traps, the main difference being that NBTI stress activates defects with larger time constants. Most importantly, the capture and emission times of the defects are uncorrelated, revealing for the first time explicitly that individual defects constitute the recovery of NBTI.



$$\tau_\mathrm{c} = \tau_0\, \mathrm{e}^{\beta \Delta E_\mathrm{B}}\, \frac{N_\mathrm{v}}{p}$$

$$\tau_\mathrm{e} = \tau_0\, \mathrm{e}^{\beta \Delta E_\mathrm{B}}\, \mathrm{e}^{\beta \Delta E_\mathrm{T}}\, \mathrm{e}^{xF/V_\mathrm{T}}$$

$$\tau_0^{-1} = N_\mathrm{v}\, v_p^\mathrm{th}\, \sigma_p\, \mathrm{e}^{-x/x_{p,0}}$$

Fig. 1: The standard model for RTN and $1/f$ noise was suggested by Kirton and Uren and is based on a LRME process. The oxide trap level $E_\mathrm{T} = E_\mathrm{T0} - \mathrm{q}\varphi_\mathrm{s} + \mathrm{q}xF$ is assumed to lie within the silicon bandgap where $\Delta E_\mathrm{T} = E_\mathrm{T0} - E_\mathrm{V0}$ determines the magnitude of the emission time constant τ_e. ΔE_B is the LRME barrier, F the modulus of the oxide field, $\beta^{-1} = \mathrm{k_B}T$, $V_\mathrm{T} = \mathrm{k_B}T/\mathrm{q}$, x the distance of the defect into the oxide, v_p^th the thermal hole velocity, φ_s the surface potential, $x_{p,0}$ from a simple WKB approximation, and σ_p the capture cross section.

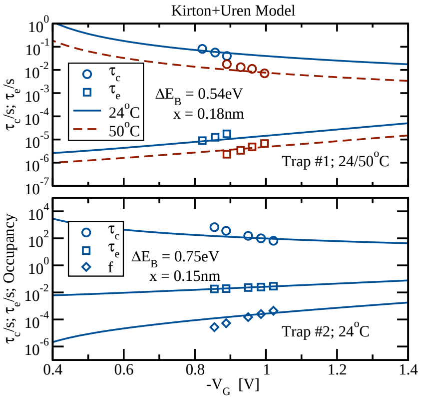


Fig. 2: Comparison of experimental data for two selected (because difficult to model) defects in a SiON pMOS with $t_\mathrm{ox} = 1.8\,\mathrm{nm}$. While the Kirton/Uren model qualitatively fits the bias dependence it cannot properly capture the strong bias dependence of both τ_c and τ_e at the same time. Fits of similar quality can be found in literature [7, 15] and are not a peculiarity of the selected defect.

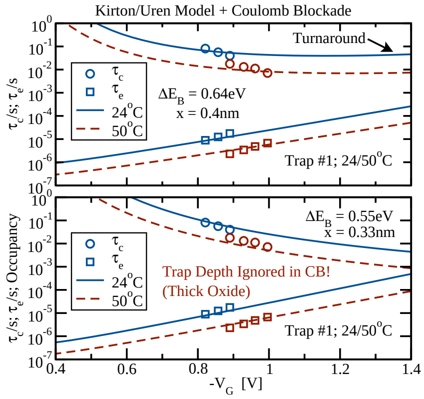
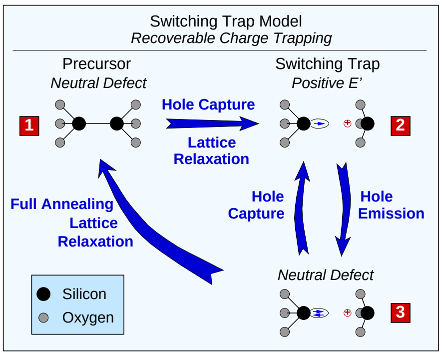
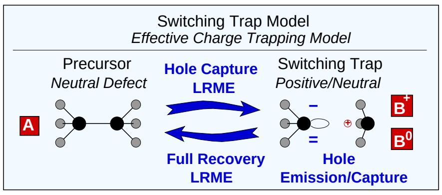


Fig. 3: Top: No improvement is obtained in the Kirton/Uren model when the Coulomb blockade (CB) is considered since in a thin oxide the CB energy begins to increase with increasing $|V_\mathrm{G}|$ even for a repulsive defect located inside the oxide [16]. **Bottom**: Only when the charge is (wrongly) treated as an interface state, the CB energy improves the quality of the fit.





$$\tau_\mathrm{c} = \tau_0\, \mathrm{e}^{\beta \Delta E_\mathrm{B}}\, \frac{N_\mathrm{v}}{p}\, \mathrm{e}^{-F^2/F_\mathrm{c}^2}\, \underbrace{\mathrm{e}^{-\beta \Delta E_\mathrm{T}}\, \mathrm{e}^{-xF/V_\mathrm{T}}}_{\text{Only for } E_\mathrm{T} < E_\mathrm{V},\ \text{otherwise } 1}$$

$$\tau_\mathrm{e} = \frac{\mathrm{e}^{\beta \Delta E_\mathrm{B}}}{\nu} \left(1 + \frac{p}{N_\mathrm{v}}\, \mathrm{e}^{\beta \Delta E'_\mathrm{T}}\, \mathrm{e}^{xF/V_\mathrm{T}}\right)$$

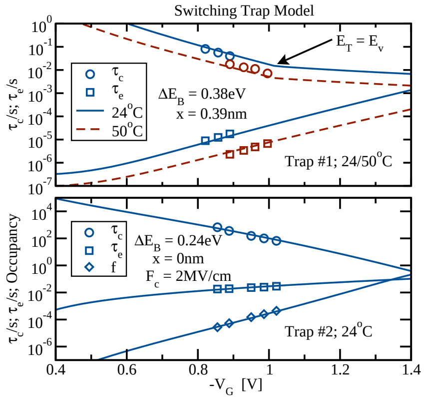


Fig. 4: Left: The Harry-Diamond-Labs (HDL) switching trap model [17] used previously to model the recoverable charge trapping component of NBTI (stage one) [4]. It assumes that a defect is created from the precursor state 1 by capturing a hole via a field-assisted LRME mechanism [23, 24]. Once the defect is created, its charge state quickly follows the Fermi-level, being positive in state 2 and neutral in state 3. Full annealing of the defect is possible from the neutral state 3 only. **Middle**: Under the assumption that the transition rates between states 2 and 3, k_{23} and k_{32}, are much larger than k_{12} and k_{31}, effective rates $k_\mathrm{AB} \approx k_{12}$ and $k_\mathrm{BA} \approx k_{31}/(1 + k_{32}/k_{23})$ can be defined. The factor $1/(1 + k_{32}/k_{23})$ gives the occupancy of state B according to Fermi-Dirac statistics and considerably increases the charge detrapping time $\tau_e = 1/k_\mathrm{BA}$ when the defect is positively charged. The parameters are the precursor energy level E_T which is assumed to lie below the valence band edge ($\Delta E_\mathrm{T} < 0$), the energy level of the created defect $\Delta E'_\mathrm{T} = E'_\mathrm{T0} - E_\mathrm{V0}$, assumed to be inside/close to the valence band edge ($\Delta E'_\mathrm{T} > 0$), F_c the reference field for the field-assisted LRME process, $\nu = 10^{13}\,\mathrm{s}^{-1}$ the phonon frequency. **Right**: The switching trap model accurately predicts the temperature and bias dependence of the capture and emission time-constants. When the precursor level E_T crosses the valence band edge E_V, a transition to the standard bias dependence $1/p$ is observed. Such a behavior has been experimentally observed [25].

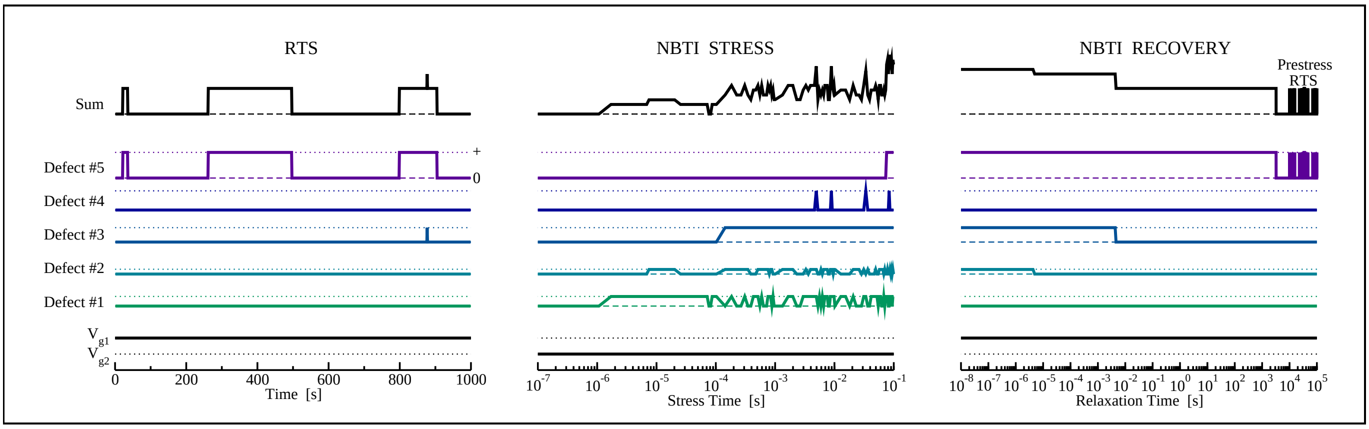


Fig. 5: Simulated RTN, NBTI stress, and NBTI recovery behavior of a nano-scale device using the stochastic solution algorithm (SSA) [26]. Results are obtained from a stochastic version of the switching trap model of Fig. 4. The model parameters are taken from an extensive calibration to ultrathin SiON devices [4] but the number of defects was reduced to 5, qualitatively demonstrating the response in a nano-scale device. **Left**: At the threshold voltage (V_g1), the RTN is dominated by defect #5 with the occasional contribution from defect #3. Defects #1, #2, and #4 remain positively charged within the 'simulation/experimental' window. **Middle**: During NBTI stress (V_g2), the capture times are dramatically reduced by the higher (more negative) gate voltage and the defects #3 and #5 become predominantly positively charged ($\tau_\mathrm{c} \ll \tau_\mathrm{e}$). Defects #1, #2, and #4 start producing RTN with $\tau_\mathrm{c} < \tau_\mathrm{e}$. **Right**: During NBTI recovery (back at V_g1), trapped charge is subsequently lost and the quasi-equilibrium behavior is gradually restored.

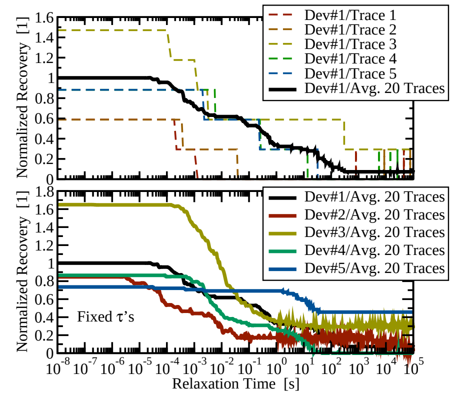


Fig. 6: Top: In the switching trap model, each individual recovery trace shows random step-like transitions. Only when recovery traces of the **same** device are averaged, the time-constants can be extracted. **Bottom**: Application of the above procedure to different devices reveals that each device has a different individual set of defects with **different** time-constants.

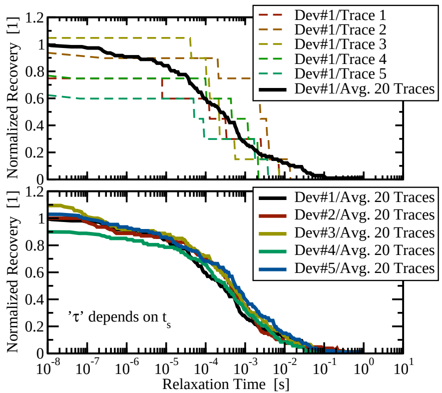


Fig. 7: Top: As with the switching trap model, the stochastic RD model also shows random step-like transitions. When recovery traces of the **same** device are averaged, the macroscopic recovery behavior [20] $1/(1+\sqrt{t_s/t_r})$ is obtained. **Bottom**: Since no dispersion is considered in the RD model, averaging recovery traces of **different** devices results in the **same** normalized behavior.

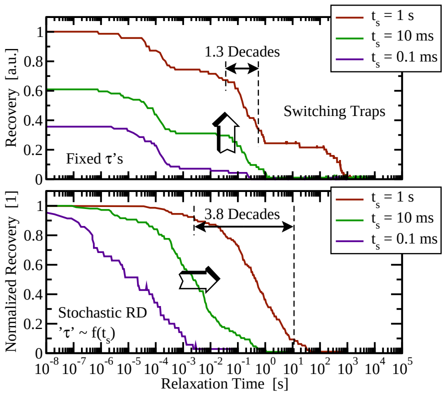


Fig. 8: Top: In the switching trap model, additional traps contribute with increasing stress time, while the time-constants of the individual traps remain fixed. **Bottom**: The effective time-constant of the (stochastic) RD recovery is proportional to the stress time and thus moves with time. The single RD step has a width (90% to 10%) of about 3.8 decades while each switching trap covers 1.3 decades.

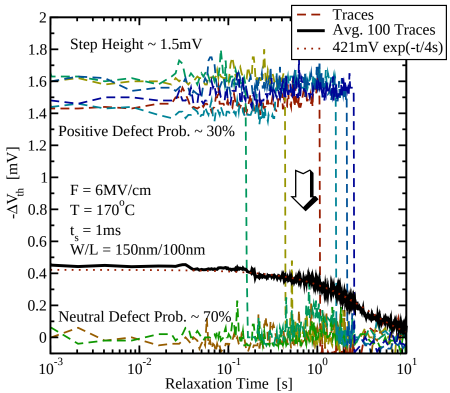


Fig. 9: Experimental recovery traces of a defect with a characteristic step-height of 1.5 mV. The probability that the defect is positively charged after a stress at 5 MV/cm for 1 ms is about 30%, implying that 70% of the traces show no signal in this window (2 examples shown). Averaging results in the expected $\exp(-t/\tau)$ behavior. Note that a 1 ms stress introduces a positive charge which recovers only after 4 s.

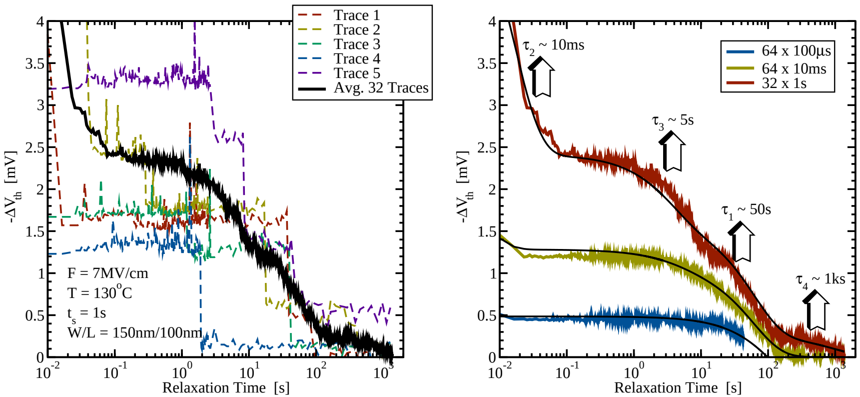


Fig. 10: Left: After a stress time of 1 s a larger number of defects become positively charged. Averaging of the random signal in the recovery traces results in clearly visible steps, cf. Fig. 6. We remark that the 32 consecutive stresses of 1 s introduce an increasing permanent contribution which does not recover within the experimental window [4, 22] and is not further studied here. **Right**: The evolution of the averaged recovery traces with increasing stress time shows that defects created after longer stress times can emit their hole prior to defects created earlier. For example, the first defect to become charged (visible after 100 μs) has an emission time constant of about 50 s (τ_1) while only after 1 s of stress the defects with emission time constants 10 ms and 5 s are created. This demonstrates that *capture and emission times in nano-scale devices are essentially uncorrelated*.

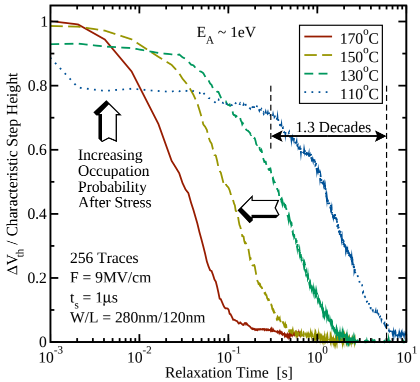


Fig. 11: Averaged recovery traces due to a single defect at four different temperatures. The data are normalized to the temperature-independent step-height for this trap (1.35 mV) and show clearly temperature dependent time constants. With increasing temperature the occupation probability of the defect right after stress increases. Above 150 °C nearly all defects are initially positively charged. Note how the 1 μs stress creates charge that is discharged only after 1 s.

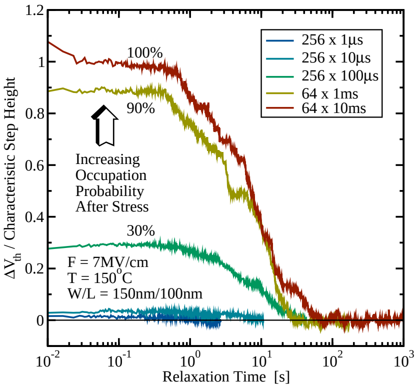


Fig. 12: Averaged recovery traces due to a single defect as a function of increasing stress time. At a short stress time of 100 μs the defect has an occupation probability of about 30% which only increases to 100% after 10 ms of stress. The time constants of all studied traces are found to be independent of the stress time, *ruling out a diffusion-controlled mechanism*, compare with Fig. 8.

References

[1] A. Asenov *et al.*, T-ED **50**, 839 (2003).
[2] E. Simoen *et al.*, Mat.Sci.Eng.B **91-92**, 136 (2002).
[3] B. Kaczer *et al.*, in *IRPS* (2009), pp. 55–60.
[4] T. Grasser *et al.*, in *IRPS* (2009), pp. 33–44.
[5] K. Ralls *et al.*, PRL **52**, 228 (1984).
[6] M. Kirton *et al.*, Adv.Phys. **38**, 367 (1989).
[7] J. Kolhatkar *et al.*, in *IEDM* (2004), pp. 759–762.
[8] H. Reisinger *et al.*, T-DMR **7**, 119 (2007).
[9] K. Jeppson *et al.*, JAP **48**, 2004 (1977).
[10] M. Alam, in *IEDM* (2003), pp. 345–348.
[11] A. McWhorter, Sem.Surf.Phys 207 (1957).
[12] M. Weissman, Rev.Mod.Phys **60**, 537 (1988).
[13] M. Schulz, JAP **74**, 2649 (1993).
[14] A. Palma *et al.*, PRB **56**, 9565 (1997).
[15] N. Zanolla *et al.*, in *ULIS* (2008), pp. 137–140.
[16] M. Lu *et al.*, PRB **72**, 235417 (2005).
[17] A. Lelis *et al.*, T-NS **41**, 1835 (1994).
[18] A. Karwath *et al.*, APL **52**, 634 (1988).
[19] D. Varghese *et al.*, in *IEDM* (2005).
[20] T. Grasser, in *IRPS* (2008), (Tutorial).
[21] H. Reisinger *et al.*, in *IIRW* (2009).
[22] V. Huard *et al.*, MR **46**, 1 (2006).
[23] S. Makram-Ebeid *et al.*, PRB **25**, 6406 (1982).
[24] S. Ganichev *et al.*, Phys.SS **39**, 1703 (1997).
[25] N. Lukyanchikova *et al.*, APL **73**, 2444 (1998).
[26] D. Gillespie, J.Comp.Phys. **22**, 403 (1976).