A study of NBTI by the statistical analysis of the properties of individual defects in pMOSFETS

H. Reisinger, T. Grasser* and C. Schlünder
Infineon Technologies AG, D-81730 Munich, Germany, *Institute for Microelectronics, TU Wien, Austria
Tel: +49 89 234 49210, Fax: +49 89 234 9555496, Email: hans.reisinger@infineon.com

ABSTRACT

Capture and emission of positive charge in individual defects has been studied using small area pMOSFETs. Analysis techniques similar to the ones used in RTS studies of nMOSFETs have been employed, with the difference that the capture process has been stimulated by stress pulses. We found that the recoverable part of the NBTI effect can be fully explained by capture and emission of these defects. As a consequence, a couple of previously used approaches trying to explain NBTI have to be discarded. The individual capture and emission time constants are widely spread over many orders of magnitude and are thermally activated with activation energies between 0.5 and 1.0 eV.

I. INTRODUCTION

After 30 years of NBTI research [1] there is still no consensus regarding the physical processes involved in NBTI. Measurement techniques employed so far have been the measurement of the VT-degradation, charge pumping measurements and some surface analysis like ESR [2]. In this work we investigate NBTI by the analysis of the temperature and gate-field dependence of the single, discrete events constituting NBTI degradation and recovery in small MOSFETs which are containing only a handful of electrically active defects. Some useful numbers regarding small FETs are compiled in table 1.

FET-name and dimension	wide		narrow		minimal	
	W	L	W	L	W	L
	10	**0.1**	**0.2**	**0.12**	**0.11**	**0.1**
number of carriers in channel at Vg=Vt-200mV	**15000**		**370**		**170**	
#Nit at DNIT=1E11/cm^2	**1000**		**24**		**11**	
ΔDNit causing a ΔVT=50mV	**4.9 x 10^{11}**		**4.9 x 10^{11}**		**4.9 x 10^{11}**	
makes Δ#NIt	**4900**		**120**		**50**	
ΔVT caused by a single trapped carrier	**0.01mV**		**0.43mV**		**1. mV**	

Table 1 – pFET geometries and useful numbers. Oxide thickness is 2.2nm. Measured specific capacitance in inversion is 1.3µF/cm^2. Real measured step-heights for the minimal FET (cmp. underlined value) range from 0.5mV to 5mV.

The analysis techniques we use are similar to those known from the Random Telegraph Signal (RTS) studies of the 1980s [3]. These RTS studies have been done for nMOSFETs only [4], and the FETs were kept in a quasi steady state. In contrast, our method uses multiple high-gate-field charging pulses to excite the defect into a charged state and then monitors the decay of the excited state. As such, our approach is similar is similar to deep-level transient spectroscopy (DLTS) studies [5]. The discrete nature of the discharging events in small pFETs has been recently noted [6]. What is new in this work is that we are repeating the same stress test over and over again in order to perform a statistical analysis of the capture- and emission-times. This way we are able to characterize the NBTI-relevant parameters of each single defect. Capture time and emission time, corresponding to stress and recovery, can be extracted as a function of gate voltage and temperature. Focus of this paper is primarily to show experiments, method, phenomena, the scope of the method, and a comparison to wide-FET results. All the implications of the results and all conclusions are drawn just on the basis of the experimental facts without involving a model or assumptions of any kind. An analysis of data with the goal to develop a microscopic physical model and identify the nature of the defects will be presented elsewhere [7].

II. SAMPLES AND EXPERIMENT

We use pMOSFETs with a 2.2 nm nitrided gate oxide. The FET geometry was chosen small enough to conveniently resolve the effect from a single carrier in the channel and large enough to have at least a handful of active defects (cmp. table 1). Stress pulses ranging from 200ns to 100s were applied to the gate. *ΔVT* corresponding to a preset drain current (typ. 1µA*W/L) was directly recorded after the pulse. We used our fast feedback loop described in [8]. This method saves the effort of recalculating *ΔVT* from *ΔID*. Measuring just the drain current steps *ΔID* would work as well but the step height of an individual defect would vary with the readout condition (i.e. gate voltage, temperature). *VT*-steps down to a size of 0.2mV could be analyzed. Each pulse and readout was done repetitively (32 to 256 times) to allow a statistical analysis. Full recovery was awaited after each stress pulse which made the experiments very time consuming.

III. COMPARISON OF WIDE AND NARROW FETS

The purpose of Figs. 1 and 2 is to show that NBTI degradation and recovery in wide and narrow FETs are qualitatively the same. The degradation in the narrow FETs is about 30% higher than in the wide FET which is presumably due to edge/stress effects. Despite this slight increase in magnitude, no difference in the recovery characteristics could be observed. It should be mentioned at this point that the wide W=10µm FETs used to reduce noise in conventional NBTI measurements are not the ones relevant in real circuits. Instead in SRAMs the minimal FETs are used and in logic circuits narrow FETs with at least sub-µm width. In order to see the thermal activation there is also a direct comparison of stress at 25°C and 170°C for the same electric field. At stress-time=100s the NBTI effect is enhanced by a factor 2 between 25° and 170°C. At short stress times below 10ms there seems to be no thermal activation. In the past this behavior gave rise to the assumption that the short time effects are not due to NBTI but to quantum mechanical tunneling which is supposed to be nearly independent of temperature. Our data will show, however, that all time constants - be they smaller than 1ms or longer - are strongly thermally activated with activation energy E_A≈1eV. We will explain this phenomenon of apparently missing thermal activation in Fig.1 in section VI. As seen in Fig. 2a there are no continuous changes in *VT*, all changes consist of discrete steps from individual defects. Only after averaging over a large gate-area the curves get smooth as shown in Figs. 2b and c. Each individual defect causes a specific step-height (cmp. Fig. 4), which allows to distinguish between individual defects as long as

978-1-4244-3922-5/09/$25.00 ©2009 IEEE

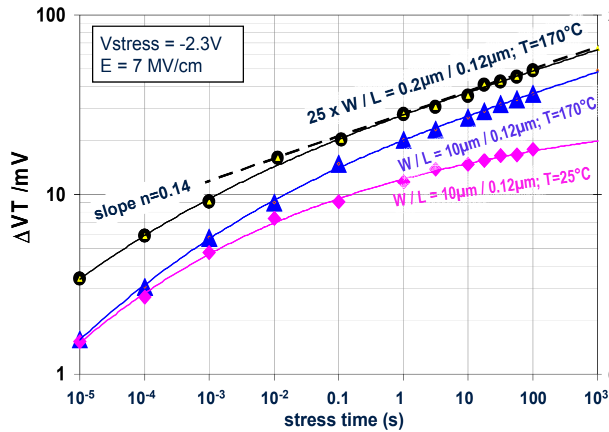

Fig. 1 – Comparison of degradation curves for a narrow and a 10µm wide FET at 170°C. The narrow FET data is from summing up 25 single 0.2µm FET data, corresponding to a total W=5µm (cmp. Fig2b). The wide FET is measured at 25°C also to see the thermal activation. Note that seemingly there is no thermal activation for short stress times.

not too many defects (<10) are active at the same time. To improve this distinguishability we continue with a minimal FET from here. All variations of parameters were done on the same FET which was always fully recovered between the repeated stress tests.

IV. Phenomenological model

All conclusions of this study will be based of experimental facts and phenomena found in this work without invoking any assumptions. We are not able to attribute our findings to a certain kind of oxide defect. On the other hand we are able to definitely exclude the validity of two approaches used in the past in conjunction with NBTI: **(1)** The Shockley-Read-Hall (SRH) recombination has been used to treat capture and emission. In SRH-theory non-equilibrium minority carriers (electrons or holes) with thermal velocity only are captured by deep states. There is no barrier, the quasi-Fermi energy of the recombining minority carriers is lower than kT, these carriers are not a Fermi-gas but a classical gas. In contrast in an inversion layer all subband-kinetic-energies are >kT [9], we have a real Fermi-gas. In addition, between inversion layer and the defect states we have a barrier with a height >>kT. **(2)** Elastic tunneling has also been used to explain some features of NBTI. In a 2nm oxide the widest barrier possibly ruling detrapping is 1nm, the maximum barrier height must be less than 4.5eV (1/2 oxide gap). Thus characteristic times would be below a µs (from WKB, and experimental [10]) and cannot explain the observed times >1000s. As the barrier height is >>kT, there would be not the strong temperature dependence observed by the experiment for all defects.
For an understanding of our results it will be useful to have a look onto the model by Kirton/Uren [4]. It has been used to interpret random telegraph signals (RTS) as due to capture and emission into/out of interface defects. The model is sketched in Fig. 3. Most crucial in this model is a structural relaxation of the oxide matrix upon capture and emission. Charging / discharging the defect causes an elastic switching between 2 meta-stable states of the oxide matrix. The charging/distortion results in a rearrangement of lattice atoms, involving overcoming of an energy barrier. Since heavy particles are involved the energy barrier cannot be overcome by tunneling but only by thermal activation. We want to stress again by introducing the Kirton/Uren model we do not want to anticipate any conclusion about the physical nature of the defects. Instead we just want - without any assumptions - to stick to the experimental facts

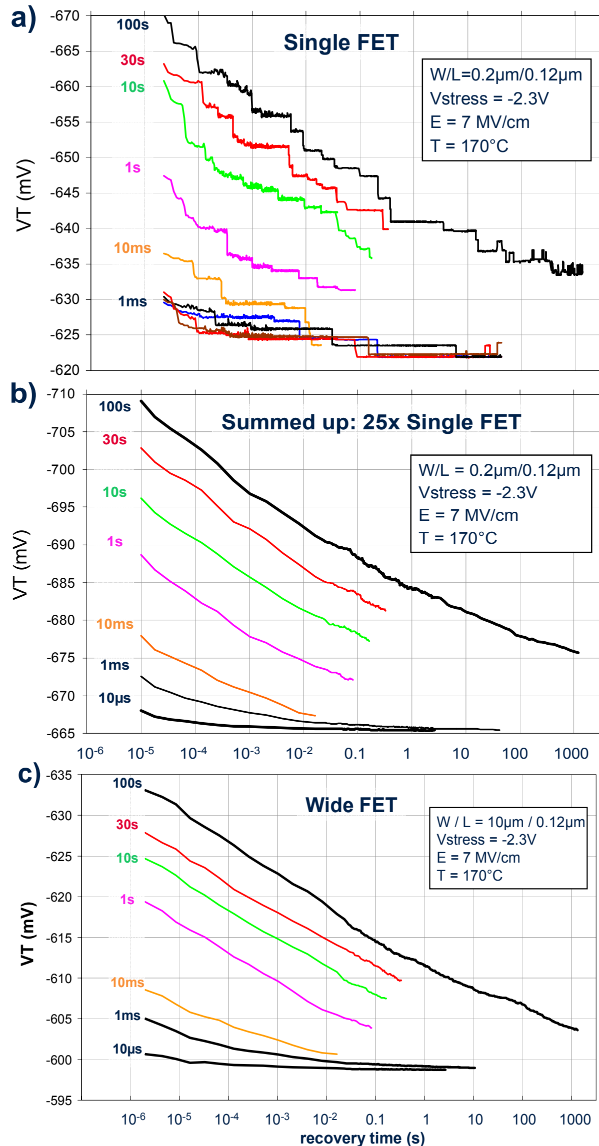

Fig. 2 – Comparison of recovery curves (same set of data as Fig.1) for narrow and wide FET. Fig. a) shows one of the 25 measured narrow FETs as an example. Labels at left hand side give the stress-time ts corresponding to each curve. Curves for short stress times <=1ms were continued until full recovery (≈10s) and done repetitively, see 4 consecutive traces for *ts*=1ms in Fig.a.

(1) Degradation and recovery are explainable by capture and emission of charges in/out of defects
(2) Capture and emission needs thermal activation across an energy barrier (i.e. requiring phonons). E_A is around 1eV. Vgate at capture is different from Vgate at emission, thus E_A s for capture and emission are different. These E_A s determine the capture/emission probability.
(3) Capture and emission depend on electric field, which actually means that these transitions are equivalent to a forward- and backward-electrochemical reaction.
(4) The equilibrium occupancy level of a charged defect is given by its energy and the Fermi-level. Note that τ_c and τ_e are determined at different fields; so the ratio $\tau_e/(\tau_c+\tau_e)$ does **not** determine the defect energy E_T, like it does in RTS experiments.
(5) In addition to the transmission probabilities the capture and emission rates are also determined by the density of states in the

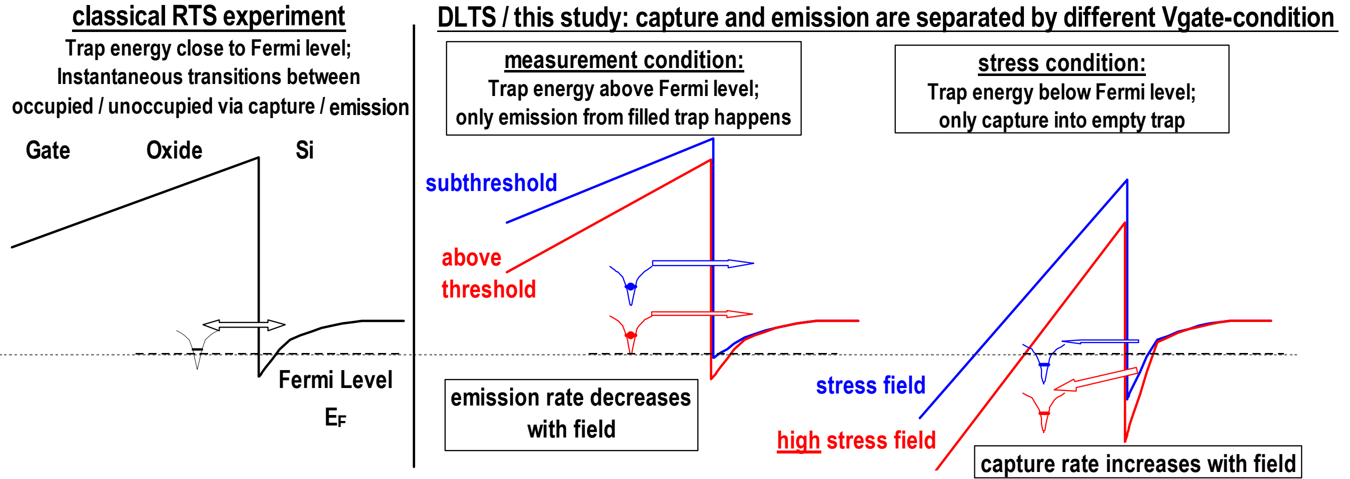

Fig. 3 – Schematic band-diagrams, energy vs. z-direction. Each diagram contains a single trap which is shown for two different fields for the NBTI-side. We do not make any assumption about the nature of the transition processes. They are not due to elastic tunneling but thermally activated and have to be treated as an electrochemical reaction. Arrows denote the direction of charge transfer. Diagrams show the conduction band and an electron-trap (nFET) for the RTS experiment and the valence band (upside down, increasing energy always in up-direction) and a hole-trap (pFET) for the NBTI experiment.

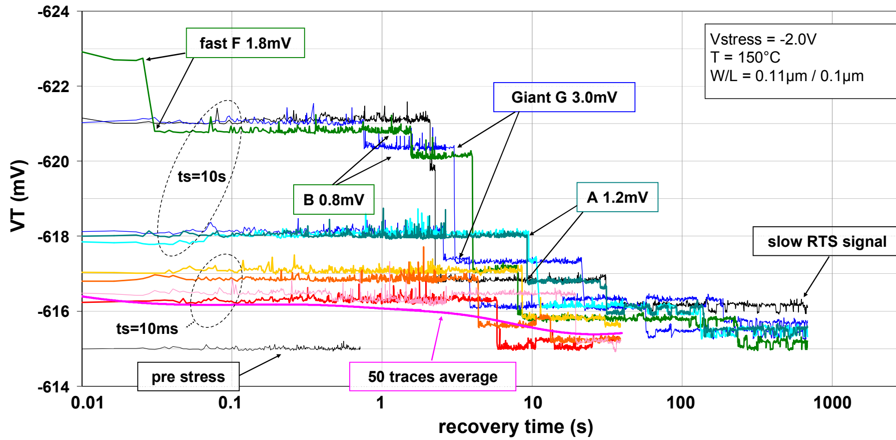

Fig. 4 – An example of the discrete steps from emission of positive charge from defects during recovery. Stress pulses are applied repetitively (typically 100 times). Some traces, 4 after 10ms stress-pulses (encircled) and 6 after 10s stress (bluish) were chosen as examples. Individual defects can be distinguished by their step heights. Trap names (A, F, G) etc are used in some of the following Figs.

inversion layer.

(6) Electric field dependence of capture and emission have the direction as expected from the model in Fig. 3.

V. STATISTICAL ANALYSIS OF A SINGLE DEFECT

An example for the defects' response to stress pulses is shown in Fig. 4. After very short stress pulses and/or very low stress-fields no defect becomes charged, i.e. we do not observe any capture/emission event. For the *ts*=10ms traces in Fig. 4 we have the case that only one defect causing a "giant" step, with an emission time constant of 5s, is active (cmp. the encircled "10ms" curves in Fig. 4). For this stress condition (Vgate=-2V/T=150°C/ts=10ms) in the *tr*=1s...10s regime the emission from this defect can be studied undisturbed by the other defects. The stress field has led to capture of a hole by pulling its energy level E_T (see Fig. 3) below E_F. In case the stress pulse (10ms) is longer than the capture time constant of this defect it is most likely that the capture-attempt is successful, i.e. it gets occupied during the stress pulse. After the stress is released E_T is above E_F again, and it's equilibrium occupancy level is close to zero. Then the defect will be discharged by **emission** (like radioactive decay) at a random time around the characteristic emission time τ_e.

Averaged over many emissions of a given individual defect it's occupancy will decrease with recovery time *tr* like

$$Occ = \exp(-tr/\tau_e) \quad (1)$$

A statistical analysis of τ_e is shown in Fig. 5. The same kind of statistical behavior describes the **capture** of a hole during stress. Occupancy during stress time *ts* increases like

$$Occ = 1 - \exp(-ts/\tau_c) \quad (2)$$

For the *ts*=10ms traces in Fig. 4 – where a single defect is unperturbed from other ones – the exponential decay behavior (eq.1) can be directly seen simply by averaging over a number of traces. For more complicated traces individual defects can be distinguished from each other by their different step heights. The switching times of steps of equal heights have to be analyzed like done in Fig. 5. A successful straight-line fit to an exp(-tr/τ_e) distribution is a proof that only one defect is contained in the set of data of the same step-height and thus for a correct analysis. Fig. 5 shows that the emission time constants can be determined with high precision (cmp. error bars in Fig. 9). This kind of precision is needed when temperature and gate voltage dependencies are to be extracted. The capture times can be determined from the number of "successful" capture events

after repetitive trials, i.e. stress pulses. A precondition for an accurate determination of τ_C from- eq.2 is that the stress pulse length *ts* is comparable (at least within a factor of 10) to τ_C as it is shown in the example of Fig. 6. For $ts >> \tau_C$ or $ts << \tau_C$ *Occ* approaches 100% or 0%, respectively, without providing much information.

VI. FIELD AND TEMPERATURE EFFECTS

The time constants τ_C and τ_e are individual for each defect. τ_C and τ_e are determined by the energy-level of the defect and the height of thermal barrier which are in turn a function of the electric field. Fig. 7 shows the dependence of τ_e on the gate voltage *VG* during recovery at VG around threshold. When *VG* is moved from inversion towards depletion τ_e slightly decreases due to the increasing density of empty final states (cmp. Fig. 3), as expected.
The capture time as a function of the stress-field is plotted in Fig. 8 as a power law.
An analysis of the thermal activation energies E_A of capture and emission is shown in Fig. 9. The measured E_As are much higher than values in literature. Actually these high E_As are the ones corresponding to the physical processes in NBTI, while the ones measured conventionally in a wide FET are just artifacts from intermixing E_As of capture and emission. An explanation will be given in Fig. 10:
The NBTI capture/emission is modeled by a simplified equivalent circuit which does not correctly describe all phenomena of degradation and recovery but is close enough to the experimental phenomena for serving to explain the temperature effects: The model is just a simple circuit consisting of an RC-ladder containing a few (four) distinct time constants τ distributed evenly on a log-time-scale. Applying voltage to the RC-ladder or grounding it charges respectively discharges the capacitors, meaning degradation or recovery. The total charge on all C's corresponds to *ΔVT*. An increase of temperature means a decrease of the time constants τ of the RC-ladder. We assume that the change in the τs is the same for all RCs, for example we let the τ's decrease by a factor of 10 when increasing the temperature by ≈30°C as it is the case for $E_A \approx$1eV. If we have distinct τ's of the RCs separated by 3 decades in time, then these RCs cause distinct bumps in the recovery-traces. A 30°C temperature increase shifts these bumps to shorter times as shown in the upper part of Fig. 10. However, in reality we do not have distinct time constants but in fact we have time constants smeared out almost evenly on a log time scale over a range far below to far above

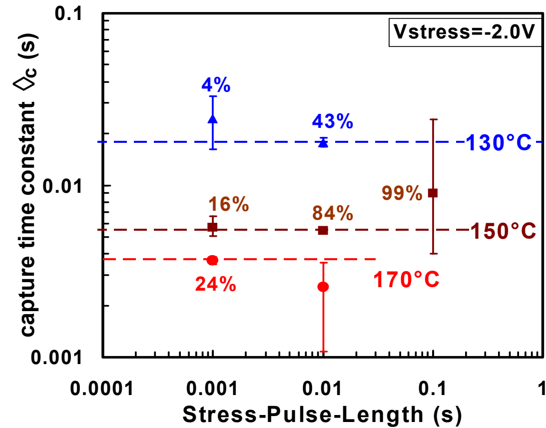

Fig. 6 – Example for the statistical analysis of emission times from repetitive trapping attempts (by stress pulses) for a single defect at 3 temperatures. Labels denote the ratio number of trapping event to number of attempts. Error bars correspond to the statistical error which is reasonable small between 20% and 80%.

the experimentally accessible range which is roughly 1µs to 10^6s. It is clear that shifting such a quasi-continuous RC-ladder up or down, which means multiplying all time constants with the same factor, has no effect as long as the RC-ladder is extended to very short and very long time constants. Thus, while the individual defects, corresponding to individual RC's, are strongly thermally activated no thermal activation is seen in the short-term "macroscopic" data. This is demonstrated once more in Fig. 10, lower part. Compared to the model used in [6] our RC-ladder model is over-simplified. It does not consider that time constants τ_Cs for charging (degradation) are different from the τ_es for discharging (recovery). But as long as both the τ_Cs and the τ_es have the **same** thermal activation this oversimplification does not change the results of our considerations. In reality and in a practical NBTI degradation experiment the τ_Cs and the τ_es do not have the same but a similar E_A. So a wide-FET-experiment "sees" a finite non-zero thermal activation which, however, does not reflect the thermal activation of the real physical processes. Our value for the recovery-E_A (see Fig. 9) of 1eV is in full agreement with the results in [11] where very fast temperature

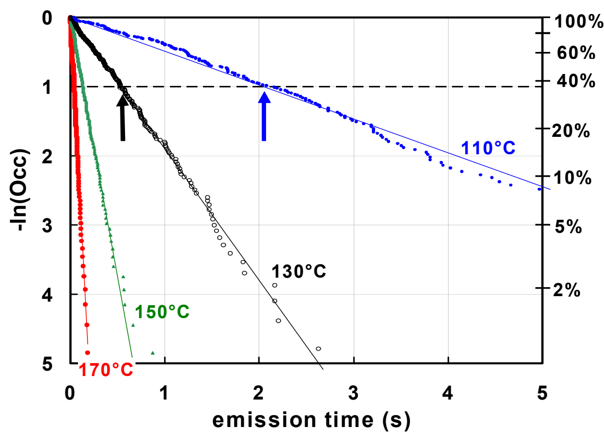

Fig. 5 – Example for the statistical analysis of emission times of a single defect at 4 temperatures. Each dot corresponds to one relaxation trace. Arrows mark the time constants τ_e. The straight line behavior ensures that only single defect (single time constant) is contained in each curve.

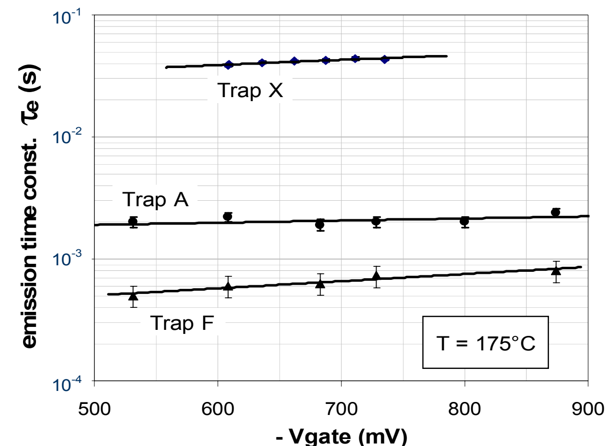

Fig. 7 – Emission time constants as a function of VG applied during recovery, for 3 defects with τ_e from sub-ms to 50ms emission rate is only slightly dependent on VG; the direction of the slope is explained by Fig. 3.

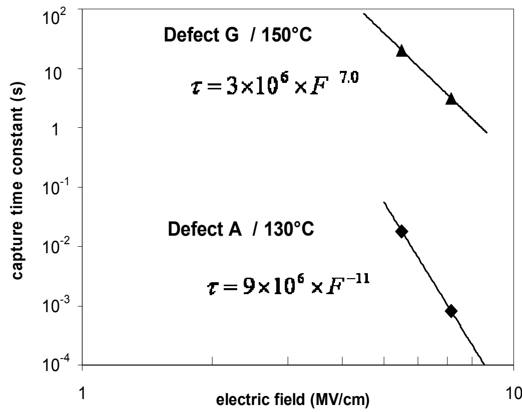

Fig. 8 – Capture times as a function of the stress field for two defects. The equations give the exponents of a fit to a power law in electric field.

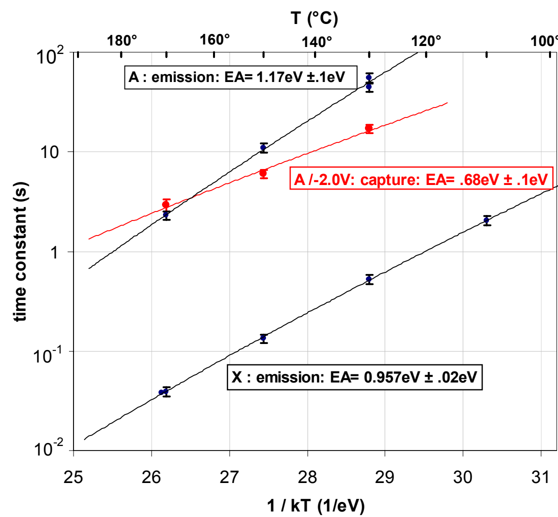

Fig. 9 – Arrhenius plot for defect A (capture and emission) and trap X (emission only). EAs can be extracted with considerable precision, and are higher than the ones observed in conventional experiments (see Fig. 10 for explanation).

changes have been done during recovery. A rough estimation of the distribution of defect properties is shown in Fig. 11. As expected from previous recovery measurement in wide FETs the time constants are varying over many orders of magnitude. The problem of wide-FET-experiments is that they do not allow an assignment of individual or even a class of τ_cs to a class of τ_es. So the new kind of information provided in Fig. 11 is that an individual defect with a capture time constant (at a given T and field) for example of 10ms might have a similar emission time, a much lower emission time of 10µs or also a much longer emission time of 10ks. For practical simulation purposes a correlation plot like Fig. 11 must be replaced by a plot having the single defects replaced by a probability distribution function. It would be sufficient to divide Fig. 11 into one-decade-wide squares representing classes of defects and containing the density of defects of each class. To create such a plot, which would be specific for a given technology, actually means time consuming experiments which were beyond the scope of this paper.

VII. Conclusion

We have employed new single-defect-DLTS experiments similar to those used to study classical RTS in nMOSFETs. Field- and

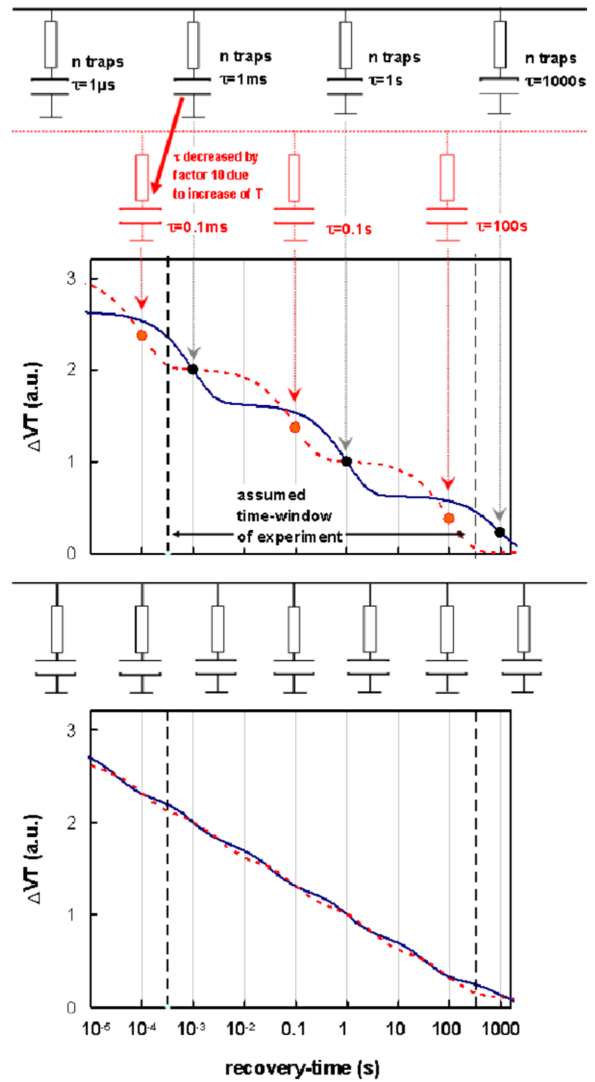

Fig. 10 – Simplified simulation of NBTI using the equivalent circuit (EC) from [6]. EC is plotted on top of each simulation graph. The τ's are aligned with the X-axis **Upper part:**. τ's are widely separated (one in 3 decades). Black solid EC and black solid recovery curve correspond to low T. A factor 10 decrease of all τ's (by increased temperature, corresponds to red dashed EC and red dashed recovery curve) shifts the distinct bumps in the recovery curves one decade left. **Lower part**: Same procedure done as in upper part, only one EC shown. The density of time constants is increased by a factor of 2 only (2 per 3 decades). The curves becomes completely featureless, thus any thermal activation has **no** effect at all.

temperature-dependence of the properties of individual defects could be characterized. We could show that NBTI in a regime up to typical VT-drift criterions is fully explainable as trapping/detrapping in defects with a wide with a wide distribution of time constants, as proposed in [6]. The experimental facts - without any further assumptions - allow the following conclusions:

(1) The recoverable part of NBTI-degradation and recovery can be fully understood as an electrochemical forward/backward reaction.

(2) Both forward- and backward-reaction are thermally activated; forward and backward activation energies depend on gate bias and are not equal in general.

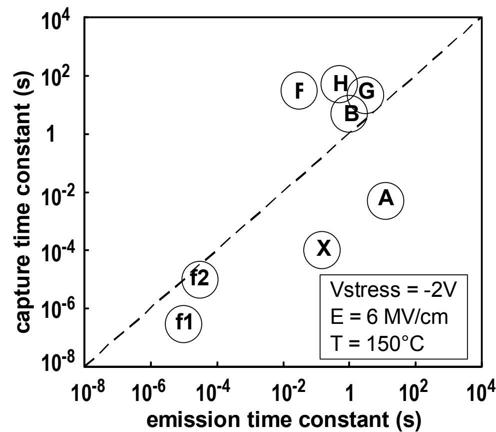

Fig. 11 – Plot showing the correlation τ_c / τ_e for the characterized defects. For a reliability model a more complete characterization would be desirable.

(3) In contrast to the **seemingly** temperature independent fast NBTI precursor (see Fig. 1) in conventional experiments (thought to be by elastic tunneling in the past) also the very fast part (200ns to 1ms) of degradation is thermally activated.

(4) Parameters relevant for the NBTI effect are capture- and emission-time constants τ_c and τ_e of the defects.

(5) defects were found to be pre-existing, there was no evidence for generation of defects due to stress up to moderate (<50mV) ΔVTs

(6) Defects investigated so far have been found to act (nearly) independently of each other.

(7) Defects have a wide distribution of capture times, <µs to >10^6s.

(8) The τ_c and τ_e distributions are continuous, for the defects τ_e may range from much larger to much smaller than τ_c .

(9) For any given defect the emission time constant τ_e (recovery) is independent of stress time in contrast to the prediction of the reaction/diffusion model [12] still favored by some research groups According to the reaction/diffusion theory diffusion would make recovery a function of stress time.

Thus some of the NBTI models/concepts discussed so far are to be discarded or questionable:

(1) Diffusion of Hydrogen is not involved in recovery; though it is most plausible that H may be a part of the defect and play a role in the electrochemical reaction.

(2) No contribution from T-independent QM-tunneling of light particles has been found.

(3) Because τ_e is independent of τ_c there is no microscopic version of the empirically observed concept of "universal recovery".

Even though the microscopic origin of the defects is not known and not taken into account, the knowledge of the defect-parameters will allow a complete and simple mathematical modeling of the degradation due to any stress / recovery sequence (e.g. DC+AC etc.).

REFERENCES

[1] J.H. Stathis a. S. Zafar, "The Negative Bias Temperature Instability in MOS Devices: A Review", MR Vol.46, no.2-4, 2006, p. 270

[2] J.P. Campbell and P.M. Lenahan, " NBTI: an atomic-scale defect perspective", Proc. IRPS 2006, pp. 442-447

[3] K. S. Ralls , W.J. Skocpol, L.D. Jackel, R.E. Howard, L.A. Fetter, R.W. Epworth, and D.M. Tennant, "Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1/*f*?) Noise", Phys. Rev. Lett. 52, pp.228-231, 1984

[4] M.J. Kirton and M.J. Uren, "Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/f) noise", Adv. Phys., vol. 38, pp. 367-468, 1989

[5] A. Karwath and M. Schulz, "Deep level transient spectroscopy on single, isolated interface traps in feiled-effect transisors", Appl. Phys. Lett. 52, p.634,1988

[6] B. Kaczer, T. Grasser, J. Martin-Martinez, E. Simoen, M. Aoulaiche, Ph. J. Roussel, G. Groeseneken, "NBTI from the perspective of defect states with widely distributed time scales", Proc. IRPS 2009, pp. 55-60.

[7] T. Grasser, H. Reisinger, W. Goes, Th. Aichinger, Ph. Hehenberger, P.-J. Wagner, M. Nelhiebel, J. Franco, and B. Kaczer, "Switching Oxide Traps as the Missing Link Between Negative Bias Temperature Instability and Random Telegraph Noise", IEDM 2009 (accepted for publication)

[8] H. Reisinger, O. Blank, W. Heinrigs, W. Gustin, and C. Schlünder, "A comparison of very fast to very slow components in degradation and recovery due to NBTI and bulk hole trapping to existing physical models", IEEE TDMR, Vol. 7, No. 1, p. 119 (2007)

[9] T. Ando, A.B. Fowler a. F. Stern, "Electronic properties of two-dimensional systems", Rev. Mod. Phys. 54, pp.437-672. 1982

[10] L. Lundkvist, I. Lundström and C. Svensson, "Discharge of MNOS Structures", Solid State Electronics 16, pp. 811-823, (1973)

[11] T. Aichinger, M. Nelhiebel, T. Grasser, "Unambiguous Identification of the NBTI Recovery Mechanism using Ultra-Fast Temperature Changes", Proc. IRPS 2009, pp. 2-7

[12] K.O. Jeppson and C.M. Svensson, "Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices", J. App. Phys. Vol. 48, No.5, 1977, pp. 2004-14

QUESTIONS AND ANSWERS

Q1: On the recovery curve, what is the measurement instrument used to measure the ΔVT versus recovery time?

A1: The measurements were done using a home-made instrument. It uses a fast feedback-loop as described in Reisinger et al, TDMR Vol.7, pp.119-129, March 2007, Vol.7. VT corresponds to a constant ID of typ. 1µA*W/L.

Q2: How do you derive the capture and emission time of the traces?

A2: Each emission time constant is determined from typically 100 recovery traces as shown in Fig.5. Each dot is from one recovery trace and is from a VT-step down, of a specific step-size, corresponding to a specific defect. Capture events happen during stress in strong inversion. In strong inversion the contribution of a single carrier to the drain current is small, thus the signal due to a single capture is below noise level. Therefore the capture time constants are derived from the statistical analysis of the capture-success-rate, corresponding to a given stress-pulse-width. Typ. 100 capture attempts (i.e. pulses) are analysed, as explained in Fig.6.