The statistical analysis of individual defects constituting NBTI and its implications for modeling DC- and AC-stress

Hans Reisinger, Tibor Grasser * , Wolfgang Gustin and Christian Schlünder
Infineon Technologies AG
München, Germany
Email: Hans.Reisinger@infineon.com

* Institute for Microelectronics, TU Wien
Wien, Austria

***Abstract* — The physical origin of the Negative Bias Temperature Instability (NBTI) is still under debate. In this work we analyze the single defects constituting NBTI. We introduce a new measurement technique stimulating a charging of these defects. By employing a statistical analysis of many stochastic stimulation processes of the same defect we are able to determine the electric field and the temperature dependence of these defects with great precision. Based on our experiments we present and verify a new, physics-based, quantitative model allowing a precise prediction of NBTI degradation and recovery. This model takes the stress history into account and also provides a prediction for degradation due to AC-NBTI and an understanding of the special features seen in conjunction with AC-NBTI.**

Keywords-NBTI; recovery; random telegraph noise; AC-stress;

INTRODUCTION

30 years of research [1] still did not bring about any consensus regarding the physical origin of NBTI. So far only ESR experiments [2] have provided some insight into the microscopic processes. The majority of electrical experiments has been done on large pFETs and thus averages over many defects (see table 1). We will show that this averaging obscures the properties of the individual defects and leads to misinterpretations, for example of the thermal activation. In this work we analyze the properties of the single, discrete events constituting NBTI degradation and recovery in small FETs. We have developed analysis techniques based on those known from the Random Telegraph Signal (RTS) studies of the 1980s [3, 4]. RTS is caused by capture and emission of charge in oxide traps. To our knowledge RTS studies have been done for nMOSFETs only [4], the FETs kept in a quasi steady state. In contrast, our method uses multiple high-gate-field charging pulses to excite the defect state to a charged state. This technique is similar to the deep level transient spectroscopy (DLTS) [6], but it is essential to vary the length of the charging pulses, because of the wide distribution of capture time constants. This is why we term the technique time dependent defect spectroscopy (TDDS). Stimulated by recent work on small pFETs [5, 7] we employ for the first time a highly accurate extraction of NBTI-relevant defect parameters. Capture and emission time constants τ_c and τ_e, (corresponding to stress and recovery) have been seen before [5, 8]. However, to extract temperature- and field-dependencies of those τ_cs and τ_es, and thus to gain insight in physical processes, a highly accurate determination of averaged time constants is required. Thus we extract averaged values for τ_c and τ_e from repetitive measurements. Typically 100 or more excitations and measurements are done for each τ_c and τ_e determination for a given defect and condition. Main parameters are the gate voltage, the length of the excitation pulse and the temperature. The purpose of the present work is to give an overview over the implications regarding our new understanding of the NBTI phenomena and a modeling of NBTI which includes recovery and AC-stress. A detailed description of a mathematical model which can be used in a degradation simulator and the corresponding parameter extraction as well as a description of our AC-stress experiments (effect of frequency, measuring delay, etc.) both are beyond the scope of this paper and will be presented in subsequent publications.

We will begin with a detailed description of our new measurement technique in section I followed by a short discussion of our sample properties and geometry dependent parameters in section II. Section III will discuss the basic properties of our physical model. Since the exact nature of the involved defects cannot be determined from an electrical characterization the model will remain phenomenological. A real microscopic, physical defect model as well as more and more detailed extractions of the defect properties are found in [9, 10]. The statistical analysis and defect parameter extraction will be demonstrated in section IV. Our mathematical model in section V, based on the defect analysis, will be used to discuss the missing thermal activation for NBTI short-term stress in VI. From section VII on we switch back to data from large FETs, treating practical modeling and model verification. AC-NBTI features will be shown and explained in section VIII. We conclude with a short discussion of permanent degradation in IX followed by the conclusions section X.

I. EXPERIMENTS

Our defect spectroscopy experiments were done in the following way: A stress pulse was applied to the gate in order to occupy (i.e. to

FET-name and dimension (in µm)	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	
number #Nit at a density DNit=1E11/cm^2	1000		24		11	
ΔDNit causing a ΔVT=50mV (in cm^{-2})	4.9 x 10^{11}		4.9 x 10^{11}		4.9 x 10^{11}	
makes a number Δ#NIt	4900		120		50	
ΔVT caused by a single trapped carrier (at interface)	0.01mV		0.43mV		1.0 mV	

Table 1 Geometries of the used pFETs and some 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 0.2mV to 5mV. DNit denotes the charged interface-states density.

charge positively) individual defects. After the end of the charging pulse the threshold voltage *VT* was continuously recorded for a time of typically 1000s. In order to determine different capture time constants the stress pulses ranged from 200ns to 100s. *VT* was directly recorded using our fast feedback loop [11]. In contrast to a direct measurement of the steps in the drain current *ID* with a fixed gate voltage *Vgate* our method is somewhat simpler because it saves the effort of recalculating the *VT*-step from the *ID*-step (which varies with *ID*). Capture and emission are stochastic processes. Thus each pulse/readout was done repetitively (32 to 256 times) to allow a statistical analysis. A rough determination of individual capture and emission times (within a factor 2, like in Fig. 3, for example) could be done by recording just a few traces. However, if precise parameter dependencies (on *Vgate*, or T, cmp. Fig. 8) are required a statistical analysis of **many** traces has to be done. To ensure the same pre-pulse condition for each pulse/readout sequence full recovery had to be awaited after each stress pulse (cmp. Fig. 5). A measuring sequence for a given gate voltage/temperature, just varying the stress pulse width from µs to 10s, takes a time of several days. It is worth noting that all information, also about capture, is gathered during recovery. Capture events occur in the stress period and in principle could be sensed from *ID*-steps occurring during stress. However, the total number in the channel of a minimal FET during recovery is 200 (cmp. table 1). At stress, in strong inversion this number is higher by a factor of >10. Thus a single carrier has an effect in *ID* smaller by more than a factor 10, and is below the noise level in general.

II. Samples

We use production-quality pFETs with a 2.2nm nitrided oxide. Such oxides around 2nm have been the subject of many NBTI-investigations of different research groups (see references in [11]). To anticipate criticism claiming our samples would exhibit an extraordinarily high defect level or defects of a special kind Fig.1 shows an inter-technology comparison of a couple of oxides. This comparison is done by a normalization of the degradation *ΔVT* resulting in the surface density of charged defects (see right hand scale of Fig. 1). It includes a series of "clean", non-nitrided oxide with different thicknesses as a reference and different samples published in the literature with a moderate nitridation level of about 6%. Fig. 1 shows that our samples' defect density is at the same level or better than samples studied in other NBTI work.

We have chosen the geometries to be small enough to conveniently resolve the effect from a single carrier in the channel

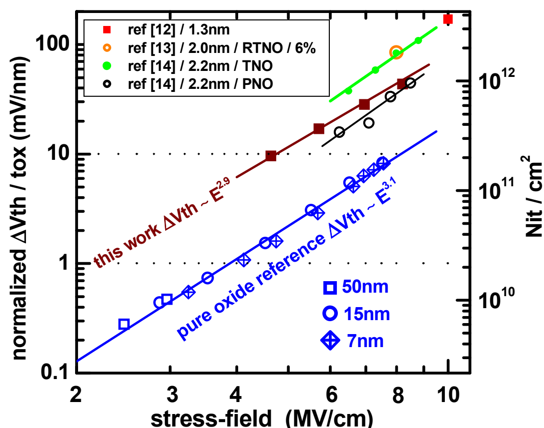

Figure.1. Comparison of different technologies from different sources. Plotted is the normalized *ΔVT* after a stress time of 10ks at 175°C and the corresponding surface charge density (right hand scale). Data from [12, 13, 14] have been extrapolated to 10ks. Oxide parameters are given in legend, as far as available. The pure oxide reference contains thicknesses from 7nm to 50nm.

(typical *ΔVT*=1mV, see table 1) and large enough to have at least a handful of active defects (≈20 average step-height defects for *ΔVT*=20mV). An overview of geometry-dependent sample properties is given in table 1. The data in table 1 were calculated using the simple charge-sheet model and thus give only average step heights in *VT*. Real individual step heights are long known to be larger or smaller than this average by factors of 10 [4]. The origin and size of step height variations in FETs of a similar geometry as the ones used here is explained in [15]. The step height of each individual defect is its "fingerprint", allowing to distinguish between the defects. In order to increase this difference in the individual step heights all our measurements were done in saturation. During the measurements, i.e. the recovery traces (cmp. Fig. 3) the drain current *ID* was kept constant while recording the gate voltage *VT*. A typical level for *ID* was 1µA**W/L*. For our "minimal FETs" the precision we achieved is sufficient to determine minimum *ΔVT* step heights of 0.2mV. This value is well below the average expected step height of 1mV (see table 1). Based on the analysis of our single defects we will draw conclusions about the NBTI phenomena observed in large FETs. So it has to be made sure that in fact these single defects are responsible for NBTI and **do not** just represent a special kind of defects playing a minor role for "real" NBTI. For this purpose Fig. 2 shows the comparison of a wide FET with a narrow FET. In order to average over the statistical variation of degradation in the small FETs 25 of these small FETs have been measured sequentially and summed up. An example of recovery traces, for a single small FET and the sum over 25 small FETs is shown in Fig. 3. All these traces are showing the clear, distinct steps in *VT* from single defects. The

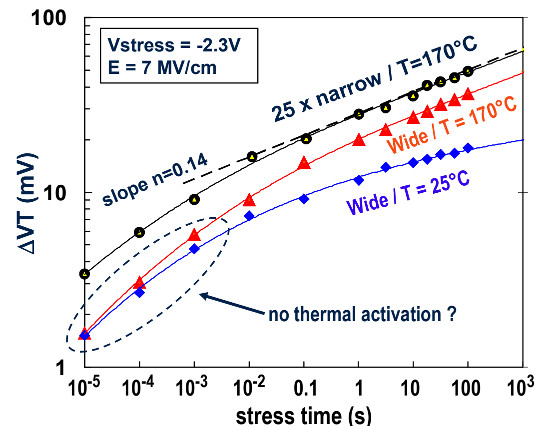

Figure 2. Comparison of degradation for a narrow (0.2µm) and a 10µm wide FET at 170°C. The narrow FET data is from summing up 25 single 0.2µm wide FET data. For the wide FET the thermal activation, and the "missing" thermal activation for µs to ms times are shown.

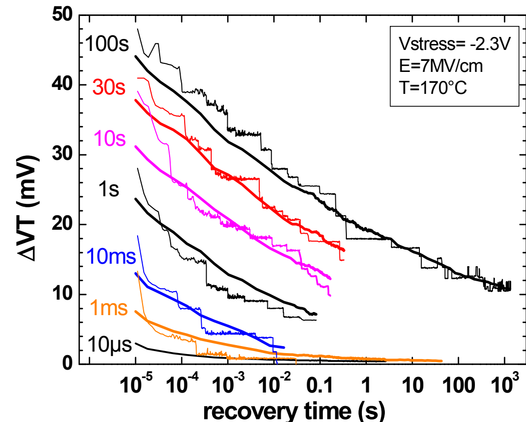

Figure 3. Comparison of the recovery traces for a single narrow (traces with steps) and the average over 25 different 0.2µm wide FETs (smooth traces). Labels denote the stress time preceding the recovery trace. The recovery behavior of the wide devices (see Fig. 2 in ref.[17]) is identical to the narrow FETs.

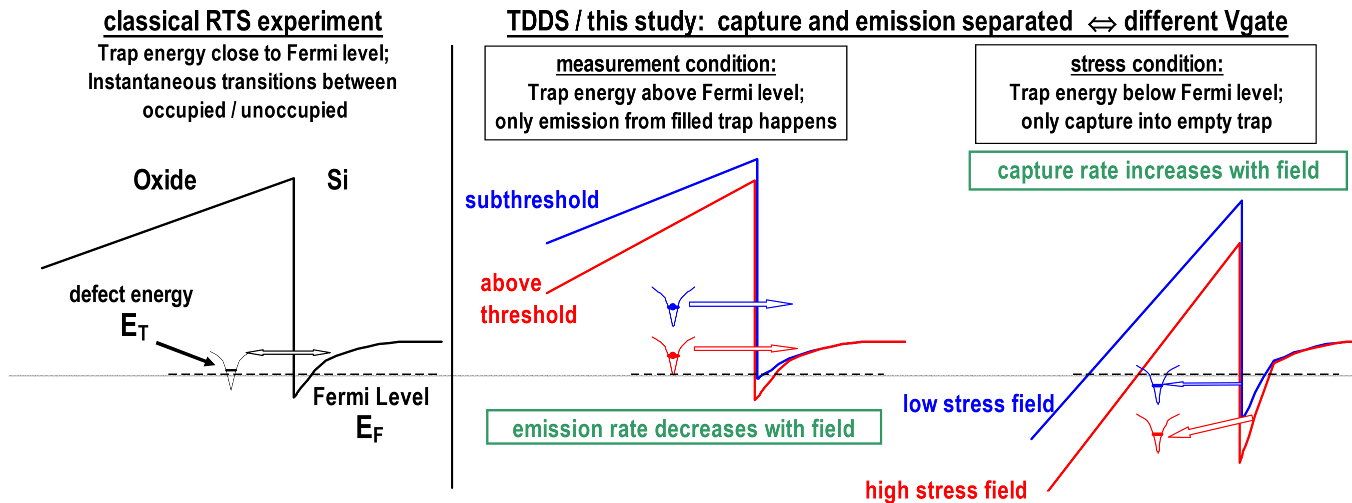

Figure 4. Schematic band-diagrams, showing the inversion layer and defect levels in the oxide in the vicinity of the substrate. The diagram is meant to be polarity independent, showing the conduction band in an nMOSFET **or** the valence band of a pMOSFET when flipped upside down. Increasing potential energy corresponds to the upward direction. States above / below E_F are empty / occupied, respectively. Each of the 3 diagrams contains the same **single trap** shown for **two different fields** (red/blue) for the TDDS-side. Arrows just indicate the charge carrier transitions between an initial and a final state. These arrows are **not supposed** to indicate that the charge transfer process is due to elastic or inelastic tunneling. In contrast the charge transfer is thermally activated for all cases, including RTS. RTS and TDDS do not differ with respect to sample type, nature of traps or physical processes involved. The main difference for TDDS is the stimulation of capture at a field higher than the field for emission, and the way of data analysis. **Capture rate** increases with increasing field, due to an increasing transition probability and an increasing density of initial states available in the inversion layer. **Emission rate** increases with decreasing field, due to an increasing density of final, empty states. Note that emission to the gate instead to the inversion layer is possible but neglected for simplicity.

agreement of the NBTI degradation and recovery in wide and narrow FETs is quite satisfactory. Degradation in the narrow FETs is 30% higher than in the wide FET, presumably due to edge/stress effects (STI). Such a geometry dependence of NBTI has been reported previously [16].

III. Phenomenological Model

For a basic understanding of our results it is useful and necessary to reconsider the experiments studying random telegraph signals or noise (RTS or RTN), especially the model by Kirton/Uren [4]. This model has been successful to explain the basic phenomena seen in the RTS experiments of nFETS, done mainly in the 1980ies. This model also explains the main phenomena seen in our experiments, and thus of NBTI in general. A major issue of the model is its failure to properly predict the bias dependence of capture and emission, see [10] for a discussion. The essential assumptions of the model are that RTN signals are generated by capture and emission of charges in oxide defects. Capture and emission involve a structural relaxation of the oxide matrix surrounding the defect. It is this structural relaxation which comprises the potential barrier ruling the transition of charge. Thus a transition requires, like in a chemical reaction, an overcoming of a barrier by **heavy** particles. As a consequence the transitions are thermally activated. Since the energy level of the charged defects is field dependent the reaction rate for both directions, capture and emission of charge, is gate voltage dependent. The experiments in this work and also low temperature experiments [20] have clearly shown that tunneling of light particles (electrons or holes), which is nearly temperature independent, does not contribute to NBTI. At this point we want to mention that Shockley-Read-Hall (SRH) recombination also has been tried (unsuccessfully) to explain the NBTI temperature dependence. However, the SRH model has been developed for interaction of semiconductor bulk traps with carriers which are not separated spatially from these traps. It never has been intended to explain the interaction of charge carriers in the silicon space charge layer with oxide defects and thus totally fails to explain even qualitatively any of the phenomena corresponding to oxide traps. A model sketching the band diagrams corresponding to capture and emission is given in Fig. 4. We want to stress that Fig. 4 is simplified and meant only to illustrate the basic properties of capture and emission of charge. It does not show the energy barriers corresponding to structural relaxation in any way. A model which treats the microscopic physical processes is presented in [10].

IV. Statistical Analysis of an Individual Defect

Before we come to the statistical analysis of single defects some introducing and clarifying remarks are in order:

(1) As indicated in Fig. 4 all defects have an equilibrium occupation probability of 100% (E_T below E_F) during stress and 0% (E_T above E_F) during recovery. This is a good approximation as long as E_T is not within an energetic distance of <50mV to the Fermi level E_F. This is actually happening for some defects (see RTN signal in Fig. 5), but is unlikely, and will be neglected for the sake of simplicity.

(2) When switching from stress (defect 100% occupied) to recovery it takes a while (i.e. the individual, *Vgate* and T-dependent emission time constant τ_e) to reach the new 0% equilibrium occupation level.

(3) The same is true vice versa when switching from recovery to stress.

(4) Capture and emission are stochastic processes. Like for radioactive decay only probabilities for capture and emission during a certain time can be given.

(5) A given defect always is either occupied or un-occupied. Only its average occupancy (averaged over many capture or emission processes) can be given by eq. 1 and 2.

(6) RTS is a static experiment (see Fig. 4, left side) and analyses the mark-space ratio for a defect having - on the average - its equilibrium occupancy, which is determined by the Fermi-Dirac distribution function. Thus RTS is able to determine (without assumptions or model) a defect energy for a given *Vgate*.

(7) For our TDDS experiment capture and emission are happening at different *Vgate* values. So - in contrast to RTS - the defect energy **cannot** be determined directly from an experiment at a single field. A determination of the defect energy requires the measurement of the field dependence of capture and emission.

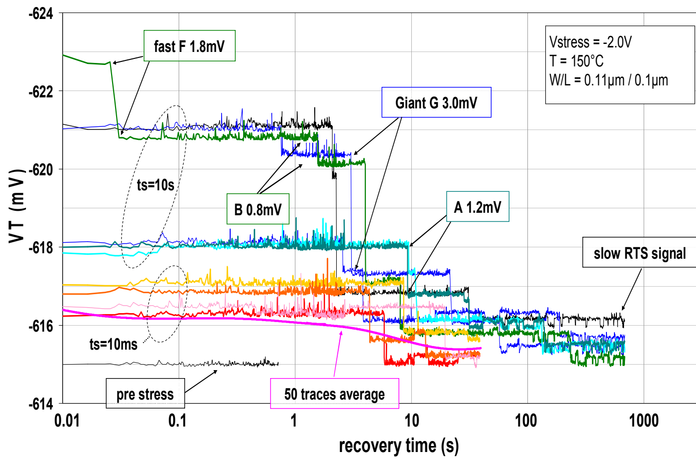

Figure 5. An example of steps from emission of positive charge from defects during recovery, all from the same device. Stress pulses are applied repetitively (typ. 100 times). Some traces, 4 after 10ms stress-pulses (reddish color) and 6 after 10s stress (bluish) were chosen as examples. Individual defects can be distinguished by their step heights, allowing to name the traps, A, F, G, etc. The smooth trace is an average over 50 traces showing the exp(-t/τ_e) (eq. 2) behavior.

(8) Major drawback of the RTS method is that the energy of a defect causing an RTS signal has to be in the vicinity of the Fermi-level. This only happens per chance, and thus allows an analysis of only a fraction of the defects. In contrast our TDDS allows the analysis of **all** the defects contributing to NBTI, provided they are within our experimental window for τ_c and τ_e. This window is from µs to >1000s, which is much wider than for RTN as shown in Fig. 9 of ref. [10].

An example for the responses of roughly 10 defects to stress pulses is given in Fig. 5. Let us first examine the recovery traces with stress time *ts*=10ms. As seen in Fig. 5 for *ts*=10ms only a single defect with a emission time around 10s happens to be active. During the stress the oxide field pulls its energy level E_T far below E_F (see Fig. 4). Please note that the valence band in Fig. 4 is flipped upside down in order to have rising energy up and occupied levels below E_F. For the given stress parameters the defect's capture time constant is <10ms, thus its occupation probability is nearly unity after stress. After the stress has been released (corresponding to recovery time=0 in Fig.5) E_T is above E_F again, the defect's equilibrium occupancy

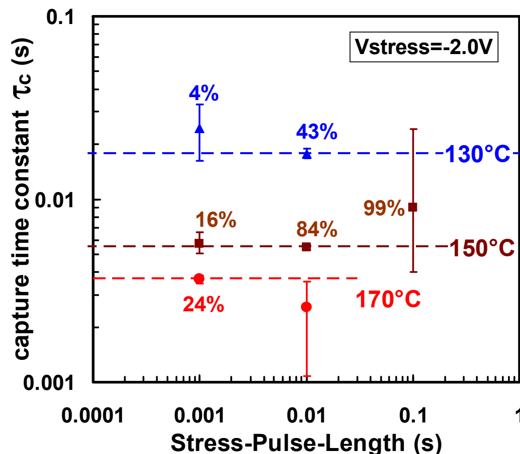

Figure 6. Example for the statistical analysis of capture time constants from repetitive capture attempts (by doing stress pulses of 1ms, 10ms and 100ms) for a single defect at 3 temperatures. Labels denote the capture probability, equal to the ratio r=number capture events / number of attempts. note that r is the quantity which is directly measured. As seen the statistical errors (see error bars) in the determined τ_c can be minimized, or at least kept small by choosing the stress pulse length suitable to have the capture probability between 15% and 90%.

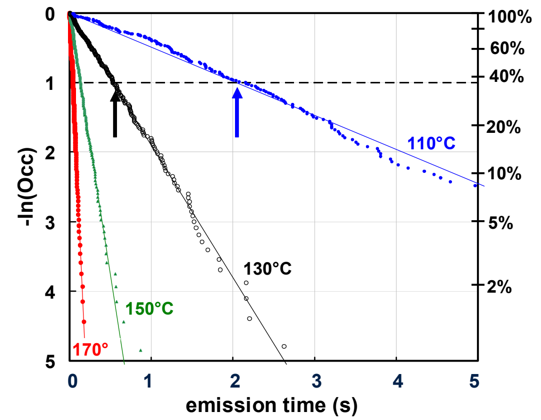

Figure 7. Example for the statistical analysis of emission times of a single defect at 4 temperatures. Plotted is the decreasing occupancy of a defect, starting with 100% just after capture, i.e. at zero emission time. Arrows mark the time constants τ_e. The straight line behavior ensures that only a single defect is contained in each curve. Each dot corresponds to an emission event in a recovery trace like in Fig. 5.

level ≈0). It will be discharged by emission at a random time around a characteristic emission time τ_e. After applying stress, capture of (positive) charge will happen in the same stochastic manner around a characteristic capture time τ_c. Averaged either over **many** capture or emission events of the same defect under the same condition (as the 50-traces average in Fig. 5 shows, or all our single-defect data) **or** a single stress pulse or recovery trace averaged over **many** equivalent defects (in a large FET) the average occupancy or occupation probability *P* during stress time *ts* and recovery time *tr* will be described by the following equations:

$$(1) \qquad P(ts) = P(ts=0)\times\left[1-\exp(ts/\tau_c)\right]$$

describes the increase in occupancy due to capture during stress, and

$$(2) \qquad P(tr) = P(tr=0)\times\exp(tr/\tau_e)$$

describes the decay of occupancy due to emission during recovery. For the sake of simplicity the Fermi distribution function is neglected and assumed to be 1 during stress and 0 during recovery. Eq. (1) and (2) gives average occupancies, any given defect always is either occupied or empty.

An example for the statistical determination of **capture** time constants according to eq. (1) is shown in Fig. 6 and an example for the statistical determination of **emission** time constants according to eq. (2) is shown in Fig. 7. In Fig. 7 the temperature dependence of τ_e of a given defect is investigated. Each emission time of this defect, extracted from a recovery trace, is plotted as a dot, in the standard way, allowing to extract τ_e corresponding to the slope of a straight line. The fit to the straight line also is a proof that the defects behave in the expected statistical way (eq. 2) and that **only one** defect is contained in the distribution. Mixing another defect into the distribution, by chance with a similar step-height, hardly to distinguish, would lead to a deviation from the straight line. In order to minimize the statistical error in the determination of τ_c the stress pulse width *ts* has to be chosen to have a similar value as τ_c. As seen in Fig. 6 *ts* increasing in steps of a factor 10 delivers a satisfactory accuracy for all τ_cs. Fig. 8 shows extracted τ_cs and τ_es for a couple of defects as an Arrhenius plot. Our statistical analysis allows a very precise determination of activation energies E_A. E_As are higher than the ones extracted from conventional experiments, in agreement with the findings from ultra-fast temperature changes [18]. We will discuss these E_As again in section VI . It should be noted, that the thermal activation of the single defects has **exactly** an Arrhenius

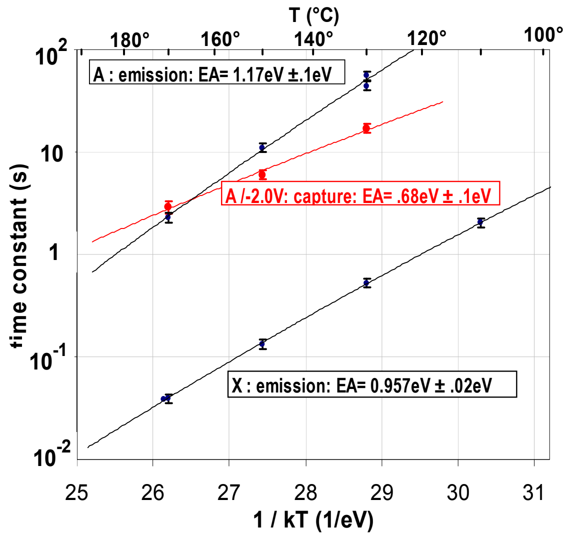

Figure 8. Arrhenius plot for defect A as well as capture and emission for trap X. E_A values are higher than in conventional experiments. Note the considerable accuracy in determining τ values. Each point is extracted from 256 emission events of the given defect.

behavior, in contrast to a conventional determination of thermal activation done on wide FETs [19].

This work shows only one example for parameter extraction, namely the T-dependence determination in Figs. 6, 7, 8. In an analogous way also the *Vgate* dependence of τ_cs and τ_es can be determined. These results are presented in [10].

V. New NBTI Model

With the findings so far we can be sure that the model of distributed time constants as proposed in [7] correctly describes the recoverable part of NBTI. Some considerations regarding permanent NBTI will be discussed in section IX. The model is equivalent to eq. (1) and (2), its equivalent circuit is shown in Fig. 9. Each defect is represented by a capacitor C which can be charged or discharged from a signal line applying stress- or recovery-voltage. The charge or voltage on C is equivalent to the individual *ΔVT* caused by the defect. The complete model for a FET thus consists of just one RC-element for each defect. The value of C for each defect will be roughly the same and corresponds to the value of *ΔVT* produced by this defect. The values for the charging resistor R_c and the discharging resistor R_e determine the capture and emission time constants R_cC and R_eC and are widely distributed. All the time constants are a function of *Vgate* and temperature. We will employ this NBTI model in the following sections.

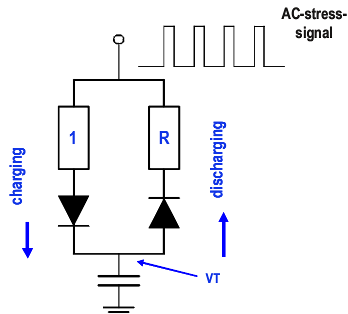

Figure 9. A single RC-element representing one given defect with a capture and emission time constants τ_c and τ_e having a ratio τ_e / τ_c =R. The RC-element is an equivalent to the equations (1) and (2). The voltage *VT* on C corresponds to the *ΔVT* produced by this defect. A real FET contains a large number of these defects (see Fig. 13) , the individual *VT*s have to be summed up. The signal applied to the upper terminal indicates charging of the RC-element by an AC stress.

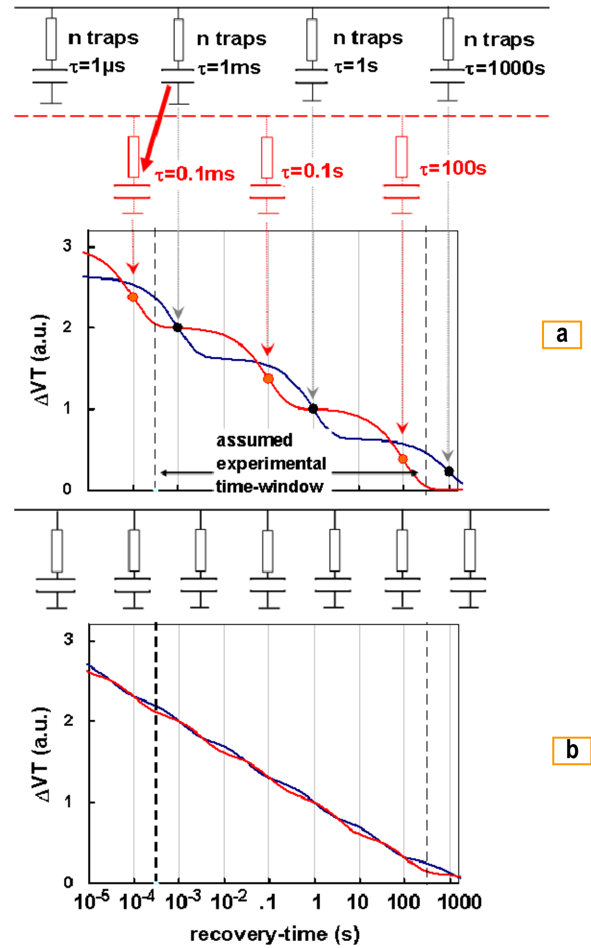

Figure 10. Simplified simulation of NBTI using the equivalent circuit from the top of the Figure. The τ's are aligned with the X-axis. When the τ's are widely separated (only one in 3 decades in Fig. a) distinct bumps are produced. A change of τ (by increased temperature, indicated by the red arrow) shifts the distinct bumps in the recovery curves. With densely, equally distributed RC's, ranging from ns to 1E6s (Fig. b) it is obvious that a shift left/right of this ladder has **no effect** in the experimental time window.

VI. The Missing-Thermal-Activation Riddle

In numerous studies by different research groups (see references in [11]), including our own [11] an apparently T-independent short-term degradation has been found. Fig. 2 shows a further example for this behavior. The standard explanation for this behavior has been that - apart from "real" NBTI - there is a special class of oxide defects which is charged and discharged by elastic quantum mechanical tunneling (QM) in and out of traps. The barrier ruling this QM-tunneling is in the order of the Si/SiO_2 valence band offset. Thus the barrier is 4eV>>kT which means that the effect is temperature independent. Typical contributions from this assumed QM-tunneling to the total *ΔVT* have been found to be in the regime of 10 to 20% [11]. In strong contrast, **all** the capture/emission events we analyzed, for <1µs up to ks, were strongly thermally activated (cmp. Figs. 6, 7, 8 and Fig. 8 in ref. [10]). In total we analyzed the thermal activation of roughly 100 defects and since this seemingly non-thermally-activated contribution is actually existing in our samples (see Fig. 2), we would by no means have overlooked such a class of not-thermally-activated defects. The solution of the riddle is given in Fig. 10: A simplified equivalent circuit describing charging and discharging of the defects is shown on top of Fig. 10. Only a few defects with time constants separated by 3 decades in times are taken into account. For the sake of simplicity τ_cs and τ_es are assumed to be equal, which is not correct but is good enough for the purpose. The time constants of the RC-elements are aligned with the

time axis in Fig. 10. Now let us consider an increase of temperature by 100°C: For an E_A=1eV this means that both τ_C and τ_e of any given defect (i.e. any of the drawn RC-elements) are accelerated by a factor of roughly 1000. Thus each of the RC-elements in Fig. 10 (each representing an individual defect or a class of defects) is getting faster, that means is shifted left by 3 decades. So, given that **all C-values are equal -** meaning the same number of traps in each class - the RC ladder just has shifted, but its effect has not changed: Were the 1ms traps were at low temperature, at the higher temperature the 1s-traps take over their "task" and so on. The apparent degradation behavior **did not change** due to the increase in temperature, the apparent activation energy is zero. Thus the strongly (a factor of 1000 for a change of 100°C) activated behavior of the individual defects is not seen in the large FETs due to the mixing of all the defects. The above postulation that all the C-values must be equal is equivalent to a uniform, flat distribution of time constants in log(t). This flat distribution is in turn equivalent to a degradation curve *ΔVT* vs *ts* which is linear in log(*ts*). This is in fact a standard observation in NBTI. To summarize, we think all conclusions about non-thermally-activated behavior for short times in large FETs were wrong. A convincing further proof for this assumption is - except our experimental results - the low temperature experiment in [20] showing that all threshold shift vanishes below T=150K.

VII. Practical Modeling

A distribution map of defect properties based on few defects is shown in Fig. 11. Clearly this map does not contain a number of defects large enough to be representative for a "real", say 0.5µm-wide FET. A 0.5µm wide FET would have filled ≈250 defects at a *ΔVT*=50mV (cmp. table 1). To fill this map with >250 measured defects in order to get a modeling-relevant spectral-density-map would be a task taking at least months. However, with our new understanding of trap properties, we now have the justification to switch back to a wide FET (with >4000 defects) and can extract a full defect density map just from measuring full-recovery traces from a single FET. This justification is mainly based on the fact that the number of defects and their properties do not change during stress (see also points 2, 3, 4, 5 in the conclusions).

The principles of extracting such a map, of gathering all the information needed, are outlined in Figs. 12 and 13. The "difference"-curve actually is a recovery curve containing **all** defects having a capture time constant **between** 10ms and 100ms. All these defects then are separated in one-decade-wide bins of different emission time constants as shown in Fig. 13. The filling level of the bins is given simply in mV-units, corresponding to the *ΔVT* generated by each bin. The density levels from Fig. 12 (denoted as labels in Fig. 12 and 13) contain the information to fill just one slice in Fig. 13. The other slices of Fig. 13 are filled in an analogous way. When all the squares in the defect density map Fig. 13 are filled this map comprises a complete set of parameters allowing a complete and straightforward modeling of NBTI for a given stress voltage and temperature. It includes recovery and the response of the sample to AC-stress or to any arbitrary sequence of stress-recovery cycles. An example for a calculated degradation curve and a recovery-curve in comparison to the experimental curves is given in Fig. 14. As seen the fit is perfect. This is not a surprise, however. Our model and the way of parameter extraction actually very much resembles a Fourier transformation: A set of data in the time-domain, (example Fig. 12) is just converted into a set of coefficients in the frequency domain (example Fig. 13), i.e. the set of "amplitudes" in the 2-dimensional spectral plot. Fig. 14 is quasi a re-conversion into the time-domain. If done correctly this re-conversion, like a Fourier transformation exactly reproduces the experimental data as demonstrated in Fig. 14. To prove the claim that the model is able to predict recovery and

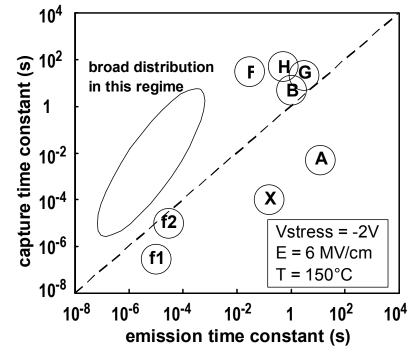

Figure 11. Correlation of τ_c/τ_e of a couple of selected defects from the statistical analysis of single defects (cmp. Fig. 5). τ_e may be larger or smaller than τ_c by orders of magnitude.

degradation for any arbitrary stress signal another model verification is shown in Fig. 15. Note that the stress condition is different from the example in Figs. 12, 13 and 14. For the model verification a series of stress/recovery sequences, with increasing times, has been applied to a sample and then simulated. As seen the quality of fit is very good. Unfortunately the log-time axis obscures all the µs to 1s time effects which are occurring for the long times also. To re-gain some of this information the experimental points have been chosen to be equidistant on a log-time scale.

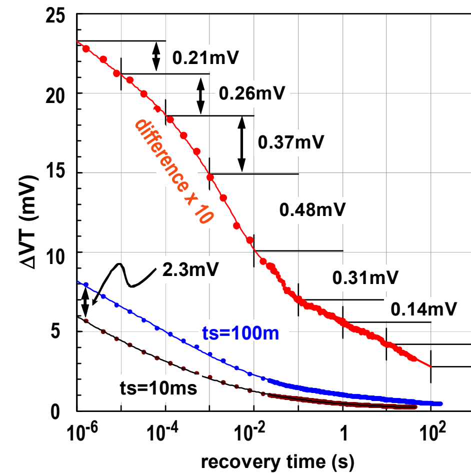

Figure 12. Example for extraction of spectral trap densities from recovery traces: The difference between ts=10ms and ts=100ms traces is due to the class of defects with 10ms<τ_c<100ms. So this class can be separated and then divided into "bins" of different τ_e s (i.e. recovery times). Experimental parameters given in Fig. 14.

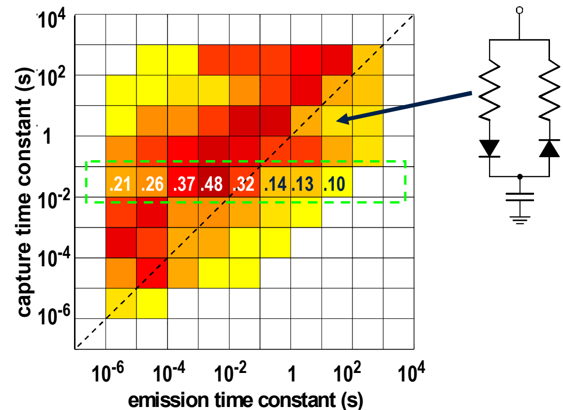

Figure 13. Complete spectral density map. The τ_c-slice analyzed in Fig. 12 is marked by a frame and labeled with the numbers from Fig. 12. Spectral density is simply given in *ΔVT*-units (=mV). Experimental parameters given in Fig. 14. Each square corresponds to a class of defects with given τ_C and τ_e as indicated by the RC-element.

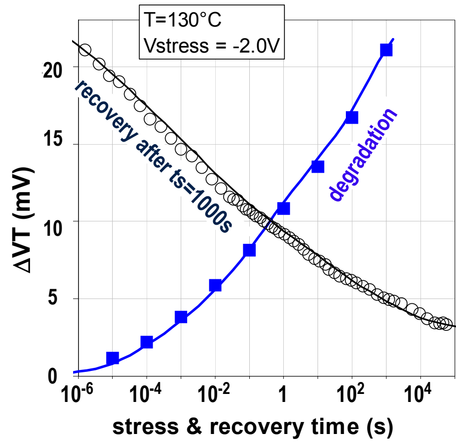

Figure 14. Using the parameter-table given in Fig. 13 a simulation of degradation to ts=1000s and subsequent recovery is done. Dots are experimental points; lines are calculated values.

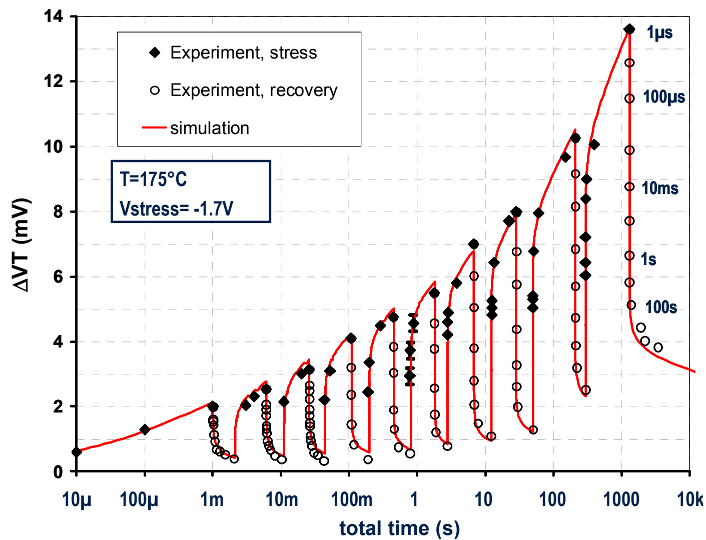

Figure 15. Example for model verification. Vstress is close to the maximum operating voltage of 1.6V. A stress/recovery sequence with increasing times was chosen. Agreement between experiment and model is in the sub-mV regime. Obviously all the experimentally observed time constants from µs to >10^3s are correctly captured by the simulation. Readouts during the stress phases (filled dots) have been measured with a measuring delay of 1µs and a total interruption of stress of 10µs. Measuring delay as well as this interruption are also considered in the simulation. The dents in the simulation, seen between 2ms and 5ms, are due to this interruption. Labels at the rightmost recovery branch denote the recovery time for each point, valid for each of the recovery branches.

A more detailed discussion of a fully automated extraction of the model parameters, of the hi-quality noise-free measurements needed, and of the required approximations needed for a fast and effective calculation of degradation, for example in a simulator like RelXpert cannot, due to limited space, be presented in this paper. These topics as well as gate voltage and temperature dependence will be discussed in a separate publication.

VIII. AC-Stress

Existing studies about AC-NBTI, e.g. [21] are showing a relation of *ΔVT* vs. duty cycle having the typical S-like shape. The reason for this shape could not been explained yet. Fig. 16 shows measured AC-stress data in comparison with a very simple simulation. The simulation perfectly fits the experimental data, proving that the AC-behavior can be fully understood by our model. The fit-curve in

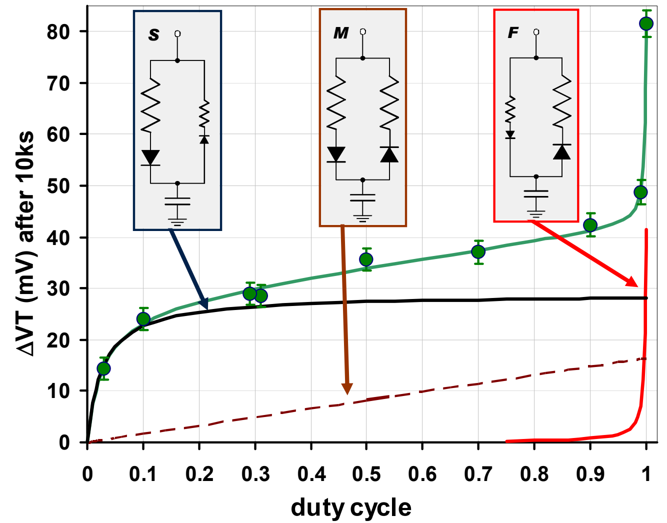

Figure 16. Degradation under AC stress at -2.7V at 175°C / 100kHz as a function of duty cycle. The full line trough the data is a best fit. It is composed as a sum from 3 contributions (the 2 full and the dashed lines), corresponding to 3 types of equivalent circuits S, M, F (=defects) as described in the text. The magnitudes of the 3 contributions are roughly equal.

Fig. 16 is "assembled" from a sum of 3 contributions, corresponding to 3 classes of defects. The 3 classes differ by the ratio between capture- and emission time constant. Class "S" has an emission time constant much slower than capture $\tau_e >> \tau_c$. Looking at the equivalent circuit it is clear that this is the behavior of an electronic peak detector. The C in the peak detector is almost fully charged as soon as the duty cycle *D* is just above zero. Class "F" has capture much slower than emission, $\tau_c >> \tau_e$. It is the opposite of the peak detector. *D* has to be very close to 100% to get a little charge into C, depending on the value of the ratio τ_e/τ_c. Finally class "M" has an emission time equal to capture time. It gives the linear response seen in Fig. 16 and is responsible for the linear center part of the experimental curve. The assumption that there are just 3 classes of defects is unrealistic. However, the shape of the curve in Fig. 16 is mainly determined by the contributions by the "extreme" classes "F" and "S". Thus the consideration of just 3 classes gives a perfect fit. It is noteworthy that mainly the ratio τ_c / τ_e determines the shape of the curve, while the absolute values are not important, as long as these time constants are longer than the AC-period. This explains a good deal of the observed independence of AC-degradation on frequency. Given that the measuring delay is much shorter than the stress-time (by more than 5 orders of magnitude) there also is no influence of the measuring delay. In our experiment the *ΔVT* measurement is always done after the completion of the stress-phase of an AC stress cycle. Measurement and stress signal are synchronized. Like for DC-stress we define the short (2µs) recovery phase before *VT* is determined as the measurement delay.

AC-stress measurements are easy to do, and will provide a lot of information about the sample. An experimental setup with a short measuring delay should be used, and the measurement-timing should be synchronized with the AC-signal. In this work we could only briefly sketch our experimental AC-results, the modeling and the derivations done. It is clear that the calculation of the response of an RC-element with e.g. a 1000s time constant, for a 10 year AC stress, with a total of 10^{15} AC-periods will need a couple of appropriate approximations in the calculation. We will provide a more detailed study of AC-NBTI, including data for more technologies in a subsequent publication.

IX. PERMANENT DEGRADATION

This study exclusively deals with the recoverable part of NBTI. Our mathematical model (represented by the spectral distribution in Fig. 13), however, is general and may include a permanent part without any changes. In our "single-defect" experiments, though the total stress-times applied to individual samples have been >10000s, there has been no evidence for a permanent degradation. In long stress experiments, reaching a high level of degradation (above 10^{12}cm^{-2} cmp. Fig. 1), we actually find experimental evidence for permanent degradation, and believe that it is really existing. So far we found it impossible to clearly distinguish a defect with a long emission time >10^5s, which appears in a distribution like the one of Fig. 13, from a defect which might appear in a distinct τ_e -column located at τ_e>>10^5s. Permanent defects may be different from the recoverable ones, and may have a different physical origin. An experimental proof for permanent defects, let alone a determination of their field- and temperature-dependencies so far just failed because it would require >10^7s-times, while most experimenters are limited to times somewhat above 10^5s.. Several studies have claimed to separate recoverable from permanent, or fast from slow components in NBTI-degradation [22]. We think that any distinction between fast and slow is still arbitrary and the distinction recoverable / permanent not yet justified by experimental data. There might also be the possibility that a potentially permanent part of NBTI is due to a damage of the gate oxide from highly-energetic, bond-breaking carriers like in time dependent dielectric breakdown tests. This would mean that additional oxide defects - the permanent ones - are generated at high field only. This effect would be different from NBTI. Due to a higher field acceleration these high field effects, causing TDDB, would not contribute to degradation at use voltage. We do not claim that this is true or proven, but it is a hypothesis which is worth to be considered.

X. CONCLUSIONS

Our new defect spectroscopy technique has been able to resolve some open questions in NBTI. We draw the following conclusions. All these conclusions are based on the experimental observations only and do not require further assumptions or models.

(**1**) The recoverable part of NBTI is completely driven by pre-existing defects. For a moderate degradation below a $5*10^{11}$cm^{-2} charge defect density no generation of defects has been observed.

(**2**) Each defect is characterized by a capture and an emission time constant τ_c and τ_e, both a function of temperature an electric field.

(**3**) τ_c and τ_e have a wide distribution and are not correlated (see Figs. 11, 13).

(**4**) Recovery of a given defect is only a matter of defect properties, absolutely independent of the previous stress-time and stress-voltage. This fact rules out that any diffusion is involved in degradation or recovery [23].

(**5**) Defects act uncorrelated with each other. This is plausible, assuming a density of 10^{12}cm^{-2} their typical distance is 10nm which is more than the screening length in the inversion layer [24]. Exceptions will be shown in ref. [10]

(**6**) All processes, also short term 1μs to 1ms, are thermally activated with activation energy typically 1eV. As explained in section VI the seemingly T-independent degradation at short times is an artifact.

(**7**) The new understanding of NBTI leads to a straightforward modeling, especially of AC-NBTI, based on a simple equivalent circuit like in Fig. 9 and a set of empiric parameters.

(**8**) If a defect density map like the one in Fig. 13 describes degradation and recovery it is clear that the concept of "universal recovery" cannot be correct but at the most an approximation.

REFERENCES

[1] J. H. Stathis and 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, P. M. Lenahan, A. T. Krishnan, and S. Krishnan, "Identification of the atomic-scale defects involved in the negative bias temperature instability in plasma-nitrided p-channel metal-oxide-silicon field-effect transistors", J. Appl. Phys., Vol. 103, no. 4, 2008, pp. 044505-1-11

[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] V. Huard, C.R. Parthasarathy, and M. Denais, "Single-Hole Detrapping Events in pMOSFETs NBTI Degradation, 2005 IIRW final report, p. 5

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

[7] 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.

[8] H.C. Ma, J.P. Chiu, C.J. Tang, T. Wang and C.S. Chang, "Investigation of Post-NBT Stress Current Instability Modes in HfSiON Gate Dielectric pMOSFETs", Proc. IRPS 2009, pp. 51-54

[9] 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 technical digest 2009 , pp. 729-732

[10] T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Goes, B. Kaczer, "The Time Dependent Defect Spectroscopy (TDDS) for the Characterization of the Bias Temperature Instability", IRPS 2010 (in press)

[11] 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)

[12] C. Shen, M.-F. Li, C.E. Foo, T, Yang, D.M. Huang, A. Yap, G.S. Samudra, Y.-C. Yeo, "Characterization and Physical Origin of Fast Vth Transient in NBTI of pMOSFETs with SiON Dielectric", IEDM 2006, p.333

[13] V. Maheta; C. Olsen, K. Ahmed, and S. Mahapatra, "The Impact of Nitrogen Engineering in Silicon Oxynitride Gate Dielectric on Negative-Bias Temperature Instability of p-MOSFETs: A Study by Ultrafast On-The-Fly IDLIN Technique", IEEE TED, Vol.55, No. 7, p. 1630

[14] S. Mahapatra, K. Ahmed, D. Varghese, A. E. Islam, G. Gupta, L. Madhav, D. Saha and M. A. Alam, "On the Physical Mechanism of NBTI in Silicon Oxynitride p-MOSFETs: Can Differences in Insulator Processing Conditions Resolve the Interface Trap Generation versus Hole Trapping Controversy?, IRPS 2007, p. 1

[15] B. Kaczer, Ph. J. Roussel, J. Franco, R. Degraeve, L.-A. Ragnarsson, E. Simoen, G. Groeseneken, T. Grasser, H. Reisinger, "Origin of NBTI Variability in Deeply Scaled pFETs", IRPS 2010 (in press)

[16] G. Math, C. Benard, J. Ogier, D. Goguenheim, "Geometry effects on the NBTI degradation of PMOS transistors", 2008 IIRW final report, pp. 60-63

[17] H. Reisinger, T. Grasser and C. Schlünder, "A study of NBTI by the statistical analysis of the properties of individual defects in pMOSFETs", 2009 IIRW final report, pp. 60-63

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

[19] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, M. Goodwin, "Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification", IRPS 2005, pp. 381-387

[20] R.G. Southwick III, W.B. Knowlton, B. Kaczer, and T. Grasser, "On the thermal activation of negative bias temperature instability", 2009 IIRW final report, p. 36

[21] R. Fernadez, B. Kaczer, A. Nackeaerts, R. Rodriguez, M. Nafria, G. Groeseneken, "AC NBTI studied in the 1 Hz - 2 GHz range on dedicated on-chip CMOS circuits", IEDM technical digest 2006, pp.337-340

[22] V. Huard, C. Parthasarathy, N. Rallet, C. Guerin, M. Mammase, D. Barge, C. Ouvrard, "New characterization and modeling approach for NBTI degradation from transistor to product level", IEDM 2007, p. 797

[23] 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

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