Recent Advances in Understanding the Bias Temperature Instability

T. Grasser*, B. Kaczer°, W. Goes*, H. Reisinger•, Th. Aichinger†, Ph. Hehenberger◇,
P.-J. Wagner*, F. Schanovsky◇, J. Franco°, Ph. Roussel°, and M. Nelhiebel‡

* Christian Doppler Laboratory for TCAD at the ◇Institute for Microelectronics, TU Wien, Austria
° IMEC, Leuven, Belgium • Infineon, Munich, Germany † KAI, Villach, Austria ‡ Infineon, Villach, Austria

Abstract

Our understanding of the bias temperature instability (BTI) has been plagued by disagreements related to measurement issues. Although even in the early papers on BTI the existence of recovery was acknowledged and discussed [1, 2], for unknown reasons this had little impact on the way we used to think about the phenomenon until recently [3–7]. Even after the re-discovery of recovery [3], it took a few years until it was fully appreciated that any measurement scheme conceived so far considerably interferes with the degradation it is supposed to measure, often accelerating its recovery. Nonetheless, this experimental nuisance has led researchers to think in more detail about the problem and has thus opened the floodgates for fresh ideas [6–11]. Some of these ideas together with the experimental data supporting them are reviewed in the following.

Introduction

Already the first papers on BTI list charge trapping in oxide defects alongside the creation of interface states as the likeliest cause of the degradation [1, 2]. Unfortunately, the role of charge trapping was completely swept under the carpet by the apparent success of reaction-diffusion (RD) theory in explaining the negative bias temperature instability (NBTI) [2, 12–14]. The most recent version of RD theory postulates NBTI as being due to a superposition of elastic hole trapping, which saturates within about a second, and an H_2 diffusion-limited creation of interface states, which dominates degradation and recovery [15, 16]. During the last couple of years evidence has piled up that clearly demonstrates that NBTI recovery cannot be the diffusion-limited process suggested by the RD model: *(i)* RD theory predicts a relatively short recovery phase lasting about 4 decades. In particular, a 50% recovery level is expected as soon as the relaxation time equals the stress time, $t_r^{50\%} = t_s$. In reality, recovery covers many more decades in time and follows a power-law with an exponent around -0.1 (about $\log(t_r)$-like), see Fig. 1. *(ii)* RD-predicted recovery is due to the back-diffusion of neutral H_2, and as such bias-independent. In fact, however, experimental recovery is *strongly bias-dependent*, particularly for V_G switches towards accumulation. This bias dependence occurs well beyond the alleged saturation of the hole trapping component of about 1 s and must thus be an intrinsic feature of the dominant recoverable component, see Fig. 2. *(iii)* The duty-factor (DF) dependent degradation predicted by RD theory follows $DF^{1/3}$, while experimentally a rotated S-shape is observed, with a very sharp drop of the degradation even for small deviations from the DC case [17–19]. In particular, for DF$= 0.5$ (AC), RD predicts 80% of the DC degradation level while experiments give a value typically smaller than 50%, see Fig. 3. *(iv)* Recovery is independent of the hydrogen passivation degree of the interface (hydrogen budget) [20–22].

The Link to Random Telegraph Noise

A detailed look under the hood of BTI degradation can be provided by studying small area devices, where recovery proceeds in discrete steps, see Fig. 4. These discrete steps are due to the emission of a single hole and consistent with a first-order reaction-rate but not with a diffusion-limited process [23–27]. It is intriguing to realize that all the properties of these discrete steps are fully consistent with charge trapping observed in the context of random telegraph noise (RTN) and $1/f$ noise [19, 24–29]. Just like in RTN, the step-heights can be considerably larger than expected by the simple charge-sheet approximation as a consequence of the discrete random dopant locations in the channel [30]. Even for short stress times, emission events with emission times 10^5 times larger than the stress time are observed, homogeneously distributed over the logarithmic time axis, see Fig. 5. The step-heights are exponentially distributed [29], see Fig. 6, which results in a non-negligible probability that defects have step heights well beyond 30 mV, a typical lifetime criterion.

The recently developed *time dependent defects spectroscopy* (TDDS), see Figs. 7 and 8, allows for the analysis of these steps as a function of stress/recovery bias conditions and temperature [19, 27] and has revealed a number of interesting facts about the hole trapping component: *(i)* Both capture and emission time constants are widely distributed as well as temperature dependent. No defects with temperature-independent capture time constants – hinting at an elastic tunneling process – could be found in the experimental window covering the micro- up to the kilosecond regime. This has been recently confirmed to be the same for PBTI in nFETs [31]. *(ii)* The capture time constants show a very strong field dependence. Similarly, the bias dependence of the emission time constant around V_{th} may be either weak or strong, depending on the configuration of the defect. *(iii)* Charge capture is dominated by pre-existing defects. *(iv)* The existence of metastable states becomes obvious due to disappearing defects, anomalous and temporary RTN (cf. Fig. 9), and probably also by the bias dependencies stronger than expected in a simple model [27]. *(v)* Finally, the defects responsible for the recoverable component of BTI are identical to those causing RTN. Conceptually, the difference is as follows: in an RTN experiment only a limited number of defects having capture and emission time constants within the experimental window are visible. Due to the strong bias dependence of the capture time constant, many more defects contribute to BTI. Consequently, NBTI can be

considered the non-equilibrium response of these defects while RTN is a consequence of their quasi-stationary behavior [29].

Charge Trapping Mechnism

All these results clearly demonstrate the central role of charge trapping in NBTI [7, 11, 19, 23–28], which is in line with the standard interpretation of PBTI in high-κ devices [32]. However, the usual theory invoked to explain such phenomena is due to McWhorter [33], who attributed the wide dispersion in the capture and emission time constants required to explain RTN and $1/f$ noise to defects located at *various depths* into the oxide. While it has been realized early on that elastic tunneling is unable to explain the universally observed strong temperature- and bias-dependence of the capture time constant [34–36], Campbell *et al.* [37] explicitly demonstrated that *modern oxides are simply too thin* to support any significant distribution of time constants based on distance alone.

Thus, *rather than an elastic tunneling* process, *inelastic processes* have been applied to explain BTI data. The most consistent explanation is provided by non-radiative multi-phonon (NMP) theory [36, 38–40], which has recently been applied to BTI data [27]. In NMP theory, charge capture and emission is only possible when the sum of the electronic and vibrational energies of the defect are conserved. *Conservation of this total energy* introduces thermal barriers into the charge exchange process, thereby explaining the temperature dependence with the relatively large activation energies around 1 eV. In addition, since the electronic contribution to the total energy is either due to an electron at the defect site or in the substrate, application of an electric field leads to a strong bias-dependence of these thermal barriers [27, 41].

Based on the standard model for E' centers developed in the context of irradiation damage [42, 43], we have suggested a defect model that is consistent with the experimental NBTI data [28], see Fig. 10. In order to explain the extremely strong bias dependence as seen in Figs. 11 and 12 for two typical defects, a metastable positive state had to be introduced [27]. The extended model also allows for the description of various anomalies observed in the spectral map obtained by the TDDS: *(i)* Defects can disappear from the spectral map for random amounts of time, with the inactive defect being electrically neutral. This implies that such a defect has an electrically neutral metastable state. *(ii)* A small number of defects produce temporary RTN (tRTN) after having been positively charged (cf. Fig. 9). This implies that the defect switches back-and-forth between its positive and metastable neutral state prior to annealing. These switches cannot be explained by transitions between the positive and the neutral precursor state, as the associated capture time constant is relatively large, particularly at low bias. This tRTN corresponds to the anomalous RTN under constant bias observed previously in standard RTN experiments [44].

The Two-Stage Model for NBTI

Our previously suggested two-stage model [28] for NBTI employs a simplified version of the defect model described above and is evaluated against some crucial NBTI benchmarks in Fig. 13. The experimentally observed behavior is fully reproduced, be it the asymmetry between stress and relaxation, the acceleration of recovery after positive bias pulses, or rapid switches in the temperature.

Of particular importance is the dependence of the recovery on positive bias pulses: in a standard two-state defect model, when the defect level is lowered below the Fermi-level, hole-emission from the defect will be *bias-independent*, as the hole simply 'bubbles-up'. In our multi-state defect model, defect annealing proceeds via its neutral metastable charge state. As a consequence, when the defect is positively charged, e.g. around the threshold voltage, recovery is inhibited. Quite contrary, when the defect is electrically neutralized by switching to positive biases, it has a much higher probability of being annealed. As such, this multi-state model can explain the density-of-states introduced in the bandgap. In other words, positive charge is not just simply stored in the oxide, but injection of positive charge creates a defect. This defect can be charged and discharged and anneals preferably from its neutral metastable state.

At a first glance, the present model appears to be at odds with the occasionally observed temperature-independence of degradation and recovery [8, 15, 16]. However, this temperature-independence is an artifact related to the relatively homogeneous density-of-states, an illusion thus visible only in the macroscopic response of large-area devices, see Fig. 14.

Similar conclusions hold for on-the-fly measurements. There, the degradation of the drain current is measured relative to the first measurement point, taken for instance at 1 μs. In a hypothetical scenario where all defects have the same activation energy, the window of observable defects is simply shifted to shorter times at higher temperatures while the number of defects stays the same. Thus, even with large individual activation energies, a temperature independent degradation would be observed in large area devices. In reality, the activation energies of the defects are distributed, resulting in a small overall macroscopic activation energy (0.1 eV), which is thus considerably smaller than the 'physical' microscopic activation energies of the individual defects (0.5–1.2 eV) [19, 27].

Conclusions

A number of important conclusions may be drawn: *(i)* BTI up to at least 1 ks is *dominated by inelastic charge exchanges between preexisting defects and the substrate*. *(ii)* RTN and the recoverable component of BTI are caused by the same defects, with *RTN* being their *quasi-equilibrium* and *BTI* their *non-equilibrium response*. *(iii)* In *small-area devices* the stochastic nature of charge exchange makes the *lifetime a stochastic variable* [29], see Figs. 15 and 16. Since every device will have a different distribution of defects, exact knowledge of the distribution of the time constants has to be acquired.

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement n°216436 (ATHENIS) and from the ENIAC project n°820379 (MODERN).

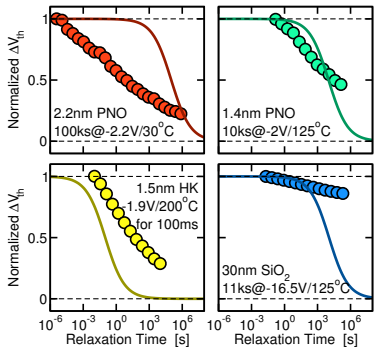

Fig. 1: Recovery at $V_G = V_{th}$ for four different technologies always follows a $\log(t_r)$-like behavior. Symbols are data, lines are the RD prediction. RD recovery occurs dominantly over 4 decades in time, in stark contradiction to experimental data.

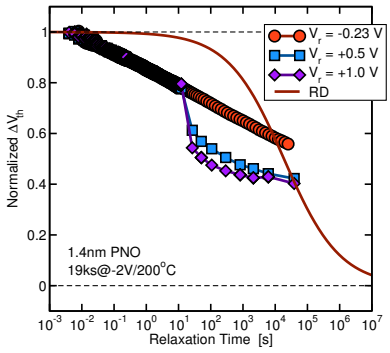

Fig. 2: Short switches to positive biases after 10 s of normal recovery at $V_G = V_{th}$ considerably accelerate recovery. The RD model is based on neutral H_2 diffusion and thus oblivious to such bias changes.

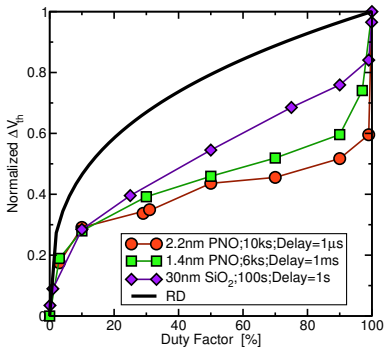

Fig. 3: DF dependence of the degradation normalized to the AC (50%) case. The RD model cannot reproduce the typical S-shape found in all three technologies. The reduction from the DC case is very sensitive to the measurement delay.

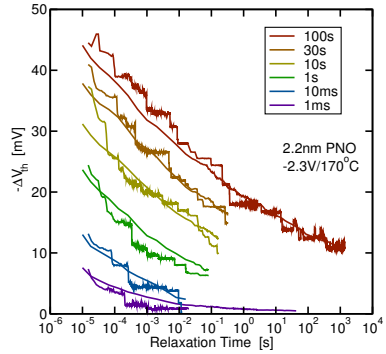

Fig. 4: Comparison of the recovery traces for a single narrow device (traces with steps) and the average over 25 different FETs (smooth traces) [19]. Each step is due to the discharge of a single hole. Emission events are equally distributed from the microsecond to the kilosecond regime.

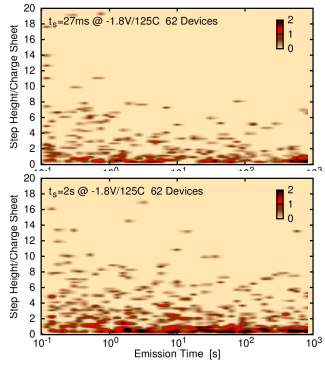

Fig. 5: Normalized step heights (0.6 mV) vs. emission time after a stress time of 27 ms and 2 s. Even after a 27 ms stress, emission events with 1 ks are generated. Increasing stress time homogeneously increases the number of emission events.

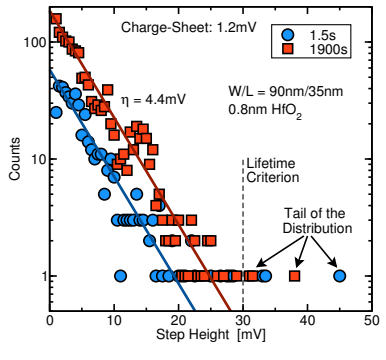

Fig. 6: Exponential distribution of step heights for two different stress times. Note that η is about 4 times larger than the charge-sheet value [29]. Even for short stress times, defects from the tail of the exponential distribution can produce threshold voltage shifts larger 30 mV, a typical lifetime value.

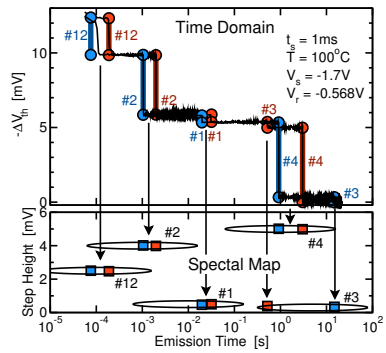

Fig. 7: The TDDS: Two typical ΔV_{th} recovery traces of a small-area pMOSFET. The measured data are given by the noisy black lines (top). The thick blue and red lines together with the symbols mark the emission times and step heights, unambiguous fingerprints of each defect building the spectral map (bottom).

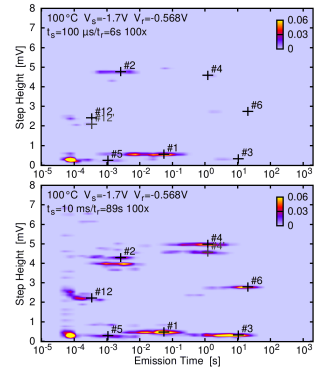

Fig. 8: Two TDDS spectral maps at two stress times, $t_s = 1$ ms (top) and $t_s = 10$ s (bottom). With increasing stress time, the number of defects in the map increases. The width of each cluster is given by the exponential distribution of τ_e and the extracted defects are marked by '+'.

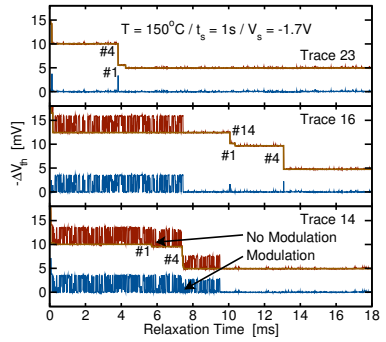

Fig. 9: In addition to the discrete recovery steps, the traces can contain temporary RTN (tRTN), which disappears after a certain amount of time, see three selected traces above. The extracted step heights (brown) are subtracted from the raw data (red) to yield the noise trace (blue).

References

[1] D. Breed, APL **26**, 116 (1975). **[2]** K. Jeppson *et al.*, JAP **48**, 2004 (1977). **[3]** S. Rangan *et al.*, *IEDM* (2003), p. 341. **[4]** M. Denais *et al.*, *IEDM* (2004), p. 109. **[5]** V. Huard *et al.*, MR **45**, 83 (2005). **[6]** B. Kaczer *et al.*, *IRPS* (2005), p. 381. **[7]** C. Shen *et al.*, *IEDM* (2006), p. 333. **[8]** H. Reisinger *et al.*, *IRPS* (2006), p. 448. **[9]** D. Ang, EDL **27**, 412 (2006). **[10]** T. Grasser *et al.*, *IEDM* (2007), p. 801. **[11]** V. Huard *et al.*, *IEDM* (2007), p. 797. **[12]** M. Alam, *IEDM* (2003), p. 345. **[13]** S. Chakravarthi *et al.*, *IRPS* (2004), p. 273. **[14]** S. Mahapatra *et al.*, ME **80**, 114 (2005). **[15]** A. Islam *et al.*, *IEDM* (2009), p. 733. **[16]** S. Mahapatra *et al.*, T-ED **56**, 236 (2009). **[17]** R. Fernandez *et al.*, *IEDM* (2006), p. 1. **[18]** T. Grasser *et al.*, *IRPS* (2007), p. 268. **[19]** H. Reisinger *et al.*, *IRPS* (2010), p. 7. **[20]** T. Aichinger *et al.*, APL **96**, (133511 (2010). **[21]** T. Aichinger *et al.*, *IRPS* (2010), p. 1063. **[22]** V. Huard, *IRPS* (2010), p. 33. **[23]** V. Huard *et al.*, *IIRW* (2005), p. 5. **[24]** B. Kaczer *et al.*, *IRPS* (2009), p. 55. **[25]** H. Reisinger *et al.*, *IIRW* (2009), p. 30. **[26]** T. Grasser *et al.*, *IEDM* (2009), p. 729. **[27]** T. Grasser *et al.*, *IRPS* (2010), p. 16. **[28]** T. Grasser *et al.*, *IRPS* (2009), p. 33. **[29]** B. Kaczer *et al.*, *IRPS* (2010), p. 26. **[30]** A. Asenov *et al.*, T-ED **50**, 839 (2003). **[31]** M. Toledano-Luque *et al.*, *WoDiM* (2010), p. 1. **[32]** C. Young *et al.*, T-ED **56**, 1322 (2008). **[33]** A. McWhorter, S.Surf.Phys 207 (1957). **[34]** P. Dutta *et al.*, PRL **43**, 646 (1979). **[35]** M. Weissman, RMP **60**, 537 (1988). **[36]** M. Kirton *et al.*, Adv.Phys. **38**, 367 (1989). **[37]** J. Campbell *et al.*, *IRPS* (2009), p. 382. **[38]** K. Huang *et al.*, Proc.R.Soc.A **204**, 406 (1950). **[39]** D. Lang *et al.*, PRL **35**, 1525 (1975). **[40]** C. Henry *et al.*, PRB **15**, 989 (1977). **[41]** W. Fowler *et al.*, PRB **41**, 8313 (1990). **[42]** A. Lelis *et al.*, T-NS **41**, 1835 (1994). **[43]** J. Conley *et al.*, APL **67**, 2179 (1995). **[44]** M. Uren *et al.*, PRB **37**, 8346 (1988). **[45]** T. Aichinger *et al.*, *IRPS* (2009), p. 2.

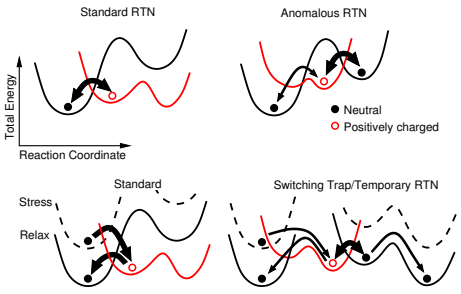

Fig. 10: Schematic configuration coordinate (CC) diagram model for a general defect [27]. In an amorphous oxide, each defect will have different configurational potentials which determine the non-radiative multiphonon (NMP) process. Depending on the bias conditions, the relative position of the potentials changes, causing a strong field dependence of the transition rates [41]. Thicker arrows indicate larger transition probabilities.

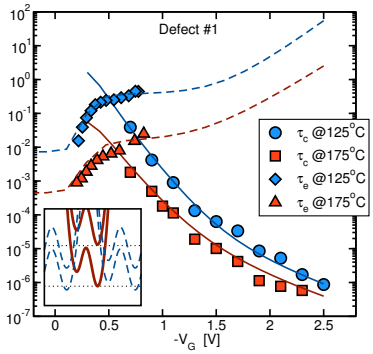

Fig. 11: Simulated capture and emission time constants for defect #1 of Fig. 8 compared to the experimental values. The CC diagram shown in the inset determines the switching trap behavior, where for small V_G defect annealing via the neutral metastable state is favored, resulting in a dramatic drop of the emission time constant.

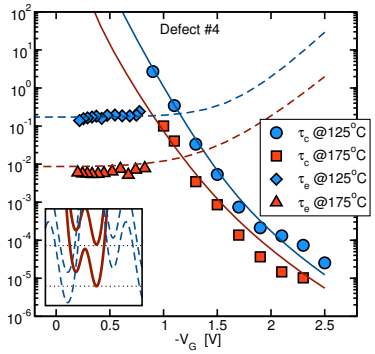

Fig. 12: In contrast to defect #1 in Fig. 11, the neutral metastable minimum of defect #4 is too high to be significantly populated and no switching trap behavior can be observed. This can be clearly seen in the CC diagram shown in the inset. As a consequence, the emission time constant is insensitive to the gate bias even for small V_G.

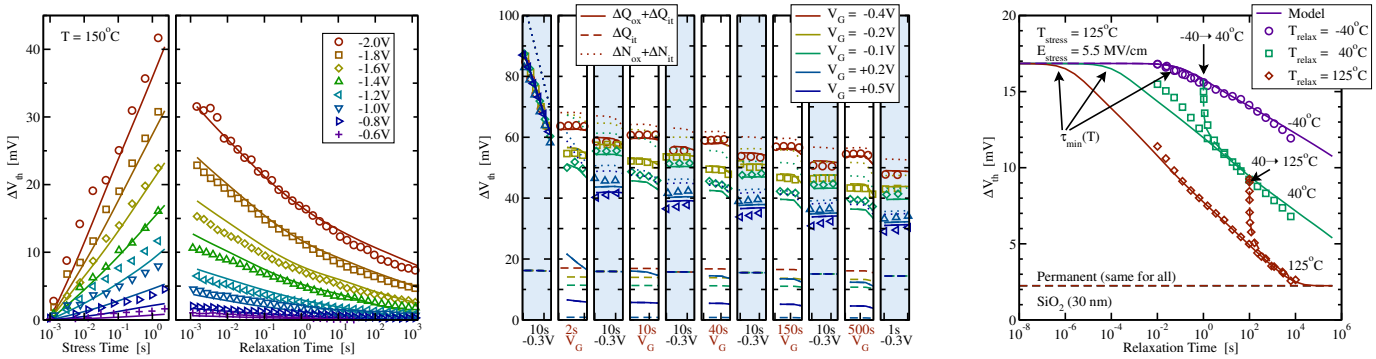

Fig. 13: The recently developed two-stage model [28] can already explain a number of crucial features of NBTI: **Left**: Stress and recovery are asymmetric, with recovery taking much longer than degradation, shown at 150°C for 8 different stress voltages for the 1.4 nm SiON devices. The asymmetry is properly reproduced by the model, resulting in a good fit during *both* stress and recovery. **Middle**: Five devices were brought to the same level of degradation ($t_s = 6000$ s, $T = 125$°C, $V_s = -2$ V). First, recovery was monitored for 10 s at $V_G = V_{th} = -0.3$ V (panel 1). Then the gate voltage was switched for 2 s to 5 different values, including more negative and more positive values (panel 2). When possible ($V_G > 0$ V), the change in the drain current was converted to $\Delta V_{th}(V_G)$ which was found to be clearly different from $\Delta V_{th}(V_{th})$ recorded at V_{th}. Next, V_G was switched back to V_{th} (panel 3), were a clear impact of the intermediate bias switch is observed. This procedure was repeated for increasing durations of bias switches (10, 40, 150, 500 s). Simulation results are given by the lines, which show very good agreement with the data, capturing both the occupancy effect (evenly numbered panels) and the acceleration/retardation of recovery as a response to the gate bias. **Right**: Application of a recently developed poly-heater technique [45] allows for high-precision switching of the device temperature within 1 s. Identically processed devices were stressed for the same time at 125°C, while prior recovery the temperature is quickly switched to −40°C or +40°C. Finally, in a fourth device the temperature is first switched to -40°C, then after 1 s recovery to +40°C, and finally after 100 s to 125°C. Excellent agreement of the model with data is again obtained.

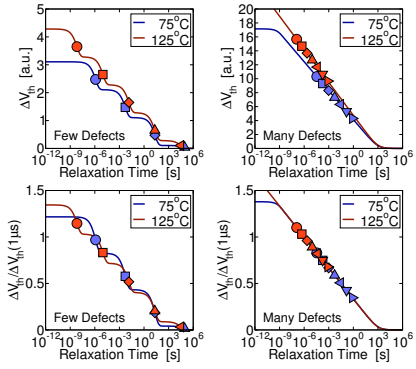

Fig. 14: In small area devices with only a few defects (left), the temperature dependence in the recovery remains visible after scaling the traces at a some t_r time, e.g. the first measurement taken at 1 μs. In large area devices (right), a defect with a higher activation energy seamlessly takes the place of a defect with a lower activation energy, resulting in *apparently* temperature-independent behavior.

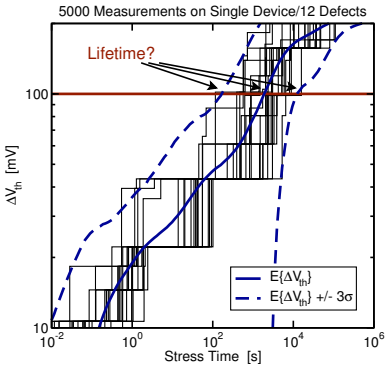

Fig. 15: Simulated repetitive degradation measurements on the *same* small area device containing 12 defects gives different results for each run due to the stochastic nature of charge trapping (thin lines). As a consequence, the *lifetime becomes a stochastic variable*. When the $\pm 3\sigma$ band (dashed lines) around the expectation value is considered (thick lines), the lifetime is reduced by about one order of magnitude.

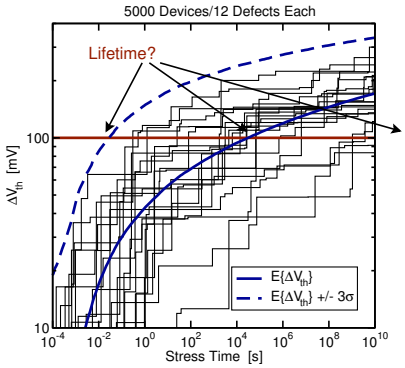

Fig. 16: In order to qualify a technology, degradation measurements have to be repeated on *different* devices (thin lines). Since each small-area device will have a different number of defects with different time constants, an estimate for the *lifetime can no longer be given* without detailed knowledge of the distribution of the time constants. This will *require a radical change in the qualification procedures*.