Physical Modeling of NBTI: From Individual Defects to Devices

G. Rzepa*, W. Goes*, G. Rott†, K. Rott†, M. Karner‡, C. Kernstock‡, B. Kaczer§, H. Reisinger†, and T. Grasser*
*Institute for Microelectronics, TU Wien, Vienna, Austria
†Infineon, Munich, Germany
‡Global TCAD Solutions, Vienna, Austria
§imec, Leuven, Belgium

***Abstract*—Given the rapid recovery of the degradation induced by bias-temperature stress, the understanding and modeling of NBTI has been a challenge for nearly half a century. With the introduction of the time-dependent defect spectroscopy (TDDS), NBTI could be studied at the single defect level, confirming that it is dominated by a collection of first-order reactions rather then the previously invoked reaction-diffusion mechanism. The most intriguing feature of these first-order processes is the wide distribution of their time constants, which can be visualized in capture/emission time (CET) maps. In the following we clarify the microscopic link between individual defects seen in TDDS studies and the response of a large ensemble visible in the CET maps. In particular, we show how the distribution of the individual defect parameters can be extracted from measurements on large-area devices.**

Introduction

The degradation of transistor parameters due to negative bias- and temperature-induced instabilities (NBTI) [1] has traditionally been described using some reaction-diffusion models [2, 3]. However, a number of studies have revealed inconsistencies in these models [4–7] and it was demonstrated that NBTI is consistent with a collection of reaction-limited processes [4, 7–11], such as charge trapping and reaction-limited generation of interface states. Based on these observations we have developed the two-stage model [6], which describes charge-trapping using an empirically extended nonradiative multiphonon (NMP) theory rather than the previously envoked elastic tunneling processes [12, 13]. More recent investigations of individual defects in nano-scale devices using time dependent defect spectroscopy (TDDS) [14] have revealed that the behavior of oxide defects can only be explained by metastable states, which led to the NMP four-state model for the recoverable component. Studies using capture/emission time (CET) maps [15, 16] have illustrated a *wide distribution of the time constants* of the defects, consistent with TDDS data. Still, the actual distributions of the *individual defect parameters* which eventually underly these CET maps are difficult to obtain as single-defect TDDS studies are very time-consuming. Here we demonstrate how the wide distributions of the defect parameters can be extracted from large-area devices, which enables us to provide the microscopic link between detailed defect models and the empirical CET maps.

Microscopic Defect Models

Based on our previous studies, we consider two types of defects, namely the NMP four-state model, which is used to model the "recoverable" defects inside the oxide and a simple double-well model which accounts for the more "permanent" defects probably located at the interface between the oxide and the substrate, possibly P_b centers [17]. Both models describe the transitions between the defect states using bias-dependent configuration coordinate (CC) diagrams [18]. As all defects are different, each of them will have a differently shaped CC diagram, see Fig. 1.

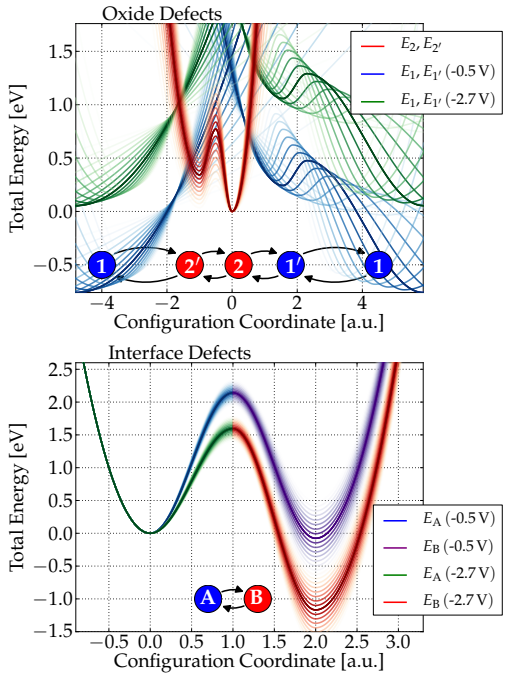

Fig. 1: The extracted distribution of the model parameters depicted in CC diagrams. The opaque lines represent the mean values and the fade-out illustrates up to one σ of the distributions for $V_G = -0.5$ and -2.7 V. **Top**: The NMP four-state model considers two neutral states 1 and 1′ and two positive states 2 and 2′. The states 1′ and 2′ are metastable and their energy minimas are above the respective energy minimas of the stable states. For simplicity this CC diagram only illustrates the situation for carrier exchange with the valence band of the substrate. An increase of the gate voltage shifts the energies according to the difference of the electrostatic potential at the defect site compared to the surface potential. Thereby, the energy barrier from state 1 to 2′ becomes smaller which results in a decreasing τ_c. At the same time the barrier from state 2′ to 1 increases, which causes an increase in τ_e. The resulting shift is depicted for an exemplary interface distance $x_t = 1.0\,\mathrm{nm}$. **Bottom**: The double well model for simulation of interface defects. Once in state B, the charge state is determined using an amphoteric SRH model.

While the NMP four-state model directly provides the charge of the oxide defects, the charge state of the defects described by the double-well model is determined using an amphoteric SRH model [19] in state B. In order to describe the ensemble of defects, the parameters which determine the CC diagrams are assumed to be *independent and normally distributed*, with the mean and variance adjusted during model

978-1-4799-5288-5/14/$31.00 © 2014 IEEE

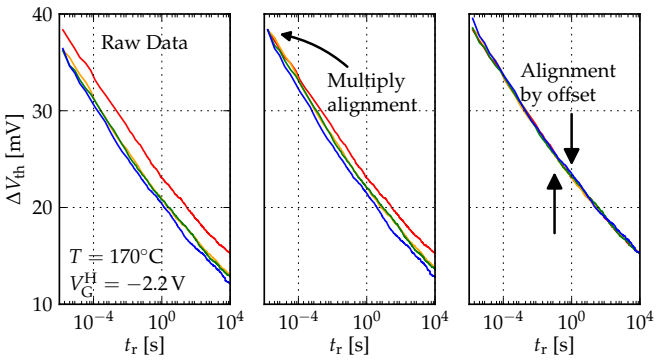

Fig. 2: **Left**: Recovery traces of different devices of the same technology after identical stress. Variability appears to be mostly due to the permanent component, as a correction using a multiplicative shift (**middle**) yields a much larger error than a correction with an additive offset (**right**) does.

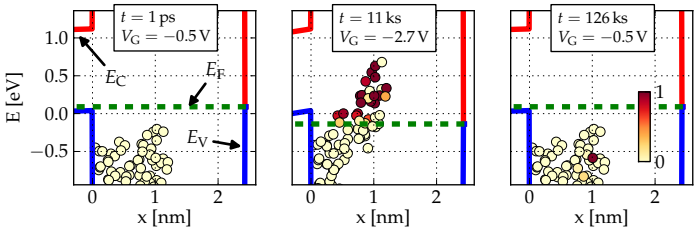

Fig. 3: Interface distance and trap level of 80 randomly selected oxide defects depicted with the band diagram. Their color indicates their charge in multiples of the elementary charge. **Left**: At the initial equilibrium of the simulation all defects with a trap level below the Fermi level are neutral. **Middle**: Due to the electrostatic potential at the defect site the trap levels are raised with increasing interface distance and increasing V_G. The traps which are shifted above the Fermi level can eventually become charged. The time constants are mostly determined by the relaxation energies rather than the trap level and the depth. **Right**: Applying low gate voltage again shifts back the trap levels and the defects gradually discharge, depending on the respective relaxation energies.

calibration. We assume all defects to be distributed uniformly in space, but only in the oxide half which is next to the substrate as we neglect the transition rates between gate and oxide in this work [18].

Experimental

For model calibration we use our long-term NBTI data [16] previously obtained on pMOSFETs with a 2.2 nm thick SiON oxide for different stress voltages ($-V_G = 1.2, 1.7, 2.2, 2.7, 3.2\,V$), temperatures ($T = 125$ and $170°C$), and stress and recovery times in the range $1\,\mu s$–$100\,ks$. The short measurement delay combined with the subsequent stress-recovery measurements with increasing stress times allows for indirect measurement of the degradation during stress. A measurement with many stress-recovery cycles for a certain stress voltage and temperature can be performed on a single device, but as model calibration has to cover various stress conditions, many devices have to be measured. A serious issue here is the device-to-device variability of the threshold voltage shifts (ΔV_{th}), as even under nominally identical stress conditions, slightly different ΔV_{th} recovery traces are observed, see Fig. 2. To first order, the deviation appears to be mostly due to a $\pm 15\%$ variability of the permanent component, which can be partially accounted for during the calibration to give the average behavior of the applied technology.

Even though the time constants of oxide and interface defects typically differ substantially, there is an overlap of the contributions of these two defect types. For improved distinction of the corresponding "permanent" and "recoverable" components we propose to use additional 'accumulation pulse' measurements. This setup is similar to conventional stress-recovery measurements but it introduces a short pulse into accumulation before the regular recovery phase. The duration of these accumulation pulses can be varied to stimulate discharge of the bias-dependent recoverable defects and thereby allows for a better analysis of the remaining permanent defects.

Parameter Extraction

The degradation of larger devices comprises contributions of many individual defects. Therefore, the corresponding microscopic simulation has to be based on a large number of defects, where each defect is described by its individual parameters. Still, these individual parameters were not assigned during model calibration directly for each defect. Instead, the *distributions* of these parameters were extracted. Based on these distributions *one set* of individual defects was obtained and applied to *all simulations* in this work. In accordance with Fig. 1 the extracted distribution of the trap levels and their position dependent shift is depicted in Fig. 3, together with the related charging and discharging processes during stress and recovery.

The actual defect density and parameter extraction was done by matching the simulated ΔV_{th} to the measured ΔV_{th}. Therefore, the simulated degradation, based on one set of defects, has to respond to various stress conditions in accordance with the corresponding measurements. In order to obtain reliable parameters the measurements and simulations were done for various stress voltages, temperatures and stress durations and additionally the previously explained 'accumulation pulse' setup was applied (Fig. 4). As this procedure yields individual defects it allows for a very detailed analysis. Beside the investigation of the stress dependent capture and emission time constants of single defects, CET maps can be computed directly from the defects. For these computations the capture time constants during stress $\left(\tau_c^H\right)$ and the emission time constants during recovery $\left(\tau_e^L\right)$ are of importance, as they define where a defect contributes to CET maps. Furthermore, the actual average contribution of a defect to ΔV_{th} depends on its equilibrium occupancy difference a [18], and its step height (Fig. 5). With the equilibrium occupancy $f\left(V_G\right)$ the equilibrium occupancy difference evaluates to [18]

$$a = f(V_G^H) - f(V_G^L) = \frac{\tau_e^H}{\tau_e^H + \tau_c^H} - \frac{\tau_e^L}{\tau_e^L + \tau_c^L}.$$

CET maps can be directly extracted from the *experimental data* within the corresponding measurement window by taking the mixed partial derivative of $\Delta V_{th}\left(t_s, t_r\right)$ [20]. Therefore, a direct comparison between model and data is possible (Fig. 6), giving excellent agreement. Moreover, the activation energies of single defects can be calculated, as the temperature dependent rates are available. Fig. 7 shows an activation energy map which was computed similar to CET maps.

DISCUSSION

In our previous studies we have observed that while individual defects show a strong decrease of τ_c for increasing gate voltages, the distribution in CET maps remain aligned to the $\tau_c^H = \tau_e^L$ axis [18]. Our calibrated model now allows for a clarification of this puzzle: as shown in Fig. 5, the distribution in the CET map is mainly governed by the equilibrium occupancy difference, which broadens equally towards higher and lower capture times for higher stress voltages. This makes sense as with increasing stress voltage new defects are shifted into the 'active region' [18] in the band diagram (see Fig. 3). These defects can become charged now, but their τ_c are usually very large as these defects have low trap levels which typically corresponds to larger energy barrier for the capture process (see Fig. 1). Therefore, they contribute mostly to the upper part of the distribution in CET maps. Together with the commonly observed shift towards shorter τ_c for increasing stress voltages this results in a $\tau_c^H = \tau_e^L$ centered broadening of CET maps. Actually, the equilibrium occupancy difference determines the shape of the distribution as long as there are still defects whose trap levels are below the Fermi level. For very large stress voltages most defects do already contribute and an significant overall shift towards lower τ_c can be observed in CET maps. These measurable effects are correctly described by the NMP four-state model and are a further verification for its validity.

Another issue is that due to the overlap of the distributions, the recoverable and the permanent components are difficult to separate in conventional ΔV_{th} measurements, see Fig. 5. By using the 'accumulation pulse' measurements, which exploits the different bias dependences of the two defect types, the confidence in the extraction can be improved. Interestingly, probably because of the extended stress and recovery times, only little modification of the parameters is required to correctly reproduce also the pulsed data with the same parameter set.

CONCLUSIONS

We have demonstrated that the NMP four-state model for oxide defects combined with a simple double-well model for the more permanent part accurately reproduces all important characteristics of NBTI, starting from the nano-scale up to large-area devices. We have presented a link between the response of a large number of single defects and the directly accessible CET maps, demonstrating the usefulness of our approach. Most importantly, knowledge of these distributions allows to predict device-to-device variability in nano-scale devices.

ACKNOWLEDGEMENTS

This work has received funding from the the European Community's FP7 n°619234 and the Intel Sponsored Research Project n°2013111914.

REFERENCES

[1] D. Schroder and J. Babcock, "Negative Bias Temperature Instability: Road to Cross in Deep Submicron Silicon Semiconductor Manufacturing," *J.Appl.Phys.*, vol. 94, no. 1, pp. 1–18, 2003.

[2] K. Jeppson and C. Svensson, "Negative Bias Stress of MOS Devices at High Electric Fields and Degradation of MNOS Devices," *J.Appl.Phys.*, vol. 48, no. 5, pp. 2004–2014, 1977.

[3] M. Alam and S. Mahapatra, "A Comprehensive Model of pMOS NBTI Degradation," *Microelectronics Reliability*, vol. 45, no. 1, pp. 71–81, 2005.

[4] V. Huard, M. Denais, and C. Parthasarathy, "NBTI Degradation: From Physical Mechanisms to Modelling," *Microelectronics Reliability*, vol. 46, no. 1, pp. 1–23, 2006.

[5] H. Reisinger, O. Blank, W. Heinrigs, A. Mühlhoff, W. Gustin, and C. Schlünder, "Analysis of NBTI Degradation- and Recovery-Behavior Based on Ultra Fast V_{th}-Measurements," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, pp. 448–453, 2006.

[6] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel, "Understanding Negative Bias Temperature Instability in the Context of Hole Trapping," *Microelectronic Engineering*, vol. 86, no. 7-9, pp. 1876–1882, 2009.

[7] D. Ang, Z. Teo, T. Ho, and C. Ng, "Reassessing the Mechanisms of Negative-Bias Temperature Instability by Repetitive Stress/Relaxation Experiments," *IEEE Trans.Dev.Mat.Rel.*, vol. 11, no. 1, pp. 19–34, 2011.

[8] V. Huard, C. Parthasarathy, and M. Denais, "Single-Hole Detrapping Events in pMOSFETs NBTI Degradation," in *Proc. Intl.Integrated Reliability Workshop*, pp. 5–9, 2005.

[9] T. Wang, C.-T. Chan, C.-J. Tang, C.-W. Tsai, H. Wang, M.-H. Chi, and D. Tang, "A Novel Transient Characterization Technique to Investigate Trap Properties in HfSiON Gate Dielectric MOSFETs-From Single Electron Emission to PBTI Recovery Transient," *IEEE Trans.Electron Devices*, vol. 53, no. 5, pp. 1073–1079, 2006.

[10] J. Zou, R. Wang, N. Gong, R. Huang, X. Xu, J. Ou, C. Liu, J. Wang, J. Liu, J. Wu, S. Yu, P. Ren, H. Wu, S. Lee, and Y. Wang, "New Insights into AC RTN in Scaled High-κ/Metal-gate MOSFETs under Digital Circuit Operations," in *IEEE Symposium on VLSI Technology Digest of Technical Papers*, pp. 139–140, 2012.

[11] T. Grasser, K. Rott, H. Reisinger, P.-J. Wagner, W. Goes, F. Schanovsky, M. Waltl, M. Toledano-Luque, and B. Kaczer, "Advanced Characterization of Oxide Traps: The Dynamic Time-Dependent Defect Spectroscopy," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, pp. 2D.2.1–2D.2.7, Apr. 2013.

[12] M. Kirton and M. Uren, "Noise in Solid-State Microstructures: A New Perspective on Individual Defects, Interface States and Low-Frequency (1/f) Noise," *Adv.Phys.*, vol. 38, no. 4, pp. 367–486, 1989.

[13] H. Reisinger, T. Grasser, and C. Schlünder, "A Study of NBTI by the Statistical Analysis of the Properties of Individual Defects in pMOSFETs," in *Proc. Intl.Integrated Reliability Workshop*, pp. 30–35, 2009.

[14] T. Grasser, H. Reisinger, P.-J. Wagner, W. Goes, F. Schanovsky, and B. Kaczer, "The Time Dependent Defect Spectroscopy (TDDS) for the Characterization of the Bias Temperature Instability," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, pp. 16–25, May 2010.

[15] H. Reisinger, T. Grasser, K. Ermisch, H. Nielen, W. Gustin, and C. Schlünder, "Understanding and Modeling AC BTI," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, pp. 597–604, Apr. 2011.

[16] T. Grasser, P.-J. Wagner, H. Reisinger, T. Aichinger, G. Pobegen, M. Nelhiebel, and B. Kaczer, "Analytic Modeling of the Bias Temperature Instability Using Capture/Emission Time Maps," in *Proc. Intl.Electron Devices Meeting (IEDM)*, pp. 27.4.1–27.4.4, Dec. 2011.

[17] A. Stesmans, "Dissociation Kinetics of Hydrogen-Passivated P_b Defects at the (111)Si/SiOt Interface," *Physical Review B*, vol. 61, no. 12, pp. 8393–8403, 2000.

[18] T. Grasser, "Stochastic Charge Trapping in Oxides: From Random Telegraph Noise to Bias Temperature Instabilities," *Microelectronics Reliability*, vol. 52, pp. 39–70, 2012.

[19] Y. Yang and M. White, "Charge Retention of Scaled SONOS Nonvolatile Memory Devices at Elevated Temperatures," *Solid-State Electron.*, vol. 44, pp. 949–958, 2000.

[20] H. Reisinger, T. Grasser, W. Gustin, and C. Schlünder, "The Statistical Analysis of Individual Defects Constituting NBTI and its Implications for Modeling DC- and AC-Stress," in *Proc. Intl.Rel.Phys.Symp. (IRPS)*, pp. 7–15, May 2010.

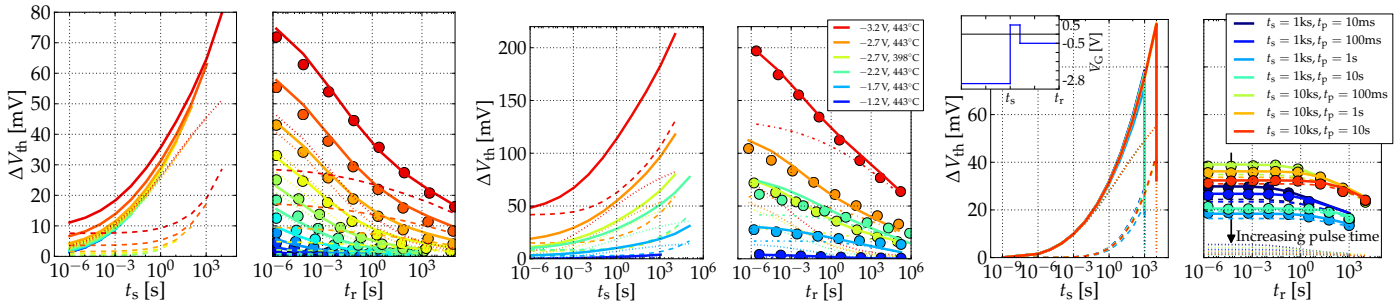

Fig. 4: Comparison of simulated threshold voltage shifts (solid lines) to the experimental data (dots) during stress (t_s) and recovery (t_r). The contribution of the NMP four-state defects (dotted lines) and of the defects described by the double-well model (dashed lines) are also depicted separately. No measurement data is available during the stress phases, but due to the short measurement delay of 1 μs and repeated stress/recovery cycles the degradation during stress is covered indirectly. **Left**: Details of an exemplary combination of stress voltage ($V_G = -2.7\,V$) and temperature ($T = 125°C$). The stress-recovery measurements and simulations were performed on the same device subsequently with increasing stress durations. However, for the sake of convenience they are plotted on top of each other in the stress and recovery part of the figure. As the ΔV_{th} of the permanent component does not recover substantially even after large recovery times there can be an offset at the start of the subsequent stress-recovery cycles. **Middle**: The last stress-recovery cycle for various stress voltages and temperatures. **Right**: Accumulation pulse setups for different stress times t_s and pulse durations t_p, all measured and simulated at $V_G = -2.8\,V$ and $T = 125°C$ with the gate voltage during the accumulation pulse $V_{G,p} = +0.5\,V$. Compared to the setup without an accumulation pulse the recoverable component (dotted lines) is substantially reduced while the permanent component (dashed lines) is less affected.

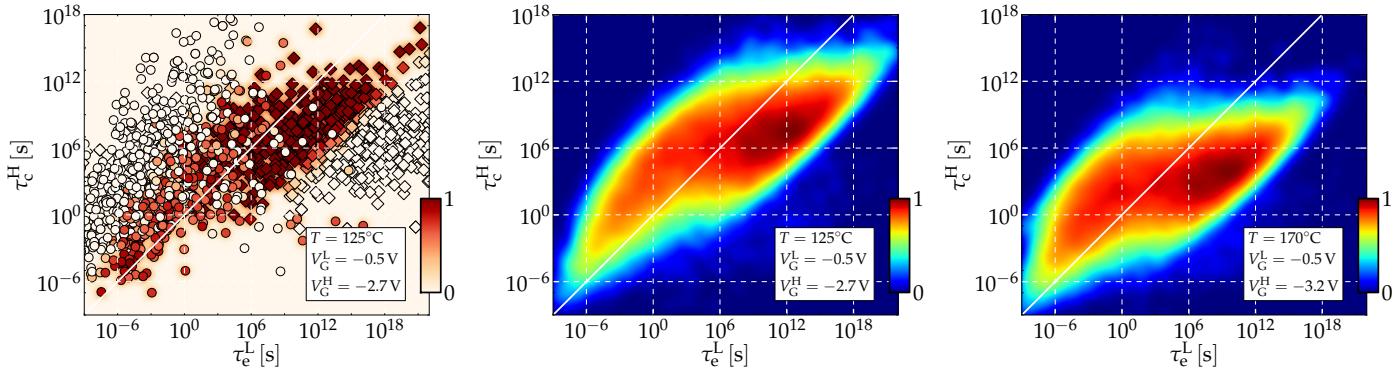

Fig. 5: Based on individual defects CET maps can be computed. τ_c^H and τ_e^L define where a defect contributes in CET maps and the average contribution of a defect to V_{th} is given by the equilibrium occupancy difference $\left(f\left(V_G^H\right) - f\left(V_G^L\right)\right)$ and the step height [18] which depends on the interface distance. **Left**: Randomly selected oxide (circle) and interface defects (diamonds) are plotted according to their τ_c^H and τ_e^L. An overlap of these defect types can be observed around $\tau_e^L \approx 10^6$. The separation of the recoverable and permanent component was improved during model calibration by using the accumulation pulse setup, which exploits the different bias dependency of the two defect types. The color indicates the equilibrium occupancy difference of the defects multiplied by their step height and normalized by the largest contribution. The equilibrium occupancy difference determines the shape of the distribution, as it is zero for many oxide defects with large τ_c^H at the given stress voltage. **Middle** and **right**: Computation of all defects yields CET maps for the corresponding stress condition. The trend towards shorter capture and emission time constants for increasing temperature and gate voltage is in accordance with the underlying ΔV_{th} measurements. Additionally the stress dependent broadening of the distribution due to the equilibrium occupancy difference towards larger τ_c can be observed.

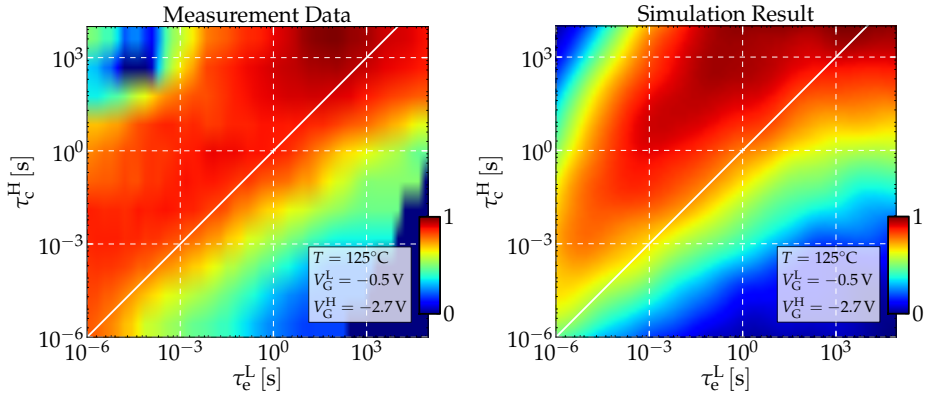

Fig. 6: Comparison of the measured (**left**) and simulated (**right**) CET maps within the measurement window. As expected, the important characteristics are reproduced: the center of the distribution is above the $\tau_c^H = \tau_e^L$ axis for short emission times and starts to cross the axis towards larger emission times.

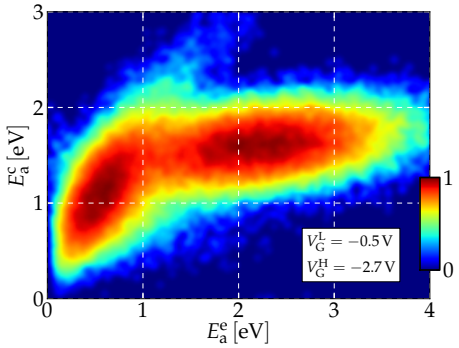

Fig. 7: Simulated activation energy map derived from individual defects similar as described for CET maps. The bivariate Gaussian distribution centered around $E_a^e = 0.5\,eV$ belongs to the NMP four-state model and the distribution with larger E_a^e belongs to the double well model.