Bistable Defects as the Cause for NBTI and RTN

W. Goes[1,a], F. Schanovsky[1,a], H. Reisinger[2,b], and B. Kaczer[3,c], T. Grasser[1,a]

[1]Insitute for Microelectronics, TU Wien, Austria
[2]Infineon Technologies AG, Munich, Germany
[3]IMEC, Leuven, Belgium

[a]{goes|schanovsky|grasser}@iue.tuwien.ac.at, [b]kaczer@imec.be, [c]hans.reisinger@infineon.com

Keywords: noise, negative bias temperature instability, charge trapping, defects

Abstract. Over the last few decades convincing evidence has been collected demonstrating that the oxide reliability is most seriously affected by hole trapping into defects. Recently, valuable information has been delivered by a newly developed measurement technique called time-dependent defect spectroscopy (TDDS), which allows to analyze the behavior of single defects. It indicates the existence of additional metastable defect configurations which are necessary to explain various features seen in TDDS. In this study, it will be shown that these bistable defects may also be the origin of noise phenomena, such as temporary and anomalous random telegraph noise observed in MOSFETs.

Introduction

As MOSFETs have been scaled into the nanometer regime, discrete fluctuations in the terminal currents have become increasingly important. This phenomenon is known as random telegraph noise (RTN) and has been intensively investigated over the years [1, 2, 3, 4, 5]. According to the current understanding, the origin of RTN lies in defects which are capable of exchanging charge carriers with the substrate via quantum mechanical tunneling. Even though this general picture is commonly accepted, the physical details of the underlying mechanism are still under debate.

Independently of the progress in this field, major advances have been made regarding the negative bias temperature instability (NBTI) [6, 7, 8, 9]. While in the past chemical reactions at the semiconductor-dielectric interface controlled by the diffusion of hydrogen have been made responsible for this phenomenon, the latest findings indicate that the device degradation is dominated by charge trapping into defects. It has therefore been suggested that the physical cause of RTN and NBTI can be ascribed to the same trapping mechanism. This speculation has been supported by a number of similarities in the defect properties [6, 10].

Recent Experimental Findings in NBTI research

Until lately, NBTI has primarily been studied by simply monitoring the degradation and recovery of large-area devices. Hence, the data contains the collective behavior of thousands of defects which obscures the details of the charge trapping process. However, the reduction in device dimensions has come to a point where single charging or discharging events appear as steps in the recorded recovery traces. Just like in RTN, the difficulty to assign a single trapping event to a certain defect, has hampered the analysis of experimental data [11]. Recently, a new measurement technique, called time-dependent defect spectroscopy (TDDS), has been suggested to overcome this problem [12, 13]. It exploits the fact that a trapped charge impacts the current percolation path of the drain current and thus the threshold voltage differently depending on where it is located on the gate area. As a consequence, the charging of a certain

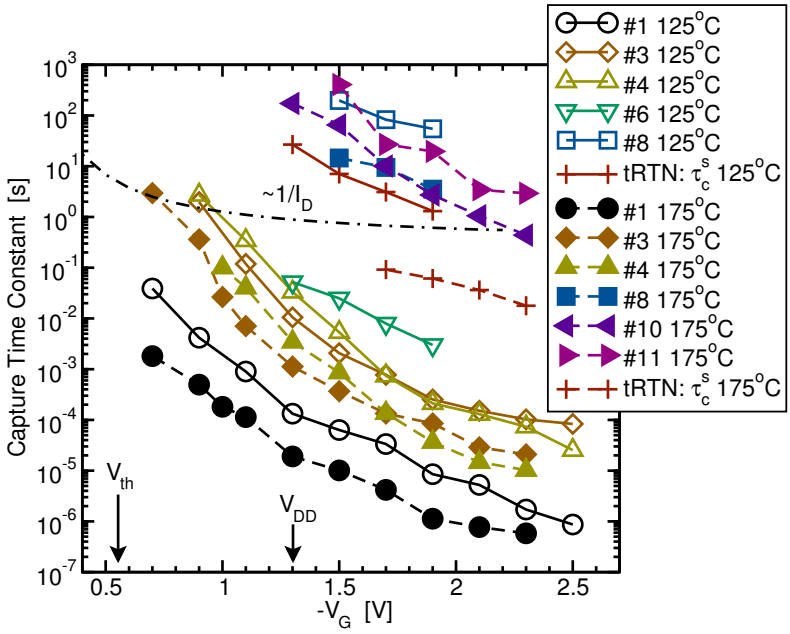

Fig. 1: Capture time constants τ_c as a function of V_G for a number of defects at different temperatures extracted from a single device. Open and closed symbols mark measurements conducted at 125 °C and 175 °C, respectively. The τ_c curves show a strong field acceleration and temperature activation. However, the observed field acceleration does not follow the $1/I_D \sim 1/p$ dependence (dot-dashed line) predicted by the conventional Shockley-Read-Hall (SRH) model.

defect always results in similar step heights. This fact can thus be used to relate a certain emission event to the corresponding defect. This method offers two substantial advantages:

- Provided that either the capture or the emission times or the step height of two defects differ, TDDS can capture a multitude of traps in a single measurement run. This fact immensely increases the versatility of the method, compared to the conventional RTN analysis.

- TDDS allows to study the "non-equilibrium" equilibrium response of defects. At the beginning of the relaxation phase, the system is perturbed by the removal of the gate bias. This triggers the occurrence of several discharging events, which are recorded and analyzed. In conventional RTN experiments, by contrast, the noise is measured at a constant gate bias, corresponding to equilibrium conditions. As a consequence, the capture and emission time constants have not to be within the same order of magnitude so that the measurement window of TDDS ranges from threshold to close to oxide breakdown and is thus considerably widened compared to RTN analysis.

TDDS has lead to several essential findings [13], which will be outlined in the following:

(*i*) The defects exhibit a strong, nearly exponential voltage dependence of τ_c as shown in Fig. 1 Empirically, it can be described by $\exp(-c_1 F_{ox} + c_2 F_{ox}^2)$. However, it differs from defect to defect, implying that it is related to certain defect properties.

(*ii*) The time constant plot 1 show a marked temperature dependence, which becomes obvious by the downward shift of the τ_c curves for higher temperatures. The activation energies E_a, extracted from Arrhenius plots are about $0.5-1.0$ eV. The investigated defects exhibit two distinct types of behavior with respect to τ_e (cf. Fig. 2):

(*ii*) τ_e of most defects is unaffected by changes in V_G ("normal" behavior).

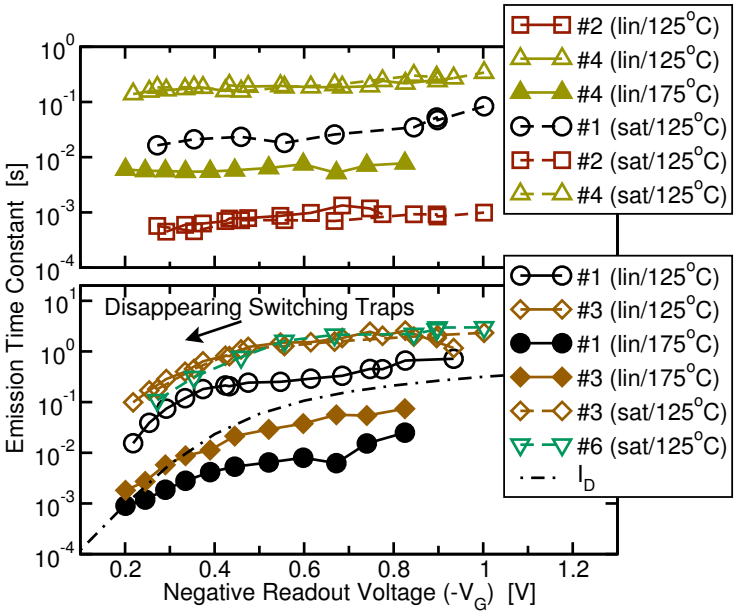

Fig. 2: Emission time constants τ_e as a function of V_G for various defects at different temperatures extracted from a single device. The two distinct types of field dependences (upper and lower panel) suggest the existence of two types of defects present in the oxide. However, defect #1 shows a different field dependence in the linear and the saturation regime. This implies that the electrostatics within the device are responsible for the two distinct field dependencies. It is noteworthy that the drop in τ_e goes hand in hand with the decrease of the interfacial hole concentration p.

(*iv*) Besides, there exists a few defects whose τ_e drops towards to lower V_G ("anomalous" behavior).

(*v*) The τ_e of both types shows a temperature activation with a large spread ($E_a = 0.6 - 1.4\,\mathrm{eV}$).

Interestingly, several TDDS recovery traces display RTN only after stress [13]. The noise at one recovery trace is physically linked to defects — in this case hole traps — which continuously exchange charge carriers with the substrate for a limited amount of time. After a while, the RTN signal vanishes and does not reappear during the remaining recovery time. The termination of the noise signal is ascribed to hole traps which change to their neutral charge state and remain therein. The associated time constants for this event are stochastically distributed. In [13], this kind of noise has been termed temporary RTN (tRTN) due to its limited lifetime.

A similar phenomenon called anomalous RTN (aRTN) has been discovered in the early studies of Kirton and Uren [2]. Therein, electron traps have been observed, which repeatedly produce noise for random time intervals. During the interruptions of the signal, the defects dwell in their negative charge state so that no noise signal is generated. The behavior of these traps has also been interpreted by the existence of a metastable defect state.

Physical Modeling

McWhorter Model. In the past, several models have been proposed to describe RTN in MOSFETs. Most of them are generalized variants of the Shockley-Read-Hall (SRH) model, which has been successfully applied to bulk traps. For instance, the McWhorter model [1] extends the SRH approach by a WKB factor in order to account for the tunneling effect. Unfortunately, it suffers from several shortcomings, which cannot be reconciled with the experimental findings: First, it predicts τ_c and τ_e to be temperature-independent in contrast to

observation (*ii*) and (*v*). Second, it predicts a $1/p$ dependence of τ_c in inversion. In Fig. 1 it is clearly demonstrated that this behavior is incompatible with point (*i*). Third, τ_e is predicted to exponentially depend on F_{ox}, which is neither the case for (*iii*) nor for (*iv*). Furthermore, no explanation can be offered for either anomalous or temporary RTN. Most importantly, the maximum time constants in the McWhorter model are limited by the oxide thickness. As such, it cannot explain time constants larger than 1 ms for devices with a 2 nm thick oxide. This is in contrast to experimental results, in which τ_e extends well into the kilosecond regime (cf. Fig. 1).

Kirton Model. Motivated by the theory of non-radiative multiphonon (NMP) processes [14], Kirton and Uren introduced bias-independent thermal barriers into the cross sections of τ_c and τ_e. This modification yields the required temperature activation and also allows for larger time constants. Despite these improvements, however, the time constant plots in Fig. 1 and 2 cannot be fitted. This is due to the fact that the slope of τ_c is determined by the trap depth x, which can only vary within a small range for modern thin gate dielectrics. Furthermore, the field dependences of capture and emission process are stringently coupled so that the Kirton model cannot reproduce the observed field dependence of τ_c and τ_e at the same time. Therefore, Schulz *et al.* [3] suggested that the charge carriers have to surmount a Coulomb barrier (CB) during the trapping process. Irrespective of the physical correctness of this assumption, the CB has only been applied successfully to MOSFETs with a thick oxides [5, 15]. Most importantly, the Kirton model can neither give an explanation for two different field dependences with respect to point (*iii*) and (*iv*) nor for the observed curvature of τ_c according to point (*i*).

NMP Model. The concept of NMP processes has further been pursued on a theoretically more profound level in the work of Palma *et al.* [4], who accounted for the fact that the thermal barriers are subject to a strong field dependence. This variant of the NMP model results in a linear dependence of τ_c on F_{ox}, which still cannot explain all experimental features, such as the curvature after point (*i*).

Advanced NMP Model. Apparently, the NMP process indeed forms the basis for charge trapping within a MOSFET. However, the NMP model must account for metastable defect configurations in order to explain the full range of experiments [12]. More precisely, the extended model relies on a type of trap, which is characterized by a bistability in both charge states. The configuration coordinate (CC) diagram of such a defect is illustrated in Fig. 3. Therein, the bistability is reflected in the double wells of the adiabatic potential energy curves. The charge exchange transfer between the substrate and the defects is described by the NMP transitions[4] $T_{1\leftrightarrow 2'}$ and $T_{2\leftrightarrow 1'}$ and strongly depends on the applied gate bias. As indicated in Fig. 3, V_G alters the relative position of the "neutral" and the "positive" adiabatic potential energy curves and thus determines their intersection point and their corresponding NMP barrier heights. For instance, an increased $|V_G|$ raises the neutral adiabatic potential energy curves and reduces the barrier for the charge transfer reaction $T_{1\rightarrow 2'}$. As a result, the transition probability for $T_{1\rightarrow 2'}$ is strongly enhanced. In contrast to the charge transfer reactions, the transitions $T_{1\leftrightarrow 1'}$ and $T_{2\leftrightarrow 2'}$ are purely thermally activated and therefore do not vary with the applied gate bias.

The multitude of transition possibilities results in quite complicated defect kinetics consistent with the defect behavior seen in RTN and TDDS measurements. The noise in conventional RTN studies is produced by defects switching back and forth between the states 1 and $2'$. These transitions are field and temperature dependent charge transfer reactions $T_{1\leftrightarrow 2'}$, which

[4]The abbreviation $T_{i\rightarrow j}$ denotes transitions from state i to j while $T_{i\leftrightarrow j}$ stands for the bidirectional analogs. Furthermore, $T_{i\rightarrow j\rightarrow k}$ stands for a chain of two transitions $T_{i\rightarrow j}$ and $T_{j\rightarrow k}$.

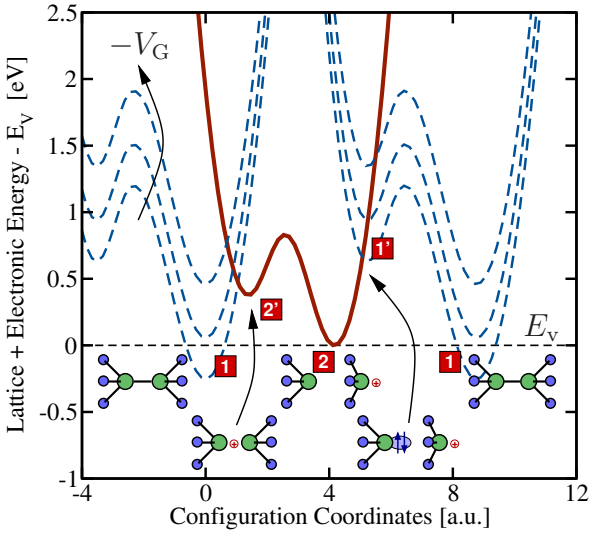

Fig. 3: Schematics of the configuration coordinate diagram for a bistable defect. The solid red and the blue dashed line represent the adiabatic potential energy curves for a defect in its positive and neutral charge state, respectively. The energy minima correspond to stable defect configurations and are labeled by numbers. Metastable states are marked by additional primes. Note that a change in the charge state of the defect is connected to a hole capture or emission event, respectively. The stick-and-ball models below display a defect in its various stable and metastable configurations. A possible candidate for such a bistable defect might be the well-known E' center frequently invoked in the context of noise in MOSFETs [16, 17].

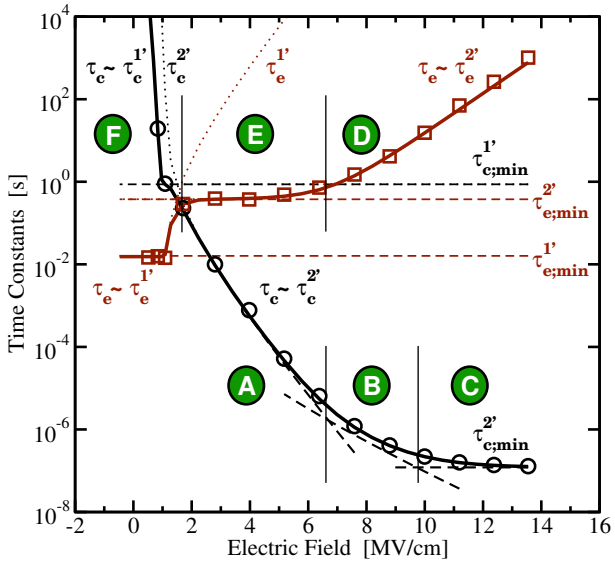

Fig. 4: Calculated hole capture and emission time constants as a function of the electric field in the oxide. The different regimes of τ_c (A, B, and C) and τ_e (D, E, and F) are labeled by the green circles with the capital letters. τ_c^i and τ_e^i denote the hole capture and emission over the intermediate state i, respectively.

are recognized as single steps in the RTN measurements. As already pointed out before, the TDDS time constant plots of Fig. 1 and 2 are not compatible with this simple picture. However, they can be explained by two consecutive transitions. Again, the defects undergo a field and temperature dependent charge transfer reaction $T_{1\to 2'}$, followed by a further thermal transition to the final state 2. Note that this two-step process requires the reverse barrier $\varepsilon_{2'1}$ to be small compared to $\varepsilon_{2'2}$.

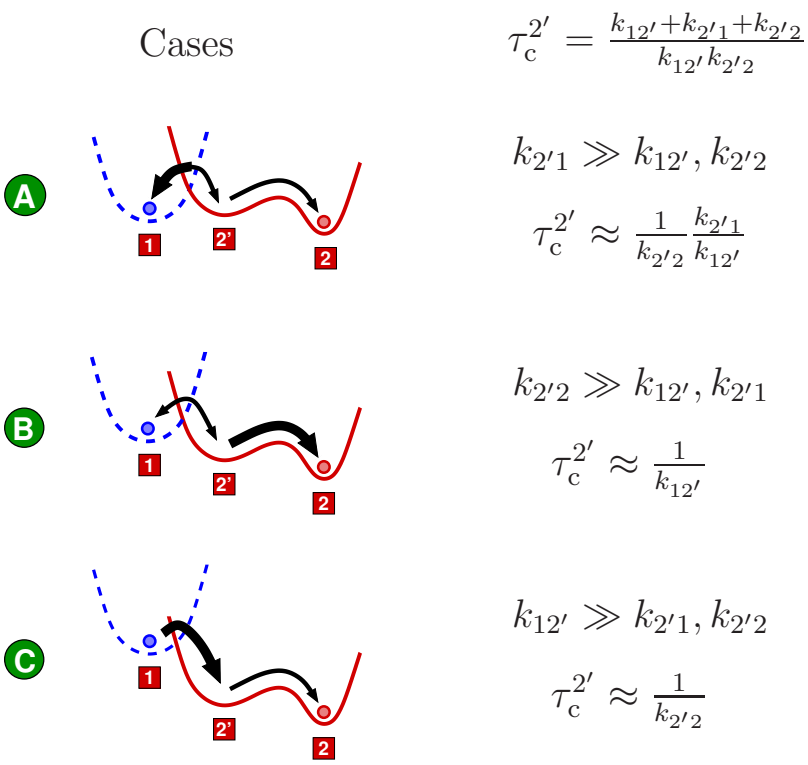

Fig. 5: Three regimes (A, B, and C) for the mean hole capture $\bar{\tau}_{2'}$ over the intermediate state $2'$. Its expression is derived from homogeneous continuous-time Markov chain theory using the concept of the first passage time.

The introduction of the intermediate state $2'$ yields a curvature in τ_c. An analytical expression for τ_c has been derived from homogeneous continuous-time Markov chain theory using the concept of first passage time [12]:

$$\tau_c \sim \tau_{c;min}^{2'} + \frac{N_2}{p}\exp\Big(-\frac{xR}{1+R}\frac{F_{ox}}{V_T}\Big) + \tau_{c;min}^{2'}\frac{N_1}{p}\exp\Big(-\frac{xF_{ox}}{V_T}\Big) \quad (1)$$

N_1 and N_2 are prefactors, V_T stands for the thermal voltage, R is the ratio of the vibrational frequencies at the minima of the neutral and the positive adiabatic potential energy curves, and $\tau_{c;min}^{2'}$ denotes the time constant for the transition $T_{2'\to 2}$. For extremely high F_{ox}, the first term in the above equation corresponds to the regime C of Fig. 4 and Fig. 5. There, the transition rate[5] $k_{12'}$ outbalances $k_{2'1}$ and $k_{2'2}$ so that the time constant $\tau_c^{2'}$ is governed by the slow transition $k_{2'2}$. For high F_{ox} (regime B), the second term of the above equation becomes dominant. In this regime the transition $T_{2'\to 2}$ proceeds much faster than $T_{1\to 2'}$ and $T_{2'\to 1}$. When a defect undergoes the transition from state 1 to $2'$, it immediately changes to state 2. As a consequence, the field dependence is given by the forward mode of the transition $T_{1\leftrightarrow 2'}$ alone. When F_{ox} is further reduced (regime A), the largest contribution to τ_c comes from the third term, which gives the steepest slope of τ_c. In this regime, $T_{2'\to 1}$ becomes dominant so that after a defect has overcome the large transition barrier $\varepsilon_{12'}$, it returns to state 1 with a high probability. Therefore, the forward as well as the reverse mode of the transition $T_{1\leftrightarrow 2'}$ enter the field dependence. The transitions between these three regimes are smooth so that a curvature appears in the time constant plots of τ_c, in agreement with point (i).

In practice, only little relevance is attached to τ_e at high electric fields. Nevertheless, it will be briefly discussed in the following as it provides valuable information on the microscopic defect physics. For extremely high F_{ox} (regime D), the field-dependent NMP barrier $\varepsilon_{2'1}$ is so

[5]Rates are denoted as k_{ij}, where the i and j correspond to the initial and the final state of the transition $T_{i\to j}$, respectively. The same applies to ε_{ij} and its associated barrier height.

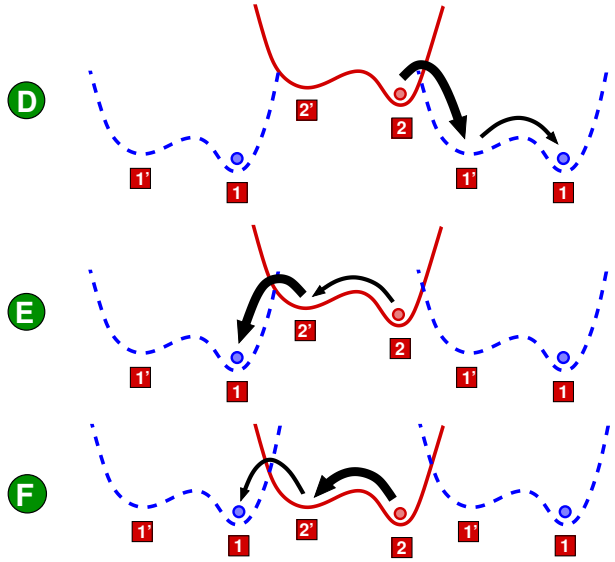

Fig. 6: The same as in Fig. 5 but for the three regimes (D, E, and F) of the mean hole emission time.

large that $T_{2'\to1}$ becomes the decisive factor and governs the field dependence of τ_e. Note that this part of the τ_e curve is only predicted by the new model but has not been experimentally assessed by the methods available. However, the experimentally relevant part of τ_e lies in regime E, where no field dependence can be observed. This is due to the fact that the NMP transition $T_{2'\to1}$ occurs on much smaller timescales than $T_{2\to2'}$ over the thermal barrier $\varepsilon_{22'}$. In regime F, the defect undergoes the transition $T_{2\to1'\to1}$ over the metastable state $1'$ due to a reduced NMP barrier $\varepsilon_{2'1}$. Since $T_{2\to1'}$ proceeds much faster than the field-independent transition $T_{1'\to1}$, the associated emission time constant in this regime does not depend on F_{ox}. Note that this does not hold for the transition regime which shows a strong field sensitivity caused by the field-dependent NMP transition $T_{2\to1'}$.

A comparison between simulation and measurement data is presented in Fig. 7. Good agreement has been achieved for both τ_c and τ_e. This is considered a strong indication that the extended NMP model correctly captures the field as well as the temperature dependence of τ_c and τ_e in agreement with points (*i*), (*ii*), (*iii*), and (*v*). This model is therefore also capable of explaining the RTN generated by the "normal" type of defects. In case the states 1 and 2 in the CC diagram are at the same energy level, the τ_c and τ_e are of the same order of magnitude and an RTN signal is produced by the transitions $T_{1\leftrightarrow2}$.

The extended NMP model also has an additional source of noise, which stems from defects switching back and forth between states 2 and $1'$. The associated charge transfer reactions $T_{2\leftrightarrow1'}$ do not involve any intermediate states and are therefore simple NMP processes. As such, they are described by conventional NMP theory. Note that the transitions $T_{2\leftrightarrow1'}$ require the energy minima of state 2 and $1'$ to be at approximately the same level when a small recovery voltage is applied to the gate (cf. insert of Fig. 9). Defects with this property are held responsible for tRTN. The investigated devices are stressed at a high V_G so that the defects are forced from state 1 into one of the states 2 or $1'$. During this step, the defects undergo the transition $T_{1\to2}$ over the intermediate state $2'$. The other pathway $T_{1\to1'}$ is assumed to go over a large barrier $\varepsilon_{11'}$. Therefore the transition $T_{1\to1'}$ proceeds on much larger timescales than $T_{1\to2'\to2}$ and can therefore be neglected. After stressing, the recovery traces are monitored at a low V_G or F_{ox}, respectively, at which the energy minima 2 and $1'$ more or less coincide and noise is produced. However, state 1 is thermodynamically preferred due to its lower energetical position compared to 2 and $1'$, respectively. When the defect returns to its initial state 1, the RTN signal disappears

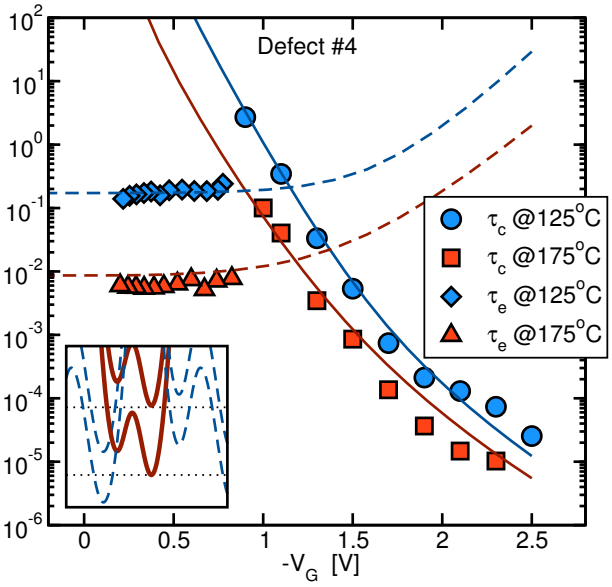

Fig. 7: Capture (solid lines) and emission (dashed lines) times as a function of the gate bias for a defect producing normal RTN. The symbols stand for the measurement data and the lines show the simulation results of our improved NMP model. The simulated time constants are in remarkable agreement with the experimental data. Most notably, the simulations correctly reproduce the field and temperature dependence of τ_c and τ_e. The insert (bottom left) shows that the energy minima $2'$ and 1 differ by at least a few tenth of an electron Volt. This eventually characterizes this defect as a trap producing "normal" RTN.

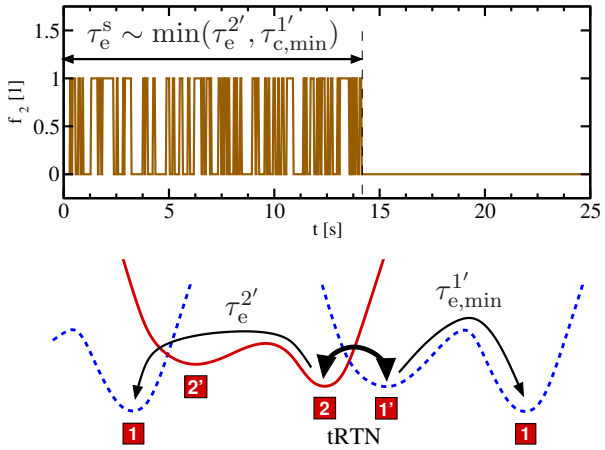

Fig. 8: **Top**: Hole occupancy during tRTN. At $t = 0$ the stress voltage has been removed and the defect is in its positive state 2. After a time τ_e^s the defect ceases to produce noise. **Bottom**: Configuration coordinate diagram for a tRTN defect. The thick arrow indicates the fast switches between the states 2 and $1'$ related to the occurrence of noise. The possibilities to escape from these states are shown by the thin arrows.

with the time constant τ_e^s. The corresponding transition could be either $T_{2\to 2'\to 1}$ or $T_{1'\to 1}$ (cf. Fig. 8). The latter is only the case for a small group of defects which are characterized by a strong sensitivity of their τ_e at low voltages (cf. Fig. 9). Their thermal barrier $\varepsilon_{1'1}$ falls below $\varepsilon_{22'}$ so that they escape out of state 2 over the metastable state $1'$. However, this pathway is hampered by the slow transition $T_{2\to 1'}$ at high F_{ox}. Therefore, a reduction of the electric field facilitates the charge transfer reaction $T_{2\to 1'}$ and allows $T_{2\to 1'\to 1}$ (regime F in Fig. 4). This coincides with the point at which τ_e in Fig. 4 drops from $\tau_e^{2'}$ to $\tau_e^{1'}$. This field dependence can thus be used as a fingerprint of the "anomalous" defects.

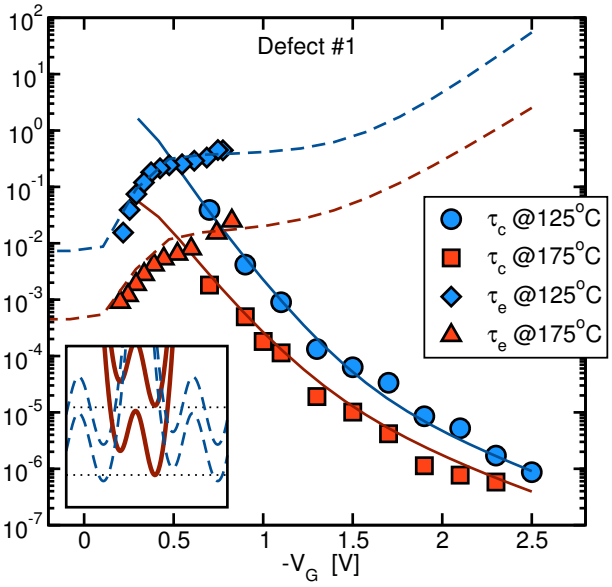

Fig. 9: The same as for Fig. 7 but for an "anomalous" defect. Compared to the defect #4 in Fig. 7, the present defect (#1) shows a strong voltage/field dependence of τ_e at low V_G or F_{ox}.

One has to consider that single trapping events between the states 2 and 1 can only be detected when the separating NMP barriers $\varepsilon_{21'}$ and $\varepsilon_{1'2}$ are sufficiently large. If this is not the case, the measurement equipment cannot resolve the single transitions and thus only records averaged occupation values of the state 2 or 1′, respectively. Furthermore, the NMP barriers must not be too large since otherwise trapping events will occur outside the time window of measurements.

The extended NMP model is also capable of explaining aRTN as will be demonstrated in the following. [6] Just as in the case of tRTN, the noise signal is generated by charge transfer reactions between the states 2 and 1′. This implies that the same requirements for the NMP barriers $\varepsilon_{21'}$ and $\varepsilon_{1'2}$ must hold as before. The recurring pauses of the noise signal (see Fig. 10) originate from transitions into the metastable state 2′, which is electrically indistinguishable from state 2. They correspond to the time during which the defect dwells in this state and no charge transfer reaction can take place. Thereby it has been implicitly assumed that the NMP transition $T_{2'\to 1}$ occurs on larger time scales than the return to the state 2 through the transition $T_{2'\to 2}$. The slow capture time constant τ_c^s in Fig. 10 defines the mean time interval during which noise is observed. Its value is given by the inverse of the transition rate $k_{22'}$. The slow emission time constant τ_e^s ends the noiseless periods and is determined by $1/k_{2'2}$. One has to consider that, in principle, aRTN also appears for defects with only one neutral state. However, such a defect could not explain the possible occurrence of tRTN and would thus restrict the generality of the model.

Conclusion

We argue that two well-known reliability issues, namely NBTI and RTN, are just two facets of the same physical trapping process. The underlying mechanism has been described in an extended NMP model using bistable defect configurations. This new model explains the various findings of recent TDDS measurements as well as the puzzling phenomena of anomalous and temporary RTN.

[6] Even though the extended NMP model is explained for a hole trap, the same concept can also be employed for electron traps observed by Kirton/Uren [2].

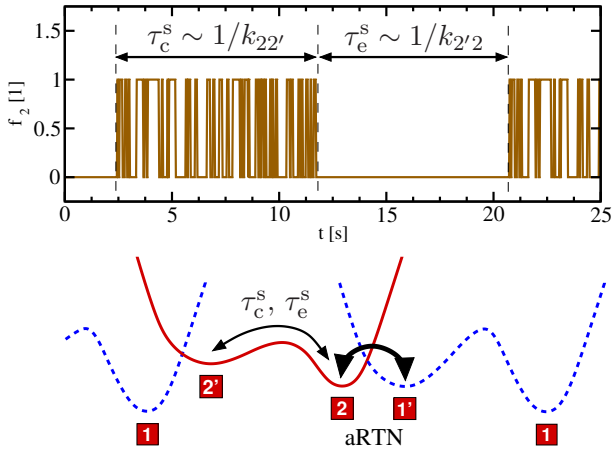

Fig. 10: **Top**: Hole occupancy during aRTN. **Bottom**: Configuration coordinate diagram for an aRTN defect. Since this defect is a hole trap, the red solid and the blue dashed line correspond to the positive and neutral charge state, respectively. The double-sided thick arrow is associated with tRTN while the thin one represents the transitions to and from the metastable state $2'$.

Acknowledgment

This work has received funding from the MORDRED project (EU Project NMP3-SL-2011-261868).

References

[1] A. McWhorter, Sem.Surf.Phys 207 (1957).

[2] M. Kirton and M. Uren, Adv.Phys. **38**, 367 (1989).

[3] M. Schulz, J.Appl.Phys. **74**, 2649 (1993).

[4] A. Palma *et al.*, Phys.Rev.B **56**, 9565 (1997).

[5] M. Lu and M. Chen, Phys.Rev.B **72**, 235417 (2005).

[6] B. Kaczer *et al.*, in *Proc.IRPS* (2009), pp. 55–60.

[7] H. Reisinger, T. Grasser, W. Gustin, and C. Schlünder, in *Proc.IRPS* (2010), pp. 7–15.

[8] T. Grasser *et al.*, in *Proc.IEDM* (2010), pp. 82–85.

[9] M. Alam, H. Kufluoglu, D. Varghese, and S. Mahapatra, Microelectron.Reliab. **47**, 853 (2007).

[10] B. Kaczer *et al.*, in *Proc.IRPS* (2010), pp. 1095–1098.

[11] T. Nagumo *et al.*, in *Proc.IEDM* (2009), pp. 759–762.

[12] T. Grasser *et al.*, in *Proc.IRPS* (2010), pp. 16 –25.

[13] T. Grasser, H. Reisinger, P.-J. Wagner, and B. Kaczer, Phys.Rev.B **82**, 245318 (2010).

[14] W. Fowler, J. Rudra, M. Zvanut, and F. Feigl, Phys.Rev.B **41**, 8313 (1990).

[15] T. Grasser *et al.*, in *Proc.IEDM* (2009).

[16] A. Lelis, T. Oldham, H. Boesch, and F. McLean, IEEE Trans.Nucl.Sci. **36**, 1808 (1989).

[17] D. Fleetwood *et al.*, IEEE Trans.Nucl.Sci. **49**, 2674 (2002).