Advanced Modeling of Oxide Defects for Random Telegraph Noise

W. Goes*, F. Schanovsky*, T. Grasser*, H. Reisinger°, and B. Kaczer†

* Insitute for Microelectronics, TU Wien, Austria ° Infineon Technologies AG, Munich, Germany † IMEC, Leuven, Belgium

***Abstract*— The results from a recently developed measurement technique, called time-dependent defect spectroscopy (TDDS), have shed new light on reliability issues, such as random telegraph noise (RTN) and the negative bias instability (NBTI). It has been found that established models fail to explain these findings. A refined charge trapping model is suggested by assuming additional metastable defect configurations. Thereby, we can give an explanation for the new TDDS findings while remaining consistent with results obtained from conventional RTN analysis.**

I. 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 [1–5]. 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) [6–9]. 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 is natural to assume that the physical cause of both phenomena can be ascribed to the same trapping mechanism. This speculation has been supported by a number of similarities in the defect properties [6]. In this paper, a new promising model [10] proposed in the context of NBTI will be investigated in the light of RTN.

II. 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 physical process of charge trapping. 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. It rests on the observation that each step in a recovery trace can be identified with one trap based on its height. 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.
- Since the capture (τ_c) and emission (τ_e) time constants are not extracted from the noise signal as in typical RTN measurements, they do not have to be within the same order of magnitude. As a result, the measurement window is considerably wider, ranging from threshold to close to the oxide breakdown.

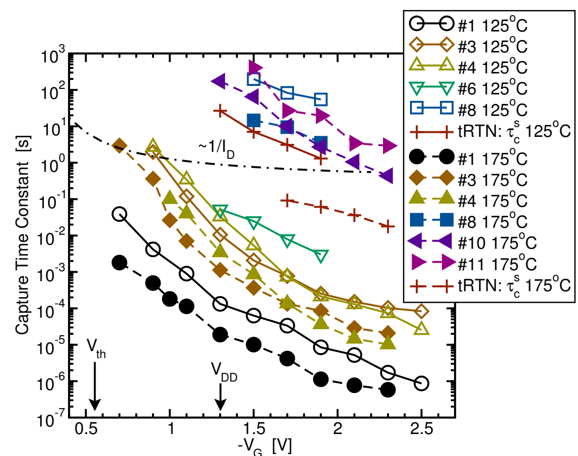

Fig. 1. Capture time constants τ_c as a function of V_G for a number of defects at different temperatures extracted from a single device. Open and closed symbols marks measurement conducted at 125 °C and 175 °C, respectively. The τ_c curves show a strong field acceleration and temperature activation.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

TDDS has lead to several essential findings [12], which will be outlined in the following. (*i*) The plot in Fig. 1 reveals that the defects exhibit a strong, nearly exponential voltage dependence of τ_c. Empirically, it can be described by $\exp(-c_1 F_\mathrm{ox} + c_2 F_\mathrm{ox}^2)$. However, it differs from defect to defect, implying that it is related to certain defect properties. (*ii*) The time constant plots 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 $0.6\,\mathrm{eV}$. The investigated defects exhibit two distinct behaviors with respect to τ_e (cf. Fig. 2): (*iii*) While a number of defects are unaffected by changes in V_G ("normal" behavior), (*iv*) others exhibit a marked sensitivity to V_G ("anomalous" behavior). (*v*) The τ_e of both types shows a temperature activation with a large spread ($E_\mathrm{a} = 0.6 - 1.4\,\mathrm{eV}$).

Astonishingly, several TDDS recovery traces display RTN only after stressing [12]. The noise at one recovery trace is physically linked to defects — in this case hole traps — which continuously exchange charge carriers with the substrate. After a while, the RTN signal vanishes and does not reoccur during the remaining measurement 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 [12], 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

978-1-4577-0191-7/11/$26.00 ©2011 IEEE

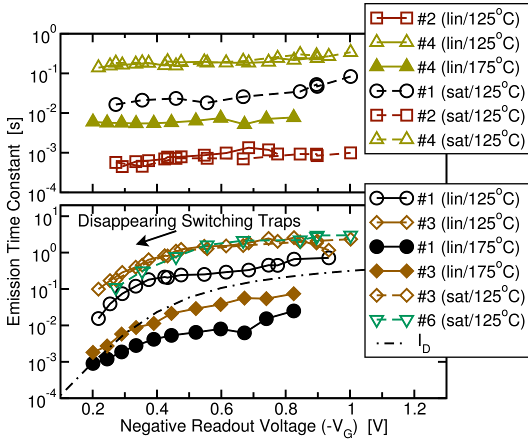

Fig. 2. Emission time constants τ_e as a function of V_G for various defects at different temperatures extracted from a single device. The two distinct field dependences (upper and lower panel) suggest the existence of two types of defects present in the oxide. However, the defect #1 shows different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

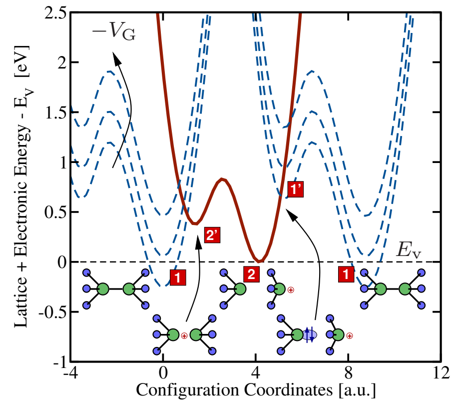

Fig. 3. Schematics of the configuration coordinate diagram for a bistable defect. The solid red and the blue dashed line represent the adiabatic potential energy curves for a defect in its positive and neutral charge state, respectively. The energy minima correspond to stable defect configurations and 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 [15, 16].

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.

III. Physical Modeling

A. 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, the field dependence of τ_c goes with $1/p$ 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 $1\,\mathrm{ms}$ for devices with a $2\,\mathrm{nm}$ thick oxide. This is in contrast to experimental results, in which τ_e extends well into the kilosecond regime (cf. Fig. 1).

B. Kirton Model

Motivated by the theory of non-radiative multiphonon (NMP) processes [13], 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, but also allows larger time constants. Despite these improvements, 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, 14]. 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*).

C. 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*).

D. 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 [10]. 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[1] $T_{1\leftrightarrow 2'}$ 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 $|V_\mathrm{G}|$ 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

[1] 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 $T_{j\rightarrow k}$.

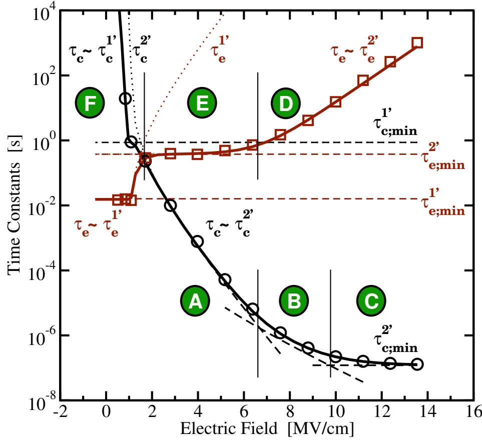

Fig. 4. Calculated hole capture and emission time constants as a function of the electric field in the oxide. The different regimes of τ_c (A, B, and C) and τ_e (D, E, and F) are labeled by the green circles with the capital letters. τ_c^i and τ_e^i denote the hole capture and emission over the intermediate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

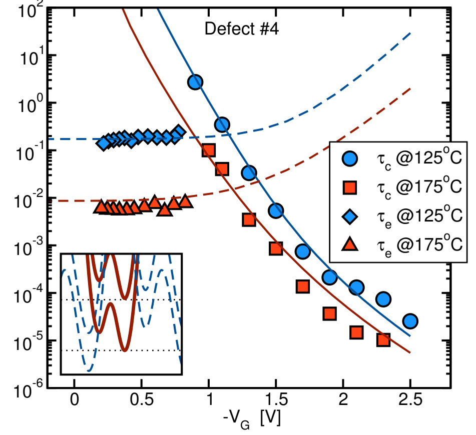

Fig. 5. Capture (solid lines) and emission (dashed lines) times as a function of the gate bias for a defect producing normal RTN. The symbols stand for the measurement data and the lines show the simulation results of our improved NMP model. The simulated time constants are in remarkable agreement with 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.

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 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 $T_{1\rightarrow 2'}$, 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 $\varepsilon_{2'2}$.

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 [10]:

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

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 $T_{2'\rightarrow 2}$. For extremely high F_{ox}, the first term in the above equation corresponds to the regime C of Fig. 4. There, the transition rate[2] $k_{12'}$ 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 exponential term of the second term becomes dominant, resulting in a weak exponential field dependence. 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. 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*).

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

In practice, only little relevance is attached to the τ_e at high electric fields. Nevertheless, it will be shortly discussed in the following. 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 transition $T_{2'\rightarrow 1}$ occurs on much smaller timescales than $T_{2\rightarrow 2'}$ over the thermal barrier $\varepsilon_{22'}$. For high F_{ox} (regime D), the field-dependent NMP barrier $\varepsilon_{2'1}$ becomes large so that $T_{2'\rightarrow 1}$ becomes the decisive factor and governs the field dependence of τ_e.

A comparison between simulation and measurement data is presented in Fig. 5. Good agreement has been achieved for both τ_c and τ_e. This is considered as 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 the 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 $T_{1\leftrightarrow 2}$.

The extended NMP model also has an additional source of noise, which stems from defects switching forth and back between states 2 and $1'$. The associated charge transfer reaction $T_{2\leftrightarrow 1'}$ does not involve any intermediate states and are therefore simple NMP processes. As such, they are described by the conventional NMP theory. Note that the transitions $T_{2\leftrightarrow 1'}$ require the energy minima of state 2 and $1'$ to be on approximately the same level, i.e. a small recovery voltage is applied to the gate (cf. insert of Fig. 7). However, this is only the case for a group of defects whose energy minima 1 and $1'$ are energetically not far separated.

The noise generated by those defects is held responsible for tRTN. In TDDS, 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 $T_{1\rightarrow 2}$ over the intermediate state $2'$. The other pathway $T_{1\rightarrow 1'}$ is assumed to go over a large barrier $\varepsilon_{11'}$. Therefore the transition $T_{1\rightarrow 1'}$ proceeds on much larger timescales than $T_{1\rightarrow 2'\rightarrow 2}$ 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

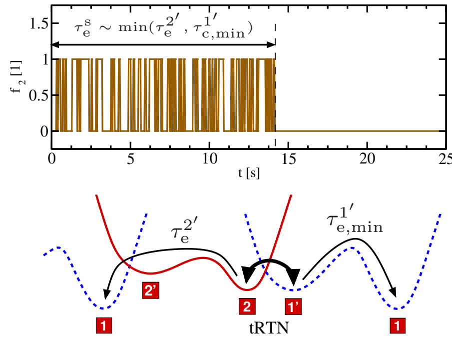

Fig. 6. **Top**: Hole occupancy during tRTN. At $t = 0$ the stress voltage has been removed and the defect is in its positive state 2. After a time $\tau_\mathrm{e}^\mathrm{s}$ the defect ceases to produce noise. **Bottom**: Configuration coordinate diagram for a tRTN defect. The thick arrow indicates 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.

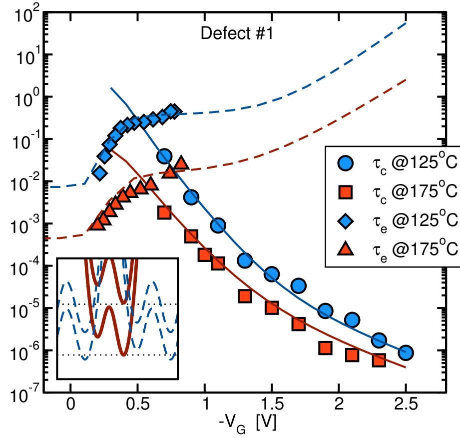

Fig. 7. The same as for Fig. 5 but for an "anomalous" defect. Compared to the defect #4 in Fig. 5, the present defect (#1) shows a strong voltage/field dependence of τ_e at low V_G or F_ox.

$1'$ 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 with the time constant $\tau_\mathrm{e}^\mathrm{s}$. The corresponding transition could be either $T_{2\to 2'\to 1}$ or $T_{1'\to 1}$ (cf. Fig. 6). 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. 7). 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 $\tau_\mathrm{e}^{2'}$ to $\tau_\mathrm{e}^{1'}$. This field dependence can thus be used as a fingerprint of the "anomalous" defects.

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. Just as in the case of tRTN,

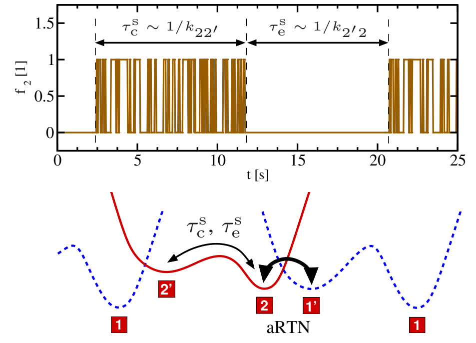

Fig. 8. **Top**: Hole occupancy during aRTN. **Bottom**: Configuration coordinate diagram for an aRTN defect. Since this defect is a hole trap, the red solid and the blue dashed line correspond to the positive and neutral charge state, respectively. The double-sided thick arrow is associated with tRTN while the thin one represents the transitions to and from the metastable state $2'$.

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. 8) 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 $\tau_\mathrm{c}^\mathrm{s}$ in Fig. 8 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 $\tau_\mathrm{e}^\mathrm{s}$ ends the noiseless periods and is determined by $1/k_{2'2}$. 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.

IV. 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.

Acknowledgment

This work has received funding from the ENIAC MODERN project n° 820379.

References

[1] A. McWhorter, Sem.Surf.Phys 207 (1957).
[2] M. Kirton *et al.*, 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 *et al.*, Phys.Rev.B **72**, 235417 (2005).
[6] B. Kaczer *et al.*, *Proc.IRPS* (2009), pp. 55–60.
[7] H. Reisinger *et al.*, *Proc.IRPS* (2010), pp. 7–15.
[8] T. Grasser *et al.*, *Proc.IEDM* (2010), pp. 82–85, (invited).
[9] M. Alam *et al.*, Microelectron.Reliab. **47**, 853 (2007).
[10] T. Grasser *et al.*, *Proc.IRPS* (2010), pp. 16 –25.
[11] T. Nagumo *et al.*, *Proc.IEDM* (2009), pp. 759–762.
[12] T. Grasser *et al.*, Phys.Rev.B **82**, 245318 (2010).
[13] W. Fowler *et al.*, Phys.Rev.B **41**, 8313 (1990).
[14] T. Grasser *et al.*, *Proc.IEDM* (2009).
[15] A. Lelis *et al.*, IEEE Trans.Nucl.Sci. **36**, 1808 (1989).
[16] D. Fleetwood *et al.*, IEEE Trans.Nucl.Sci. **49**, 2674 (2002).