Temperature dependence of the emission and capture times of SiON individual traps after positive bias temperature stress

M. Toledano-Luque[a)]
IMEC, Kapeldreef 75, B-3001 Leuven, Belgium and Departamento Física Aplicada III, Universidad Complutense de Madrid, Ciudad Universitaria, E-28040 Madrid, Spain

B. Kaczer, Ph. Roussel, and M. J. Cho
IMEC, Kapeldreef 75, B-3001 Leuven, Belgium

T. Grasser
Institute for Microelectronics, TU Wien, Austria

G. Groeseneken
IMEC, Kapeldreef 75, B-3001 Leuven, Belgium and Department of ESAT, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium

(Received 10 August 2010; accepted 6 December 2010; published 25 January 2011)

The authors study the statistical properties of individual defects in *n*-type metal-oxide-semiconductor field-effect transistors (nMOSFETs) using time dependent defect spectroscopy. This technique is based on the analysis of quantized threshold voltage transients observed on nanoscaled *p*-type metal-oxide-semiconductor field-effect transistors (pMOSFETs) after negative stress and provides the characteristic emission and capture times of individual traps. To complement to previous studies, the authors apply the methodology to SiON nMOSFETs and positive bias temperature stress. The authors demonstrate that the relaxation transients are due to the collective behavior of individual traps. Furthermore, a strong temperature dependence is observed for both emission and capture times. This is incompatible with elastic tunneling theory which is used in trap characterization techniques such as charge pumping, and also in simulations of erase and program transients of nonvolatile memories. The calculated thermal activation energies for both times are in the order of 0.6 eV and are close to the values obtained for SiON pMOSFETs when negatively stressed. © *2011 American Vacuum Society.* [DOI: 10.1116/1.3532947]

I. INTRODUCTION

As the dimensions of the metal-oxide-semiconductor field-effect transistors shrink toward the nanometer scale, the collective behavior of individual defects is discretized. One of the most evident effects is the random telegraph noise (RTN), i.e., the fluctuation in the drain current caused by the charge and discharge of individual traps. This effect has fascinated device physicists since it has been reported for the first time in 1984.[1] Recently, RTN has also caught the attention of reliability engineers because of the following: (1) RTN can affect the operation of flash memories[2] and (2) RTN has been linked to the temporal component of negative bias temperature instability (NBTI).[3,4] This RTN to NBTI correlation has been established after a thorough study of the relaxation transients observed on nanoscaled *p*-type metal-oxide-semiconductor field-effect transistors (pMOSFETs) following negative bias temperature stress (NBTS). In this work, we focus on ($L\times W = 70\times 90$ nm^2) nanoscaled SiON *n*-type metal-oxide-semiconductor field-effect transistors (nMOSFETs) after positive gate bias stress in order to confirm the relation between the relaxation curves after positive gate bias temperature stress (PBTS) and RTN.

In this article, we also study the trapping and detrapping mechanisms as a function of the temperature. In literature, elastic tunneling is frequently used to model the erase and program transients in nonvolatile memories[5] and the depth localization of gate oxide traps.[6,7] To bring insight about the correctness of the model, the study of the trapping and detrapping mechanisms as a function of the temperature is compulsory since elastic tunneling predicts a weak dependence with this parameter.

This article is organized as follows. In Sec. II, we show experimental evidence of the link between RTN and positive bias temperature instability (PBTI) from the study of relaxation curves obtained on multiple devices. In Sec. III, we focus on a particular device and the methodology to determine the emission and capture times of a single trap is explained. In Sec. IV, the temperature dependence of these two characteristic times is determined. Finally, the main ideas of the article are wrapped up in Sec. V.

II. PBTI RELAXATION: NONEQUILIBRIUM CASE OF RTN

From the quantized recovery behavior observed in nanoscaled pMOSFETs following negative gate bias temperature stress, the recoverable component of NBTI was explained as dynamic nonequilibrium of RTN.[8] This quantized recovery behavior is also observed after positive bias stress in nMOSFETs. Figure 1 shows the recovery of the threshold voltage ΔV_{TH} following positive stress in five 70×90 nm^2 nMOS-

[a)]Electronic mail: toleda@imec.be

Author complimentary copy. Redistribution subject to AIP license or copyright, see http://jvb.aip.org/jvb/copyright.jsp

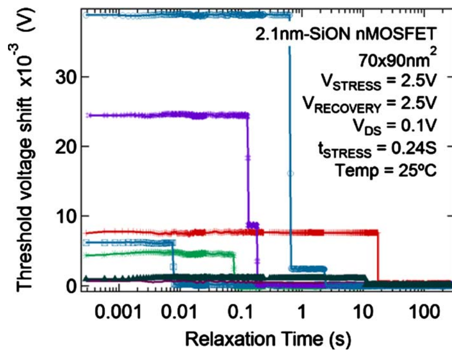

FIG. 1. (Color online) V_{TH} transients of multiple nanoscaled (70×90 nm^2) 2.1 nm SiON nMOSFETs stressed at 25 °C and 2.8 V for 241 ms. Every defect presents its characteristic emission time and step height ΔV_{TH}. Note that the step heights of some defects exceed 10 mV, and the emission times can reach 10 s.

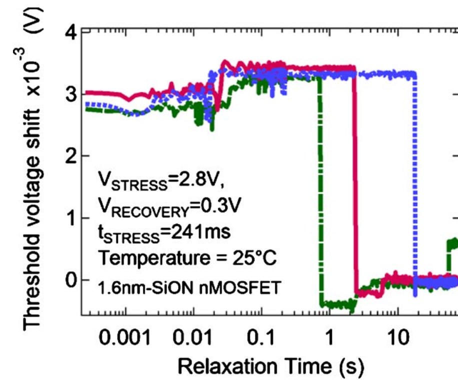

FIG. 3. (Color online) Three V_{TH} transients of a single 1.6 nm SiON nMOSFET stressed at 25 °C and 2.8 V for 241 ms. The V_{TH} transients show a discrete step behavior due to an electron emission from an individual trap.

FETs with 2.1 nm SiON. The procedure to obtain these curves is based on the fast V_{TH}-evaluation methods developed for bias temperature instability.[9] In contrast to pMOSFETs, the discrete V_{TH} drops are due to the electron emission from individual traps. The large step heights are caused by the random dopant fluctuations in the nMOSFET channel[2] that generate different conduction paths between drain and source. These paths can be modified by the charge and discharge of traps, thus modifying the drain current I_D. The change of drain current can be in turn transformed into a V_{TH} shift when taking the $I_D - V_G$ curve of the fresh device as a reference.[6]

The five relaxation curves plotted in Fig. 1 were selected from 295 nMOSFETs after positive bias temperature stress. Under the stress condition of the experiment ($V_{stress} = 2.5$ V, $t_{stress} = 0.24$ s), 37 out of the 295 devices presented discrete drops of the threshold voltage larger than 3 mV. As we can see in Fig. 2, every device presents its characteristic relaxation trace, and several devices show a step down exceeding 10 mV. The statistical properties of the step heights and emission times will be presented in a separate paper. However, it is more important to note that the average of the

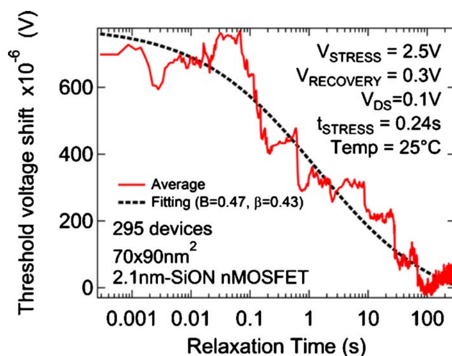

FIG. 2. (Color online) Averaged relaxation traces of 295 (70×90 nm^2) 2.1 nm SiON nMOSFETs after positive bias temperature stress. The obtained curve resembles the recovery behavior of a large device (Ref. 10), demonstrating the link between detrapping from individual traps and the recovery component of PBTI.

relaxation traces obtained in multiple nanoscaled devices resembles the relaxation curve of large devices. Figure 2 shows the averaged relaxation curve of the 295 devices along with a fit with the generalized relaxation curve $1/[1 + B(t_{relax}/t_{stress})^{\beta}]$ for a large device.[7] The relaxation curves obtained from large devices (1×1 μm^2) of the same wafer yield similar shape as the averaged transients obtained from nanoscaled ones (not shown). These facts confirm the close link between the charge and discharge of traps and the recovery behavior observed after positive bias stress. As in the NBTI case, the collective behavior of multiple individual traps is responsible for the recovery component after PBTI.

III. EMISSION AND CAPTURE TIMES AT ROOM TEMPERATURE

Based on the quantized transient observed on nanoscaled devices after stress, Grasser *el al.*[11] developed a new methodology to analyze the emission and capture kinetics of individual defects. This technique, named time dependent defect spectroscopy (TDDS), was applied to pMOSFETs. In this section, we make use of this technique to study an individual defect in an nMOSFET to prove that the emission and capture times follow the same kinetics as the defects in pMOSFETs. For completeness, the procedure to determine the emission and capture times is described.

Figure 3 shows three typical V_{TH} transients after stressing a single 70×90 nm^2 1.6 nm SiON nMOSFET device at 25 °C and 2.8 V for 241 ms (i.e., 0.5 nm thinner than the physical thickness of the SiON layer used in Sec. V). Remarkably, similar curves are obtained when the experiment is repeated on the same device. An abrupt V_{TH} shift of about 3 mV is observed for the three traces at the start of the relaxation period, indicating the presence of a single active trap. To simplify the analysis, a device with only one trap active under the conditions of the experiment was chosen. In Fig. 3, we can also observe that the emission times range from ~ 1 to ~ 10 s. Since the emission of the electron is a stochastic process, a large number of traces have to be recorded for a reliable characterization.

Author complimentary copy. Redistribution subject to AIP license or copyright, see http://jvb.aip.org/jvb/copyright.jsp

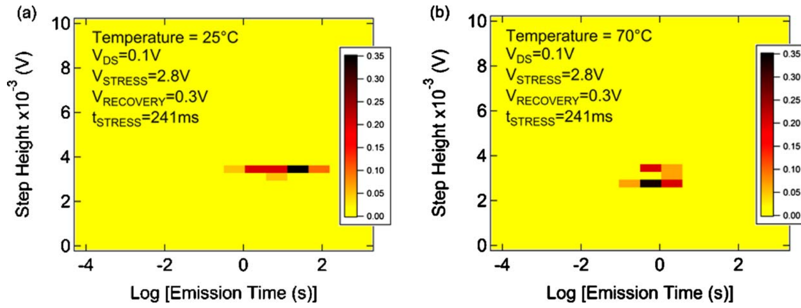

FIG. 4. (Color online) 2D histograms (TDDS spectra) of the step heights and emission times extracted from 20 recovery traces at (a) 25 °C and (b) 70 °C. Note the shift to shorter emission times when the temperature increases, proving the temperature dependence of the emission time.

In Fig. 4(a), the threshold voltage shift and the emission time of 20 traces are binned into a two-dimensional histogram normalized to the number of traces. Note that the emission times are binned on logarithmic scale. The trap is unambiguously identified by its step height, in this case, 3.64 ± 0.02 mV. It is worth highlighting that all the traces registered under these conditions presented this characteristic step; therefore, the trap studied is efficiently charged under these stress conditions. In Fig. 4, we also observe that the emission times are distributed over about 2 decades.

Figure 5 shows the histogram of the emission times t_{emission}. The emission times can be fitted to an exponential distribution with the maximum likehood method and the histogram when plotted on logarithmic scale can be fitted to the following equation:

$$f_{\text{emission}} = \frac{t_{\text{emission}}}{\tau_{\text{emission}}}\exp\left(-\frac{t_{\text{emission}}}{\tau_{\text{emission}}}\right), \quad (1)$$

where τ_{emission} is the expected value of the emission time $\langle t_{\text{emission}}\rangle$. From the fit, we obtain a τ_{emission} equal to 10.7 ± 1.1 s at 25 °C. Note that the emission time is registered at $V_{\text{recovery}}=0.3$ V.

In order to obtain the capture time, the procedure was repeated on the same device with decreasing stress time $t_{\text{stress}}=\{241,189,136,28,14\}$ ms. As shown in Fig. 6, the cumulative probability of charging the trap P_C provided that τ_{capture} (at V_{STRESS}) $\sim t_{\text{stress}} \ll \tau_{\text{emission}}$ (at V_{stress}) decreases with decreasing stress time and is described by Eq. (2), where τ_{capture} is the mean value of the capture time $\langle t_{\text{capture}}\rangle$ at V_{stress},

$$P_C = 1-\exp\left(-\frac{t_{\text{stress}}}{\tau_{\text{capture}}}\right). \quad (2)$$

The obtained capture time τ_{capture} for this characteristic trap was 37 ± 14 ms at 25 °C and a stress voltage V_{stress} of 2.8 V. Note that conversely to RTN measurements, the mean emission time is determined at V_{recovery} and the mean capture time at V_{stress}.

Once the capture time is known, the spatial position of the trap could be calculated by means of elastic tunneling. This theory is frequently used for spatial trapping profiling by charge pumping,[6,7,12,13] and trap spectroscopy by charge injection and sensing (TSCIS),[14] and also to describe the operation of charge trap memories,[5,15] and the effect of PBTS.[16] Conventionally, the depth of traps is calculated by means of the following: (1) a Wentzel-Kramers Brillouin (WKB)-approximation for determination of the tunneling distance,[17] and (2) quantum considerations to calculate the lowest subband energy level in the inversion layer of the MOSFET channels. Following these models, the filling function F is given by

$$F(t_{\text{stress}}) = 1-\exp(c_n t_{\text{stress}}), \quad (3)$$

where t_{stress} is the charging pulse and c_n is the capture rate, given by

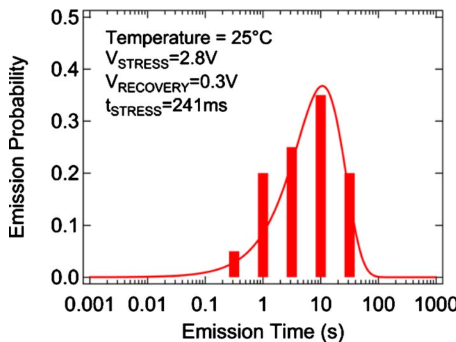

FIG. 5. (Color online) Histogram of the emission times t_{emission} extracted from 20 V_{TH} transients of a single device stressed at 25 °C and 2.8 V for 241 ms. Note that the emission times are binned on a logarithmic scale.

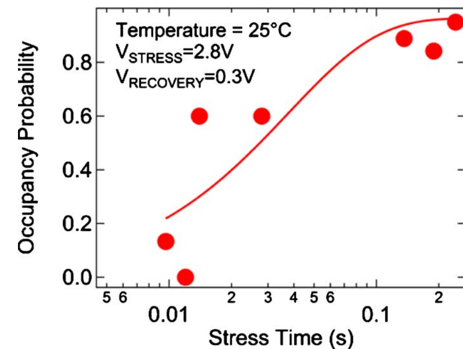

FIG. 6. (Color online) Trap occupancy P_C as a function of the stress time t_{stress} for a single device stressed at 25 °C and 2.8 V. The mean capture time obtained from the fit was 32 ms.

Author complimentary copy. Redistribution subject to AIP license or copyright, see http://jvb.aip.org/jvb/copyright.jsp

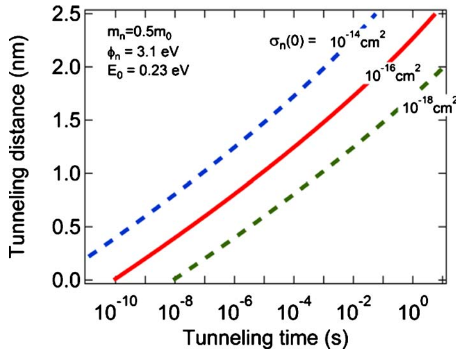

FIG. 7. (Color online) Tunneling distance as a function of the charging time for different capture cross sections. The parameters used for the simulations are the usual for SiON layers.

$$c_n(x) = \sigma_n(0) v_{\mathrm{th}} n_{\mathrm{ef}} \exp\left\{ -\frac{2\sqrt{2m_e}}{\hbar}\frac{2}{3}\frac{t_{\mathrm{SiON}}}{V_{\mathrm{SiON}}}\left[\phi_{\mathrm{eff}}^{3/2} - \left(\phi_{\mathrm{eff}} - \frac{xV_{\mathrm{SiON}}}{t_{\mathrm{SiON}}} \right)^{3/2} \right] \right\}, \quad (4)$$

where x is the physical distance from the interface Si/SiON.

The parameters used to assess the electron capture rate c_n were as usual for SiON layers: $m_e = 0.5m_0$, $\Phi_{\mathrm{eff}} = \Phi_n - E_0$, where $\Phi_n = 3.1$ eV is the energy barrier Si/SiON, and $E_0 = 0.23$ eV is the lowest sub-band energy position in the Si-inversion layer under the conditions of the experiment. It is worth noting that there exists disagreement in literature about the value for the capture cross section $\sigma_n(0)$. One can find values which extend from 10^{-18} to 10^{-14} cm^2.[18–21] Figure 7 shows the tunneling distance as a function of the charging time for $\sigma_n(0) = 10^{-14}$, 10^{-16}, and 10^{-18} cm^2. It is important to note that depending on the chosen capture cross section, the tunneling distance sweeps four orders of magnitude (!) in time. In the particular case of our experiment with a capture time of 32 ms, the trap could be placed in the range of 1.2–2.2 nm from the substrate interface. Taking into account that the thickness of the SiON layer is 1.6 nm, the maximum expected tunneling distance should be about half of the thickness of the SiON layer,[22] in our case 0.8 nm, by far shorter than the distance predicted by elastic tunneling. An unrealistic capture cross section of 5×10^{-20} cm^2 would have to be used to obtain that value. On the other hand, if the trap was placed closer to the gate, the probability to tunnel directly to the gate would exponentially increase and the trap would emit before it could be observed in our experiment.

In line with elastic tunneling, capture and emission times are expected to be weakly dependent on temperature. Therefore, the concluding test to corroborate or rule out the validity of the elastic tunneling model is to perform the experiment at different temperatures.

IV. TEMPERATURE DEPENDENCE OF THE EMISSION AND CAPTURE TIMES

In order to get insight into the temperature dependence of the emission and capture times, the procedure described in the previous section was applied to the same device with different temperatures (55, 70, and 100 °C).

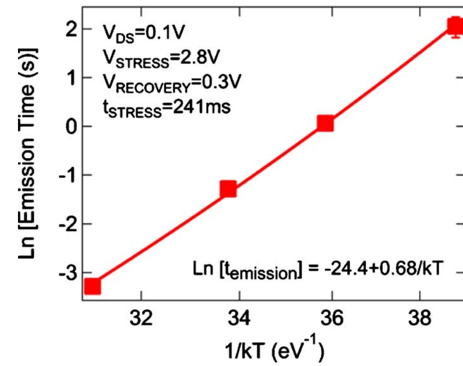

FIG. 8. (Color online) Arrhenius plot of τ_{emission}. The activation energy for the emission time obtained from the fit is 0.68 eV.

Figure 4(b) shows the two-dimensional (2D) histogram for the step heights and emission times obtained under the same conditions as in the previous section, except for the temperature, which was 70 °C. Interestingly, the uniform cluster observed at 25 °C splits up in two clusters and shifts to lower emission times. This feature has already been observed in pMOSFETs (Ref. 11) and was explained by the electrostatic interaction with another charge trap in the same percolation path.

More interesting is the dependence of the emission and capture times on temperature. In Fig. 4, we can observe how the emission time shifts horizontally two orders of magnitude with 45 °C temperature difference. The Arrhenius plot of the mean emission time τ_{emission} shown in Fig. 8 demonstrates a strong temperature dependence of this parameter. The emission time goes from 10.7 s at 25 °C to 37 ms when the temperature increases to 100 °C. This indicates without any doubt that we are dealing with a thermally activated process. Note that the extracted activation energy for the emission time is 0.68 eV, a value close to the values obtained in pMOSFET after NBTI stress.[4]

Figure 9 shows the Arrhenius plot for the capture time τ_{capture}. Note that the capture time τ_{capture} could only be determined at 25 and 50 °C due to the reduced window for the stress time t_{stress} in our experiment. As discussed previously, the expected capture time can be calculated provided that τ_{capture} (at V_{stress}) $\sim t_{\mathrm{stress}}$. For higher temperature, the occu-

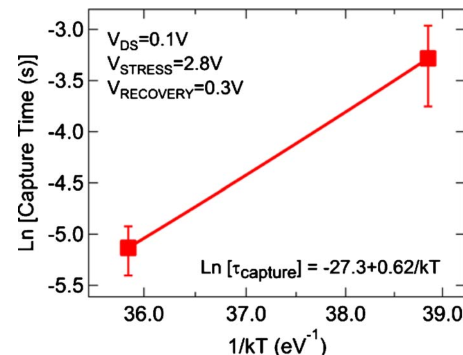

FIG. 9. (Color online) Arrhenius plot of τ_{capture}. The activation energy for capture time obtained from the fit is 0.62 eV.

Author complimentary copy. Redistribution subject to AIP license or copyright, see http://jvb.aip.org/jvb/copyright.jsp

pancy P_C was 1 for all the stress times, in other words, the expected capture time was much shorter than the stress time, so the capture time could not be determined. In Fig. 8, we can observe an important reduction of the capture time with temperature despite the temperature difference being only 25 °C. The capture time $\tau_{capture}$ was 37 ± 14 ms at 25 °C and decreased to 6 ± 1 ms when the temperature was increased to 50 °C. As shown in Fig. 9, the extracted thermal activation energy for the capture time of this trap was 0.62 eV. Again, this value is close to the activation energies for capture times obtained for pMOSFETs after NBTI in Ref. 4.

We therefore conclude that electron trapping and detrapping in SiON are thermally activated processes. Elastic tunneling would follow a weak temperature dependence, which is incompatible with the experimental data presented in this article. Moreover, the capture time obtained experimentally is by far larger than what can be expected by elastic tunneling theory. Therefore, a more complete model has to be considered. Based on the link between the recovery component after PBTS and RTN, it is natural to apply the nonradiative mulpihonon theory used for RTN.[23] This theory explains the necessary thermal dependence and the large capture and emission times obtained in this experiment. In Ref. 8, an improved version of this model has been proposed after a thorough study of the statistical properties of the individual defects in pMOSFETs following NBTI stress. The similarities observed between both experiments lead us to conclude that an analogous model can be applied to nMOSFETs following PBTI stress. Therefore, traps with similar properties are responsible for the recoverable component in pMOSFETs and nMOSFETs after NBTI stress and PBTI stress, respectively.

V. SUMMARY

In summary, we have successfully applied a recently developed methodology for the study of individual traps in positively stressed nanoscaled SiON nMOSFETs. From this study, we conclude that the emission and capture times of individual traps are thermally activated, in agreement with the findings presented on SiON pMOSFETs after negative bias temperature stress. Therefore, traps with similar properties are responsible for the relaxation curves in pMOSFETs and nMOSFETs under NBTI stress and PBTI stress, respectively. This result disagrees with the direct tunneling theory ubiquitously used in different trap characterization techniques and simulations. Proper modeling of erase and program transients of nonvolatile memories and depth localization of oxide traps can be only be achieved by considering the thermal activation of the trapping and detrapping mechanisms.

ACKNOWLEDGMENTS

This work was performed under the IMEC core partner affiliation program. M.T. stay was supported in part by the Spanish Ministry of Education and Science under Contract No. TEC2010-18051 and the grant program "José Castillejo" (Grant No. JC2009/00052). The authors would like to thank the IMEC DRE and NVM groups, and AMSIMEC for helpful discussions and input throughout this work.

[1]K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant, Phys. Rev. Lett. **52**, 228 (1984).
[2]A. Ghetti, C. Monzio Compagnoni, F. Biancardi, A. L. Lacaita, S. Beltrami, L. Chiavarone, A. S. Spinelli, and A. Visconti, Proceedings of IEDM, 2008 (unpublished), pp. 835–838.
[3]B. Kaczer, T. Grasser, J. Martin-Martinez, E. Simoen, M. Aoulaiche, Ph. J. Roussel, and G. Groeseneken, Proceedings of IRPS, 2009 (unpublished), pp. 55–60.
[4]T. Grasser, B. Kaczer, W. Goes, Th. Aichinger, Ph. Hehenberger, and M. Nelhiebel, Proceedings of IRPS, 2009 (unpublished), pp. 33–44.
[5]P. C. Arnett, J. Appl. Phys. **46**, 5236 (1975).
[6]D. Heh, C. D. Young, G. A. Brown, P. Y. Hung, A. Diebold, E. M. Vogel, J. B. Bernstein, and G. Bersuker, IEEE Trans. Electron Devices **54**, 1338 (2007).
[7]M. Toledano-Luque, R. Degraeve, M. B. Zahid, L. Pantisano, E. San Andrés, G. Groeseneken, and S. De Gendt, IEEE Trans. Electron Devices **55**, 3184 (2008).
[8]B. Kaczer, T. Grasser, Ph. J. Roussel, J. Franco, R. Degraeve, L.-A. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, Proceedings of IRPS, 2010 (unpublished), pp. 26–32.
[9]B. Kaczer, T. Grasser, Ph. J. Roussel, J. Martin-Martinez, R. O'Connor, B. J. O'Sullivan, and G. Groeseneken, Proceedings of IRPS, 2008 (unpublished), pp. 20.
[10]T. Grasser, W. Goes, V. Sverdlov, and B. Kaczer, Proceedings of IRPS, 2007 (unpublished), pp. 268–280.
[11]T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Goes, and B. Kaczer, Proceedings of IRPS, 2010 (unpublished), pp. 16–25.
[12]C. E. Weintraub, E. Vogel, J. R. Hauser, N. Yang, V. Misra, J. J. Wortman, J. Ganem, and P. Masson, IEEE Trans. Electron Devices **48**, 2754 (2001).
[13]G. T. Sasse and J. Schmitz, IEEE Trans. Electron Devices **55**, 881 (2008).
[14]R. Degraeve, M. Cho, B. Govoreanu, B. Kaczer, M. B. Zahid, J. Van Houdt, M. Jurczak, and G. Groeseneken, Proceedings of IEDM, 2008 (unpublished), pp. 775–778.
[15]E. Vianello *et al.*, Proceedings of IEDM, 2009 (unpublished), pp. 83–86.
[16]G. Reimbold, J. Mitard, X. Garros, C. Leroux, G. Ghibaudo, and F. Martin, Microelectron. Reliab. **47**, 489 (2007).
[17]E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970), Vol. 1, Chap. 7, p. 113.
[18]R. Degraeve *et al.*, IEEE Trans. Electron Devices **51**, 1392 (2004).
[19]J. G. Simmons and L. S. Weisse, Solid-State Electron. **16**, 53 (1973).
[20]G. Groeseneken, H. E. Maes, N. Beltran, and R. F. De Keersmaecker, IEEE Trans. Electron Devices **31**, 42 (1984).
[21]D. Bauza, Solid-State Electron. **46**, 2035 (2002).
[22]D. Ielmini, A. S. Spinelli, A. L. Lacaita, and A. Modelli, IEEE Trans. Electron Devices **49**, 1955 (2002).
[23]M. J. Kirton and M. J. Uren, Appl. Phys. Lett. **48**, 1270 (1986).

Author complimentary copy. Redistribution subject to AIP license or copyright, see http://jvb.aip.org/jvb/copyright.jsp