ELSEVIER

Microelectronics Reliability 44 (2004) 1879–1884

MICROELECTRONICS RELIABILITY

www.elsevier.com/locate/microrel

Statistical simulation of gate dielectric wearout, leakage, and breakdown

A. Gehring* and S. Selberherr

Institute for Microelectronics, Vienna University of Technology
Gußhausstraße 27–29, A-1040 Vienna, Austria

Abstract

We present a set of models for the simulation of gate dielectric leakage, wearout, and breakdown. The leakage model accounts for direct and trap-assisted tunneling through the dielectric layer. Wearout is caused by the leakage-induced creation of neutral defects at random positions in the dielectric layer, which, if occupied, degrade the threshold voltage of the device. Gate dielectric breakdown is triggered by the formation of a conductive path through the insulator. To allow trap characterization and for the simulation of fast transients the modeling of trap charging and decharging processes is outlined. The models have been implemented into a three-dimensional device simulator and are used for the characterization of ZrO_2-based dielectrics and for the study of gate leakage and wearout effects in standard CMOS inverter circuits.

1. Introduction

Shrinking of gate dielectric thicknesses of submicron CMOS transistors demands the use of alternative gate dielectrics such as ZrO_2. These materials, however, suffer from high defect densities which degrade the device performance. Therefore, the gate dielectric reliability becomes a crucial issue not only for non-volatile memories but also for logic applications. However, state-of-the-art device simulators do not properly account for this effect. While the current transport through high-κ dielectric layers either by direct [1] or defect-assisted tunneling [2] has been studied intensely applying sophisticated methods, modeling of dielectric breakdown has been investigated only recently [3]. Since the processes of leakage, trap creation, and dielectric breakdown are physically directly related, we propose a set of models which directly link the simulation of direct and trap-assisted leakage current with the creation and occupation of traps and the occurrence of breakdown.

The random trap placement leads to a statistical fluctuation of the threshold voltage and the gate current density. We find that this variation is reduced for thinner dielectrics. The implementation of these models into the general-purpose device simulator MINIMOS-NT [4] also allows to investigate the effect of gate leakage and wearout on circuits such as a standard CMOS inverter. We see no significant degradation of circuit performance due to gate leakage, while dielectric wearout leads to a shift of the transfer characteristics and dielectric breakdown deteriorates the slew rate of the circuit.

The paper is structured as follows. In Section 2 the models for leakage, wearout, and breakdown are described. The trap-assisted tunneling model is enhanced to describe transient trap charging in Section 3. In Section 4 statistical simulations with random trap placement are carried out, and Section 5 shows the effect of gate leakage and dielectric wearout on the transfer characteristics of inverter circuits. A summary and conclusions are presented in Section 6.

*Corresponding author. Tel.: +43-1-58801/36016; fax:+43-1-58801/36099. *E-mail address:* Gehring@iue.tuwien.ac.at.

0026-2714/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.microrel.2004.07.101

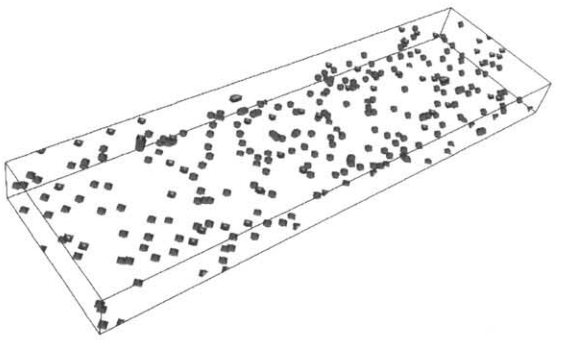

Figure 1. Random trap distribution in a MOSFET dielectric layer simulated by MINIMOS-NT.

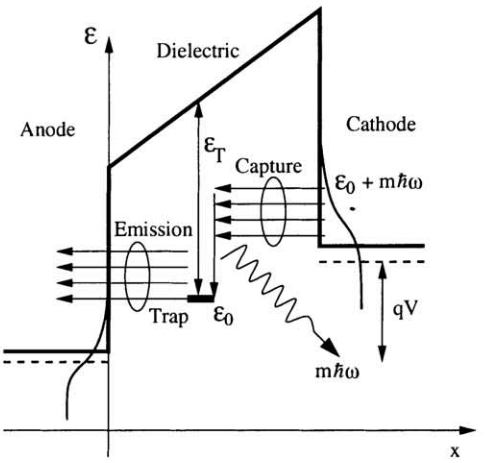

Figure 2. Trap-assisted tunneling transition by inelastic phonon emission.

2. Leakage, Wearout, and Breakdown

We distinguish three processes which happen sequentially and finally trigger breakdown. Starting from a fresh dielectric layer with low trap concentration or no defects at all, the direct tunneling current gives rise to the creation of neutral defects. These defects cause trap-assisted tunneling, leading to two detrimental effects. First, some of the existing traps become occupied by electrons, which degrades the threshold voltage of the device. Second, new defects are created in the dielectric layer. The location of the traps is assumed to be random within the layer, while a constant energy level and a specific charge state (positive or negative) is assumed. Finally, if a conductive path through the dielectric is formed, a localized breakdown occurs and the current density increases according to the conductivity of the dielectric layer.

2.1. Modeling of Leakage

Gate leakage is modeled as the sum of direct and trap-assisted tunneling. Assuming a fresh and defect-free dielectric layer, only direct tunneling is present which is modeled following the commonly applied Tsu-Esaki approach [5] where the gate current density is given as

$$J = \frac{4\pi m_{\mathrm{eff}} \mathrm{q}}{h^3} \int_{\mathcal{E}_{\min}}^{\mathcal{E}_{\max}} TC(\mathcal{E}_x) N(\mathcal{E}_x)\, \mathrm{d}\mathcal{E}_x \,. \quad (1)$$

The symbol $TC(\mathcal{E}_x)$ in this expression denotes the quantum-mechanical transmission coefficient $TC(\mathcal{E}_x)$ which can be computed by several methods [6]. We have used a numerical WKB method to allow for conduction band discontinuities as encountered in the modeling of high-κ dielectrics, since this method represents a sound compromise between accuracy and computational effort.

The supply function $N(\mathcal{E})$ is calculated from Fermi-Dirac distributions at both sides of the dielectric.

$$N(\mathcal{E}_x) = \mathrm{k_B} T \ln \left(\frac{1 + \exp\left(-\dfrac{\mathcal{E}_x - \mathcal{E}_{\mathrm{F},1}}{\mathrm{k_B} T} \right)}{1 + \exp\left(-\dfrac{\mathcal{E}_x - \mathcal{E}_{\mathrm{F},2}}{\mathrm{k_B} T} \right)} \right) \quad (2)$$

The current through the dielectric layer gives rise to the creation of neutral defects which are randomly placed in the dielectric layer, as shown in Fig. 1 for a three-dimensional simulation.

The defects give rise to additional trap-assisted tunneling which is modeled via inelastic phonon-assisted transitions [7, 8]. Fig. 2 shows the basic trap-assisted tunneling process through the gate dielectric. Electrons are captured from the cathode, relax to the energy of the trap $\mathcal{E}_0$ by phonon emission with energy $m\hbar\omega$, and are emitted to the anode. The trap-assisted tunneling current is found by integration over the dielectric thickness

$$J_{\mathrm{t}} = \mathrm{q} \int_0^{t_{\mathrm{diel}}} \frac{N_{\mathrm{T}}(x)}{\tau_{\mathrm{c}}(x) + \tau_{\mathrm{e}}(x)} \mathrm{d}x, \quad (3)$$

where $N_{\mathrm{T}}(x)$ is the trap concentration and $\tau_{\mathrm{c}}(x)$ and $\tau_{\mathrm{e}}(x)$ denote the capture and emission times calculated from

$$\tau_{\mathrm{c}}^{-1}(z) = \int_{\mathcal{E}_0}^{\infty} c_n(\mathcal{E}, x) T_{\mathrm{l}}(\mathcal{E}) f_{\mathrm{l}}(\mathcal{E})\, \mathrm{d}\mathcal{E} \quad (4)$$

$$\tau_{\mathrm{e}}^{-1}(z) = \int_{\mathcal{E}_0}^{\infty} e_n(\mathcal{E}, x) T_{\mathrm{r}}(\mathcal{E}) (1 - f_{\mathrm{r}}(\mathcal{E}))\, \mathrm{d}\mathcal{E} \,. \quad (5)$$

In these expressions, c_n and e_n denote the capture and emission rates, f_l and f_r the Fermi distributions,

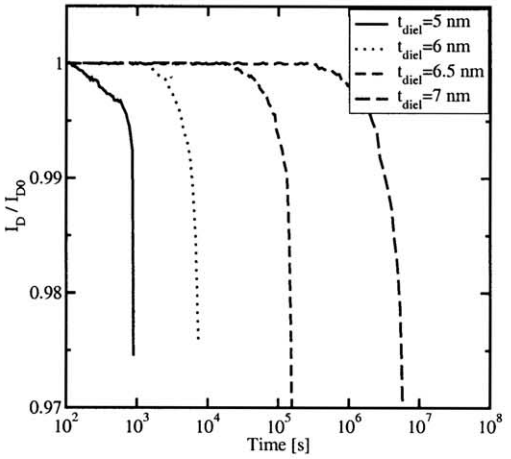

Figure 3. Drain current as a function of time for different dielectric thicknesses.

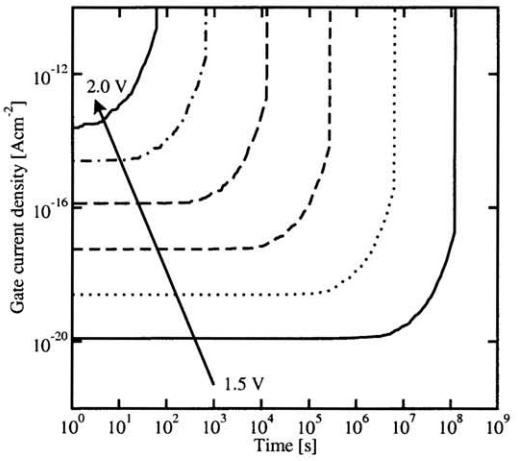

Figure 4. Dielectric breakdown for a 3 nm SiO_2 layer as a function of gate bias.

and T_l and T_r the transmission coefficients from the left and right side of the dielectric, respectively. The capture and emission processes are described by their respective probabilities as suggested by Herrmann and Schenk [7], and the transmission coefficients were evaluated by the same numerical WKB method as used for the direct current calculation.

2.2. Modeling of Wearout

While the neutral defects cause trap-assisted tunneling and gate leakage, only the occupied traps lead to threshold voltage degradation and wearout of the gate dielectric. This is modeled by an additional space charge component

$$\rho(x) = Q_T N_T(x) f_T(x) \tag{6}$$

where f_T denotes the trap occupancy and Q_T the trap charge state. Due to this charge the threshold voltage of the device degrades over time. This is shown in Fig. 3 for devices with different dielectric thicknesses. Here, an initial trap concentration of $2 \times 10^{18}\,\mathrm{cm}^{-3}$ and a trap energy level of 2.75 eV was assumed. It can be seen that the wearout process is characterized by a region of slow threshold shift where chare in the dielectric builds up. As soon as a certain number of traps in the dielectric is reached, the threshold voltage is suddenly reduced and dielectric breakdown follows.

Note that the assumption of inelastic phonon-assisted tunneling implies that, depending on the bias conditions, only a fraction of the traps in the dielectric layer might be occupied [9].

2.3. Modeling of Breakdown

The neutral and charged defects create percolation paths in the dielectric, which eventually connect the gate with the substrate [3]. The traps are placed randomly with a fixed trap energy, as shown in Fig. 5. The defect concentration N_T is assumed to be proportional to the total injected charge Q_i via

$$N_T = C Q_i^{\alpha}\,, \tag{7}$$

as proposed by Degraeve *et al.* [10], who found values of $C = 5.3 \times 10^{-19}\,\mathrm{cm}^{-1.88}\mathrm{As}^{-0.56}$ and $\alpha = 0.56$ for dielectric thicknesses between 7.3 and 13.8 nm. Since no measurements for state-of-the-art logic devices have been available, these values were used for the thinner dielectric layers, too.

As soon as a percolation path through the dielectric is created, the dielectric layer loses its insulating behavior and the current suddenly increases. The gate current density is shown in Fig. 4 for a 3 nm layer of SiO_2 as a function of time for different gate voltages assuming an initial trap concentration of $10^{16}\,\mathrm{cm}^{-3}$.

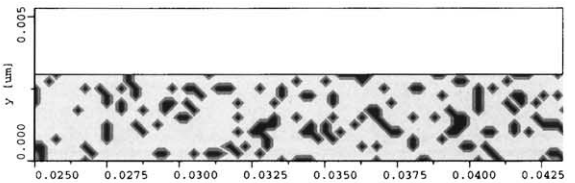

Figure 5. Two-dimensional cut through the dielectric layer showing the random trap placement (darks spots).

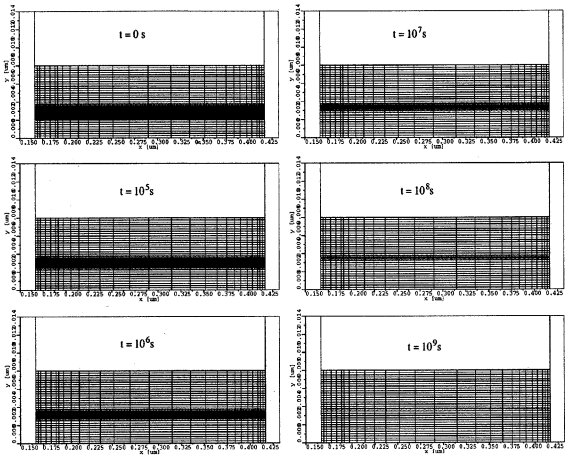

Figure 6. Trap occupancy $f_T(x)$ in the dielectric layer during a decharging process as a function of time.

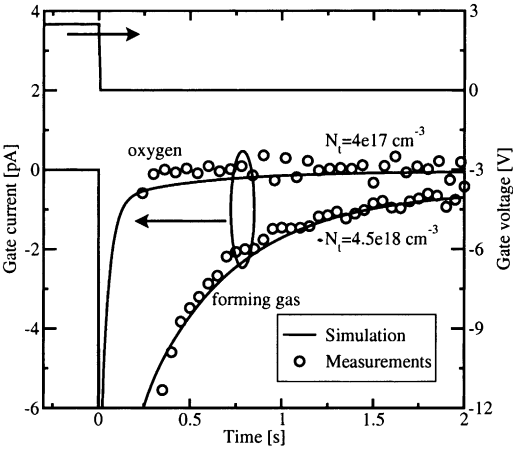

Figure 7. Transient trap charging currents for a ZrO_2 layer [12].

The time-to-breakdown strongly decreases and the gate leakage strongly increases with higher gate bias. However, the gate current density after breakdown can no more be described by a tunneling process. Measurements indicate that the gate current after breakdown is related to the gate voltage by a simple power law $I = KV_G^p$, where the parameter K reflects the size of the breakdown spot, and the parameter p is in the range of $2-5$ [11]. This expression can be used to account for the effect of dielectric breakdown in circuit simulations.

3. Modeling of Transient Trap Charging

To predict the transient behavior of fast switching processes, the charging and decharging dynamics of the traps must be considered. The concentration of occupied traps at position x and time t is generally described by the rate equation

$$N_T(x)\frac{df_T(x,t)}{dt} = N_T(x)\frac{1-f_T(x,t)}{\tau_c(x,t)} - N_T(x)\frac{f_T(x,t)}{\tau_e(x,t)}$$

where τ_c and τ_e describe the capture and emission time of the trap. For the stationary case the time derivative on the left-hand side is zero and (3) can be derived, while for the transient case, the time constants must be evaluated in each time step. The occupancy function can be calculated iteratively by

$$f_T(x,t_i) = A_i + B_i f_T(x,t_{i-1}) \tag{8}$$

where A_i and B_i depend on the capture and emission times at the time step t_i by [9]

$$A_i = \frac{\tau_c^{-1}(z,t_i)\Delta t_i}{1+C_i} \qquad B_i = \frac{1-C_i}{1+C_i}$$
$$C_i = \frac{\tau_m^{-1}(z,t_i)\Delta t_i}{2}.$$

Here, $\Delta t_i = t_i - t_{i-1}$ and t_i denote the discretized time steps. Once the time-dependent occupancy function in the dielectric is known, the tunnel current through one of the interfaces at time t_i is

$$J_{l,r}(t_i) = q\int_0^{t_{diel}} N_T(x)\tau_{l,r}^{-1}(x,t_i)\,dx\,. \tag{9}$$

In this expression the symbols l and r denotes the considered interface (left or right) and the time constants τ_l and τ_r are calculated from

$$\tau_{l,r}^{-1}(x,t_i) = \tau_{cl,r}^{-1}(x,t_i) - f_T(x,t_i)\left[\tau_{cl,r}^{-1}(x,t_i) + \tau_{el,r}^{-1}(x,t_i)\right].$$

Note that the current through the two interfaces is, in general, not equal. Only after the trap charging processes are finished, the capture and emission currents at the interfaces are in equilibrium [13]. The trap occupancy during such a decharging step is shown in Fig. 6.

This model can be applied to the characterization of traps in the dielectric layer. Fig. 7 shows the step response of two MOS capacitors with ZrO_2 dielectrics annealed in reducing and oxidizing conditions [14]. The gate voltage is first fixed at a value of 2.5 V to achieve a steady initial trap occupation.

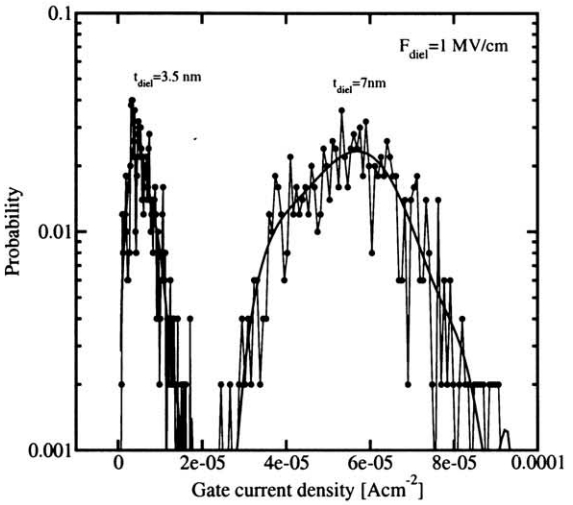

Figure 8. Probability distribution of the gate leakage current in a 3.5 nm dielectric (left) and a 7 nm dielectric right) at an electric field of 1 MV/cm.

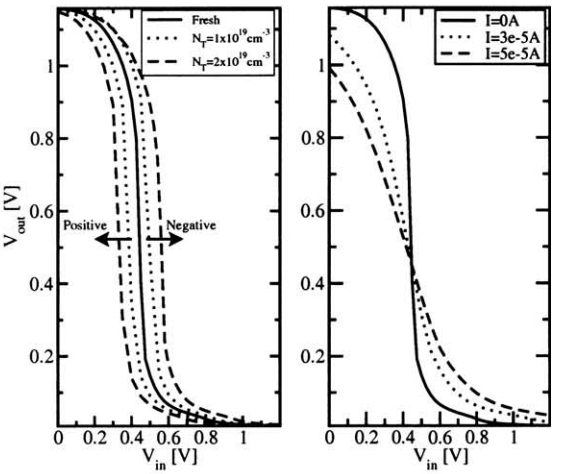

Figure 9. Transfer characteristics of a CMOS inverter for positive and negative trap charges (left) and for the case of dielectric breakdown with different leakage currents.

Then, the gate voltage is turned off and the resulting gate current is measured over time. The resulting transient gate current peak exceeds the static gate current by orders of magnitude. Furthermore, especially for the oxide annealed in forming gas atmosphere, the gate current decays very slowly with a time constant in the order of a second. This may be caused by a different trap distribution in the oxide or even different trap energy levels which lead to a different time constant for the decharging process [12].

4. Statistical Simulations

The random trap creation in the dielectric layer imposes an element of uncertainty in the gate current calculation. To investigate the resulting gate current densities arising from this random trap placement, we performed a set of 500 simulations for devices with a 3.5 nm and a 7 nm dielectric layer at a gate voltage of 3.5 V and 7 V, respectively. Source, drain, and bulk contacts were grounded, so the electric field in the dielectric layer is the same in both structures. The probability distribution of the resulting gate current density is shown in Fig. 8. It can be seen that the thinner dielectric layer has a much smaller variance as the thicker layer. This can be attributed to the fact that the traps can be placed in a wider region for the thicker dielectric, while they are confined to a thinner region in the 3.5 nm dielectric. This observation is in qualitative agreement with results of Larcher [2] from a one-dimensional study adopting random trap placement and Gaussian distributed trap energy levels.

5. Circuit Simulations

To investigate the effect of the presented models on CMOS circuits, a CMOS inverter was simulated using the mixed-mode capabilities of MINIMOS-NT. Following a recent work of Rodriguez *et al.* [15], nMOS and pMOS devices with a gate length of 130 nm and an oxide thickness of 1.5 nm were assumed. The supply voltage was set to 1.2 V. First, the effect of gate leakage was studied.

The nMOS tunneling current exceeds the pMOS tunneling current by orders of magnitude due to the fact that at low positive bias, the pMOS tunneling current is composed of holes tunneling from the substrate to the gate, facing a high energy barrier and high effective mass in the oxide [16]. The nMOS, on the other hand, is biased in inversion and leads to tunneling current orders of magnitude higher. However, even for a gate current density in the order of 100 Acm^{-2}, the effect on the inverter characteristics was hardly visible.

The trap charge, on the other hand, strongly degraded the threshold voltage as shown in the left part of Fig. 4, where positively charged traps shift the transfer characteristics to the left (decreasing threshold voltage) and negatively charged traps shift the transfer characteristics to the right (increasing threshold voltage).

Finally, the effect of dielectric breakdown was investigated by including an additional current source between the gate and drain contacts of the transistors. This leads to a strong degradation of the inverter characteristics as shown in the right part of Fig. 4 for different cur-

rent values. This is in qualitative agreement to results presented in [15]

6. Summary and Conclusions

We presented a set of models for the description of leakage, wearout, and breakdown of dielectric layers suitable for two- and three-dimensional device simulation. Leakage is modeled by a combination of direct and trap-assisted inelastic tunneling and leads to the creation of neutral traps in the dielectric, which take part in the trap-assisted tunneling process and, if occupied, cause a threshold voltage shift. Over time, the random creation of defects in the dielectric layer eventually results in the formation of a conducting path. For fast transient processes, it is necessary to model charging and decharging processes. The model was used to investigate the probability distribution of gate leakage currents and to study its effect on CMOS inverters. It was found that with reduced oxide thickness, the statistical variance of the gate leakage is reduced. The leakage has only minor effects on the transfer characteristics of CMOS inverters, but the threshold voltage shift due to occupied traps and the high gate current after breakdown may be a showstopper for the use of these devices in circuits.

Acknowledgments

The good cooperation with Francisco Jiménez-Molinos and Stefan Harasek is gratefully acknowledged.

References

[1] Y.-Y. Fan, Q. Xiang, J. An, L. F. Register, and S. K. Banerjee, "Impact of Interfacial Layer and Transition Region on Gate Current Performance for High-K Gate Dielectric Stack: Its Tradeoff With Gate Capacitance," *IEEE Trans.Electron Devices*, vol. 50, no. 2, pp. 433–439, 2003.

[2] L. Larcher, "Statistical Simulation of Leakage Currents in MOS and Flash Memory Devices with a New Multiphonon Trap-Assisted Tunneling Model," *IEEE Trans.Electron Devices*, vol. 50, no. 5, pp. 1246–1253, 2003.

[3] J. H. Stathis, "Reliability Limits for the Gate Insulator in CMOS Technology," *IBM J.Res.Dev.*, vol. 46, no. 2/3, pp. 265–286, 2002.

[4] Institut für Mikroelektronik, Technische Universität Wien, Austria, *MINIMOS-NT User's Guide*, 2002.

[5] R. Tsu and L. Esaki, "Tunneling in a Finite Superlattice," *Appl.Phys.Lett.*, vol. 22, no. 11, pp. 562–564, 1973.

[6] A. Gehring, H. Kosina, T. Grasser, and S. Selberherr, "Consistent Comparison of Tunneling Models for Device Simulation," in *Proc. 4th European Workshop on Ultimate Integration of Silicon*, Udine, Italy, 2003, pp. 131–134.

[7] M. Herrmann and A. Schenk, "Field and high-temperature dependence of the long term charge loss in erasable programmable read only memories: Measurements and modeling," *J.Appl.Phys.*, vol. 77, no. 9, pp. 4522–4540, 1995.

[8] F. Jiménez-Molinos, A. Palma, F. Gámiz, J. Banqueri, and J. A. Lopez-Villanueva, "Physical Model for Trap-Assisted Inelastic Tunneling in Metal-Oxide-Semiconductor Structures," *J.Appl.Phys.*, vol. 90, no. 7, pp. 3396–3404, 2001.

[9] A. Gehring, F. Jiménez-Molinos, H. Kosina, A. Palma, F. Gámiz, and S. Selberherr, "Modeling of Retention Time Degradation Due to Inelastic Trap-Assisted Tunneling in EEPROM Devices," *Microelectron.Reliab.*, vol. 43, no. 9-11, pp. 1495–1500, 2003.

[10] R. Degraeve, G. Groeseneken, R. Bellens, J. L. Ogier, M. Depas, P. J. Roussel, and H. E. Maes, "New Insights in the Relation Between Electron Trap Generation and the Statistical Properties of Oxide Breakdown," *IEEE Trans.Electron Devices*, vol. 45, no. 4, pp. 904–911, 1998.

[11] J. H. Stathis, B. P. Linder, R. Rodríguez, and S. Lombardo, "Reliability of Ultra-Thin Oxides in CMOS Circuits," *Microelectron.Reliab.*, vol. 43, no. 9-11, pp. 1353–1360, 2003.

[12] A. Gehring, S. Harasek, E. Bertagnolli, and S. Selberherr, "Evaluation of ZrO_2 Gate Dielectrics for Advanced CMOS Devices," in *Proc. European Solid-State Device Research Conf.*, Estoril, Portugal, 2003, pp. 473–476.

[13] F. Jiménez-Molinos, A. Palma, A. Gehring, F. Gámiz, H. Kosina, and S. Selberherr, "Static and transient simulation of inelastic trap-assisted tunneling," in *Proc. 14^{th} Workshop on Modeling and Simulation of Electron Devices*, Barcelona, Spain, October 2003, pp. 65–68.

[14] S. Harasek, H. D. Wanzenböck, and E. Bertagnolli, "Compositional and Electrical Properties of Zirconium Dioxide Thin Films Chemically Deposited on Silicon," *J.Vac.Sci.Technol.A*, vol. 21, no. 3, pp. 653–659, 2003.

[15] R. Rodríguez, J. H. Stathis, B. P. Linder, R. V. Joshi, and C. T. Chuang, "Influence and Model of Gate Oxide Breakdown on CMOS Inverters," *Microelectron.Reliab.*, vol. 43, no. 9-11, pp. 1439–1444, 2003.

[16] J. Cai and C.-T. Sah, "Gate Tunneling Currents in Ultrathin Oxide Metal-Oxide-Silicon Transistors," *J.Appl.Phys.*, vol. 89, no. 4, pp. 2272–2285, 2001.