Modeling of Wearout, Leakage, and Breakdown of Gate Dielectrics

A. Gehring and S. Selberherr

Institute for Microelectronics, TU Vienna, Gusshausstr. 27–29, A-1040 Vienna, Austria
Phone +43-1-58801-36016, FAX +43-1-58801-36099, e-mail: `gehring@iue.tuwien.ac.at`

Abstract

We present a set of models for the simulation of gate dielectric wearout, leakage, and breakdown. 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. Leakage is due to direct and trap-assisted tunneling through these defects. Finally, gate dielectric breakdown is triggered by the formation of a conductive path through the insulator. To allow the trap characterization and for the simulation of fast transients the modeling of trap charging and decharging processes is outlined. The model has been implemented into a three-dimensional device simulator and is 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. Therefore, the gate dielectric reliability becomes a crucial issue not only for non-volatile memories but also for logic applications. 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]. The processes of leakage, trap creation, and dielectric breakdown are usually modeled independently, although they are physically directly related. Thus, we use 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 implementation of these models into the device simulator MINIMOS-NT [4] allows to investigate the effect of gate leakage and wearout on circuits such as a standard CMOS inverter.

2 Leakage, Wearout, and Breakdown

We distinguish three processes which happen sequentially and finally trigger breakdown. Starting from a fresh dielectric layer with a low trap concentration, the direct tunneling current gives rise to the creation of neutral defects. These defects cause trap-assisted tunnel-

ing, leading to two 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 and Wearout

Gate leakage is modeled as the sum of two processes, 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

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

The transmission coefficient $TC(\mathcal{E}_x$ is computed by a numerical WKB method to allow for conduction band discontinuities as encountered in the modeling of high-κ dielectrics. Quantization effects in the MOSFET inversion layer are neglected. The supply function $N(\mathcal{E})$ is calculated from Fermi-Dirac distributions at both sides of the dielectric. 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. 2.1 for a three-dimensional simulation.
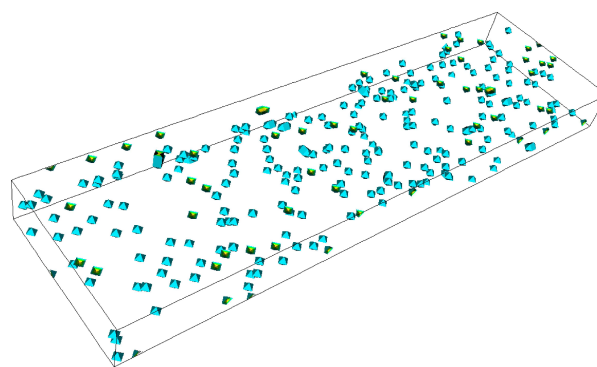

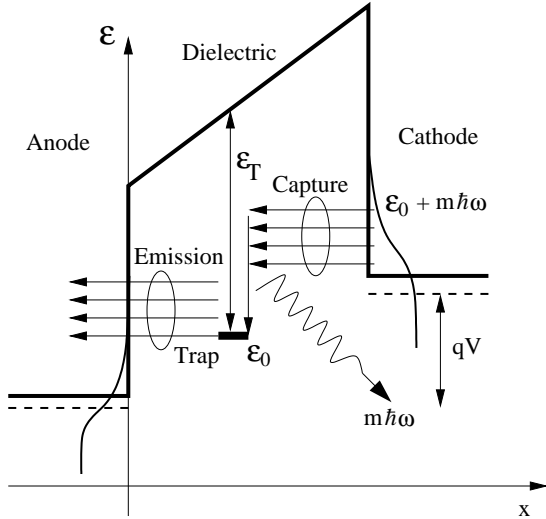**Figure 1:** Random trap distribution in a MOSFET dielectric layer simulated by MINIMOS-NT.

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

The defects give rise to additional trap-assisted tunneling which is modeled via inelastic phonon-assisted transitions [6, 7]. 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

$$J_{\mathrm{t}} = \mathrm{q} \int_0^{t_{\mathrm{diel}}} \frac{N_{\mathrm{T}}(x)}{\tau_{\mathrm{c}}(x) + \tau_{\mathrm{e}}(x)} \mathrm{d}x, \qquad (2)$$

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

$$\tau_{\mathrm{c}}^{-1}(z) = \int_{\mathcal{E}_0}^{\infty} c_n(\mathcal{E}, x) T_{\mathrm{l}}(\mathcal{E}) f_{\mathrm{l}}(\mathcal{E}) \, \mathrm{d}\mathcal{E} \qquad (3)$$

$$\tau_{\mathrm{e}}^{-1}(z) = \int_{\mathcal{E}_0}^{\infty} e_n(\mathcal{E}, x) T_{\mathrm{r}}(\mathcal{E})(1 - f_{\mathrm{r}}(\mathcal{E})) \, \mathrm{d}\mathcal{E} \; . \qquad (4)$$

In these expressions, c_n and e_n denote the capture and emission rates, f_l and f_r the Fermi distributions, 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 [6], and the transmission coefficients were evaluated by a numerical WKB method. Fig. 3 shows a comparison with experimental data for MOS capacitors [8], where the transition from the trap-assisted tunneling regime at low bias to the Fowler-Nordheim tunneling regime at high bias is clearly visible.

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
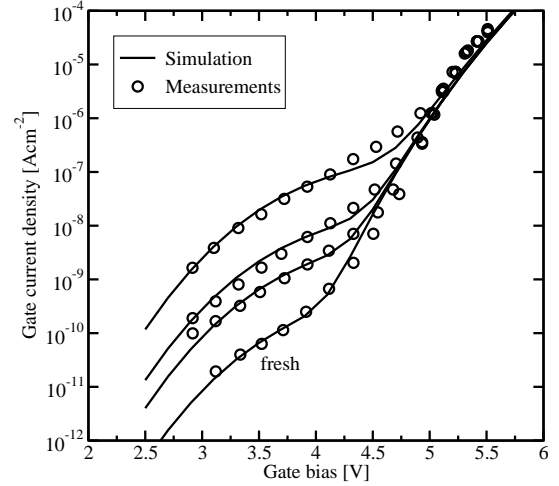

Figure 3: Gate current density for different stress times [8] for t_{diel}=5.5 nm. The model parameters are $\mathcal{E}_{\mathrm{T}}$=2.7 eV, $S\hbar\omega$=1.3 eV, and N_{T}= 9.0 × 10^{17} cm^{-3}, 1.0 × 10^{17} cm^{-3}, 3.0 × 10^{16} cm^{-3}, and 3.0 × 10^{15} cm^{-3} (from top to bottom).

dielectric. This is modeled by an additional space charge $\rho(x) = Q_{\mathrm{T}} N_{\mathrm{T}}(x) f_{\mathrm{T}}(x)$ in the Poisson equation, where f_{T} denotes the trap occupancy and Q_{T} the trap charge state. Note that the assumption of phonon-assisted tunneling implies that, depending on the bias conditions, only a fraction of the traps in the dielectric layer may really be occupied [9].

2.2 Modeling of Breakdown

The neutral defects create percolation paths in the dielectric, which eventually connect the gate with the substrate [3]. In MINIMOS-NT the traps are placed randomly, and the defect concentration N_{T} is assumed to be proportional to the total injected charge Q_{i} via

$$N_{\mathrm{T}} = C Q_{\mathrm{i}}^{\alpha} \; , \qquad (5)$$

as proposed by Degraeve *et al.* [10], who found values of $C = 5.3 \times 10^{-19}$ cm$^{-1.88}$As$^{-0.56}$ and $\alpha = 0.56$ for dielectric thicknesses between 7.3 and 13.8 nm. This is shown in Fig. 2.2 for a 3 nm layer of SiO$_2$ at different time steps.
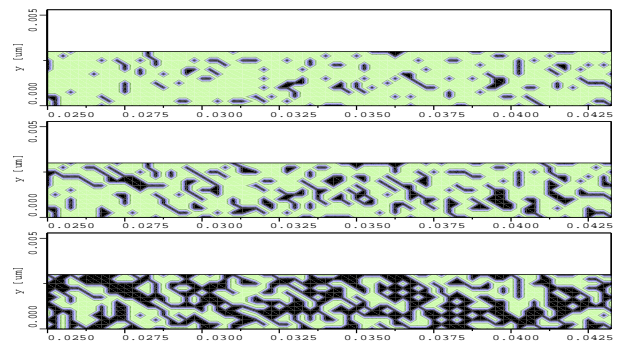

Figure 4: Defect distribution in a 3 nm SiO$_2$ layer at different time steps.
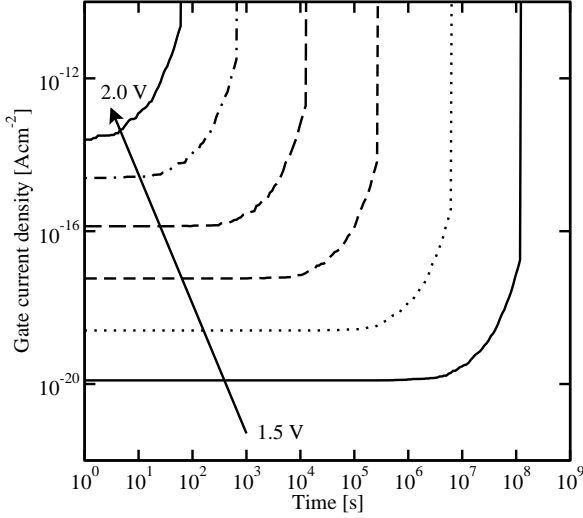
Figure 5: Dielectric breakdown for a $3\,\mathrm{nm}$ SiO$_2$ layer as a function of gate bias.

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. 5 for a $3\,\mathrm{nm}$ layer of SiO$_2$ as a function of time for different gate voltages assuming an initial trap concentration of $10^{16}\,\mathrm{cm}^{-3}$. 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 $I = KV_\mathrm{G}^p$, where the parameter K reflects the size of the breakdown spot, and the parameter p is in the range of $2-5$ [11].

3 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

$$N_\mathrm{T}(x)\frac{\mathrm{d}f_\mathrm{T}(x,t)}{\mathrm{d}t} = N_\mathrm{T}(x)\frac{1-f_\mathrm{T}(x,t)}{\tau_\mathrm{c}(x,t)} - N_\mathrm{T}(x)\frac{f_\mathrm{T}(x,t)}{\tau_\mathrm{e}(x,t)}$$

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 (2) 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 $f_\mathrm{T}(x,t_i) = A_i + B_i f_\mathrm{T}(x,t_{i-1})$ where A_i and B_i depend on the capture and emission times at the time step t_i by [9]

$$A_i = \frac{\tau_\mathrm{c}^{-1}(z,t_i)\Delta t_i}{1+C_i} \qquad B_i = \frac{1-C_i}{1+C_i}$$

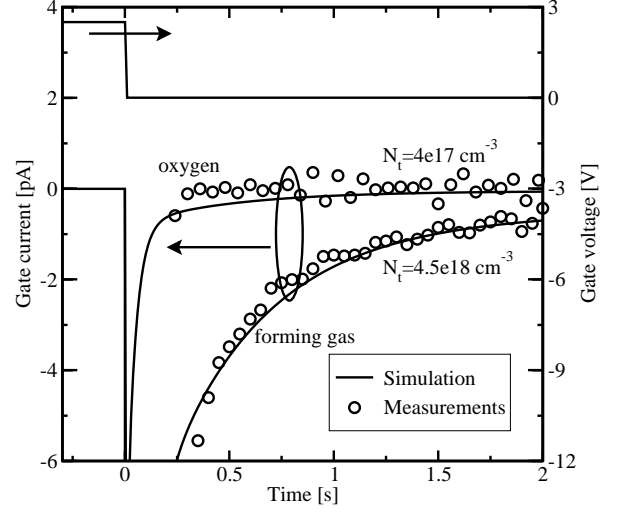$$C_i = \frac{\tau_\mathrm{m}^{-1}(z,t_i)\Delta t_i}{2}.$$

Figure 6: Transient trap charging currents for a ZrO$_2$ layer [14].

In these expressions $\Delta t_i = t_i - t_{i-1}$ 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

$$J_\mathrm{l,r}(t_i) = \mathrm{q}\int_0^{t_\mathrm{diel}} N_\mathrm{T}(x)\tau_\mathrm{l,r}^{-1}(x,t_i)\,\mathrm{d}x \qquad (6)$$

where l,r denotes the considered interface (left or right) and the time constants τ_l and τ_r are calculated from

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

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.

This model can be applied to the characterization of traps in the dielectric layer. Fig. 6 shows the step response of two MOS capacitors with ZrO$_2$ dielectrics annealed in reducing oxidizing conditions [12]. The gate voltage is first fixed at a value of $2.5\,\mathrm{V}$ to achieve a steady initial trap occupation. 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 [13]. The measurements could be fitted assuming different trap concentrations as indicated in the figure.
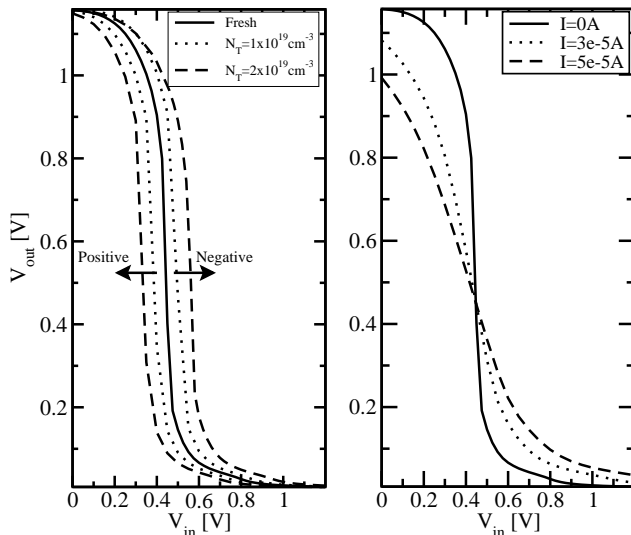
Figure 7: Transfer characteristics of a CMOS inverter for positive and negative trap charges (left) and for the case of dielectric breakdown with different leakage currents.

4 Circuit Simulations

To investigate the effect of gate dielectric leakage, wearout, and breakdown on CMOS circuits, a conventional 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 $100 \, \mathrm{Acm^{-2}}$, the effect on the inverter characteristics is hardly visible.

The trap charge, on the other hand, strongly degraded the threshold voltage as shown in the left part of Fig. 7, where positively charged traps shift the characteristics to the left (decreasing threshold voltage) and negatively charged traps shift the 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. 7 for different current values, in qualitative agreement to results presented in [15]

5 Conclusion

We presented a set of models for the description of leakage, wearout, and breakdown of dielectric layers 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. Only traps which take part in the trap-assisted tunneling process become occupied and cause threshold voltage degradation. Over time, the random creation of defects in the dielectric layer eventually results in the formation of a conducting path, which can be modeled by a current source from the source or drain to the gate electrodes. For fast transient processes, it is necessary to model charging and decharging processes. The model was used to investigate the effect of leakage, wearout, and breakdown on CMOS inverters and it was found that, while gate leakage has only minor effects on the inverter characteristics, the threshold voltage shift due to occupied traps and the high gate current after breakdown may be a show-stopper for the use of these devices in circuits. The implementation of this model into the device and circuit simulator MINIMOS-NT allows the study of leakage and degradation effects in devices and circuits based on high-κ dielectric layers.

Acknowledgments

References

[1] Y.-Y. Fan *et al.*, IEEE Trans.Electron Devices **50**, 433 (2003).

[2] L. Larcher, IEEE Trans.Electron Devices **50**, 1246 (2003).

[3] J. H. Stathis, IBM J.Res.Dev. **46**, 265 (2002).

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

[5] R. Tsu and L. Esaki, Appl.Phys.Lett. **22**, 562 (1973).

[6] M. Herrmann and A. Schenk, J.Appl.Phys. **77**, 4522 (1995).

[7] F. Jiménez-Molinos *et al.*, J.Appl.Phys. **90**, 3396 (2001).

[8] E. Rosenbaum and L. F. Register, IEEE Trans.Electron Devices **44**, 317 (1997).

[9] A. Gehring *et al.*, Microelectron.Reliab. **43**, 1495 (2003).

[10] R. Degraeve *et al.*, IEEE Trans.Electron Devices **45**, 904 (1998).

[11] J. H. Stathis *et al.*, Microelectron.Reliab. **43**, 1353 (2003).

[12] S. Harasek *et al.*, J.Vac.Sci.Technol.A **21**, 653 (2003).

[13] A. Gehring *et al.*, in *Proc. European Solid-State Device Research Conf.* (Estoril, Portugal, 2003), pp. 473–476.

[14] A. Gehring *et al.*, in *Proc. ESSDERC* (Estoril, Portugal, 2003), pp. 473–476.

[15] R. Rodríguez *et al.*, Microelectron.Reliab. **43**, 1439 (2003).

[16] J. Cai and C.-T. Sah, J.Appl.Phys. **89**, 2272 (2001).