14th Workshop on Modelling and Simulation of Electron Devices, Barcelona, Spain. October 16-17 2003.

Static and Transient Simulation of Inelastic Trap-Assisted Tunneling

Francisco Jiménez-Molinos, A. Palma, A. Gehring*, F. Gámiz, H. Kosina*, S. Selberherr*

Departamento de Electrónica y Tecnología de Computadores. Facultad de Ciencias. Avd. Fuentenueva s/n.
Universidad de Granada. 18071 Granada. Spain. e-mail: jmolinos@ugr.es

*Institute for Microelectronics, TU Vienna, Gusshausstr. 27-29, A-1040. Vienna, Austria.

Abstract *– We present an analytical model to describe static and transient trap-assisted inelastic tunneling of electrons through insulating energy barriers. The model was implemented in a device simulator in order to calculate the gate current in metal-oxide-semiconductor capacitors, the trap occupancy in the gate oxides and the charging and discharging characteristics in stressed electrically erasable programmable read-only-memories. The model shows good agreement with experimental data.*

Keywords : Stress-induced leakage current, trap-assisted tunneling, device simulation, retention time

1. INTRODUCTION

Recently, a new model for trap-assisted inelastic tunneling has been presented [1]. This model is based on the theory of non-radiative capture and emission of electrons by multiphonon processes [2]. Therefore, it does not require capture cross sections or emission rates as fitting parameters, but calculates the capture and emission probabilities for each trap, position and bias. The free parameters of the model are intrinsic trap characteristics: the trap energy level, the trap concentration and the Huang-Rhys factor.

In this paper, an adaptation of this physical model facilitating its implementation in general-purpose device simulators is shown. Because this model requires the knowledge of the wave function at the position of the trap, the main step is the transformation of the barriers in order to get the analytical expression for the electron wave functions. A fully analytical model is obtained using this approximation and it has been implemented in the general-purpose device simulator MINIMOS-NT [3]. In this way, a wide range of devices and phenomena related to trapping and detrapping processes in barriers can be studied.

The models of trap-assisted transitions are usually employed for calculating steady stress-induced leakage current (SILC) in metal-oxide-semiconductor (MOS) capacitors. A fully transient simulation implies not only the solution of the Poisson equation at each time iteration, but also the calculation of the capture and emission times, because of the change in the barrier shape and/or in the density of electrons (in a floating gate, for example). The physical model implemented to take into account the inelastic trap-assisted transitions allows this calculation because the capture and emission probabilities depend only on trap characteristics.

The present paper describes a general model suitable for the simulation of static and transient trap-assisted leakage current in MOS based devices, such as MOS capacitors or electrically erasable programmable read only memories (EEPROM), in which the barrier shape or the concentration of electrons changes over time.

2. THE MODEL

The trap-assisted tunneling model is based on [1], in which capture and emission times are calculated incorporating energy loss by phonon relaxation.

2.1. Calculation of the capture and emission times

Fig. 1 shows the basic two-step trap-assisted tunneling process through the gate oxide. Electrons are captured from the gate and lose energy due to phonon emission. The capture and emission processes are described by their respective probabilities W_c and W_e which are calculated as suggested in [4]. This requires to calculate the following transition matrix element at the trap position z_T

$$|V_e|^2 = 5pS(\hbar w)^2 \frac{a_T^2}{V} \int_{z_0 - z_T/2}^{z_0 - z_T/2} |x(z)|^2 \, dz. \qquad (1)$$

In this expression S is the Huang-Rhys factor, $\hbar w$ the phonon energy, $x(z)$ the electron wave function, $a_T = \hbar(2m_{ox}E_T)^{-1/2}$ the trap radius, z_T the respective trap cube side length and E_T the energy difference between the trap energy and the conduction band edge as shown in Fig. 1.

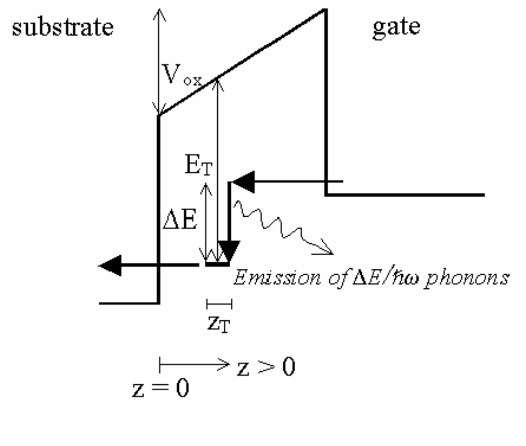

Figure 1: The trap-assisted tunneling model

Once the capture and emission probabilities have been calculated, the respective capture and emission times (t_c and t_e) can be also obtained [1].

The numerically expensive calculation of the wave functions inside the barrier is avoided by using approximate barrier shapes in the Fowler-Nordheim and direct tunneling regimes [5], leading to closed-form expressions for the matrix elements (1) and the capture and emission times.

To check the validity of the approximations, we have calculated the capture and emission times using both the numerical and the analytical model for a MOS capacitor with a gate oxide thickness of 5 nm. Fig. 2 shows that the analytical and the numerical results are in good agreement. Furthermore, the typical emission time oscillations in the Fowler-Nordheim regime can be observed.

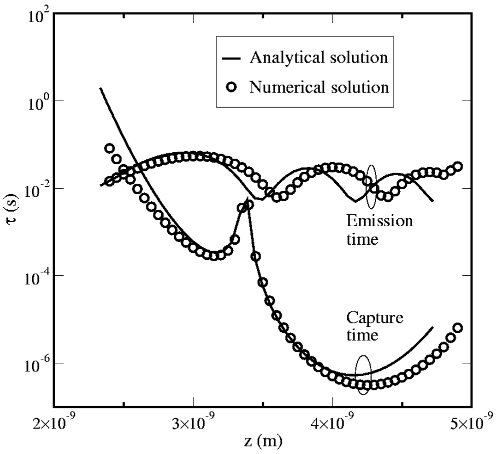

Figure 2: Capture and emission time for analytical and numerical solutions (emission takes place through a triangular barrier)

In order to calculate the tunnel current through gate oxides we have to distinguish between two cases.

2.2. Steady-state current

From the capture and emission times, the steady-state trap-assisted tunnel current is calculated by an integration across the oxide:

$$J_{TAT} = q\int_0^{t_{ox}} \frac{N_t(z)}{t_c(z)+t_e(z)}\,dz\,, \tag{2}$$

where N_t denotes the trap concentration [6]. The total tunneling current is found as the sum of the trap-assisted tunneling current and the direct tunneling current (J_t), for which we used the Tsu-Esaki expression [7] and the transmission coefficients calculated according to the WKB approximation. The total tunneling current is the sum of J_{TAT} and J_t.

2.3. Transient current

The change ratio of the concentration of occupied traps in position z is given by:

$$\begin{aligned} N_T\frac{df_T(z,t)}{dt} &= N_T(z)\left[1-f_T(z,t)\right]t_c^{-1}(z,t)- \\ &-N_T(z)\,f_T(z,t)\,t_e^{-1}(z,t)\equiv R_c(z)-R_e(z). \end{aligned} \tag{3}$$

where $f_T(z,t)$ is the occupancy function of the traps, $N_T(z)$ is their concentration and $t_c(z,t)$ and $t_e(z,t)$ are the time constants for capture and emission of electrons by a trap sited at position z. Now they depend on the time and include transitions from and to both electrodes, i.e.:

$$t_c^{-1}(z,t)=t_{ca}^{-1}(z,t)+t_{cc}^{-1}(z,t), \tag{4}$$

$$t_e^{-1}(z,t)=t_{ea}^{-1}(z,t)+t_{ec}^{-1}(z,t), \tag{5}$$

where $t_{ea}(z,t)$ and $t_{ec}(z,t)$ are the emission times to the anode and to the cathode, and $t_{ca}(z,t)$ and $t_{cc}(z,t)$ the corresponding capture times.

In the transient regime, the transition rates $R_c(z)$ and $R_e(z)$ are different. Therefore, the occupancy function can not be directly obtained from (3), as in steady state regime, but this differential equation must be solved. We should differentiate between two cases. If the capture and emission times $t_c(z,t)$ and $t_e(z,t)$ are constant over time (it requires that the potential distribution and the electron concentrations do not change), we can solve (3) analytically:

$$\begin{aligned} f_T(z,t) &= f_T(z,0)\exp\left(-\frac{t}{t_m(z,t)}\right)+ \\ &+\frac{t_m(z,t)}{t_c(z,t)}\left[1-\exp\left(-\frac{t}{t_m(z,t)}\right)\right], \end{aligned} \tag{6}$$

with $t_m^{-1}=t_c^{-1}+t_e^{-1}$.

In general, the capture and emission times change over time and a fully transient simulation has to be performed. The time constant must be evaluated for each time step and the occupancy function $f_T(z,t_i)$ calculated by using its value from the previous time step, as explained below.

We look at the change of the trap distribution at discrete times t_0 and t. Integration of (3) in the time variable between t_{i-1} and t_i and changing to discrete time steps yields

$$f_T(z,t_i) - f_T(z,t_{i-1}) \approx \\ \approx t_c^{-1}(z,t_{i-1})\Delta t_i - t_m^{-1}(z,t_{i-1})\overline{f_i}\Delta t_i, \qquad (7)$$

with $\overline{f_i} = \dfrac{f_T(z,t_i) + f_T(z,t_{i-1})}{2}$ and $\Delta t_i = t_i - t_{i-1}$. Thus it is possible to write the trap distribution over time in the following recursive manner:

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

with

$$A_i = \frac{t_c^{-1}(z,t_i)\Delta t_i}{1+C_i}, \\ B_i = \frac{1-C_i}{1+C_i}, \qquad (9) \\ C_i = \frac{t_m^{-1}(z,t_i)\Delta t_i}{2}.$$

Once the time-dependent occupancy function in the dielectric is known, the tunnel current through one of the interfaces is:

$$J_{TAT,s}(t) = q \times \\ \int_0^{t_{ox}} N_T(z)\left(t_{cs}^{-1} - f_T(z,t)\left[t_{cs}^{-1}(z) + t_{es}^{-1}(z)\right]\right)dz \qquad (10)$$

where s denotes the considered interface (either anode or cathode).

3. RESULTS

3.1. Steady-state current

Figure 3 shows the dependence on the model parameters E_T (trap energy level) and S (Huang-Rhys factor) for a fixed phonon energy of 20 meV. Figure 4 shows a comparison with experimental data taken from [8]. The fitting parameters used in the simulations, which are indicated in the figure caption, agree well with those found in [1] for other samples.

From the balance equation (3) the trap occupancy f_T can be derived which allows us to study the trap behavior in the oxide. Figure 5 shows the trap occupancy along the gate oxide of a MOS capacitor with an oxide thickness of 3 nm and the gate voltage as parameter. The fact that some trapped electrons face a trapezoidal barrier and others a triangular barrier (according to their position) for the emission process gives rise to an additional peak in the trap occupancy near the middle of the oxide.

3.2. Transient current

The resulting gate current density of a transient simulation using expression (6) is shown in figure 6, with samples of oxide thickness equal to 8.5 and 13.0 nm. Initially, the traps are assumed to be empty (for example, after a sufficient time under a flat band condition). At t = 0, the gate voltage is suddenly

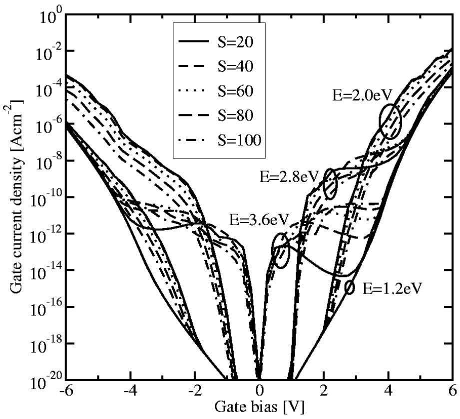

Figure 3: Influence of the trap energy level E_T and the Huang-Rhys factor (S).

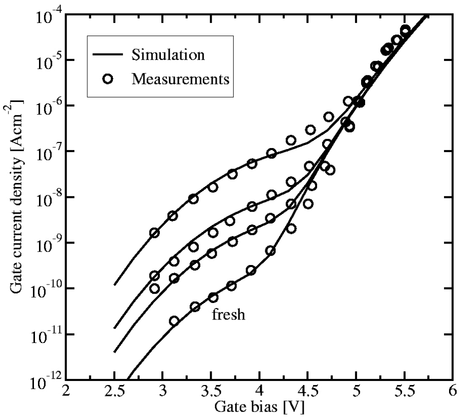

Figure 4: Comparison to measurement data of an MOS capacitor with an oxide thickness t_{ox} = *5.5 nm* [8]. The model parameters are E_T = 2.7 eV, $S\hbar w$ = 1.3 eV and a trap concentration of 9.0 × 10^{17} cm^{-3}, 1.0 × 10^{17} cm^{-3}, 3.0 × 10^{16} cm^{-3}, 3.0 × 10^{15} cm^{-3} (from top to bottom).

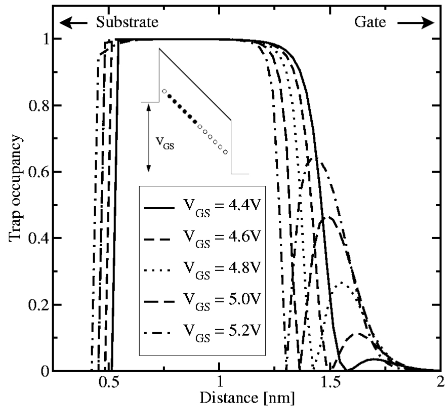

Figure 5: Trap occupancy f_T in the gate oxide as a function of the applied gate voltage V_g for a trap level of E_T = 3 eV and a gate oxide thickness of 3 nm. The approximate position of the occupied traps is depicted in the inset.

changed to a new value. It can be seen that the charging current exceeds the steady-state current by several orders of magnitude. The simulation results are compared to the measurement values taken from [9] and a good fit could be achieved using the trap parameters indicated in the figure caption.

Fig. 7 shows the gate current of a MOS capacitor when a rectangular pulse train is applied. Again, an initial flat band condition is assumed. The spikes are due to the sudden voltage change while the trap concentration remains constant. Thus, for example, in the transition from 3.0 V to 3.5 V, the barrier shape changes suddenly, and the traps are rapidly emptied because the barrier is lower and so is the occupancy factor corresponding to the new voltage. On the other hand, new traps near the cathode can now assist trap transitions and are filled. After several µs the new steady state is reached.

Finally, the model was used to simulate the discharging characteristics of an EEPROM device. Figure 8 shows the programming (left) and the retention (right) characteristics for different trap concentrations. The investigated device is sketched in the inset. Depending on the stress conditions, the retention time is significantly reduced.

4. CONCLUSIONS

A model based on first principles has been adapted and implemented into the device simulator MINIMOS-NT to study static and transient inelastic trap-assisted tunnel current in several devices.

REFERENCES

[1] F. Jiménez-Molinos et al, "*Physical Model for Trap-Assisted Inelastic Tunneling in Metal-Oxide-Semiconductor Structures*", *J. Appl. Phys.*, vol. 90, no 7, pp. 3396-3404, 2001.

[2] B. Ridley, "*Quantum Processes in Semiconductors*", Oxford University Press, 1982.

[3] Institut für Mikroelektronik, *MINIMOS-NT User's Guide, http://www.iue.tuwien.ac.at.*

[4] A. Palma et al, "*Quantum Two-Dimensional Calculation of Time Constants of Random Telegraph Signals in Metal-Oxide-Semiconductor Structures*", *Phys. Rev. B*, vol. 56, no 15, pp. 9565-9574, 1997.

[5] L. Larcher et al, "*A Model of the Stress Induced Leakage Current in Gate Oxides*", *IEEE Trans. Elec. Dev.*, vol. 48, no 2, pp. 285-288, 2001.

[6] M. Herrman 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.

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

[8] E. Rosembaum and L.F. Register, "*Mechanism of Stress-Induced Leakage Current in MOS Capacitors*", *IEEE Trans. Elec. Dev.*, vol. 44, no 2, pp. 317-323, 1997.

[9] R. Moazzami and C.Hu, "*Stress-Induced Current in Thin Silicon Dioxide Films*", IEDM p.139, 1992.

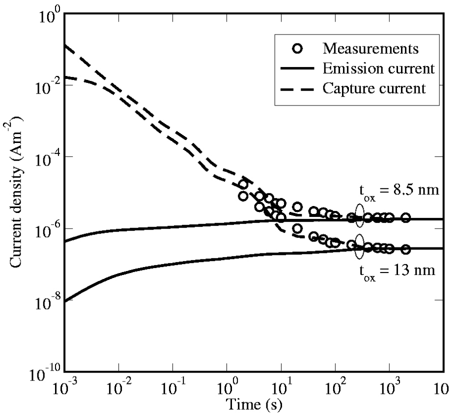

Figure 6: Transient capture and emission currents in a MOS capacitor starting with all the traps empty. Experimental data from [9]. The trap parameters are: E_T = 2.5 eV, $S\hbar w$ = 1.0 eV and N_T = 3.0 × 10^{18} cm^{-3} for the thinnest oxide and E_T = 2.5 eV, $S\hbar w$ = 1.26 eV and N_T = 1.0 × 10^{18} cm^{-3} for the other one.

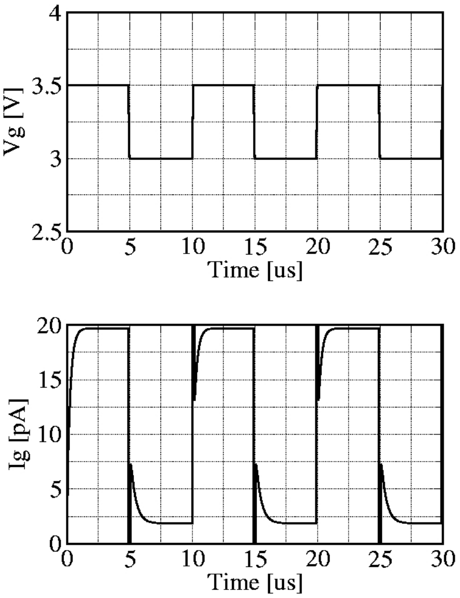

Figure 7: Transient simulation of a MOS capacitor with a gate oxide thickness of 3 nm and a trap level of E_T = 3 eV.

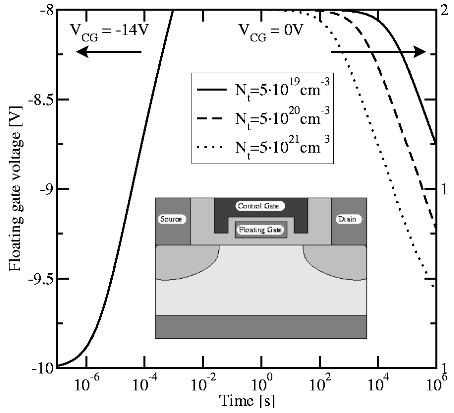

Figure 8: Programming and discharging characteristics of an EEPROM. The floating gate is charged at a control gate voltage of -14 V (left axis) and is then left floating at a control gate voltage of 0 V (right axis).