

Modeling and Simulation of Tunneling Current in MOS Devices Including Quantum Mechanical Effects

Andrea GHETTI^{†*a)}, Jeff BUDE[†], Paul SILVERMAN[†], Amal HAMAD[†],
and Hem VAIDYA[†], *Nonmembers*

SUMMARY In this paper we report on the modeling and simulation of tunneling current in MOS devices including quantum mechanical effects. The simulation model features an original scheme for the self-consistent solution of Poisson and Schrödinger equations and it is used for the extraction of the oxide thickness, by fitting CV curves, and the calculation of the tunneling current. Simulations and experiments are compared for different device types and oxide thicknesses (1.5–6.5 nm) showing good agreement and pointing out the importance of quantum mechanical modeling and the presence of many tunneling mechanisms in ultra-thin oxide MOS devices.

key words: modeling and simulation, quantum mechanical effects, tunneling current, ultra-thin oxides

1. Introduction

Sub-0.1 μm technologies require gate oxide thicknesses $t_{ox} < 3\text{ nm}$ and high doping levels [1], [2]. For such thin oxides, a significant tunneling current is expected even at normal operating conditions, and this can be viewed as one of the limiting factors of the scaling process [3]. The high dopings and strong fields confine carriers in narrow potential wells close to an inverted or accumulated silicon-oxide interface. In these conditions, they exhibit a two-dimensional (2D) behavior and must be described quantum-mechanically. Thus, proper design of CMOS device mandates the inclusion of charge quantization (CQ) effects on the potential profile and an accurate modeling of the tunneling current.

In addition, the oxide degradation process depends on the energy in the anode of injected particles [4], [5]. In order to understand the effects of different electrical stress conditions on oxide reliability, a detailed description of the many tunneling mechanisms and their energy characteristics is necessary.

Moreover, matching tunneling current simulations and experiments can be exploited for electrical characterization and physical parameter extraction (especially t_{ox} [6]), or to investigate intrinsic and stress-induced oxide defects [7].

In order to accomplish these goals an accurate de-

scription of the whole structure where the tunneling phenomenon takes place, including charge quantization effects in the accumulated or inverted cathode and anode regions, is necessary, and the absence of device dependent parameters is a desirable condition.

In this paper we report on the modeling and simulation of tunneling current in MOS devices including quantum mechanical effects. The paper is arranged as follows. Section 2 describes the modeling approach. In Sect. 3 the simulation model is applied to a number of reference devices for the purpose of validating the model using both CV and IV measurements. Then in Sect. 4 the simulation results of different device structures featuring ultra-thin (1.5–3 nm) oxides are presented. Finally Sect. 5 draws some conclusions.

2. Simulation Model

2.1 Self-Consistent Potential Profile

First, the potential profile is computed self-consistently with charge quantization effects for a one dimensional polysilicon-oxide-silicon structure. The band structure of the silicon conduction band is constituted by the usual six ellipsoidal and parabolic valleys ($m_{le} = 0.98 m_0$, $m_{te} = 0.19 m_0$, center at 0.85X), while the valence band is made of two spherical and parabolic bands (light holes $m_{lh} = 0.16 m_0$, heavy holes $m_{hh} = 0.5 m_0$). A single spherical and parabolic band is assumed for oxide. The poly is modeled as silicon but considering the appropriate work function accounting for the correct doping level and band gap narrowing [8]. Since the tunneling current is generally too small to affect the equilibrium condition it is neglected in the calculation of the potential profile, and quasi-thermal equilibrium is assumed. This is a valid assumption for our devices because, even in the worst case considered in the following, the tunneling time ($\approx 1\text{ }\mu\text{s}$) is much longer than the minimum time ($\approx 1\text{ ns}$) required for the validity of the quasi-equilibrium approximation [9]. The Fermi-Dirac statistics is adopted.

In order to account for the charge quantization effects, the potential profile is computed by solving self-consistently the Poisson and Schrödinger equations. Self-consistency is achieved through an original iteration scheme. The Poisson equation is written in the

Manuscript received November 12, 1999.

Manuscript revised February 7, 2000.

[†]The authors were with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA.

*Presently, with ST Microelectronics, 20041 Agrate Brianza, Italy.

a) E-mail: andrea.ghetti@st.com

following non linear form:

$$\begin{aligned}
 & -\nabla \cdot (\epsilon \nabla V^{k+1}) \\
 & = q \left[N_V F_{1/2} \left(\frac{q}{k_B T} (V^k - V^{k+1}) + F_{1/2}^{-1} \left(\frac{p}{N_V} \right) \right) \right. \\
 & \quad \left. - N_C F_{1/2} \left(\frac{q}{k_B T} (V^{k+1} - V^k) + F_{1/2}^{-1} \left(\frac{n}{N_C} \right) \right) \right. \\
 & \quad \left. + N(z) \right], \quad (1)
 \end{aligned}$$

where $F_{1/2}(\eta)$ is the Fermi-Dirac integral of order 1/2, $N(z)$ is the net doping, and n/p is the electron/hole concentration as provided by the Schrödinger equation.

The particular form of Eq. (1) reduces to the usual linear expression of the Poisson equation near convergence ($V^{k+1} = V^k$), but helps to reduce the numerical instability typical of self-consistent Poisson-Schrödinger iterative solution, thus speeding up convergence. For a single bias point the number of iterations needed is typically 5. The heaviest CPU burden is the computation of the Fermi-Dirac integral that can be reduced adopting a spline interpolation.

The charge distribution accounting for quantization effects is determined by solving numerically the Schrödinger equation with closed boundary conditions (BC) for the envelope function (ζ) and energy level (E_z) associated to each bound state. The Schrödinger equation is written in the framework of the effective mass approximation as:

$$-\frac{\hbar^2}{2m_z} \frac{\partial^2 \zeta}{\partial z^2} - qV(z) = E_z \zeta. \quad (2)$$

In order to compute the correct charge density in both accumulation and inversion layers using the minimum number of bound states possible, we take into account bound states up to a given threshold energy E_{cl} , above which carriers are thought to form a free gas [6] (see Fig. 1). In all the cases we found that the first 5 bound

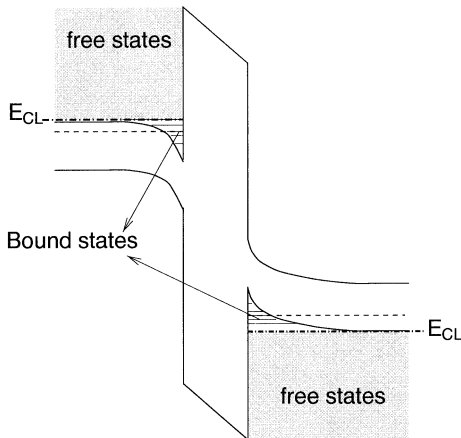


Fig. 1 Schematic representation of the simulation domain. Electrons/holes with energy above/below E_{cl} form a free gas.

states are enough to compute with good accuracy the 2D charge concentration, then E_{cl} is set equal to the sixth energy level. Thus:

$$n(z) = n_{3D}(z) + n_{2D}(z) \quad (3)$$

$$\begin{aligned}
 n_{2D}(z) &= \sum_{ij}^{E_{ij} < E_{cl}} \frac{g_j m_{d_j} k_B T}{\pi \hbar^2} \\
 &\quad \times \ln \left(\frac{1 + e^{\frac{E_F - E_{ij}}{k_B T}}}{1 + e^{\frac{E_F - E_{cl}}{k_B T}}} \right) |\zeta_{ij}(z)|^2 \quad (4)
 \end{aligned}$$

$$E_M(z) = \max(E_{cl}, E_C(z)) \quad (5)$$

$$n_{3D}(z) = N_C F_{1/2}^{(i)} \left(\frac{E_F - E_M}{k_B T}, \frac{E_{cl} - E_C(z)}{k_B T} \right) \quad (6)$$

where $E_C(z)$ is the conduction band edge, g_j is the degeneracy of the j -th valley, m_{d_j} its density of states effective mass, E_{ij} and ζ_{ij} are respectively the energy level and the corresponding envelope function of the i -th bound state in the j -th valley, and $F_{1/2}^{(i)}(z, b)$ is the incomplete Fermi-Dirac integral as defined in [10]. Hole quantization is treated in a symmetric way.

This approach gives the correct continuum and 2D density of states accounting only for the lowest quantum levels and simplifies the choice of boundary conditions for the Schrödinger equation. Quantum levels are computed exploiting the properties of Sturm sequences, and envelope functions are found by inverse iteration.

Gate depletion effects are implicitly accounted for by solving for the potential also over the gate region. Capacitance is computed differentiating charge-to-voltage characteristics.

2.2 Transmission Probability

Once the potential profile has been computed taking into account charge quantization effects, it is also used to compute the transmission probability (P_T) through the general procedure described in the following, valid for any energy barrier.

Let's consider an incoming plane wave $\zeta_i = I e^{ik_L z}$ which is partly reflected $\zeta_r = R e^{-ik_L z}$ and partly transmitted $\zeta_t = T e^{ik_R z}$ through a generic potential barrier (Fig. 2). The transmission probability is defined as the ratio of the transmitted and incoming current

$$P_T = \frac{J_{\zeta_t}}{J_{\zeta_i}} = \frac{v_{gt} |\zeta_t|^2}{v_{gi} |\zeta_i|^2} = \frac{k_R m_L |T|^2}{k_L m_R |I|^2}. \quad (7)$$

where v_g is the group velocity. The transmitted and incoming waves are related through the transfer matrix method. Following the example of [11], Eq. (2) can be solved analytically at all energies with open boundary conditions. Between each couple of mesh nodes, the potential energy ($U = -qV$) can be written as:

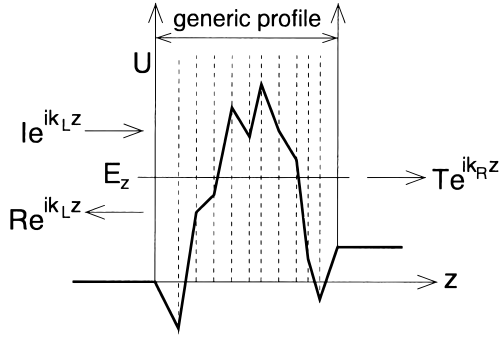


Fig. 2 Schematic representation of the potential profile for transmission probability computation.

$$U^{(i)}(z) = U_i + s_i |F_i| (z - z_i); \quad z_i < z < z_{i+1} \quad (8)$$

where

$$F_i = \frac{U_{i+1} - U_i}{z_{i+1} - z_i}; \quad s_i = \text{sign}(F_i). \quad (9)$$

Using the variable

$$u^{(i)} = s_i \left(\frac{2m_i}{\hbar^2} |F_i| \right)^{\frac{1}{3}} \left[z - z_i + \frac{U_i - E_z}{s_i |F_i|} \right] \quad (10)$$

Equation (2) becomes

$$\frac{d^2 \zeta_i}{du^{(i)2}} - u^{(i)} \zeta_i = 0, \quad (11)$$

whose solutions are, by definition, the Airy functions (Ai and Bi)

$$\zeta_i(z) = C_1^{(i)} Ai(u^{(i)}(z)) + C_2^{(i)} Bi(u^{(i)}(z)).$$

If $F_i = 0$ the solutions of Eq. (2) are:

$$\zeta_i(z) = C_1^{(i)} \cos(k_i z) + C_2^{(i)} \sin(k_i z)$$

where $k_i = \frac{\sqrt{2m_i(E_z - U_i)}}{\hbar}$ if $E_z > U_i$, or

$$\zeta_i(z) = C_1^{(i)} e^{-\alpha_i z} + C_2^{(i)} e^{\alpha_i z}$$

where $\alpha_i = \frac{\sqrt{2m_i(U_i - E_z)}}{\hbar}$ if $E_z < U_i$.

Summarizing, the solution of Eq. (2) in the range $z_i < z < z_{i+1}$ can be written as

$$\zeta_i(z) = C_1^{(i)} S_1^{(i)}(z) + C_2^{(i)} S_2^{(i)}(z), \quad (12)$$

where

$$S_1^{(i)}(z) = \begin{cases} Ai(u^{(i)}(z)) & \text{if } F_i \neq 0 \\ \cos(k_i z) & \text{if } F_i = 0 \text{ and } E_z > U_i \\ e^{-\alpha_i z} & \text{if } F_i = 0 \text{ and } E_z < U_i \end{cases}$$

$$S_2^{(i)}(z) = \begin{cases} Bi(u^{(i)}(z)) & \text{if } F_i \neq 0 \\ \sin(k_i z) & \text{if } F_i = 0 \text{ and } E_z > U_i \\ e^{\alpha_i z} & \text{if } F_i = 0 \text{ and } E_z < U_i \end{cases}$$

In order to determine the envelope function over the

whole potential barrier, boundary conditions must be enforced at each grid node [12]

$$\zeta_{i-1}(z_i) = \zeta_i(z_i)$$

$$\frac{1}{m_{i-1}} \frac{d\zeta_{i-1}(z_i)}{dz} = \frac{1}{m_i} \frac{d\zeta_i(z_i)}{dz} \quad (13)$$

that can be written in a matrix form

$$M_{i-1}(z_i) C^{(i-1)} = M_i(z_i) C^{(i)} \quad (14)$$

where

$$M_i(z) = \begin{bmatrix} S_1^{(i)}(z) & S_2^{(i)}(z) \\ \frac{1}{m_i} \frac{dS_1^{(i)}(z)}{dz} & \frac{1}{m_i} \frac{dS_2^{(i)}(z)}{dz} \end{bmatrix};$$

$$C^{(i)} = \begin{bmatrix} C_1^{(i)} \\ C_2^{(i)} \end{bmatrix}.$$

Then, the transfer matrix is

$$M = \prod_{i=1}^n M_{i-1}^{-1}(z_i) M_i(z_i) \quad (15)$$

where

$$M_0(z) = \begin{bmatrix} e^{jk_L z} & e^{-jk_L z} \\ \frac{jk_L}{m_L} e^{jk_L z} & -\frac{jk_L}{m_L} e^{-jk_L z} \end{bmatrix}$$

$$M_n(z) = \begin{bmatrix} e^{jk_R z} & e^{-jk_R z} \\ \frac{jk_R}{m_R} e^{jk_R z} & -\frac{jk_R}{m_R} e^{-jk_R z} \end{bmatrix},$$

that relates the envelope functions at both sides of the potential barrier as

$$\begin{bmatrix} I \\ R \end{bmatrix} = M \begin{bmatrix} T \\ 0 \end{bmatrix}, \quad (16)$$

from which it derives immediately

$$\frac{|T|^2}{|I|^2} = \frac{1}{|M_{11}|^2}. \quad (17)$$

When compared with the standard WKB approximation, this method gives comparable results for energies well below the energy barrier, but provides a smoother transition between tunneling and thermionic emission and accounts for possible quantum reflection above the barrier and at the anode-oxide interface.

Oxide transport is assumed ballistic. Barrier lowering is not included. Parallel momentum conservation is enforced so that $E_z = E - \hbar^2 k_{||}^2 / 2m$.

2.3 Tunneling Current Components

The tunneling current can be generally computed summing up the contribution of all k states. In the case of direct tunneling electrons (Fig. 4(a)), and because of the assumed silicon band structure, the sum over all

free states reduces to:

$$J_{3D} = \frac{qgm_d}{4\pi^3\hbar^3} \int_{E_{cl}}^{\infty} dE_{\perp} \int_0^{\infty} dE_{\parallel} \Delta f(E_{\perp} + E_{\parallel}) \cdot \int_0^{2\pi} \mathcal{P}_T(E_{\perp}, E_{\parallel}, \theta) d\theta, \quad (18)$$

where Δf is the difference of the Fermi-Dirac statistics at the two ends of the barrier.

The contribution to the tunneling current of a bound state depends on its lifetime (or decay time), that is the time spent by an electron in the bound state before tunneling away. It can be computed as [13]:

$$\tau_L = \frac{\tau(E_i)}{\mathcal{P}_T} = \frac{\int_a^b \sqrt{\frac{2m_{\perp}}{E_i - E_C(z)}} dz}{\mathcal{P}_T}, \quad (19)$$

where a and b are the classical turning points (see Fig. 3(a)). \mathcal{P}_T in Eq. (19) is computed following the general procedure illustrated in the previous subsection, but with the appropriate energy profile (profile between e and d in Fig. 3(a)) that includes most of the potential well, the oxide layer and the cathode field region. Although this algorithm considers plane waves at the left of e and at the right of d , the inclusion of

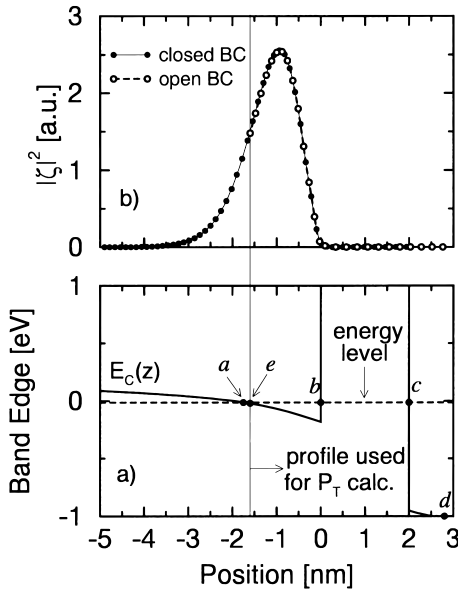


Fig. 3 (a) Potential profile for a test case device (N-MOS capacitors with $N_A = 10^{17} \text{ cm}^{-3}$, $t_{ox} = 2 \text{ nm}$ and $V_G = 1 \text{ V}$). a and b are the classical turning points. e is the first grid point whose potential is below the energy level of the bound state under consideration. The energy profile between e and d is used to compute the transmission probability in Eq. (19) for the bound state being considered. (b) Wave functions used to compute the carrier concentrations (solution of Eq. (2) with closed boundary conditions (●)) and to compute the transmission probability (solution of Eq. (2) with open boundary conditions (○)).

most of the potential well makes the wave function hitting the interface and used to compute the transmission probability (solution of Eq. (2) with open boundary conditions) the same as the wave function used to compute the carrier concentrations (solution of Eq. (2) with closed boundary conditions), as demonstrated in Fig. 3(b), thus accounting also for the non planarity of the bound state wave function. Therefore, the tunneling current from all bound states is given by

$$J_{2D} = \sum_i^{E_i < E_{cl}} \frac{qg_i m_{d_i}}{2\pi^2 \hbar^2 \tau(E_i)} \int_0^{\infty} dE_{\parallel} \Delta f(E_i + E_{\parallel}) \cdot \int_0^{2\pi} \mathcal{P}_T(E_i, E_{\parallel}, \theta) d\theta. \quad (20)$$

The corresponding hole components are computed in a symmetric way accounting for the appropriate band structure and potential profile (Fig. 4(b)).

There is no conceptual difference between conduction band and valence band electrons, apart from their different band structure. Therefore the valence to conduction band tunneling current (Fig. 4(c)) can be computed as in (18) taking into account the valence band structure and the appropriate potential profile.

Trap-assisted electron tunneling is also included in the simulation (Fig. 4(d)). It is modeled according to a two step model similar to [14]. In steady state condition, the current flowing via the traps located at z is given by the balance between the tunnel-in current ($J_{in}(z)$) and the tunnel-out current ($J_{out}(z)$). Then

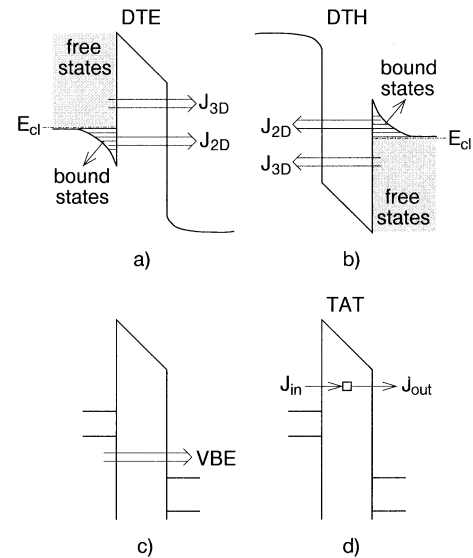


Fig. 4 Tunneling components included in the model: (a) direct tunneling of electrons from free and bound states (DTE); (b) direct tunneling of holes from free and bound states (DTH); (c) valence to conduction band electron tunneling (VBE); (d) elastic trap-assisted tunneling (TAT).

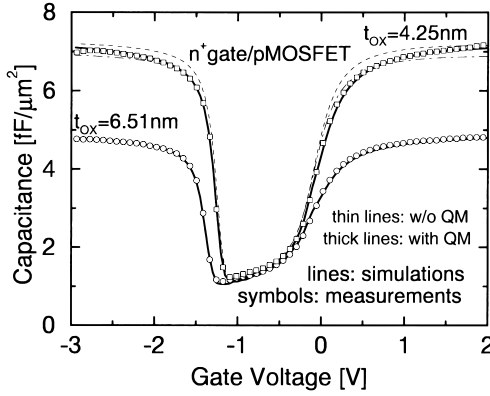


Fig. 5 Quasi static CV measurements (symbols) and simulations performed with (thick solid lines) and without (thin lines) the inclusion of CQ effects for n^+ poly P-MOS transistors. t_{ox} given by the best QM fitting is indicated along each curve. Classical simulations fitting different parts of the experimental data provide $t_{ox} = 4.4$ nm (dashed thin line) and $t_{ox} = 4.6$ nm (dot-dashed thin line).

$$J_{TAT} = \int_0^{t_{ox}} \sigma N_{tr}(z) \frac{J_{in}(z) J_{out}(z)}{J_{in}(z) + J_{out}(z)} dz, \quad (21)$$

where σ is the trap capture cross section and N_{tr} is the trap concentration. J_{in} and J_{out} are computed according to (18) or (20) depending on the initial state.

In summary, the model computes the following tunneling current components (Fig. 4): conduction band electrons (DTE) and valence band holes (DTH) tunneling from bound and free states, valence band electron (VBE) tunneling, and trap-assisted electron tunneling (TAT). This model has been implemented in the simulation program called QUASI.

3. Calibration and Validation

When the potential profile is known, the computation of the tunneling current shown in the previous section depends only on two parameters: the oxide effective mass (m_{ox}) and barrier height (Φ_B). In the framework of the present model, m_{ox} and Φ_B are treated as fitting parameters. However, they are device independent. Therefore, we must first find the numerical value of these two parameters and, above all, verify that this model is able to reproduce the measured tunneling current with a unique set of parameters when the potential profile is known from an *independent* determination.

To this purpose, we first analyzed a set of relatively thick oxide devices because of the availability of both CV and IV measurements. The poly doping profile of these devices was determined with spreading resistance measurements, while substrate doping was estimated with process simulation. These profiles were found to be in good agreement with CV data (Fig. 5). First, physical t_{ox} was extracted by fitting experimental CV data with simulations including CQ effects (QM sim-

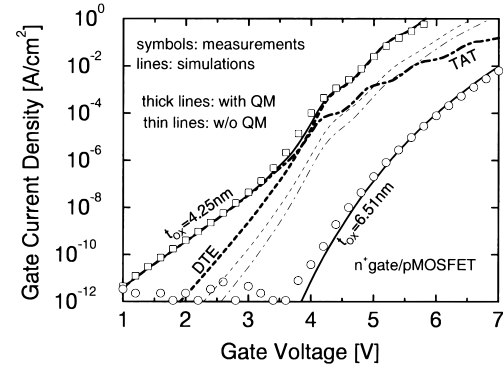


Fig. 6 Tunneling IV measurements (symbols) and simulations (lines) of the same devices of Fig. 5 in accumulation regime. Quantum mechanical simulations (thick lines) featuring t_{ox} given by the best QM fit of Fig. 5 are in good agreement with experiments. Classical DTE simulations featuring $t_{ox} = 4.4$ nm (dashed line) and $t_{ox} = 4.6$ nm (dot-dashed line) underestimate the tunneling current of more than 1 order of magnitude. The thinner device shows a native trap assisted component (TAT).

ulations). Figure 5 reports CV measurements and simulations of n^+ poly P-MOS transistors. The excellent agreement in both accumulation and inversion regimes is due to the inclusion of CQ effects for both electrons and holes, and the adoption of the Fermi-Dirac statistics. Simulations performed without CQ effects but still adopting the Fermi-Dirac statistics (classical simulations), do not match data as well and provide a larger t_{ox} .

Then, tunneling IV characteristics were simulated *with the same parameters* and adopting $m_{ox} = 0.5 m_0$ and $\Phi_B = 3.1$ eV (Fig. 6). In the case of the thickest oxide, DTE fits very well the experiments. For the 4.25 nm device, DTE simulated including CQ effects fits experimental data at high enough voltages, also reproducing the characteristic oscillations due to quantum mechanical reflection at the oxide/anode interface [15]–[17]. On the other hand, classical simulations provide a tunneling current more than 1 order of magnitude smaller.

The extra current with respect to DTE featured by experiments at low voltages is attributed to a TAT component due to native traps [18]. It was computed assuming a trap distribution uniform in space and energy, and an elastic mechanism [19]. The product σN_{tr} was used as a fitting parameter. The best fit was achieved with $\sigma N_{tr} = 4$ cm⁻¹.

4. Simulation of Ultra-Thin Oxides

In this section, the model is applied to the simulation of ultra-thin gate oxide MOS devices. Differently from the previous section, reliable CV are not available because of the large leakage current or of the small area, and several components of the tunneling current are present and can be switched on and off depending on

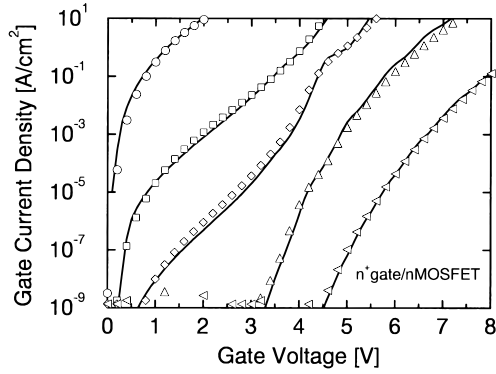


Fig. 7 Tunneling IV measurements (symbols) and simulations (lines) for n^+ poly N-MOS transistors in inversion regime. Simulations were performed adopting the t_{ox} given by ellipsometry. Oxide thicknesses are (in nm): 1.56 (\circ), 2.47 (\square), 3.27 (\diamond), 4.59 (\triangle), 6.56 (∇). Devices with t_{ox} of 2.47 nm and 3.27 nm feature a TAT component at low gate voltage as explained in [18].

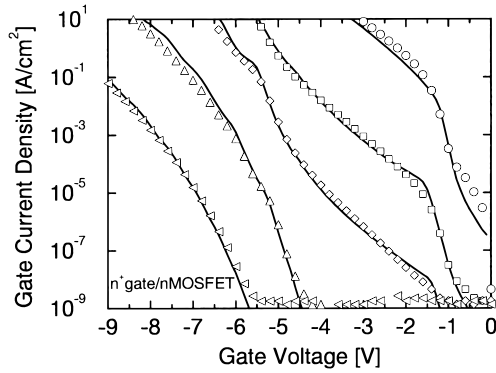


Fig. 8 Tunneling IV measurements (symbols) and simulations (lines) for the same devices, symbols and simulation parameters of Fig. 7, but in accumulation regime.

the applied voltage and the gate doping polarity.

Figures 7, 8 compare experimental data with simulations for n^+ poly N-MOS transistors featuring t_{ox} in the range 1.5–6.5 nm, in inversion and accumulation regime respectively. Here, DTE is always the dominant component. Simulations were performed adopting the average t_{ox} given by ellipsometry on a set of 64 devices. The agreement is more than satisfactory, considering the non uniformity of the thin oxide across the wafer.

If the oxide is thin enough, it is possible to measure the valence band electron tunneling current component (VBE) by biasing in inversion a N-MOS transistor. In this case, the substrate current (I_B) is made of the holes left behind by the valence band tunneling electrons [20], [21] and, therefore, it coincides with the VBE component. Figure 9 compares I_G and I_B measurements to simulations for the same device with $t_{ox} = 2.47$ nm in Figs. 7, 8. VBE simulation agrees pretty well with I_B . It was computed adopting $m_{ox} = 0.42 m_0$.

As a last example, Fig. 10 reports experiments and simulations of p^+ gate over lightly doped (10^{15} cm^{-3}) p-

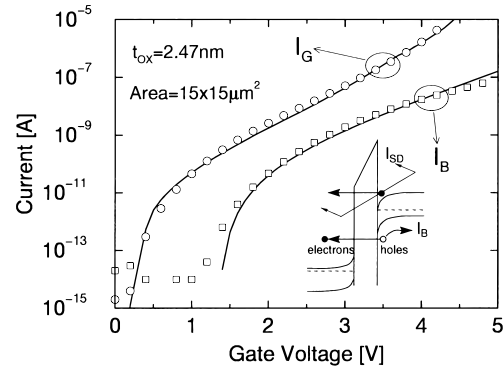


Fig. 9 I_G and I_B measurements (symbols) and simulations (lines) for the device with $t_{ox} = 2.47$ nm of Fig. 7. The simulation of I_G accounts for both DTE and TAT components. VBE simulation coincides with I_B .

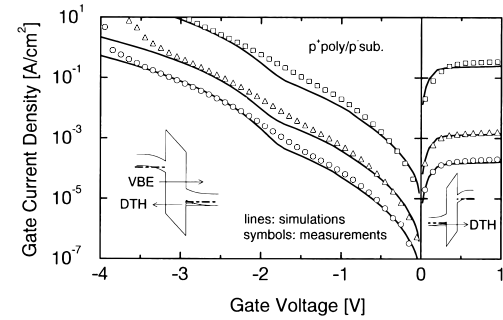


Fig. 10 Tunneling IV measurements (symbols) and QM simulations (lines) for p^+ poly over p-substrate MOS capacitors. Matching simulations with experiments resulted in the following values for t_{ox} (in nm) for the different curves (in parenthesis the corresponding value from ellipsometric measurements) from top to bottom: 1.65(1.53), 1.86(1.8), 2(1.97).

substrate capacitors with t_{ox} in the range 1.5–2 nm. In this case, the results of spreading resistance measurements and SIMS analysis indicate that the p^+ poly active doping concentration is approximately constant at $5 \times 10^{19} \text{ cm}^{-3}$. The oxide thickness was used as a fitting parameter. The values of t_{ox} providing the best fit are reported in the caption of Fig. 10 in comparison with the average values provided by ellipsometric measurements. They are well inside the spread of the t_{ox} distribution due to the non uniformity of the thin oxide across the wafer.

When the gate bias is positive, holes are injected from the gate valence band into the substrate valence band (DTH). In the voltage range shown in Fig. 10, the DTH component is almost constant because most of the applied bias drops across the substrate depletion region due to the very low doping. The DTH component was computed assuming $m_{ox} = 0.42 m_0$ and $\Phi_B = 4.8 \text{ eV}$.

On the other hand, when the applied gate voltage is negative, the gate current is the sum of the DTH and VBE components. This situation is shown in more detailed in Fig. 11, that reports all the gate

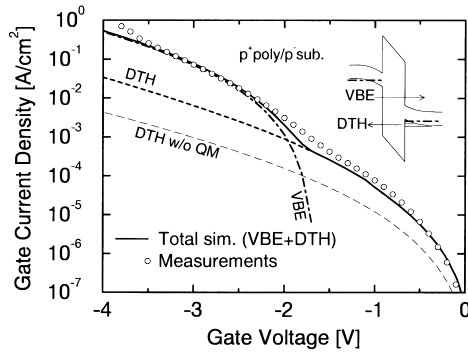


Fig. 11 Analysis of the tunneling current components for a p^+ -poly over p-type substrate MOS capacitor with an estimated $t_{ox} = 2$ nm. Measured (\circ) and simulated (solid line) gate tunneling current are reported together with the simulated DTH (dashed line) and VBE (dot-dashed) components. The thin long dashed line represents the simulated DTH component if hole quantization is neglected.

current components for the 2 nm device of Fig. 10. For $-1.5 \text{ V} < V_G < 0 \text{ V}$, substrate hole direct tunneling dominates the conduction, while VBE will eventually take over when the gate valence band faces the substrate conduction band.

Notice that the exclusion of CQ effects leads to an underestimation of DTH, pointing out the importance of accounting for CQ effects in the calculation of tunneling current even for such low doping. In the QM simulation, DTH is increased because the ground level is separated from the valence band edge, so that the effective barrier seen by most carriers is smaller.

5. Conclusion

We have presented a model for the self-consistent simulation of tunneling current and quantization effects of both electrons and holes in MOS structures. It is suitable for both accumulation and inversion layers and does not rely on device dependent parameters. We also showed that it is consistent with CV measurements and ellipsometry data.

The model provides a very good agreement with experimental tunneling IV data over a broad variety of device types and t_{ox} with a unique set of parameters, provided that an accurate modeling of quantization effects and of all tunneling mechanisms are included in the simulation.

Acknowledgement

The authors would like to thank C.T. Liu for many helpful discussions, and G. Timp and T.W. Sorsch for providing some of the devices tested.

References

[1] Semiconductor Industry Association, The National Technology Roadmap for Semiconductors, SIA, 1997.

[2] C.T. Liu, "Circuit requirement and processing challenges of thin gate dielectrics for ultra-small FET's," IEDM Technical Digest, p.747, 1998.

[3] Y. Taur, Y.-J. Mii, D.J. Frank, H.-S. Wong, D.A. Buchanan, S.J. Wind, S.A. Rishton, G.A. Sai-Halasz, and E.J. Nowak, "CMOS scaling into the 21st century: 0.1 μm and beyond," IBM J. Research and Development, vol.39, pp.245–260, 1995.

[4] J. Bude, B. Weir, and P. Silverman, "Explanation of stress-induced damage in thin oxides," IEDM Technical Digest, p.179, 1998.

[5] D.J. DiMaria, "Explanation for the polarity dependence of breakdown in ultrathin silicon dioxide films," Applied Physics Letters, vol.68, no.21, p.3004, 1996.

[6] C. Bowen, C.L. Fernando, G. Klimeck, A. Chatterjee, D. Blanks, R. Lake, J. Hu, J. Davis, M. Kulkarni, S. Hattangady, and I.C. Chen, "Physical oxide thickness extraction and verification using quantum mechanical simulation," IEDM Technical Digest, p.869, 1997.

[7] C.T. Liu, A. Ghetti, Y. Ma, G. Alers, C.P. Chang, K.P. Cheung, J.I. Colonell, W.Y.C. Lai, C.S. Lai, R. Liu, H. Vaidya, and J.T. Clemens, "Intrinsic and stress-induced traps in the direct tunneling current of 2.3–2.8 nm oxides and unified characterization methodologies of sub-3 nm oxides," IEDM Technical Digest, pp.85–88, 1997.

[8] N. Lifshitz, "Dependence of the work-function difference between the polysilicon gate and silicon substrate on the doping level in polysilicon," IEEE Trans. Electron Devices, vol.32, no.3, pp.617–621, 1985.

[9] J. Suñé and X. Oriols, "Is it reasonable to assume equilibrium conditions for the modeling of ultrathin oxide MOS devices?" Proc. SISC Conference, p.3.2, 1999.

[10] M. Goano, "Series expansion of the Fermi-Dirac integral $\mathcal{F}_j(x)$ over the entire domain of real j and x ," Solid State Electron., vol.36, no.2, p.217, 1993.

[11] W.W. Lui and M. Kukuma, "Exact solution of the Schrödinger equation across an arbitrary one-dimensional piecewise-linear potential barrier," J. Appl. Phys., vol.60, no.5, p.1555, 1986.

[12] B.R. Nag, "Boundary condition for tunneling through potential barriers in nonparabolic semiconductors," Applied Physics Letters, vol.59, pp.1620–1622, 1991.

[13] F. Rana, S. Tiwari, and D.A. Buchanan, "Self-consistent modeling of accumulation layers and tunneling currents through very thin oxides," Applied Physics Letters, vol.69, no.8, p.1104, 1996.

[14] E. Suzuki and D.K. Schroder, "Carrier conduction in ultrathin nitrided oxide films," J. Appl. Phys., vol.60, no.10, p.3616, 1986.

[15] J. Maserjian, "Tunneling in thin MOS structures," J. Vac. Sci. Technol., vol.11, no.6, p.996, 1974.

[16] G. Lewicji and J. Maserjian, "Oscillations in MOS tunneling," J. Appl. Phys., vol.46, no.7, p.3032, 1975.

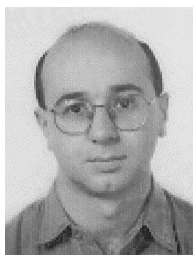
[17] J. Maserjian and N. Zamani, "Behavior of the Si/SiO₂ interface observed by Fowler-Nordheim tunneling," J. Appl. Phys., vol.53, pp.559–567, 1982.

[18] A. Ghetti, E. Sangiorgi, T.W. Sorsch, and I. Kizilyalli, "The role of native traps on the tunneling characteristics of ultra-thin (1.5–3 nm) oxides," Microelectronic Engineering, vol.48, p.31, 1999.

[19] A. Ghetti, M. Alam, J. Bude, D. Monroe, E. Sangiorgi, and H. Vaidya, "Analysis of trap-assisted conduction mechanisms through silicon dioxide films using quantum yield," IEDM Technical Digest, p.723, 1999.

[20] B. Eitan and A. Kolodny, "Two components of tunneling current in metal-oxide-semiconductor structures," Applied Physics Letters, vol.43, pp.106–108, 1983.

- [21] I.C. Chen, S. Holland, K.K. Young, C. Chang, and C. Hu, "Substrate hole current and oxide breakdown," *Applied Physics Letters*, vol.49, no.11, p.669, 1986.



Andrea Ghetti was born in Forlì, Italy, in 1968. He received the laurea degree and the Ph.D. degree in Electrical Engineering from the University of Bologna, Italy, in 1994 and 1998, respectively. In 1994 he joined the TCAD department of Intel Corporation, Santa Clara, CA, for three months as visiting student. From March 1997 to March 2000 he held a post-doctoral position at Lucent Technologies, Bell Laboratories, Murray

Hill, NJ. His research interests are in the field of device modeling, simulation and characterization with particular regard to hot carrier effects, tunneling in MOS structures and oxide reliability.



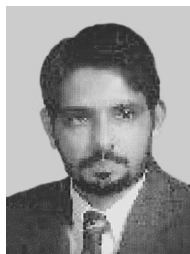
Jeff Bude was born in St. Louis, Missouri on July 26, 1966. He received the B.S. degree with highest honors in Electrical Engineering at the University of Illinois, Urbana, in 1987. He received the M.S. degree in 1989 and the Ph.D. degree in 1992, both from the University of Illinois, Urbana. In 1992 he joined Bell Laboratories in Murray Hill, NJ as a member of the technical staff in the ULSI Technology Research Department. He has been

involved in simulation and design of advanced devices and in research emphasizing hot carrier effects and reliability.



Paul J. Silverman was born in Philadelphia, PA in 1948 and holds a M.S. in physics from Purdue University. He is working on electrical device characterization and dielectric reliability testing at Bell Labs, Lucent Technologies, Murray Hill, NJ.

Amal Hamad His biography is not available.



Hem Vaidya was born in India. He received the MSEE degree from University of Central Florida in 1998. He is currently working at Lucent Technologies, Bell Laboratories working on sub-micron technology integration and EDRAM.