PAPER *Special Issue: SISPAD'99*

Simulation of Direct Tunneling through Stacked Gate Dielectrics by a Fully Integrated 1D-Schrödinger-Poisson Solver

Andreas WETTSTEIN†, **Andreas SCHENK**†, *and* **Wolfgang FICHTNER**†, *Nonmembers*

SUMMARY We compare the numerical results for electron direct tunneling currents for single gate oxides, ON- and ONO-structures. We demonstrate that stacked dielectrics can keep the tunneling currents a few orders of magnitude lower than electrostatically equivalent single oxides. We also discuss the impact of gate material and of the modeling of electron transport in silicon.
key words: stacked dielectrics, direct tunneling, Schrödinger equation

1. Introduction

Scaling the gate SiO_2 thickness below 2 nm will lead to high direct tunneling leakage currents and thus will require to use alternative materials with high ϵ like nitride, TaO_2, TiO_2, or BST. This allows to increase the physical thickness d (and thus the tunneling barrier) of the layer while keeping the same equivalent thickness d_{equiv},

$$d_{\text{equiv}} = \frac{\epsilon_{\text{ox}}}{\epsilon} d,$$

where ϵ_{ox} and ϵ are the dielectric constants of silicon dioxide and the material in question, respectively. In practice, it is also desirable to keep the excellent properties of the Si-SiO_2 interface. To combine the best of both worlds, stacked dielectrics with an oxide layer directly on the silicon substrate and a high-ϵ layer on top of it are favorable.

From a simulation point of view, the problem with stacked dielectrics is that simple WKB-based formulas for tunneling cease to work, because they rely on the assumption of a trapezoidal tunnel barrier. It is possible to generalize these models by splitting the barrier into several trapezoidal parts. However, the numerical cost increases with increasing complexity of the barrier. For complicated structures, more general approaches become more attractive as their performance disadvantage decreases.

We applied a tunneling formula based on the numerical solution of the 1D-Schrödinger equation to the problem, where all barrier shapes are admissible. We studied the tunneling currents for electrostatically equivalent oxide, ON and ONO barriers and analyzed the results.

†The authors are with the Institut für Integrierte Systeme, ETH Zürich, CH-8092 Zürich, Switzerland

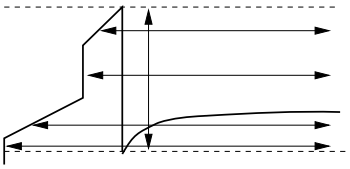

Fig. 1 Band structure dependent energy range for which the Schrödinger equation is solved (vertical arrow) and Schrödinger zone as a function of energy (horizontal arrows).

2. Theory

We start from the solutions of the 1D-Schrödinger equation

$$\left(-\frac{\partial}{\partial z}\frac{\hbar^2}{2m_z^\nu(z)}\frac{\partial}{\partial z} + V(z)\right)\psi_i^\nu = E_i^\nu \psi_i^\nu \qquad (1)$$

(z-direction perpendicular to the interface), where ν labels the silicon conduction band valleys, m_z^ν is the z-component of the effective mass and V the total potential, comprised of a band edge part and electrostatic contributions. The eigensolutions (ψ_i^ν, E_i^ν) are computed for energies from the lowest potential point in the Si up to the maximum potential in the structure. The Schrödinger equation is solved for a zone that covers the entire channel and extends up to the (energy-dependent) gate-side turning point. Fig. 1 sketches energetic and spatial regions of integration.

At the end points of the Schrödinger zone, we impose boundary conditions of the form $|\psi_i'^\nu/\psi_i^\nu| = \sqrt{2m_z^\nu|E_i^\nu - V|}/\hbar$, where the potential V and the z-component of the effective mass, m_z^ν, are taken at the respective end points of the Schrödinger zone. These boundary conditions contain the implicit assumption that outside the Schrödinger zone, V and m_z^ν keep the values they take at its out-most point. The deviation of the actual values from these assumed values represents a perturbation for the bound states. Due to this perturbation the states decay into the gate.

The decay rate is determined using Bardeen's method of the time dependent perturbation theory [1]. Fermi's Golden Rule for the decay rate $\Gamma_{j\nu\to k}$ from a state (j,ν) into a state k reads

$$\Gamma_{j\nu\to k} = \frac{2\pi}{\hbar}\left|M_{k,j\nu}\right|^2 \delta\left(E_j^\nu - E_{\text{G}}^k\right), \qquad (2)$$

where the matrix element $M_{k,j\nu}$ is given as

$$M_{k,j\nu} = \frac{\hbar^2}{2m_z(z_0)}\left[\Psi_{\rm G}^k \frac{\partial\psi_j^\nu}{\partial z} - \psi_j^\nu \frac{\partial\Psi_{\rm G}^k}{\partial z}\right]_{z=z_0}. \quad (3)$$

z_0 is the gate-side end point of the Schrödinger zone. The special aspect of Bardeen's method that shows up in this equation is that two different complete systems of states are used. The first of them is determined numerically using the Hamiltonian and the boundary conditions described above. States in this base are labeled by a subband index j and a valley index ν and can be thought to be located in the channel. The other system is the gate system of states, denoted by a subscript 'G'. For this system we make a WKB approximation assuming a potential that agrees to the correct potential in the gate and the insulator barrier. We find

$$\left[\Psi_{\rm G}^E(z_0)\right]^2 = \frac{4m_{\rm G}}{L_{\rm G}\hbar^2}\frac{m_{\rm ins}^2(E-E_{\rm G})}{(k_{\rm G}^E)^2 m_{\rm ins}^2 + (k_{\rm ins}^E)^2 p^E m_{\rm G}^2},$$

$$\frac{\partial\Psi_{\rm G}^E}{\partial z}(z_0+) = \frac{m_{\rm G}}{m_{\rm ins}} k_{\rm G}^E \sqrt{p^E}\,\Psi_{\rm G}^E(z_0),$$

where

$$k_{\rm G}^E = \sqrt{2m_{\rm G}(E-E_{\rm G})}/\hbar,$$
$$k_{\rm ins}^E = \sqrt{2m_{\rm ins}|V(z)-E|}/\hbar$$
$$p^E \approx 1 - \frac{1}{(k_{\rm G}^E)^2}\frac{\partial k_{\rm G}^E}{\partial z}.$$

$E_{\rm G}$ and $m_{\rm G}$ are the conduction band energy and the effective mass in the gate, respectively. $m_{\rm ins}$ is the effective mass in the barrier layer next to the gate and $L_{\rm G}$ is the size of the gate. We assume the spectrum in the gate to be characterized by the 1D density of states

$$D_1(E+E_{\rm G}) = \sqrt{\frac{m_{\rm G}^\nu}{2E}}\frac{L_{\rm G}}{\hbar\pi}.$$

and thus have replaced the discrete index k in Eq. (3) by a the continuous energy index E.

Then, from Eq. (2), the current density can be obtained by an integration over all states,

$$j = \frac{q\sqrt{2m_{\rm G}}L_{\rm G}}{8\pi}\sum_{i,\nu} m_{xy}^\nu \int_0^\infty dE\,\frac{\Theta(\tilde{E}-E_{\rm G})}{\sqrt{\tilde{E}-E_{\rm G}}}\times$$
$$\times\Delta f(E_i^\nu+E)\left|\frac{\Psi_{\rm G}^{\tilde{E}}}{m_z}\frac{\partial\psi_i^\nu}{\partial z} - \frac{\psi_i^\nu}{m_z}\frac{\partial\Psi_{\rm G}^{\tilde{E}}}{\partial z}\right|_{z=z_0}^2, \quad (4)$$

with $\tilde{E} = E_i^\nu + (1 - m_{xy}^\nu/m_{\rm G})E$. Δf denotes the difference of the Fermi functions in gate and channel, and m_{xy}^ν is the effective mass in the cannel in direction parallel to the interface.

3. Implementation

The numerical procedure to solve Eq. (1) starts by guessing an eigenvalue E_i^ν and inserting it into the Schrödinger equation. The resulting ordinary differential equation is then solved for the left and the right part of the Schrödinger zone using the CPM(1) method [4]. If the partial solutions can be matched at a properly chosen intermediate point, the guessed value was really an eigenvalue; otherwise, the mismatch allows to compute a better estimate [2].

In order to calculate the correct tunnel current from Eq. (1) and Eq. (4), a sufficient number of eigensolutions has to be considered. For strong negative gate biases this number can be quite large (in the hundreds), because most of the electrons injected from the gate have a high energy.

For 2D-structures, tunneling currents can be obtained by splitting the device into several slices perpendicular to the insulator-silicon interface as sketched in Fig. 2. For each slice, a 1D quantum mechanical problem is solved. The total tunneling current is obtained by summing over all slices. To enable slicing of the Schrödinger zone, we use a tensor grid in this region. This allows us to take the grid lines perpendicular to the interface to solve the quantum mechanical problem on, without the need of interpolation.

While this way we are able to simulate idealized geometries where each material layer has a uniform thickness, we are not able to treat realistic devices with nonplanar interfaces and laterally varying layer thicknesses [5], due to the intrinsic one-dimensionality of our approach. Since tunneling currents are very sensitive to the insulator thickness, our method is quantitatively valid only for simple test devices. For realistic MOSFETs it can only give qualitative information.

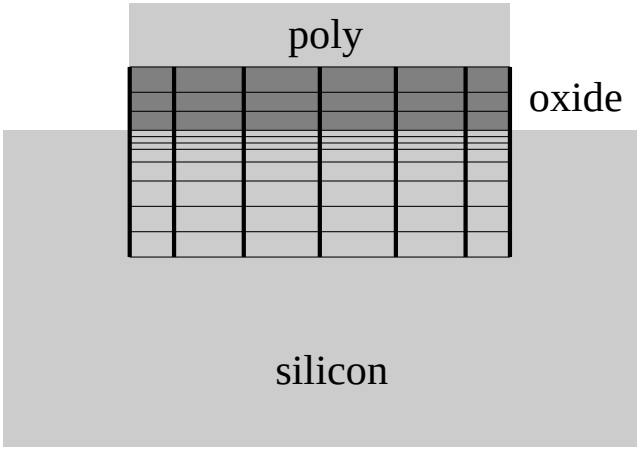

Fig. 2 2D devices are treated by solving the 1D-problem on 1D slices, indicated by the bold vertical lines of the tensor grid in the channel region.

4. Simulation Results

We present results for capacitors with n^+-poly and aluminum gates, respectively, and for an n-channel MOS-

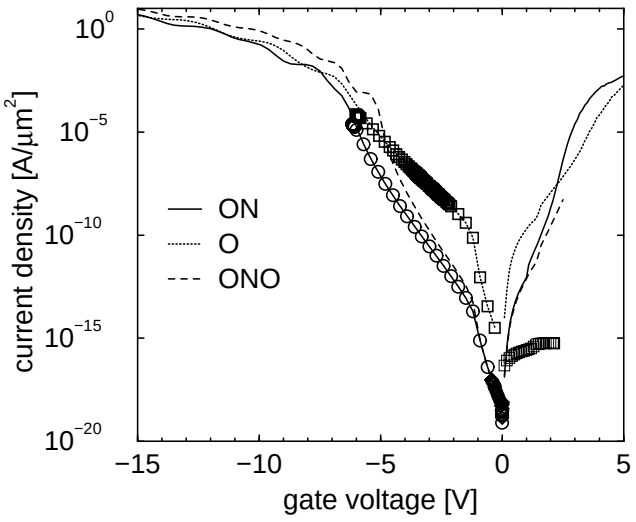

Fig. 3 Simulated gate currents of MIS capacitors with poly gate containing oxide (O), oxide-nitride (ON), and oxide-nitride-oxide (ONO) dielectrics with physical thickness as indicated in the text. The lines are results for fixed quasi Fermi potential, symbols are for full self-consistent computation.

FET with n^+-poly gate, obtained with the device simulator DESSIS$_{-ISE}$[3]. In the program, the flux of tunneling carriers (Eq. 4) across the interface between silicon and dielectric yields a local generation-recombination rate. Three different gate dielectrics are compared with each other – a 2.04 nm single SiO_2, an oxide-nitride (ON) stack with 1 nm SiO_2 and 2 nm Si_3N_4 on top, and an oxide-nitride-oxide (ONO) stack with 0.5 nm SiO_2, 2 nm Si_3N_4, and 0.5 nm SiO_2 –, all systems having an equivalent oxide thickness of 2.04 nm.

In all cases the following parameters were used: $\epsilon_{Si} = 11.7$, $\epsilon_{Ox} = 3.9$, $\epsilon_{Ni} = 7.5$, $m_{Ox} = m_{Ni} = 0.42\,m_0$, $m_{Alu} = 0.32\,m_0$, $\chi_{Si} = 4.05$ eV, $\chi_{Ox} = 0.9$ eV, $\chi_{Ni} = 1.9$ eV, $\chi_{Poly} = 4.05$ eV, $\Phi_{Alu} = 4.25$ eV, $E_{F,Alu} = 11.7$ eV, $E_{G,Ox} = 9$ eV, $E_{G,Ni} = 5$ eV. The effective masses in the poly gate were treated as in silicon. The doping concentrations were chosen as $N_D^+ = 3\times10^{20}\,cm^{-3}$ in the poly gate and $N_A^- = 5\times10^{17}\,cm^{-3}$ in silicon.

Fig. 3 shows the gate currents of the MIS capacitors with poly gate. The lines were obtained neglecting the feedback of the tunnel current on the Fermi level, whereas the symbols give the results of a fully self-consistent treatment. Both approaches give identical currents as long as the tunnel barrier is the dominant resistance in the structure. This holds for a wide range of negative and for very small positive gate voltages. For negative gate voltages the electrons are injected from the gate into the silicon. From there they can easily flow to the back contact without encountering another barrier. Therefore, the current is mainly determined by the insulator barrier and only at very strong negative voltages limitations for the electron flux in the silicon become visible (see Fig. 5). For positive gate voltages the structure is similar to a reverse biased pn-diode. Only for very small voltages the current is determined by the insulator barrier. Even for moderate positive bias the current is limited by the generation of electrons in the depletion zone. Therefore, the full self-consistent computation of the electron transport gives currents significantly lower than those obtained with the simplified treatment. However, the results of the latter are not completely meaningless. In MISFETs, electrons can be provided via source and drain contacts. Thus, tunneling currents obtained by full self-consistent calculations of MISFETs (see Fig. 7) are similar to those obtained for MIS-diodes in the simplified treatment.

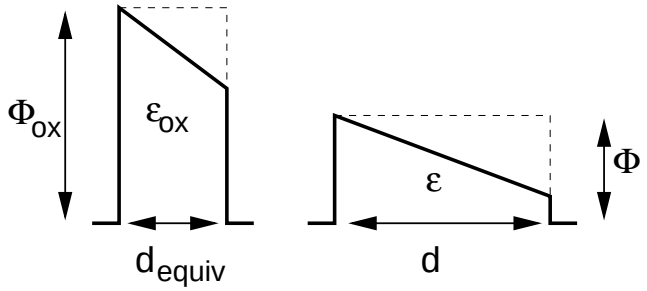

Fig. 4 Comparison of a thick, low barrier and a thin, high barrier. By looking at the area of the cut off triangles, it can be seen that the thick barrier is more reduced by an applied voltage than the thin barrier.

At low positive and negative gate voltages, the dielectrics containing nitride suppress the current at small gate voltages by 3–4 orders of magnitude with negligible difference between ON and ONO. In Fig. 3, a shoulder at -1.5 V appears because the gate Fermi level aligns with the silicon conduction band edge. The sharp structure in the case of ONO around -5.5 V originates from the resonance when the gate Fermi level crosses the triangular-shaped ONO potential well.

At high positive and negative gate voltage the advantage of the barriers containing nitride disappears. The reason is that, for a given voltage drop, the barrier reduction for a thick barrier is larger than for a thin barrier. Figure 4 illustrates this.

The effect of the two different gate materials in the case of a single oxide is demonstrated in Fig. 5. The MOS diode with poly gate yields a larger current than the diode with aluminium gate because of the smaller work function difference, which leads to a barrier height reduction of 0.2 eV. This is shown in Fig. 6 for various gate voltages. The difference is particularly pronounced before flat-band conditions are reached. Because of the different energetic distance between gate Fermi level and silicon conduction band edge, the exponential tail of the distribution function in the gate has a large effect.

Gate and drain currents versus gate voltage at a drain bias of 1 V of an n-channel MOSFET with 300 nm

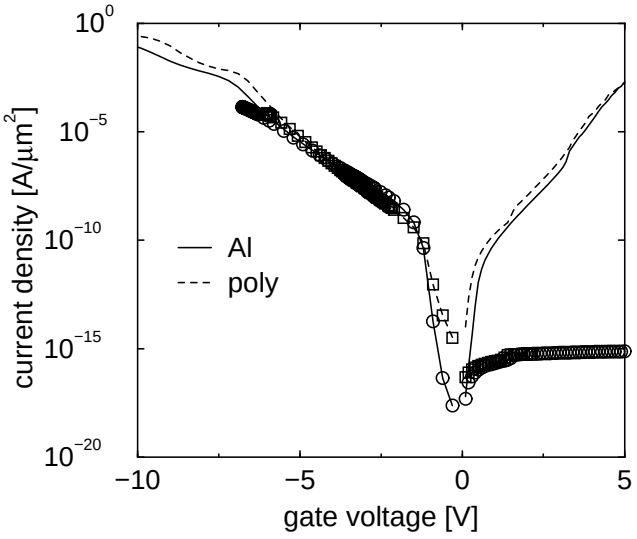

Fig. 5 Simulated gate currents of MIS capacitor with single oxide dielectric. Comparison between aluminum gate and poly gate. The lines are results for fixed quasi Fermi potential, symbols are for full self-consistent computation.

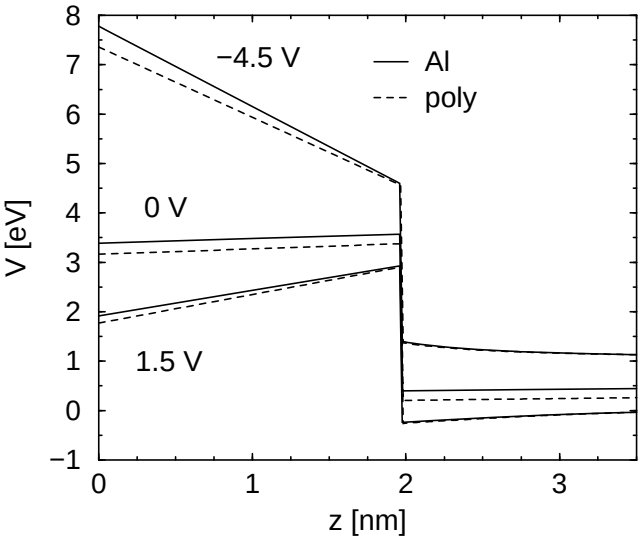

Fig. 6 Comparison of band edges in oxide (0–2 nm) and channel for devices with poly and aluminium gate

gate length, 240 nm effective channel length, 1 μm channel width, source/drain extension depths of 20 nm, and deep source/drain junction depths of 60 nm are shown in Fig. 7. The dielectrics containing nitride suppress the gate current by 3–4 orders of magnitude.

For negative gate voltage, the tunnel current in the ONO structure exceeds the current in the ON structure, whereas for positive gate voltage the situation is reversed. For moderate voltage drop, this is due to the difference in the dielectric constant of Si and nitride. As the field is stronger in the material with lower dielectric constant (SiO_2), the resistance of the overall

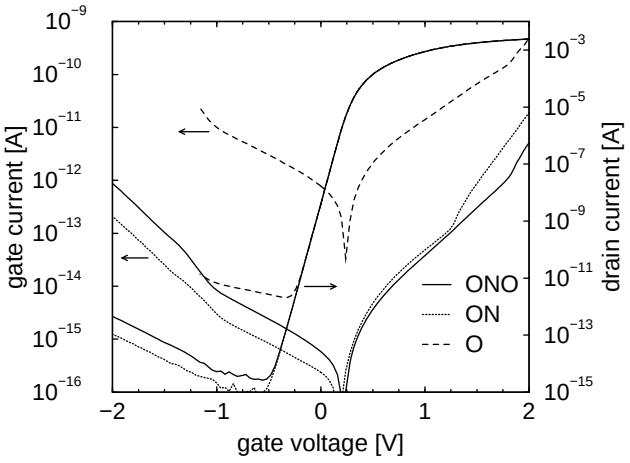

Fig. 7 Gate and drain currents versus gate voltage at a drain bias of 1 V of an n-channel MOSFET with different gate dielectrics.

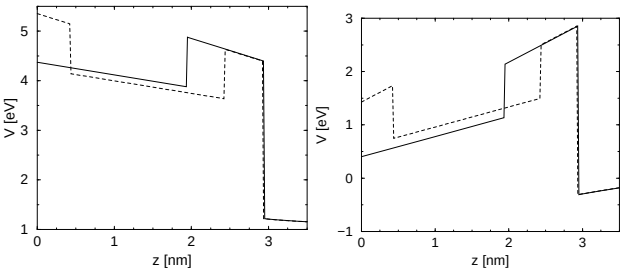

Fig. 8 Conduction band in a ON and a ONO structure, for a negative voltage (left) and for a positive voltage (right).

barrier is larger, if this material is on the side of the positively biased contact. Fig. 8 illustrates this. For bigger voltage drop or high energetic electrons the conduction band of the insulators can be reached (Fowler-Nordheim regime) and this simple picture is rendered insufficient.

Fig. 7 shows that threshold voltage and subthreshold swing are independent of the insulator system since the equivalent oxide thickness is the same in all cases and short-channel effects are not influenced by the very small gate thickness-to-length aspect ratio. For large negative gate bias a significant amount of the drain current is generated by tunneling electrons.

5. Summary

We have described a tunneling model based on the numerical solutions of the 1D-Schrödinger equation. This model is applicable for barriers of arbitrary shape. Embedding this model into the device simulator DESSIS–ISE enabled us to perform self-consistent and 2D calculations of device characteristics.

Comparing single oxide dielectrics to electrostatically equivalent ON- and ONO-stacks, we have demon-

strated that the latter can reduce the tunneling current by several orders of magnitude. We have demonstrated that also the gate material has a significant impact on tunneling at low gate voltages.

References

[1] J. Bardeen. "Tunneling from a many-particle point of view," *Phys. Rev. Letters*, vol.6, pp.57–59, 1961.

[2] J. M. Blatt. "Practical points concerning the solution of the Schrödinger equation," *Journal of Computational Physics*, vol.1, pp.382–396, 1967.

[3] ISE Integrated Systems Engineering AG. DESSIS_ISE *reference manual*, 1998.

[4] L. GR. Ixaru. *Numerical Methods for Differential Equations and Applications*. Reidel, Dordrecht; Boston; Lancaster, 1984.

[5] C. T. Liu, F. H. Baumann, A. Ghetti, H. H. Vuong, C. P. Chang, K. P. Cheung, J. I. Colonell, W. Y. C. Lai, E. J. Lloyd, J. F. Miner, C. S. Pai, H. Vaidya, R. Liu, and J. T. Clemens. "Severe thickness variation of sub-3 nm gate oxide due to Si surface faceting, poly-Si intrusion, and corner stress." In *1999 Symposium on VLSI Technology Digest of Technical Papers*, pp.75–76, 1999.

Andreas Wettstein received his Diploma in physics from the Universität Karlsruhe in 1995. Since then he is with the Institut für Integrierte Systeme, ETH Zürich.

Andreas Schenk was born in Berlin, Germany, in 1957. He received the Diploma in physics and the Ph.D. in theoretical physics from the Humboldt University Berlin (HUB) in 1981 and 1987, respectively. In 1987 he became a research assistant at the Department of Semiconductor Theory of HUB, and 1988 he joined the R&D division of WF Berlin. From 1987 till 1991 he was working on various aspects of the physics and simulation of optoelectronic devices, esp. infrared detector arrays, and the development and implementation of physical models for modeling infrared sensors. He is now with ETH, Zurich, Switzerland, as a senior lecturer at the Integrated Systems Laboratory. His main activities are in the development of physics-based models for simulation of submicron silicon devices. He is a member of the German Physical Society (DPG).

Wolfgang Fichtner received the Dipl.Ing. degree in physics and the Ph.D. degree in electrical engineering from the Technical University of Vienna, Austria, in 1974 and 1978, respectively. From 1975 to 1978, he was an Assistant Professor in the Department of Electrical Engineering, Technical University of Vienna. From 1979 through 1985, he worked at AT&T Bell Laboratories, Murray Hill, NJ. Since 1985 he is Professor and Head of the Integrated Systems Laboratory at the Swiss Federal Institute of Technology (ETH). In 1993, he founded ISE Integrated Systems Engineering AG, a company in the field of technology CAD. Wolfgang Fichtner is a Fellow of the IEEE and a member of the Swiss National Academy of Engineering.

List of Figures