Transmission Coefficient Estimation for High-κ Gate Stack Evaluation

Andreas Gehring, Hans Kosina, and Siegfried Selberherr
Institut für Mikroelektronik
Technische Universität Wien
Gusshausstrasse 27-29, A-1040 Wien, Austria

Abstract: We investigate methods to estimate the transmission coefficient of high-κ gate stacks. Based on the commonly used Tsu-Esaki model, we approximate the WKB and Gundlach methods to be feasible for device simulators. Comparisons with a rigorous solution using the transfer-matrix method show good qualitative agreement. We further use the models to analyze the trade-off between barrier height and permittivity in different high-κ dielectrics and identify ZrO_2, Al_2O_3 and possibly Ta_2O_5 as the materials with appropriate gate current blocking ability.

Key-Words: Device simulation, high-k dielectrics, tunneling current, gate stack.

1 Introduction

According to the 2001 edition of the ITRS roadmap, proper modeling of high-κ gate stacks is one of the key issues for the 65 nm technology node. This requires the calculation of tunneling currents through energy barriers which are not of triangular or trapezoidal shape. However, general purpose device simulators usually apply the WKB or Gundlach methods to approximate the gate current density, which are not valid in this regime. A rigorous solution to this problem can only be found by solving Schrödinger's equation in the insulator segment, but this approach is usually avoided due to restrictive demands on computation time. In this paper we compare WKB and Gundlach method based approximations for two-step barriers with solutions of a Schrödinger solver based on the transfer-matrix method and analyze the gate current for several high-κ dielectrics.

2 The Device

In [1] a well-tempered MOSFET with an effective channel length of 50 nm is described by an analytic doping profile incorporating LDD and halo implants (see Fig. 1). Retaining the original gate oxide thickness of 2 nm we replaced the gate oxide with a gate stack as proposed in [2] consisting of an underlying silicon layer, which is necessary due to the poor stability of high-κ insulators on silicon, and a high-κ dielectric on top.

The effective oxide thickness [3] of such a stack is $t_{eq} = t_s + t_{hk} \cdot \kappa_s / \kappa_{hk}$, where t_s and t_{hk} are the thicknesses of the SiO_2 and the high-κ layer, and κ_s and κ_{hk} their respective permittivities. Tab. 1 shows
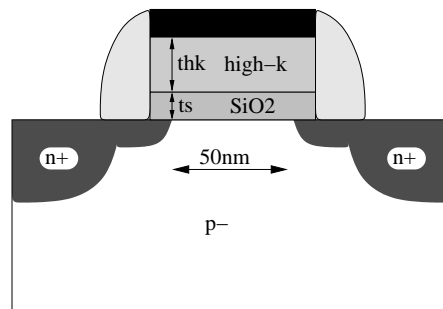

Figure 1: MOSFET with a gate stack consisting of a layer of SiO_2 and a high-κ layer on top.

material parameters of several high-κ dielectrics which may be used in gate stacks. E_g is the band gap of the material and Φ_b the energy barrier between the silicon and insulator conduction band. The values were taken from [4] [5].

Material	κ [As/Vm]	E_g [eV]	Φ_b [eV]
SiO_2	3.9	9.0	3.18
Si_3N_4	7.5	5.0	2.0
Ta_2O_5	25.0	4.4	1.4
TiO_2	40.0	3.5	1.1
Al_2O_3	7.9	5.6	3.5
ZrO_2	25.0	7.8	1.9
HfO_2	25.0	5.8	1.13

Table 1: Material parameters of oxide dielectrics.

3 Tunneling model

In general purpose device simulators like MINIMOS-NT, DESSIS and MEDICI, the tunneling current is computed using [6]:

$$J = \frac{4\pi m_{\text{ox}} q k_{\text{B}} T}{h^3} \int_{E_{\min}}^{\infty} TC(E) N(E) \, dE \qquad (1)$$

where m_{ox} is the effective electron mass in the oxide, $TC(E)$ the transmission coefficient, and $N(E)$ the supply function derived as

$$N(E) = \ln \left[\frac{1 + \exp\left(\frac{E_{f_1} - E}{k_{\text{B}} T}\right)}{1 + \exp\left(\frac{E_{f_2} - E)}{k_{\text{B}} T}\right)} \right] \qquad (2)$$

where E_{f_1} and E_{f_2} are the Fermi energies.

3.1 WKB method

Within the WKB (Wentzel-Kramers-Brillouin) approximation, the transmission coefficient reads [7]

$$TC(E) = \exp\left(-2 \int_{x_1}^{x_2} k(E, x) \, dx\right) \qquad (3)$$

with x_1 and x_2 being the classical turning points, and $k(E, x)$ the wave number. This expression can be evaluated analytically for trapezoidal or triangular barriers, leading to

$$TC(E) = \exp\left\{ -4 \frac{\sqrt{2 m_{\text{ox}}}}{3 \hbar q F_{\text{ox}}} \cdot \phi \right\} \qquad (4)$$

with

$$\phi = \begin{cases} (\Phi - E)^{3/2} & \Phi_0 < E < \Phi \\ (\Phi - E)^{3/2} - (\Phi_0 - E)^{3/2} & E < \Phi_0 \end{cases} \qquad (5)$$

where Φ and Φ_0 are the upper and lower edge of the barrier, and F_{ox} is the electrostatic field in the oxide layer. The method is robust and fast, but it does not account for oscillations in the transmission coefficient.

3.2 Gundlach method

In [8] Gundlach presents an analytical solution for the transmission coefficient of a trapezoidal barrier, yielding a transmission coefficient of

$$TC = \frac{k_2}{k_1} \frac{4}{\pi^2} \left[\left(\frac{z'}{k_1} A + \frac{k_2}{z'} B\right)^2 + \left(\frac{k_2}{k_1} C + D\right)^2 \right]^{-1} \qquad (6)$$

where k_1 and k_2 are the wave numbers next to the oxide segment. The other symbols in (6) indicate solutions of the Airy function and their derivatives, depending on the upper and lower edge of the energy barrier (cf. [8] for a detailed explanation). The Gundlach method is, due to the costly evaluation of the Airy functions, computationally more expensive than the WKB method.

3.3 Transfer-matrix method

The transfer-matrix method is based on the work of Tsu and Esaki on resonant tunneling diodes [6]. The trapezoidal energy barrier is approximated by a series of n regions with constant energy. Since the transmission coefficient through a constant potential barrier is known, the amplitudes A_n and B_n of the transmitted and reflected wave in region n can be derived by a number of consecutive 2x2-matrix multiplications:

$$\begin{pmatrix} A_n \\ B_n \end{pmatrix} = \overline{T}_{n-1} \cdot \overline{T}_{n-2} \cdot ... \cdot \overline{T}_2 \cdot \overline{T}_1 \cdot \begin{pmatrix} A_1 \\ B_1 \end{pmatrix}. \quad (7)$$

If it is assumed that there is no reflected wave in region n and the amplitude of the incoming wave A_1 is unity, the transmission coefficient can be derived as

$$TC = \frac{k_n}{k_1} \cdot \frac{m_1}{m_n} \cdot A_n^2 \qquad (8)$$

where k_i and m_i are the wave numbers and effective masses in region i. The accuracy and the computation time of this method depend on the resolution of the barrier, i.e. the number of regions n.

3.4 Approximations for TC(E)

Only the transfer-matrix method allows a rigorous solution for the transmission coefficient through a high-κ gate stack as shown in Fig. 2. However, it was shown in [4] that for $TC \ll 1$, the total transmission coefficient through two consecutive barriers can be approximated by $TC = TC_1 \cdot TC_2$ where TC_1 and TC_2 are the transmission coefficients of the two barriers. We used this approximation for the WKB and Gundlach methods and compared it to the solutions of the transfer-matrix method.
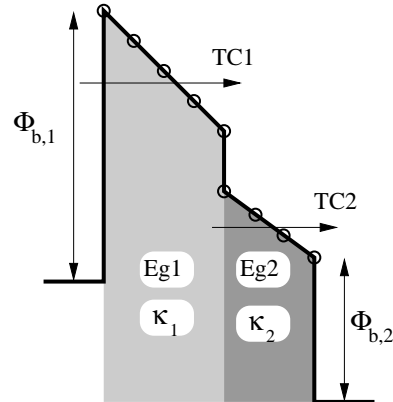

Figure 2: Barrier for two insulators with different barrier heights and permittivities.
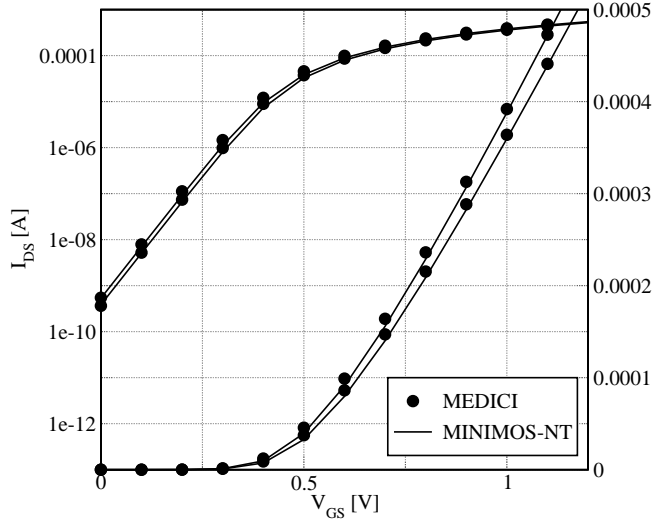
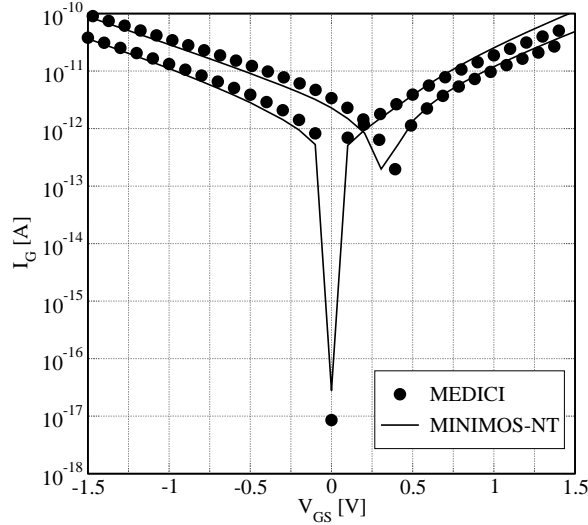Figure 3: Transfer characteristics at $V_{DS} = 0.7\,V$ and $V_{DS} = 1.2\,V$.

Figure 5: Tunneling Current for a gate stack with 2 nm equivalent oxide thickness.

Figure 4: Tunneling current using the Gundlach method $V_{DS} = 0.7\,V$ and $V_{DS} = 1.2\,V$.
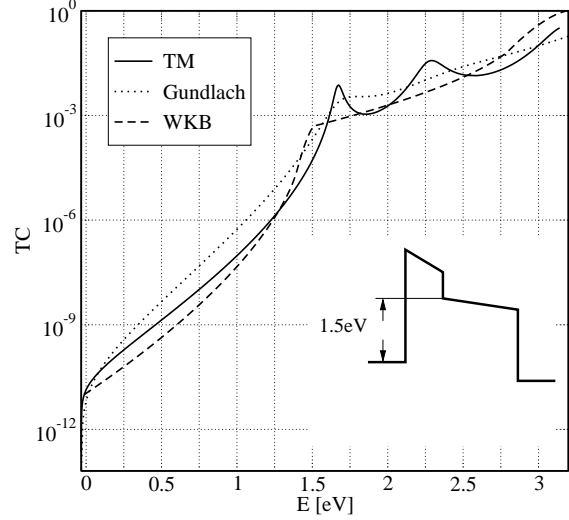
Figure 6: Transmission coefficient for a SiO2 - Si3N4 gate stack at $V_{GS} = 0.5\,V$.

4   Simulation results

The different methods have been implemented into the general purpose device simulator MINIMOS-NT and compared to the results of MEDICI for the 50 nm well-tempered MOSFET described in Section 2. Figs. 3 and 4 show the transfer characteristics and the gate current at different drain-source voltages. Excellent agreement could be achieved using the Gundlach method for a single dielectric layer and m_{ox} as fitting parameter. We then investigated the gate current behavior for a gate stack consisting of an underlying 1 nm SiO_2 layer and a high-κ dielectric on top, retaining an equivalent oxide thickness of 2 nm (see Fig. 5). It can be seen that, despite the good qualitative agreement, the WKB and Gundlach approximations differ significantly from the transfer-matrix (TM) results.
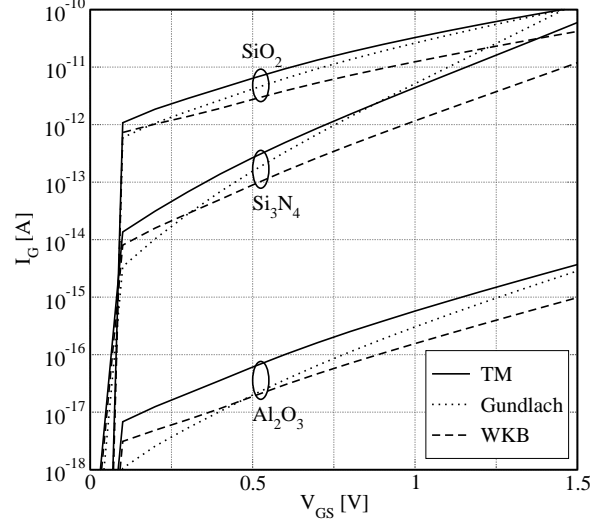
The reason for that behavior is shown in Fig. 6, which depicts the resulting transmission coefficient of the different methods for a SiO_2-Si_3N_4 stack at a gate bias of 0.5 V. The approximate shape of the barrier is sketched in the inset. It can be seen that for $E > 1.5\,eV$ resonances appear which neither the WKB nor the Gundlach method can reproduce. These resonances cause additional tunneling current which is not taken into account by the investigated approximations. In Fig. 7 we show the influence of the number of grid lines in the oxide region on the computation time and the accuracy of the transfer-matrix method. While the WKB and Gundlach methods show only a small computation time increase, the transfer-matrix method decelerates stronger with increasing resolution.
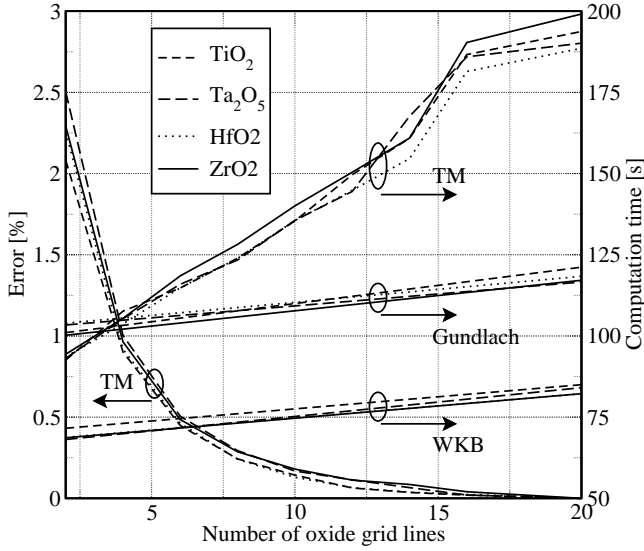
Figure 7: CPU time (right) and resolution error of the TM method (left) as function of the resolution.

Figure 8: Gate current as a function of the SiO2 layer thickness for $V_{DS} = 1\,V$ and $V_{GS} = 2\,V$.

However, for the transfer-matrix method, a few grid lines in the oxide segment suffice to yield a difference of about 1% to the result with the finest resolution using 20 grid lines. For that resolution the transfer-matrix and the Gundlach methods require approximately equal computation times, while the WKB method is still faster.

In Fig. 8 the gate current has been evaluated as a function of the high-κ layer thickness at $V_{DS} = 1\,V$ and $V_{GS} = 2\,V$. An equivalent oxide thickness of $2\,nm$ was assumed. For a SiO_2 layer thickness above approximately $1\,nm$, only ZrO_2, Al_2O_3, and Ta_2O_5 are still viable gate dielectrics, while the other materials(Si_3N_4, TiO_2, and HfO_2), due to their low barrier height, lead to even more leakage than a single $2\,nm$ SiO_2 layer.

5   Conclusions

We have presented tunneling current simulations through high-κ gate stacks using models of increasing sophistication. The approximations based on the WKB and Gundlach methods show good agreement compared to the transfer-matrix method, with the WKB method being considerably faster than the Gundlach method. We investigated several high-κ dielectrics and found that ZrO_2, Al_2O_3, and possibly Ta_2O_5 give appropriate performance. We find that for implementation into a device simulator, the WKB method is most feasible since it provides good accuracy and efficiency. However, for calibration purposes, the transfer-matrix model appears to be necessary.
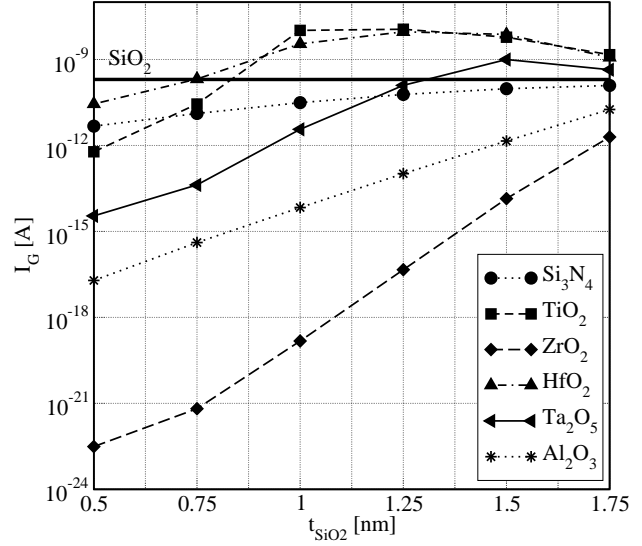
References:

[1] D. Antoniadis, I. Djomehri, K. Jackson, and S. Miller, ""Well-Tempered" Bulk-Si NMOSFET Device Home Page." http://www-mtl.mit.edu/Well/, 2001.

[2] S. Mudanai, Y. Fan, Q. Ouyang, A. F. Tasch, and S. K. Banerjee, "Modeling of Direct Tunneling Current Through Gate Dielectric Stacks," *IEEE Trans.Electron Devices*, vol. 47, no. 10, pp. 1851–1857, 2000.

[3] D. A. Buchanan, "Scaling the Gate Dielectric: Materials, Integration and Reliability," *IBM J.Res.Dev.*, vol. 43, no. 3, pp. 245–264, 1999.

[4] J. Zhang, J. S. Yuan, Y. Ma, and A. S. Oates, "Design Optimization of Stacked Layer Dielectrics for Minimum Gate Leakage Currents," *Solid-State Electron.*, vol. 44, pp. 2165–2170, 2000.

[5] T. Hori, *Gate Dielectrics and MOS ULSIs.* Springer Verlag, 1997.

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

[7] R. Shankar, *Principles of Quantum Mechanics.* Plenum Press, New York, 1994.

[8] K. Gundlach, "Zur Berechnung des Tunnelstroms durch eine trapezförmige Potentialstufe," *Solid-State Electron.*, vol. 9, pp. 949–957, 1966.