

Modeling and simulation of tunneling through ultra-thin gate dielectrics

Andreas Schenk^{a)}

Integrated Systems Laboratory, Swiss Federal Institute of Technology, Gloriastrasse 35, CH-8092 Zürich, Switzerland

Gernot Heiser

School of Computer Science and Engineering, The University of New South Wales, Sydney 2052, Australia

(Received 17 January 1997; accepted for publication 24 March 1997)

Direct and Fowler-Nordheim tunneling through ultra-thin gate dielectrics is modeled based on an approach for the transmission coefficient (TC) of a potential barrier that is modified by the image force. Under the constraint of equal actions the true barrier is mapped to a trapezoidal pseudobarrier resulting in a TC very close to the numerical solution of the Schrödinger equation for all insulator thicknesses and for all energies of the tunneling electron. The barrier height of the pseudopotential is used as a free parameter and becomes a function of energy in balancing the actions. This function can be approximated by a parabolic relation which makes the TC of arbitrary barriers fully analytical with little loss of accuracy. The model was implemented into a multidimensional device simulator and applied to the self-consistent simulation of gate currents in metal-oxide-semiconductor (MOS) capacitors with gate oxides in the thickness range 15 Å–42 Å. Excellent agreement with experimental data was obtained using a thickness-independent tunnel mass $m_{ox} = 0.42 m_0$. Thanks to the CPU-time efficiency of the method the simulation of a complete MOS-field-effect-transistor with dominating gate current becomes possible and shows the potential for further applications. © 1997 American Institute of Physics. [S0021-8979(97)05812-X]

10.1063/1.365364

I. INTRODUCTION

In modern microelectronics the transport of electrons and holes across ultra-thin dielectric barriers is of considerable interest. Well-known examples are the injection of carriers into gate oxides of metal-oxide-semiconductor-field-effect-transistors (MOSFETs)^{1,2} leading to a long-term shift of their threshold voltage (so-called degradation), the strong tunnel currents during the erase mode of electrically erasable programmable read only memories (EPROMs),³ the current-voltage characteristics of metal-insulator-semiconductor (MIS) solar cells,^{4–7} or the tunneling leakage occurring in memory cells.^{8,9} Modeling and numerical simulation of these currents rely not only on realistic distribution functions for the charge carriers, but also on a good knowledge of the quantum-mechanical transmission probability for ultra-thin barriers.

Since the pioneering work of Fowler and Nordheim^{10,11} there were several attempts to improve the calculation of currents through thin insulating layers. Gundlach¹² showed by using the exact transmission probability of a trapezoidal barrier that oscillations in the tunnel I - V curves occur. These were observed experimentally by Maserjian¹³ from Fowler-Nordheim (FN) currents through thermally grown SiO₂ in the thickness range of 2.2–4.0 nm. Contrary to the expectation that the simple one-dimensional band model breaks down when applied to extremely thin amorphous oxides (since the band structure seems to be not well defined in a few molecular layers, the interface might be smeared out, and/or thickness fluctuations¹⁴ or pinholes can dominate the I - V characteristics), Maserjian's MOS structures showed an extremely abrupt oxide-silicon interface. Quantum reflec-

tions at this interface were found to agree well with the ideal one-dimensional (1D) model.

Today, oxide thicknesses in the range of 5–10 nm are customary in manufacturing, and research has reached 1.5 nm.¹⁵ Below 3 nm, direct tunneling becomes a limiting leakage mechanism in MOS technology. Direct tunneling was confirmed for ultra-thin dry¹⁶ and chemical SiO₂ layers formed in H₂SO₄, the latter being as thin as 0.68 nm.¹⁷

The subject of this article is the numerical simulation of *transport across ultra-thin dielectrics* in the sense of either a *direct* process, which means that the whole barrier is surmounted, or of *Fowler-Nordheim tunneling*, where the final state lies in the dielectric, but the succeeding part of transport in the remainder of the barrier (e.g. drift) does *not* limit the total transition rate. Roughly speaking, such a description is justified as long as the mean free path in the dielectric is comparable to the width of the remainder of the barrier (a rough guess: 10 nm). The bulk-limited currents,^{18,19} hopping conduction and space charge-limited current, will not be considered here.

In Section II we recapitulate the transmission probability of a trapezoidal barrier, discuss the effect of the image force on ultra-thin dielectric layers, give an overview over measured values for two critical parameters—the barrier height and the tunneling mass, and touch the issue of band structure mismatch at the Si-SiO₂ interface. A new approach, called the pseudobarrier method, will be outlined in Section III and applied to the calculation of MOS gate currents in Section IV. We show that common approximations fail in the case of ultra-thin dielectrics if the image-force effect is included. The accuracy of the new analytical approach and its CPU-time efficiency will be demonstrated. Section V presents self-consistent simulation results for MOS capacitors and

^{a)}Electronic mail: schenk@iis.ee.ethz.ch

MOSFETs with tunnel oxides, and a final discussion follows in Section VI.

II. TRANSPORT MODEL

A. Transmission probability

The following derivation will be given for *electrons* only (holes are treated accordingly), and *oxide* will replace *dielectric*. In order to calculate the elastic electron tunnel current through a thin oxide layer as a barrier penetration problem following Bardeen,²⁰ Price and Radcliffe,²¹ Harrison,²² and Stratton:²³

$$j_n = \frac{qm_c^* k_B T}{2\pi^2 \hbar^3} \int_0^\infty dE \mathcal{T}(E) \times \ln \left\{ \frac{\exp\left[\frac{E_{F,S}(d) - E_c(d) - E}{k_B T}\right] + 1}{\exp\left[\frac{E_{F,M}(0) - E_c(d) - E}{k_B T}\right] + 1} \right\}, \quad (1)$$

with q the elementary charge, m_c^* the conductivity mass, k_B the Boltzmann constant, $E_{F,S}(d)$ the substrate Fermi energy at the Si-SiO₂ interface, $E_{F,M}(0)$ the gate Fermi energy, $E_c(d)$ the conduction band energy at the Si-SiO₂ interface, and E the energy, T the temperature, one has to determine the transmission coefficient (TC) $\mathcal{T}(E)$. Eq. (1) is based on the independent particle approximation, i.e. all scattering processes as well as the interaction with the environment are implicitly absorbed by the one-particle quasi-equilibrium distribution functions and by the potential model for $\mathcal{T}(E)$.²⁴ Approximating incoming and outgoing states by plane waves, the shape of the oxide barrier by a trapezoid (no fixed charges), and assuming a parabolic $E(\mathbf{k})$ relation in the barrier region as well, the 1D solution for the TC can easily be found by the common matching procedure:

$$\mathcal{T}(E) = \frac{2}{1 + g(E)}, \quad (2)$$

with

$$g(E) = \frac{\pi^2}{2} \left\{ \frac{m_{Si} k_F}{m_M k_{Si}} (Bi'_d Ai_0 - Ai'_d Bi_0)^2 + \frac{m_M k_{Si}}{m_{Si} k_F} (Bi_d Ai'_0 - Ai_d Bi'_0)^2 + \frac{m_M m_{Si}}{\lambda_0^2 m_{ox}^2 k_F k_{Si}} (Bi'_d Ai'_0 - Ai'_d Bi'_0)^2 + \frac{\lambda_0^2 m_{ox}^2 k_F k_{Si}}{m_M m_{Si}} (Bi_d Ai_0 - Ai_d Bi_0)^2 \right\}, \quad (3)$$

and

$$Ai_0 \equiv Ai\left(\frac{\Phi_B - E}{\hbar \Theta_{ox}}\right), \quad Ai_d \equiv Ai\left(\frac{\Phi_B + qF_{ox}d - E}{\hbar \Theta_{ox}}\right), \quad (4)$$

where Φ_B denotes the barrier height for electrons, m_M , m_{ox} , and m_{Si} are the effective electron masses in the three materials, respectively, k_F and k_{Si} are the momenta in

the electrodes, $\lambda_0 = \hbar \Theta_{ox} / qF_{ox}$, $\hbar \Theta_{ox} = (q^2 \hbar^2 F_{ox}^2 / 2m_{ox})^{1/3}$, and Ai , Bi are the Airy functions. In deriving Eq. (1) the explicit dependence of $\mathcal{T}(E)$ on the energy perpendicular to the tunnel current was neglected replacing $k_M(E_\perp)$ by the Fermi momentum k_F in the gate electrode.

B. Image force

Although there is still some controversy about the image force (the same holds for the value of ϵ_{ox} , the significance of a band structure in the case of only a few molecular layers, and the issue of k_\perp -conservation), it (or some proper modification of its “classical” form) should be most important for the ultra-thin oxides with large TCs, such as gate oxides of scaled MOSFETs ($d \sim 4$ nm) or the native oxides of MIS solar cells ($d \sim 0.5$ – 2 nm). In a series of papers Weinberg and Hartstein^{25–28} had explained a discrepancy between results from photon-assisted tunneling and internal photoemission experiments by a “quantum-mechanical” image force, obtained from the “classical” one by multiplication with the TC. According to this modification, a tunneling electron cannot induce an image charge, whereas electrons that are emitted over the barrier “feel” almost the total image force.

The basic question in this context is whether the dwell time of the tunneling electron²⁹ within the barrier is long enough for the completion of the spreading of the exchange-correlation hole in the interfacial plane. Hence, the traversal time must be long compared to the inverse of the plasma frequency in the cathode, which is met under the conditions of interest here. This was also shown by Puri and Schaich³⁰ who compared a 1D model system (an electron weakly coupled to one polarizable mode) with the adiabatic limit of that model (the “classical” image force). The calculated tunneling probabilities agreed perfectly, whereas the product of TC and “classical” image force disagreed significantly. Binnig *et al.*³¹ gave further confirmation by the interpretation of their vacuum-tunneling experiments. They showed that the adiabatic image potential is indispensable to correctly describe the barrier-width dependence and absolute value of the vacuum tunnel current.

On the other hand, the “classical” image force can only be used asymptotically, i.e. a few bohrs away from the image plane. In the vicinity of that plane, the classical singularity is to be removed as shown by Serena *et al.*,³² who convoluted the local density approximation (LDA) of the exchange-correlation hole density with the classical surface charge density, which is induced by a point charge on a conductor surface. As a result, image plane and potential shape were obtained self-consistently. They found the “classical” image potential for distances larger than about 5 bohrs away from the image plane. Stern³³ assumed that an always present transition layer of at least two atomic layers (about 3 Å in the Si-SiO₂ case) can be modeled by a smooth variation of the dielectric constant, which leads to a smooth image potential through the interface.

From the above findings we can conclude that the expression of the “classical” image potential can be well used for our purposes with only some failure at the edges of the barrier. For simplicity, we remove the singularity of the clas-

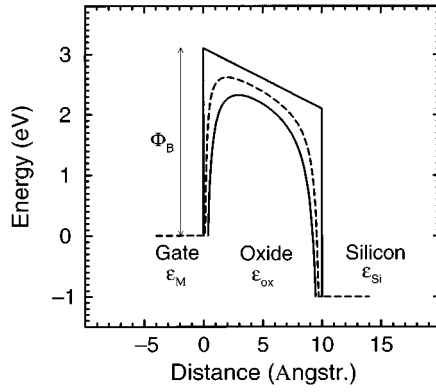


FIG. 1. Image-force effect on an idealized potential barrier due to an oxide of 1 nm thickness. Dashed curve: $\epsilon_{ox}=3.9$, solid curve: $\epsilon_{ox}=2.13$. Other parameters: $F_{ox}=10$ MV/cm, $\Phi_B=3.15$ eV, $m_{ox}=0.5 m_0$.

sical image potential in the simple way illustrated in Fig. 1, i.e. by a straight continuation of both, the band edge in the semiconductor and the gate Fermi level. Including the effect of all images in the two electrodes, the image potential is given by³⁴

$$E_{im}(x) = \frac{q^2}{16\pi\epsilon_{ox}} \sum_{n=0}^{\infty} (k_1 k_2)^n \times \left[\frac{k_1}{nd+x} + \frac{k_2}{d(n+1)-x} + \frac{2k_1 k_2}{d(n+1)} \right], \quad (5)$$

with

$$k_1 = \frac{\epsilon_{ox} - \epsilon_M}{\epsilon_{ox} + \epsilon_M} = -1, \quad k_2 = \frac{\epsilon_{ox} - \epsilon_{Si}}{\epsilon_{ox} + \epsilon_{Si}}. \quad (6)$$

Whereas $\epsilon_{ox}=3.9$ (low-frequency dielectric constant) is usually applied to thick oxides, the value for ultra-thin SiO₂ layers remains vague. Deal *et al.*³⁵ fitted the optical dielectric constant by the voltage dependence of the measured barrier height in their photoresponse measurements. A value of $\epsilon_{ox}=2.1$ was found for electron transitions from aluminum to SiO₂. In Refs. 36 and 37 an optical dielectric constant of 2.15 was assumed.

Fig. 1 demonstrates the image force effect for a 1 nm barrier at $F_{ox}=1$ V/nm. It is a common approximation to use the Wentzel-Kramers-Brillouin (WKB) probability which, however, breaks down for *all* energies in the case of very thin barriers. Alternatively, approximate expressions for the barrier lowering due to the image force are used.³⁸ Stratton²³ has considered the parabolic barrier shape being a close approximation to a very thin image force barrier. Fig. 2 compares the calculated transmission probabilities for the barrier of Fig. 1 using the WKB approach, the numerical solution of the envelope Schrödinger equation (transfer matrix method), and neglecting the image force altogether, respectively.

C. Barrier height and tunneling mass

Apart from the dielectric constant ϵ_{ox} of the barrier material, the tunneling mass m_{ox} and the barrier height Φ_B are the two remaining crucial parameters. Lenzlinger and Snow³⁹ used $\epsilon_{ox}=3.9$ for thermally grown oxides of 100 nm

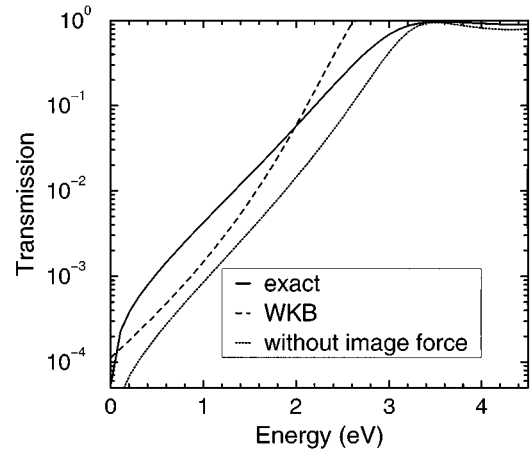


FIG. 2. Calculated transmission probabilities for a MOS structure with 1 nm oxide thickness. Parameters: $F_{ox}=10$ MV/cm, $\epsilon_{ox}=3.9$, $\Phi_B=3.15$ eV, $m_{ox}=0.5 m_0$.

thickness and fitted the tunneling mass m_{ox} from the measured FN characteristics. In the case of electron emission from silicon they obtained $m_{ox}=0.42 m_0$ fitting the field dependence, but $m_{ox}=0.96 m_0$ in fitting the temperature dependence of the FN theory. For electron emission from an aluminum gate the respective values were $m_{ox}=0.39 m_0$ (field dependence) and $m_{ox}=0.94 m_0$ (temperature dependence). The small difference between $0.42 m_0$ and $0.39 m_0$ indicated the minor importance of the true initial states in the electrodes, whereas the discrepancy between the fitted masses arising from the field and temperature dependencies was explained by a temperature-dependent barrier height. However, Av-Ron *et al.*⁴⁰ confirmed the temperature insensitivity of the barrier height by internal photoemission measurements in the range 300–675 K. Krieger and Swanson⁴¹ compared results based on the effective potential method (applying the Franz-type band model⁴²) *with* and *without* inclusion of the image-force effect, respectively. Experimental FN data were claimed to fit with equal accuracy using $m_{ox}=1.03 m_0$ (later corrected to $0.5 m_0$ Ref. 43) for the former and $m_{ox}=0.362 m_0$ for the latter case, respectively. These results were questioned by Weinberg.⁴³ Weinberg⁴⁴ determined the critical field of the FN exponent to be 2.385×10^8 V/cm for a $\langle 100 \rangle$, 2Ω cm silicon substrate and deduced $m_{ox}=0.5 m_0$ using a barrier height of 3.1 eV. This is to be compared with the self-consistent band structure calculation for α -quartz by Chelikowsky and Schlüter,⁴⁵ who obtained $m_{c,ox} \approx 0.3 m_0$ at the Γ point, and with refined calculations by Car *et al.* (cited as a private communication in Ref. 43) who found $m_{c,ox} \approx 0.5 m_0$. Various authors^{13,40,41} used a Franz-type two-band model⁴² instead of a parabolic relation in the SiO₂ gap. However, the correction becomes negligible if a two-mass Franz-type expression is used (yielding the parabolic relation with the correct hole mass at the valence band edge) due to the large hole mass of $\approx (5-10) m_0$.⁴⁵ The impression arises that, depending on the details of the potential and band models, m_{ox} finally represents the adjustable parameter of the “forgiving” model. However, Weinberg⁴³ compared the FN slope as a function

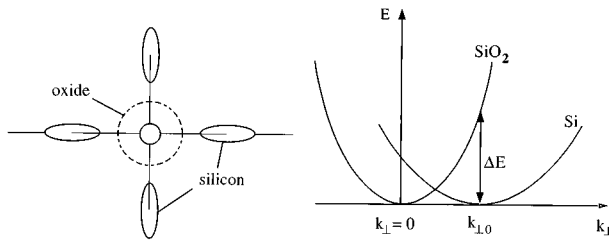


FIG. 3. Illustration of the hypothetical band structure mismatch at the Si-SiO₂ interface.

of $\Phi_B^{3/2}$ as was reported and published by different authors (for different MOS systems) and found a surprising consistency of these data when $m_{ox}=0.5 m_0$ was used for the FN slope.

Dressendorfer and Barker⁴⁶ showed by means of internal photoemission experiments that tunnel oxides as thin as 4 nm have the same Si-SiO₂ and Al-SiO₂ barrier heights as that found for thick oxides. Hiroshima *et al.*⁴⁷ confirmed the Al-SiO₂ barrier height of 3.17 eV for oxides as thin as 3.5 nm using $m_{ox}=(0.46\pm 0.05) m_0$ in fitting the electron FN current from Al to SiO₂. Heike *et al.*⁴⁸ applied beam-assisted scanning tunneling microscopy and found the SiO₂ band gap for layers with thicknesses between 1.8 nm and 4.5 nm to be independent. Recent X-ray photoelectron spectroscopy measurements by Alay *et al.*⁴⁹ revealed a constant barrier height also for holes, even for SiO₂ layers as thin as 1.6 nm.

D. Band structure mismatch

One of the unresolved problems in connection with tunneling across SiO₂ layers is the apparent absence of a band structure mismatch at the Si-SiO₂ interface. Assuming a direct, parabolic $E(k)$ relation in the SiO₂, the conditions of energy and k_{\perp} conservation would yield, for the longitudinal Si valleys (parallel to the tunnel current), i.e. those with $k_{\perp 0}=0$,

$$|k_{\parallel,ox}| = \sqrt{\frac{2m_{ox}}{\hbar^2}} \times \sqrt{\left| \Phi_B + qF_{ox}x - E_{\parallel,Si} - E_{\perp,Si} \left(1 - \frac{m_t}{m_{ox}} \right) \right|}. \quad (7)$$

For Si-SiO₂ we have $m_t < m_{ox}$ (m_t —transverse effective mass), hence the k_{\perp} conservation acts as an effective reduction of the barrier height. In the transverse valleys (perpendicular to the tunnel current) $k_{\perp 0} \neq 0$, and the momentum of electrons with $k_{\perp} \approx k_{\perp 0}$ becomes

$$|k_{\parallel,ox}| = \sqrt{\frac{2m_{ox}}{\hbar^2}} \sqrt{\left| \Phi_B + qF_{ox}x - E_{\parallel,Si} + \frac{\hbar^2 k_{\perp 0}^2}{2m_{ox}} \right|} \quad (8)$$

in the oxide. In this case an effective increase of the barrier height by $\Delta E = \hbar^2 k_{\perp 0}^2 / 2m_{ox}$ occurs, which is of the order of $\Delta E \sim 3-4$ eV. The situation is sketched in Fig. 3. However, from photoemission,³⁵ capacitance-voltage³⁵ and FN experiments⁴⁴ no essential difference was found between $\langle 100 \rangle$ and $\langle 111 \rangle$ substrate orientations. A small difference

in FN currents between the two orientations (corresponding to an increase of the barrier height by about 0.5 eV) is removed after a postmetallization 400 °C standard forming gas anneal.⁵⁰ The following possible causes for relaxation of k_{\perp} conservation were discussed: phonon-assisted tunneling, no band structure mismatch due to the amorphous nature of thermally grown SiO₂ films, impurities at the interface, and interface roughness.⁵⁰ A favored phonon-assistance of tunneling for $\langle 111 \rangle$ silicon is rather unlikely, because the measured current density is not essentially smaller as compared to $\langle 100 \rangle$ -oriented material (as it then would for a phonon-assisted process). It remains only a speculation that for some reason the transitions are phonon-assisted for *all* substrate orientations. The assumption that amorphous SiO₂ has a band structure similar to α -quartz, possibly as a result of atomic short-range order seen by the short-wavelength electrons,⁵¹ is supported by the sharp optical absorption edge of SiO₂ at a value close to the gap of α -quartz, and the sharp phonon lines observed in ballistic electron spectroscopy (see references in Ref. 51). There is no final conclusion about the apparent relaxation of k_{\perp} conservation.

III. PSEUDOBARRIER METHOD

For the purpose of device simulation both the numerical solution of the Schrödinger equation and the WKB approximation, which includes a numerical integration, are too time-consuming, if the image potential is regarded as indispensable. An analytical function $\mathcal{S}(E)$ that takes into account the image force in a better way than by the lowering of a trapezoidal barrier is desirable. This can be done by the following “pseudobarrier method.” The actual shape of the barrier potential is mapped to a trapezoid (the “pseudobarrier”) under the constraint of equal actions S :

$$S_{tra}(E) = S_{im}(E). \quad (9)$$

The action of the trapezoidal pseudobarrier $S_{tra}(E)$ has the simple form

$$S_{tra}(E) = \frac{2}{3} \left[\left(\frac{\Phi_B(E) - E}{\hbar \theta_{ox}} \right)^{3/2} - \left(\frac{\Phi_B(E) + qF_{ox}d - E}{\hbar \theta_{ox}} \right)^{3/2} \right] \times \Theta[\Phi_B(E) + qF_{ox}d - E], \quad (10)$$

whereas $S_{im}(E)$ is found by numerical integration:

$$S_{im}(E) = \sqrt{\frac{2m_{ox}}{\hbar^2}} \int_{x_l(E)}^{x_r(E)} d\xi \sqrt{\Phi_B + qF_{ox}\xi + E_{im}(\xi) - E} \quad (11)$$

($\Phi_B = \text{const} = 3.15$ eV for Al-SiO₂). Here $x_{l,r}(E)$ denote the classical turning points for a given carrier energy which follow from

$$\Phi_B + qF_{ox}x_{l,r} + E_{im}(x_{l,r}) = E. \quad (12)$$

In Eq. (10) the height $\Phi_B(E)$ of the trapezoidal barrier is used as a free parameter to balance the actions. This is done for each energy of the tunneling electrons, hence the barrier height in \mathcal{S} becomes a function of energy. $\Phi_B(E)$ results from the numerical iteration of Eq. (9). For the thickness d of the pseudobarrier in Eq. (10) we use $d = x_r(E=0)$

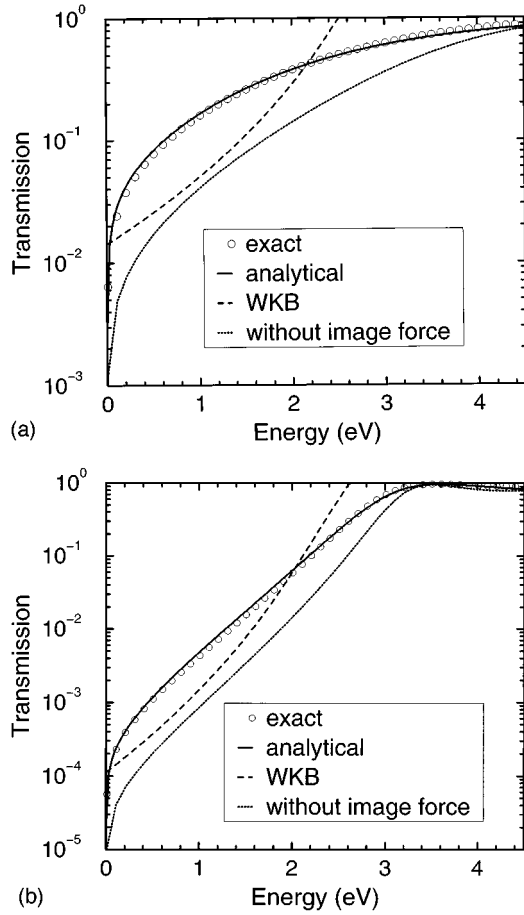


FIG. 4. Calculated transmission probabilities for oxides of 0.5 nm thickness (a) and 1 nm thickness (b). Parameters: $F_{ox}=10$ MV/cm, $\epsilon_{ox}=3.9$, $\Phi_B=3.15$ eV, $m_{ox}=0.5 m_0$.

– $x_i(E=0)$, i.e. the distance of the two turning points at the energy of the semiconductor conduction band edge at the interface (note all energies are relative to that value).

In order to further reduce CPU time consumption, $\Phi_B(E)$ is approximated by an analytical function. Thus, Eq. (9) is applied to only three energy values $E_j (j=0,1,2)$ which gives us the barrier heights for these energies. The function $\Phi_B(E)$ is then found by parabolic interpolation:

$$\Phi_B(E) = \Phi_B(E_0) + \frac{\Phi_B(E_2) - \Phi_B(E_0)}{(E_2 - E_0)(E_1 - E_2)}(E - E_0)(E_1 - E) - \frac{\Phi_B(E_1) - \Phi_B(E_0)}{(E_1 - E_0)(E_1 - E_2)}(E - E_0)(E_2 - E). \quad (13)$$

The best fit with the numerical solutions is obtained if the three energy levels are chosen in the lower half of the barrier. The accuracy of the analytical approach depends slightly on the choice of $E_{0,1,2}$. However, Fig. 4 and Fig. 5 show an excellent overall agreement for oxides with different thicknesses. One should note that this is achieved by varying only one parameter of the pseudobarrier—its height, but leaving thickness d and oxide field F_{ox} unchanged. Fig. 6 demonstrates that one and the same set $E_{0,1,2}$ also yields a comparable accuracy for different oxide fields which makes the method applicable to the simulation of I - V curves. Besides

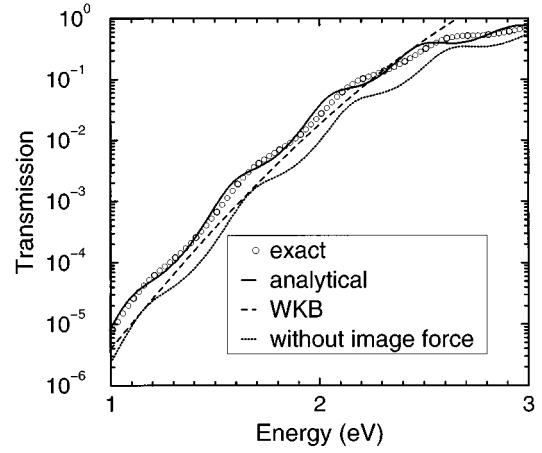


FIG. 5. Calculated transmission probabilities for a MOS structure with 5 nm oxide thickness. Quantum oscillations due to interference in the FN regime are observed. Parameters: $F_{ox}=10$ MV/cm, $\epsilon_{ox}=3.9$, $\Phi_B=3.15$ eV $m_{ox}=0.5 m_0$.

the good fit over the entire energy range there is still another advantage as compared to the WKB approximation: Only for the three energies $E_{0,1,2}$ does the action integral has to be solved numerically. After another three numerical iterations to obtain $\Phi_B(E_{0,1,2})$ the transmission probability $\mathcal{T}(E)$ of the pseudobarrier is fully analytical, and the current can be calculated by the integral (1) even in a device simulator. In contrast, the WKB approximation requires a numerical integration for *each* (!) energy of the outer integral.

IV. I - V -CHARACTERISTICS OF DIRECT AND FN TUNNELING

The calculated TC in Eq. (1) neglects electron confinement in the inversion layer and the two-dimensional (2D) character of states, which changes the pre-exponential factor of the tunneling probability. Weinberg⁴⁴ compared the pre-exponential factors of FN theory when three-dimensional (3D) or 2D states are used. He found a ratio of not more than 2.5, despite an entirely different dependence on parameters.

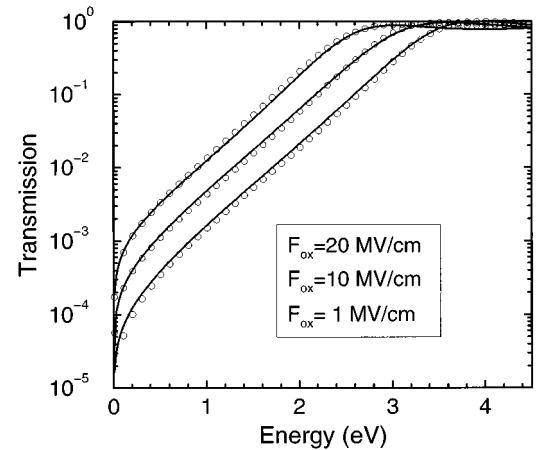


FIG. 6. Numerical (circles) and analytical (solid lines) transmission probabilities for a MOS structure with 1 nm oxide thickness at different oxide fields.

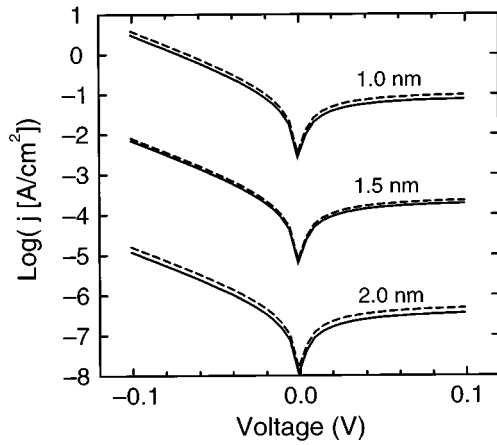


FIG. 7. I - V characteristics of an Al-SiO₂-Si(n) diode for different oxide thicknesses. Solid curves: numerical solution of the Schrödinger equation, dashed curves: pseudobarrier method. Parameters: $\Phi_B = 3.15$ eV, $m_{ox} = 0.5 m_0$, $\epsilon_{ox} = 2.13$, $m_{Si} = 0.19 m_0$, $m_M = 0.32 m_0$, $m_c^* = 0.32 m_0$, $E_{F,M} = 11.7$ eV, $E_c(d) - E_{F,S}(d) = 0.2$ eV, $E_0 = 0.1$ eV, $E_1 = 0.3$ eV, $E_2 = 0.7$ eV.

On the other hand, using simple plane waves for the calculation of the TC neglects the real structure of the semiconductor conduction bands at higher energies. The latter drawback seems to be less severe for tunneling at low energies, i.e. for ultra-thin barriers and not too high temperatures. The case of thermionic emission of hot carriers (e.g. Ning's experiment;³⁷ see also the Monte Carlo simulations in Ref. 36) requires special care of both the distribution function and the TC around 3.1 eV. In the following we will neglect the above described effects. Fig. 7 was obtained assuming a substrate doping such that $E_c(d) - E_{F,S}(d) = 0.2$ eV = const and that the voltage drops only over the barrier. In Fig. 8 the voltage range was extended to the FN regime. Again, the pseudobarrier method gives results very close to the numerical solution of the Schrödinger equation. The deviations in period and magnitude of the quantum oscillations are easily understood: Since the trapezoidal pseudobarrier has a re-

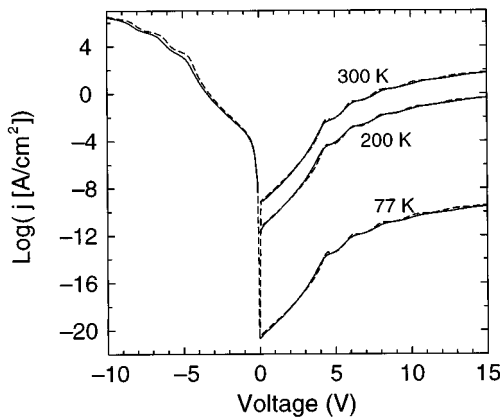


FIG. 8. I - V characteristics of an Al-SiO₂-Si(n) diode with 2.5 nm oxide thickness for different temperatures. Solid curves: numerical solution of the Schrödinger equation, dashed curves: pseudobarrier method. Parameters: $m_M = m_0$, $E_0 = 0.1$ eV, $E_1 = 0.6$ eV, $E_2 = 1.2$ eV; for other parameters see Fig. 7.

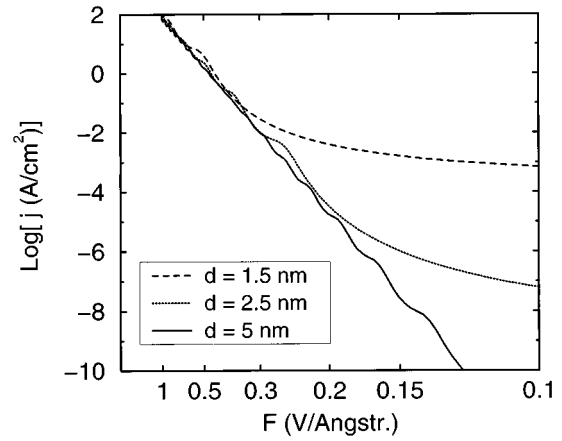


FIG. 9. Fowler-Nordheim plot for different oxide thicknesses.

duced height, it gives a slightly shorter period due to the slightly larger distance between the reflection points as compared to the original image potential barrier. Furthermore, the linear potential of the pseudobarrier produces stronger oscillations than the rounded-off potential of the original barrier.

In Fig. 9 Fowler-Nordheim plots are presented for different oxide thicknesses. When the slope is interpreted in terms of the critical field in the FN exponent $F_c = 4/(3q) \times (2m_{ox}/\hbar^2)^{1/2} \Phi_B^{3/2} = 6.83 \times 10^7 (m_{ox}/m_0)^{1/2} \Phi_B^{3/2}$ V/cm, we obtain $\Phi_B = 2.66$ eV if $m_{ox} = 0.5 m_0$ is used or, alternatively, $m_{ox} = 0.30 m_0$, if $\Phi_B = 3.15$ eV is inserted.

V. SIMULATION RESULTS

We have implemented the model in the multidimensional device simulator DESSIS-ISE,⁵² which numerically solves the drift-diffusion equations in a semiconductor device.

Assuming that the voltage of the contact, as well as the properties of the insulating barrier, are known, implementation of Eqs. (1)–(6) requires the knowledge of (i) the insulator electric field, F_{ox} , and (ii) the semiconductor Fermi level, $E_{F,S}(d)$. The latter is only defined in thermal equilibrium, while under other operating conditions it splits into the electron and hole *quasi-Fermi levels*, ϕ_n , ϕ_p , respectively. We use $E_{F,S}(d) = \phi_n$ for electron, and $E_{F,S}(d) = \phi_p$ for hole tunneling. F_{ox} is obtained by solving the Laplace equation in the insulator. The tunnel current obtained from Eq. (1) is balanced with the drift-diffusion current in the semiconductor when solving the continuity equations, resulting in a self-consistent solution.

Fig. 10 shows simulations of the tunnel currents of MOS capacitors with SiO₂ thicknesses of 15 Å, 30 Å, and 42 Å, respectively. Symbols correspond to experimental data of Fig. 4 in Ref. 15 (substrate doping and area given there) with upper limits set by the oxide breakdown. The minority carrier lifetime (an integral and process-dependent parameter for the silicon layer underneath the oxide with an extension of the order of one Debye length) was fitted to match the reverse current of the 15 Å capacitor. The value obtained is

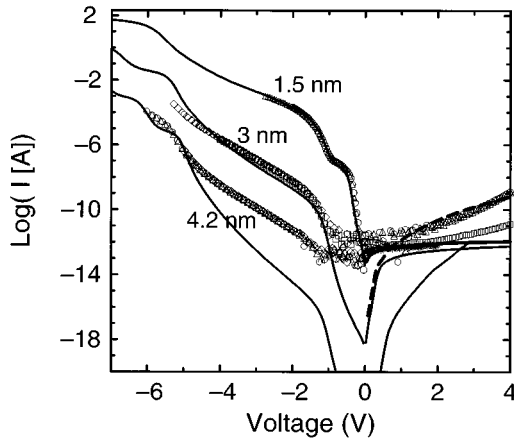


FIG. 10. Simulated (lines) vs measured (symbols) currents of MOS capacitors with 15 Å, 30 Å, and 42 Å oxide thickness, respectively. Data points correspond to those in Fig. 4 of Ref. 15.

$\tau_n = 1.7 \mu\text{s}$. Generation via interface states was found to be negligible. For the work function difference between silicon and poly gate we used -0.51 eV .

The supply of carriers at reverse bias (positive voltages) is driven by thermal generation of electrons in the silicon substrate. The resistances due to the 15 Å and 30 Å barriers are of minor importance in this branch as can be seen by the small difference between the two theoretical curves (solid lines). However, the data points of the 30 Å capacitor reveal a soft-breakdown behavior, that can be attributed to and modeled by defect-assisted tunneling in the substrate⁵³ (dashed line). In contrast, the reverse characteristic of the 42 Å capacitor is determined by the tunnel resistance of the barrier.

The forward characteristic (negative voltages) of the 15 Å capacitor shows the well-known shoulder caused by the transition from the branch of the half-sided pn junction (n side missing) to the branch of the tunnel oxide ($V < -1 \text{ V}$). The height of this shoulder depends on various parameters, e.g. the work function difference between silicon and gate. The forward current in the range $V < -1 \text{ V}$ is completely determined by the tunnel model and its parameters. The close fit was obtained with the parameter set $\Phi_B = 3.15 \text{ eV}$, $\epsilon_{ox} = 2.13$, and $m_{ox} = 0.42 m_0$ using the Fermi energy of an aluminum gate. Note that the value of m_{ox} is close to $0.39 m_0$ found by Lenzlinger and Snow³⁹ for FN emission from an aluminum gate. The inclusion of the image force effect in our simulation leads to a somewhat larger tunneling mass (see the discussion in Section VI). All parameters remained unchanged in the simulation of the 30 Å and 42 Å capacitors. It is interesting to observe that capacitors with oxides thicker than 30 Å start to reveal another tunnel mechanism in the regime of direct tunneling, most likely resonant tunneling via oxide defect states.⁹ For the 42 Å sample a large leakage current is found experimentally in the bias range -5 V to 5 V which cannot be explained by direct tunneling. Only in the FN range (at reduced tunnel length) simulated and measured currents coincide again. We conclude from the good fit in Fig. 10 that there is no need for a

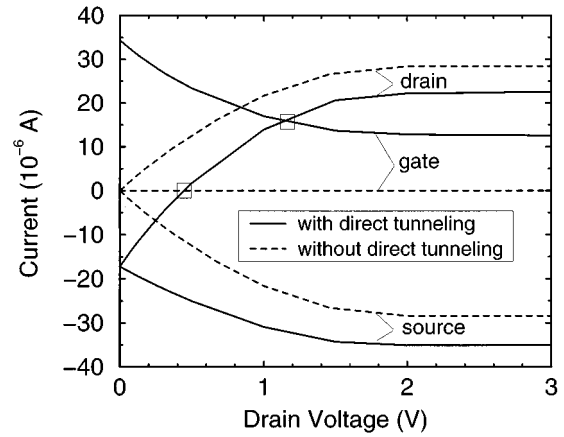


FIG. 11. Simulated drain, gate, and source currents as function of drain voltage for the n-channel MOSFET with $10 \mu\text{m}$ channel length and 1.5 nm gate oxide described in Ref. 15. The gate voltage was 2 V .

thickness-dependent tunneling mass if the image-force effect is correctly included. Such a thickness-dependent mass parameter was extracted from FN plots in Ref. 54, and ranged from $0.29 m_0$ for $d = 2.8 \text{ nm}$ to $0.36 m_0$ for $d = 6.2 \text{ nm}$.

In Fig. 11 we show simulated drain, source and gate currents as function of drain voltage ($V_g = 2 \text{ V}$) for the n-channel MOSFET with $10 \mu\text{m}$ channel length and 1.5 nm gate oxide described in Ref. 15. Without direct tunneling, the gate current is always zero (straight dashed line). Switching on the direct tunnel model, the gate current is the strongest component at $V_d = 0 \text{ V}$ balancing the sum of drain and source currents. The drain current has a negative sign because electrons are pulled out of the drain and injected into the channel. With increasing drain voltage the gate tunnel current decreases in favor of the drain current which changes its sign at about 0.4 V (left rectangle in Fig. 11). At about 1.1 V , drain and gate currents become equal (right rectangle in Fig. 11). In the saturation region the gate current is still more than one-half of the drain current.

The lateral distribution of the direct tunnel current is illustrated in Fig. 12 for a number of drain voltages

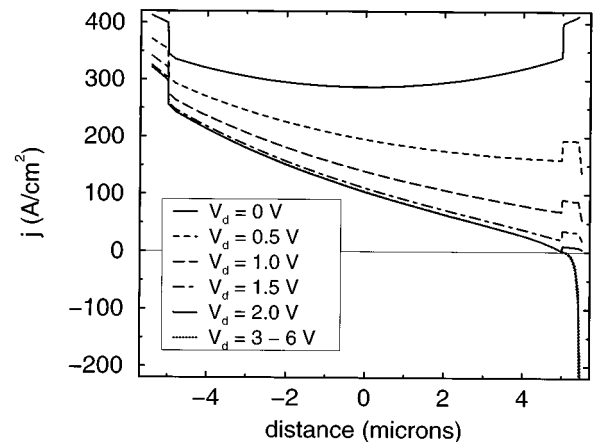


FIG. 12. Lateral distribution of the gate tunnel current along the interface of the n-channel MOSFET with $10 \mu\text{m}$ channel length and 1.5 nm gate oxide described in Ref. 15. The gate voltage was 2 V in all cases.

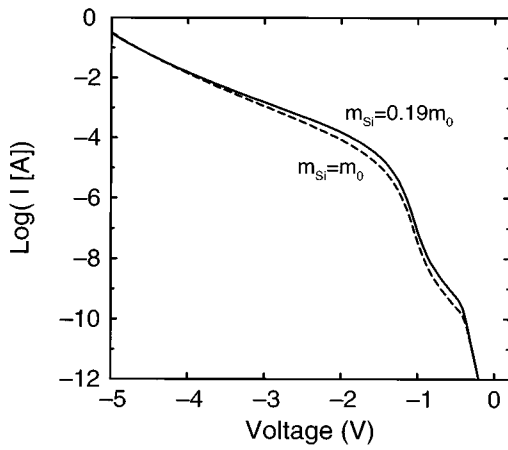


FIG. 13. Effect of the parameter m_{Si} on the forward current of the MOS capacitor with 15 Å oxide thickness.

($V_g = 2$ V). Source and drain portions are emphasized by a stronger current due to the higher electron density in these heavily doped n^+ regions. The symmetrical distribution at $V_d = 0$ V with a monotonously decreasing current density towards the center of the channel reflects the out tunneling of electrons from the channel into the gate electrode after having been injected from source and drain. With increasing drain voltage the lateral distribution becomes more and more asymmetrical. For $V_d = 2$ V = V_g the tunnel current vanishes in the vicinity of the drain edge (balance of the Fermi levels). Higher drain voltages leave the distribution unchanged, with tunnel injection from the gate into the substrate around the edge of the drain.

VI. DISCUSSION

Our primary achievements reported in this article are (i) the pseudobarrier method which allows accurate and fast computation of the transmission probability of an image potential barrier, and (ii) its self-consistent implementation in an advanced TCAD software package for the simulation of tunnel currents through ultra-thin gate oxides. The success of the pseudobarrier method is most visible for the 5 Å barrier where no other method is able to reproduce the correct TC even qualitatively (Fig. 4). We used this example for demonstration purposes only, since the actual potential induced by an SiO_x layer with an extension comparable to the transition region is not clear. However, there are other applications as in vacuum tunneling, where the barrier potential is much better defined. The success of the self-consistent implementation in a drift-diffusion simulator was shown by simulations of MOS capacitors and MOSFETs with tunnel oxides, and by the good fit to experimental data (Fig. 10).

There seems to be a contradiction between the many simplifications and assumptions on one hand, and the excellent fits presented in Figs. 4 and 10 on the other hand, that needs further explanation. Let us summarize the approximations and discuss their impact on the final results. The TC was calculated using an oversimplified band structure model for Si, which increasingly fails with rising electron energy. In the regime of direct tunneling from Si into the gate, hot

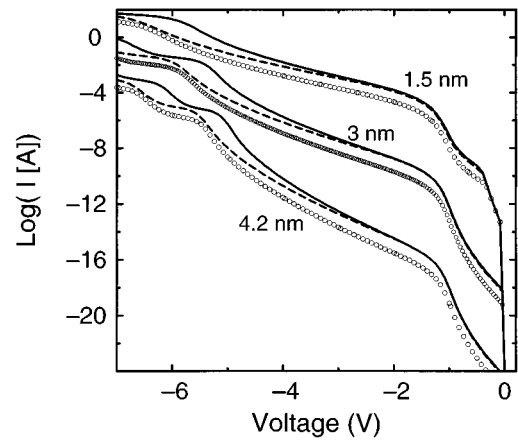


FIG. 14. Simulated forward currents of MOS capacitors with 15 Å, 30 Å, and 42 Å oxide thickness, respectively. Solid curves: numerical solution of the Schrödinger equation, open circles: without image force ($m_{ox} = 0.42 m_0$), dashed curves: without image force but with m_{ox} adjusted for best fit in the range of direct tunneling.

electrons are of minor importance, at least at moderate biases. In the regime of FN tunneling from the gate into Si, electrons with the largest tunnel probability tunnel into highly excited states at the Si-SiO₂ interface. Here, the effective mass m_{Si} is meaningless, however, as it only enters the pre-exponential factor of the TC. In Fig. 13 we changed m_{Si} from $0.19 m_0$ to m_0 and show the effect on the forward I - V -curve of a MOS capacitor with 15 Å oxide thickness. The effect is rather small, particularly in the FN regime, where the deviation of m_{Si} from its band edge value should be most pronounced.

For the ultra-thin amorphous SiO₂ layer we assumed a parabolic, direct band model and we have neglected any possible band structure mismatch. These assumptions are fully absorbed in the fit parameter “tunneling mass” m_{ox} . More surprising, m_{ox} does not depend on voltage, although for large negative bias it becomes an average over a broad range of the SiO₂ energy spectrum, ranging from less than $E_{c,ox} - \Phi_B$ up to a certain high level $E_{c,ox} + \Delta E$ in the SiO₂ conduction band.

For the gate current given by a single integral, the momentum in the gate electrode had to be replaced by some k_{\perp} independent quantity. We used the Fermi momentum here which also produces some error. In the case of metal electrodes the approximation is justified by the large Fermi momentum, whereas in the case of highly doped poly gates $k_M(E_{\perp})$ could be replaced by the thermal momentum. A variation of k_F has a similar effect on the final result as the variation of m_{Si} which is obvious from Eq. (3).

Plane waves were assumed for the states in Si in the calculation of the TC even for the case of deep inversion. The 2D density of confined states in the inversion channel⁵⁵ will change the pre-exponential factor of the TC, and the lower limit of the energy integration will shift from the conduction band edge to the lowest subband energy. However, these non-negligible effects are not visible in the reverse I - V -characteristics as long as thermal generation in the Si depletion region is the bottleneck, since in that case they

become fully absorbed in the adjusted minority carrier lifetime.

We used an idealized potential barrier with (i) the classical image potential, (ii) its divergencies removed in a simple way, and (iii) the optical dielectric constant of SiO₂. We believe that such a model is certainly better than neglecting the image force entirely. The question arises as to what extent the effect of the image force can be compensated by a change of the tunneling mass m_{ox} . We found $m_{ox}=0.42 m_0$ for the best fit and thickness independence in the regime of direct tunneling. In Fig. 14 the measured MOS capacitors were again simulated, now neglecting the image potential and adjusting m_{ox} to yield the best fit in the range of direct tunneling. The latter turned out using $m_{ox}=0.36 m_0$ for the 15 Å oxide, $m_{ox}=0.37 m_0$ for the 30 Å oxide, and $m_{ox}=0.38 m_0$ for the 42 Å oxide. This is in complete agreement with the value of $m_{ox}=0.36 m_0$ found by Nagano *et al.*¹⁶ who used the transfer matrix method but neglected the image force.

In conclusion, we arrived at a consistent picture of direct tunneling through ultra-thin gate oxides which is not essentially disturbed even by the sum of the above listed assumptions and simplifications. The inclusion of the image force effect by the pseudobarrier approach proved to be an accurate and CPU-time efficient method, superior to the WKB approximation and suitable for self-consistent numerical simulations. The latter were demonstrated to yield good results for MOS capacitors over 10 orders of magnitude in the current. The simulation of a complete MOSFET with dominating gate current has shown the potential for further applications as advanced MOSFETs, EPROMs, and MIS solar cells.

ACKNOWLEDGMENTS

The authors thank Dr. W. Grabinski (ETH Zürich) for measurements of MOS capacitors, Dr. H. Iwai (Toshiba ULSI Research Center) for his cooperation and for valuable discussions, and Dr. U. Krumbein (ETH Zürich) for support on DESSIS-ISE.

- ¹D. J. DiMaria and E. Cartier, J. Appl. Phys. **78**, 3883 (1995).
- ²Q. D. M. Khosru, N. Yasuda, K. Taniguchi, and C. Hamaguchi, J. Appl. Phys. **77**, 4494 (1995).
- ³R. B. Sethi, U. S. Kim, I. Johnson, P. Cacharelis, and M. Manley, IEEE Electron Device Lett. **13**, 244 (1992).
- ⁴M. A. Green, F. D. King, and J. Shewchun, Solid-State Electron. **17**, 551 (1974).
- ⁵V. A. K. Temple, M. A. Green, and J. Shewchun, J. Appl. Phys. **45**, 4934 (1974).
- ⁶J. Shewchun, R. Singh, and M. A. Green, J. Appl. Phys. **48**, 765 (1977).
- ⁷M. Y. Doghish and F. D. Ho, IEEE Trans. Electron Devices **ED-39**, 2771 (1992).
- ⁸S. Banerjee, D. Coleman, Jr., W. Richardson, and A. Shah, IEEE Trans. Electron Devices **ED-35**, 108 (1988).
- ⁹M. Herrmann and A. Schenk, J. Appl. Phys. **77**, 4522 (1995).
- ¹⁰R. H. Fowler and L. Nordheim, Proc. R. Soc. London, Ser. A **119**, 173 (1928).

- ¹¹L. W. Nordheim, Proc. R. Soc. London, Ser. A **121**, 626 (1928).
- ¹²K. H. Gundlach, Solid-State Electron. **9**, 949 (1966).
- ¹³J. Maserjian, J. Vac. Sci. Technol. **11**, 996 (1974).
- ¹⁴B. Majkusiak and A. Strojwas, J. Appl. Phys. **74**, 5638 (1993).
- ¹⁵H. S. Momose, M. Ono, T. Yoshitomi, T. Ohguro, S. Nakamura, M. Saito, and H. Iwai, IEEE Trans. Electron Devices **43**, 1233 (1996).
- ¹⁶S. Nagano, M. Tsukiji, K. Ando, E. Hasegawa, and A. Ishitani, J. Appl. Phys. **75**, 3530 (1994).
- ¹⁷M. Matsuda, K. Watanabe, M. Yasutake, and T. Hattori, Proceedings of the International Conference on Solid State Devices and Materials, Osaka, Japan, 1995.
- ¹⁸S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981).
- ¹⁹P. Hesto, in: *Instabilities in Silicon Devices*, edited by G. Barbottin and A. Vapaille (Elsevier Science, Amsterdam, 1986), p. 265.
- ²⁰J. Bardeen, Phys. Rev. Lett. **6**, 57 (1961).
- ²¹P. J. Price and J. M. Radcliffe, IBM J. Res. Dev. **October**, 364 (1959).
- ²²W. A. Harrison, Phys. Rev. **123**, 85 (1961).
- ²³R. Stratton, J. Phys. Chem. Solids **23**, 1177 (1962).
- ²⁴W. Pötz, J. Appl. Phys. **66**, 2458 (1989).
- ²⁵Z. A. Weinberg and A. Hartstein, Solid State Commun. **20**, 179 (1976).
- ²⁶A. Hartstein and Z. A. Weinberg, J. Phys. C **11**, L469 (1978).
- ²⁷A. Hartstein and Z. A. Weinberg, Phys. Rev. B **20**, 1335 (1979).
- ²⁸A. Hartstein, Z. A. Weinberg, and D. J. DiMaria, Phys. Rev. B **25**, 7174 (1982).
- ²⁹M. Jonson, *Quantum Transport in Semiconductors*, edited by D. K. Ferry and C. Jacoboni (Plenum, New York, 1992), p. 193.
- ³⁰A. Puri and W. L. Schaich, Phys. Rev. B **28**, 1781 (1983).
- ³¹G. Binnig, N. Garcia, H. Rohrer, J. M. Soler, and F. Flores, Phys. Rev. B **30**, 4816 (1984).
- ³²P. A. Serena, J. M. Soler, and N. Garcia, Phys. Rev. B **34**, 6767 (1986).
- ³³F. Stern, Phys. Rev. B **17**, 5009 (1978).
- ³⁴M. Kleefstra and G. C. Herman, J. Appl. Phys. **51**, 4923 (1980).
- ³⁵B. E. Deal, E. H. Snow, and C. A. Mead, J. Phys. Chem. Solids **27**, 1873 (1966).
- ³⁶G. Jin, R. W. Dutton, Y.-J. Park, and H.-S. Min, J. Appl. Phys. **78**, 3174 (1995).
- ³⁷T. H. Ning, Solid-State Electron. **21**, 273 (1978).
- ³⁸K. Kobayashi, A. Teramoto, M. Hirayama, and Y. Fujita, J. Appl. Phys. **77**, 3277 (1995).
- ³⁹M. Lenzlinger and E. H. Snow, J. Appl. Phys. **40**, 278 (1969).
- ⁴⁰M. Av-Ron, M. Shatzkes, T. H. DiStefano, and R. A. Gdula, J. Appl. Phys. **52**, 2897 (1981).
- ⁴¹G. Krieger and R. M. Swanson, J. Appl. Phys. **52**, 5710 (1981).
- ⁴²W. Franz, *Handbook on Semiconductors*, edited by S. Flügge (Springer, Berlin, 1956), Vol. 17, p. 155.
- ⁴³Z. A. Weinberg, J. Appl. Phys. **53**, 5052 (1982).
- ⁴⁴Z. A. Weinberg, Solid-State Electron. **22**, 11 (1977).
- ⁴⁵J. R. Chelikowsky and M. Schlüter, Phys. Rev. B **15**, 4020 (1977).
- ⁴⁶P. V. Dressendorfer and R. C. Barker, Appl. Phys. Lett. **36**, 933 (1980).
- ⁴⁷M. Hiroshima, T. Yasaka, S. Miyazaki, and M. Hirose, Proceedings of the International Conference on Solid State Devices and Materials, Makuhari, Japan, 1993.
- ⁴⁸S. Heike, Y. Wada, S. Kondo, and M. Lutwyche, Proceedings of the International Conference on Solid State Devices and Materials, Japan, 1994.
- ⁴⁹J. L. Alay, M. Fukuda, K. Nakagawa, S. Yokohama, and M. Hirose, in Ref. 17.
- ⁵⁰Z. A. Weinberg and A. Hartstein, J. Appl. Phys. **54**, 2517 (1983).
- ⁵¹M. V. Fischetti, S. E. Laux, and E. Crabbé, J. Appl. Phys. **78**, 1058 (1995).
- ⁵²DESSIS 3.0: manual, ISE Integrated Systems Engineering AG, Zurich, Switzerland, 1996.
- ⁵³A. Schenk, Solid-State Electron. **35**, 1585 (1992).
- ⁵⁴M. Depas, B. Vermeire, P. W. Mertens, R. L. van Meirhaeghe, and M. M. Heyns, Solid-State Electron. **38**, 1465 (1995).
- ⁵⁵F. Stern and W. E. Howard, Phys. Rev. **163**, 816 (1967).