Available online at www.sciencedirect.com
ScienceDirect
Microelectronics Reliability 46 (2006) 1939–1956
MICROELECTRONICS RELIABILITY
www.elsevier.com/locate/microrel

Introductory Invited Paper

A review of gate tunneling current in MOS devices

Juan C. Ranuárez, M.J. Deen *, Chih-Hung Chen

Department of Electrical and Computer Engineering, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1

Received 24 October 2005; received in revised form 27 December 2005
Available online 9 March 2006

Abstract

Gate current in metal–oxide–semiconductor (MOS) devices, caused by carriers tunneling through a classically forbidden energy barrier, is studied in this paper. The physical mechanisms of tunneling in an MOS structure are reviewed, along with the particularities of tunneling in modern MOS transistors, including effects such as direct tunneling, polysilicon depletion, hole tunneling and valence band tunneling and gate current partitioning. The modeling approach to gate current used in several compact MOS models is presented and compared. Also, some of the effects of this gate current in the performance of digital, analog and RF circuits is discussed, and it is shown how new effects and considerations will come into play when designing circuits that use MOSFETs with ultra-thin oxides.
© 2006 Published by Elsevier Ltd.

1. Introduction

The basic principles for scaling MOS devices, which were established in the early 1970s [1] indicate that, when reducing the lateral dimensions of MOS devices, the vertical dimensions must be scaled accordingly. Fig. 1 shows the 2004 International Technology Roadmap for Semiconductors (ITRS) trends for effective channel length (L_{eff}) and equivalent oxide thickness (EOT) scaling [2]. This reduction in the oxide thickness causes an important flow of current through the gate, which depends exponentially on the thickness of the oxide [3]; this current is caused by carriers tunneling through the insulator potential barrier, a quantum-mechanical effect that has no classical explanation.

The gate current might affect the performance of circuits that employ MOS devices, and it can be a limiting factor in device down-scaling. The feasibility of MOS devices that operate with oxides as thin as 1.5 nm was demonstrated 10 years ago [4], and it is believed now that if alternate gate dielectrics are used, the ultimate down-scaling limit will be set by other factors such as noise [5], reliability [6], drain current reduction, direct tunneling between the source and drain, on-chip interconnections, power dissipation, or even the atomic dimensions [7]. Nevertheless, given the practical feasibility of devices that have a significant gate tunneling current under normal operating conditions, knowledge of the basic mechanisms of gate tunneling current, and the modeling and circuit design issues that are involved, will probably be essential for anyone working with these devices in the next decades, either from a device-level or circuit-level point-of-view.

While the basic physical underlying mechanisms of tunneling through a thin dielectric have been known for decades [8–10], their application to the characterization, modeling and evaluation of circuit performance in a way that is useful for the purpose of circuit or system design is still an active field of research.

This review begins by presenting in Section 2 the essential physical mechanisms of tunneling in a metal–insulator–semiconductor (MIS) structure, from the classical Fowler–Nordheim regime for metal-contact MOS capacitors with thick dielectrics, to direct tunneling in modern structures with effects such as carrier quantization and tunneling from the valence band. The emphasis of this presentation is on the need for the development of simple yet accurate models of the current–voltage characteristics of the tunneling current.

In Section 3, the modeling approaches to gate tunneling current used in several industry-standard compact MOS models: BSIM4, MOS Model 11, EKV, SP and HiSIM

* Corresponding author. Tel.: +905 525 9140; fax: +905 523 4407.
E-mail address: jamal@mcmaster.ca (M.J. Deen).

0026-2714/$ - see front matter © 2006 Published by Elsevier Ltd.
doi:10.1016/j.microrel.2005.12.006

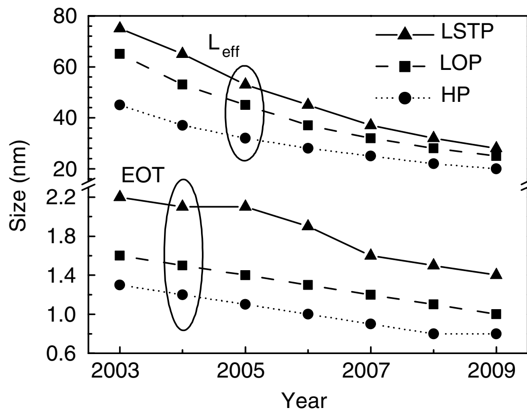

Fig. 1. Short-term ITRS 2004 projections of effective channel length (L_{eff}) and equivalent oxide thickness for (EOT) Low Standby Power (LSTP), low operating power (LOP) and high-performance logic (HP) technologies [2].

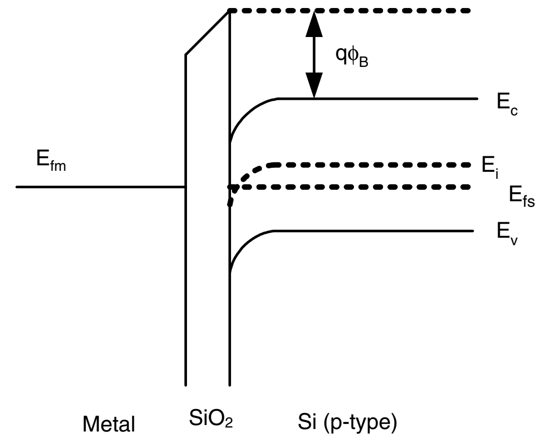

Fig. 2. Band diagram for a metal–oxide–semiconductor structure without applied voltage; the band bending is caused by the metal–semiconductor work function difference and the oxide fixed charges. E_c is the conduction band, E_i is the intrinsic Fermi level, E_{fs} is the Fermi level in the semiconductor and E_{fm} is the Fermi level in the metal.

are presented, building on the concepts presented in the preceding section.

Finally, in Section 4, the impact of the tunneling current in the performance of the circuits or systems that employ MOS devices with ultra-thin oxides is studied, from the perspectives of power consumption, circuit design, and noise, both for digital and analog/RF applications.

This review spans several decades of research and development in a very active field, from basic physics to system-level design. For this reason, some topics are not discussed in detail. No consideration is made of the particulars of tunneling phenomena in SOI devices [11], or the different technological approaches that are being considered for reduction of the tunneling from a device-level [12,13] or circuit-level point-of-view [14]. Also, the influence of the tunneling current in device parameter extraction and characterization, is not treated extensively [15–17].

2. Tunneling in metal–insulator–semiconductor structures

2.1. Fowler–Nordheim tunneling in MOS structures

The energy band diagram for a metal–oxide–p-type semiconductor, without applied voltage, is shown in Fig. 2 (it is assumed that the metal–semiconductor work function difference is negative, as is the case for example for a MOS device with an Al gate on a Si substrate). When a large positive voltage is applied to the metal with respect to the substrate, the left-hand side of the band diagram is lowered, and tunneling of electrons from the conduction band of the semiconductor into the conduction band of the oxide, through an approximately triangular barrier can occur, as shown in Fig. 3(a).

This kind of tunneling through an approximately triangular potential barrier is known as Fowler–Nordheim[1] tunneling, and a relatively simple expression for the current as a function of the applied voltage has been derived [8]:

$$J_{\text{FN}} = \frac{q^3}{16\pi^2\hbar\phi_{\text{b}}} F_{\text{ox}}^2 \exp\left[-\frac{4}{3}\frac{\sqrt{2m_{\text{ox}}^*}\phi_{\text{b}}^{3/2}}{\hbar q}\frac{1}{F_{\text{ox}}}\right] \tag{1}$$

where q is the electron charge, $\hbar$ is Planck's reduced constant, m_{ox}^* is the electron effective mass in the insulator, ϕ_{b} is the barrier height at the semiconductor–oxide interface and F_{ox} is the electric field across the oxide. The assumptions for the derivation of (1) are [8,9]:

- The electrons on the emitting electrode can be described as a free Fermi gas.
- The dependence of carrier availability for tunneling with respect to temperature is not taken into account.
- The potential barrier has triangular shape and barrier lowering due to image forces is neglected.
- The effect of the insulator can be described by a single effective mass.
- The tunneling probability takes into account only the component of the electron momentum in the direction normal to the surface.

In the Fowler–Nordheim tunneling regime, a plot of the logarithm of $J_{\text{FN}}/F_{\text{ox}}^2$ versus $1/F_{\text{ox}}$ (known as Fowler–Nordheim plot) yields a straight line. For MOS structures with a relatively thick oxide and metal gate, this model provided a very good fit to experimental data for a wide range of applied voltage, as can be seen in Fig. 4.

For modern structures however, the relatively simple Fowler–Nordheim model is not enough to account for effects such as:

- Direct tunneling through an approximately trapezoidal barrier for low applied voltages.

[1] Fowler and Nordheim [18] explained the field emission of electrons from a metal into vacuum by tunneling through a triangular potential barrier.

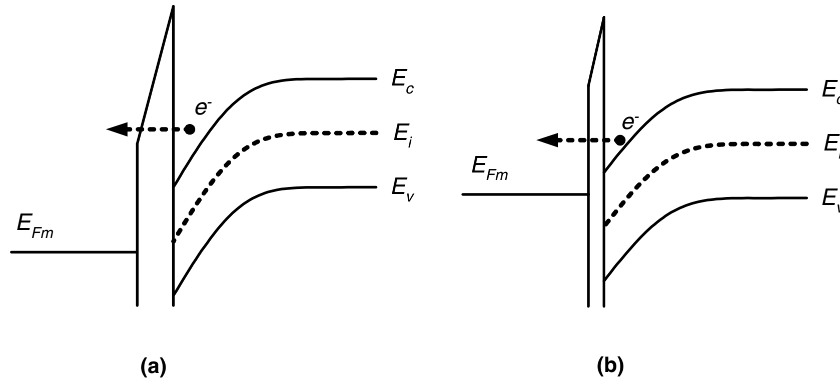

Fig. 3. (a) Tunneling through a triangular barrier into the insulator conduction band (Fowler–Nordheim tunneling). (b) Tunneling through a trapezoidal barrier (direct tunneling).

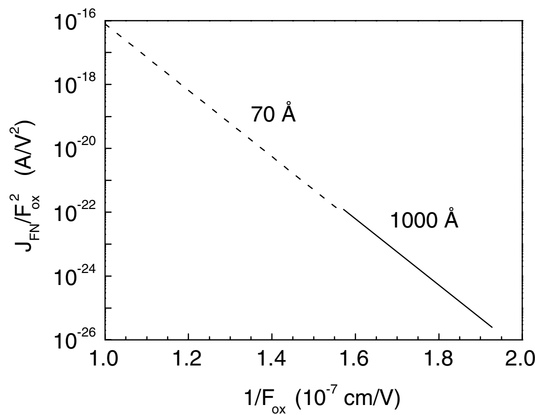

Fig. 4. Fowler–Nordheim plot of tunneling data from ⟨100⟩ silicon into thick SiO_2 oxides [9]. The two lines correspond to two sets of experimental data from different groups for two different oxide thicknesses; the slope and intersect of the Fowler–Nordheim plot are both independent of oxide thickness, as expected from (1).

- Carrier quantization in the inversion/accumulation layer.
- Tunneling mechanisms other than conduction band tunneling of electrons (valence band hole tunneling and valence band electron tunneling).
- Effect of the finite temperature on the availability of carriers for tunneling.
- Depletion effects on the polysilicon gate.

A discussion of these effects is presented in the following sections.

2.2. Direct tunneling

In the early MOS structures used for the study of tunneling currents, tunneling was observed when the applied field was high enough to cause lowering of one side of the potential barrier in the silicon–oxide interface, allowing tunneling into the conduction band of the oxide, as described in the previous section. The oxides used, in the order of 70–1000 Å [8,9], were thick enough to make direct tunneling into the conduction band of the other electrode very unlikely.

However, when the oxides have thicknesses in the order of a few nanometers, direct tunneling at lower fields is no longer negligible, and in this case the barrier is not triangular, but approximately trapezoidal (if barrier lowering due to image forces is neglected), as shown in Fig. 3(b). As can be seen in Fig. 5, the Fowler–Nordheim model is not valid when the potential across the insulator is lower than the potential barrier at the Si–SiO_2 interface.

In general, the direct tunneling current through the energy gap of an insulator or semiconductor is given by [20,21]:

$$J_T = \frac{2q}{(2\pi)^3\hbar}\int_0^{\infty}(f_1 - f_2)\left\{\int\int P \mathrm{d}k_y \mathrm{d}k_z\right\}\mathrm{d}E \quad (2)$$

where P is the tunneling probability, E is the total electron energy, k_y and k_z are the wave vectors in the plane of the barrier (perpendicular to the tunneling direction), and f_1 and f_2 are the probabilities of occupation of the states on each side of the barrier, given by the Fermi–Dirac

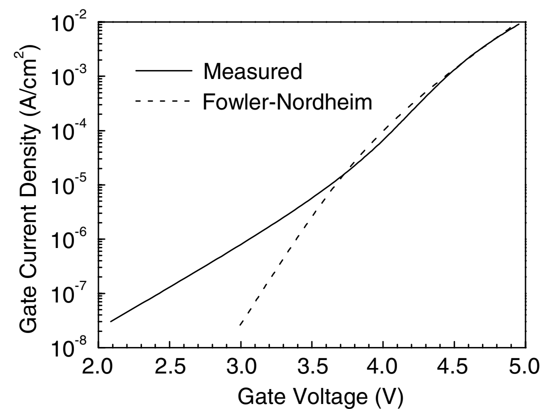

Fig. 5. Experimental gate current density versus gate voltage for a MOS device with an oxide thickness of 4.1 nm (solid line) and calculated Fowler–Nordheim current (dashed line) [19]. The Fowler–Nordheim model underestimates the gate current at low voltages.

distribution functions; the integration limits for k_y and k_z are determined from momentum and energy conservation.

Assuming that a single effective mass can be used for the electron in all three regions and integrating the Fermi–Dirac distribution function yields [22]:

$$J_{\mathrm{T}} = \frac{4\pi q m_x^* kT}{h^3} \int_0^\infty P(E_x) S(E_x) \mathrm{d}E_x \tag{3}$$

where h is Planck's constant, m_x^* is the electron effective mass in the direction perpendicular to the barrier, E_x is the electron kinetic energy in the direction perpendicular to the barrier, k is the Boltzmann constant, T is the temperature, and $S(E_x)$ is the "supply" function, which is derived from the integration of the Fermi–Dirac distribution function, and is given by [3]:

$$S(E_x) = \ln\left\{\frac{1+\exp[(E_{\mathrm{fs}}-E_x)/kT]}{1+\exp[(E_{\mathrm{fg}}-E_x)/kT]}\right\} \tag{4}$$

where E_{fs} and E_{fg} are the electron Fermi levels in the semiconductor and the gate, respectively. Note that the applied voltage is implicitly present in this equation through the difference between the Fermi levels.

The problem of determining the tunneling current then reduces to that of finding the probability of tunneling through the barrier, and solving the integral in (3). Calculation of the tunneling probability requires solving Schrödinger's equation for the electron wave function; typically this is done only in the direction perpendicular to the surface, and using the effective mass approximation, that is, assuming the electron is free and including the effect of the periodic lattice potential through the use of the effective mass. In that case, Schrödinger's wave equation in one dimension is

$$\frac{\hbar^2}{2m_x^*}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi(x) + [E-\phi(x)]\psi(x) = 0 \tag{5}$$

where x is the position perpendicular to the $Si–SiO_2$ interface, ψ is the time-independent part of the wave function and ϕ is the potential. Solving this equation exactly for an arbitrary potential distribution is a non-trivial problem. If the potential were constant, the position-dependent wave function in one dimension would be given by plane waves of the form:

$$\psi(x) = C_1\exp(+\mathrm{i}kx) + C_2\exp(-\mathrm{i}kx) \tag{6}$$

where i is the imaginary unit, k is the wave number in the direction perpendicular to the barrier, and C_1 and C_2 are constants. If the potential is not constant, but it changes "slowly" with position, then an approximate solution to Schrödinger's position-dependent equation is [23]:

$$\psi(x) = \frac{C_1}{\sqrt{k(x)}}\exp\left(+\mathrm{i}\int k(x)\mathrm{d}x\right) + \frac{C_2}{\sqrt{k(x)}} \times \exp\left(-\mathrm{i}\int k(x)\mathrm{d}x\right) \tag{7}$$

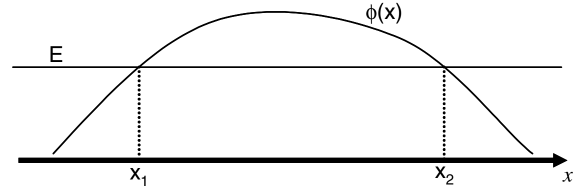

Fig. 6. Arbitrary potential barrier, showing the classical turning points.

This is known as the Wentzel–Kramers–Brillouin (WKB) approximation. The condition of a "slow" change of the potential means that the changes in the potential and its derivative on each electron wavelength, should be small compared to the energy of the particle.

If the total energy of the particle is lower than the potential ($E < \phi$), which is the case inside the potential barrier, then the wave number k in (7) is imaginary, and (7) becomes

$$\psi(x) = \frac{C_1}{\sqrt{\kappa(x)}}\exp\left(+\int \kappa(x)\mathrm{d}x\right) + \frac{C_2}{\sqrt{\kappa(x)}}\exp\left(-\int \kappa(x)\mathrm{d}x\right) \tag{8}$$

where $\kappa = \mathrm{i}k$ is real.

Using this approximation, the tunneling probability through a barrier of arbitrary shape is approximately given by [23]:

$$P \approx \exp\left(-2\int_{x_1}^{x_2}\kappa(x)\mathrm{d}x\right) \tag{9}$$

where x_1 and x_2 are the classical turning points, that is, the points at which the energy of the particle is equal to the potential, as shown in Fig. 6. For direct tunneling through a trapezoidal barrier, the classical turning points are the metal–insulator and insulator–semiconductor interfaces [24].

The integral in (9) can be computed by performing a change of variables if $\kappa(E)$ and $E(x)$ are known. For a trapezoidal barrier, if image force lowering of the barrier is neglected, $E(x)$ depends linearly on x (see Fig. 3). The relation between E and κ for an electron in free space, known as dispersion relation, is quadratic. This relation is often extended to electrons in solids by the use of the effective mass m^*:

$$\kappa(E) = \sqrt{\frac{2m^*}{\hbar^2}(E-E_{\mathrm{c}})} \tag{10}$$

where E_{c} is the conduction band energy. Since this approximation is valid only close to a band edge in the energy gap, other E–κ relations have been proposed which are more consistent with the case of two bands separated by an energy gap, as in an MOS structure. One of the most common is the so-called Franz relation, given by [10,25],[2]

[2] Nevertheless, the simple parabolic relationship is still used by many authors. Weinberg [9] showed that, in the Fowler–Nordheim tunneling regime, the simple parabolic E–κ relationship is equivalent to the Franz relation, if the effective mass value is changed accordingly.

$$\kappa(E)=\sqrt{\frac{2m_{ox}^{*}}{\hbar^{2}}(E-E_{c})\left(1-\frac{(E-E_{c})}{E_{gox}}\right)} \quad (11)$$

where E_{gox} is the insulator energy gap. Using (11), and, for example, a linear dependence of E with respect to x for an ideal trapezoidal barrier, the tunneling probability given by (9) can be calculated.

One drawback of the WKB approximation is that it neglects reflections and interference of the electron wave functions at the oxide–electrode interface, because it assumes a smoothly-varying potential, even at the edges of the barrier. Gundlach [26] solved Schrödinger's wave equation numerically for a trapezoidal potential by using Airy functions [27], and found that the current does not increase monotonically with the field, as is predicted by the WKB approximation, but it shows some periodic "oscillations", which he attributed to electron wave function reflection and interference at the barrier. This effect has also been observed experimentally [10], as shown in Fig. 7.

If when using the WKB approximation the matching of the wave functions in the abrupt potential barrier is taken into account, then a modified expression for the transmission probability can be found [20,29,32]

$$P=\frac{16}{\left[\frac{m_{ox}^{*}k_{g}}{m_{g}^{*}\kappa(x_{1})}+\frac{m_{g}^{*}\kappa(x_{1})}{m_{ox}^{*}k_{g}}\right]\left[\frac{m_{ox}^{*}k_{s}}{m_{s}^{*}\kappa(x_{2})}+\frac{m_{s}^{*}\kappa(x_{2})}{m_{ox}^{*}k_{s}}\right]}\times\exp\left(-2\int_{x_{1}}^{x_{2}}\kappa(x)\mathrm{d}x\right) \quad (12)$$

where m_g^*, m_{ox}^* and m_s^* are the effective masses in the gate, insulator and semiconductor, respectively, and k_g and k_s are the wave numbers in the gate and semiconductor, respectively. Note that this is similar to (9), but with the addition of a pre-exponential factor.

The expression in (3), along with the WKB approximation for the transmission coefficient with corrections like (12), has been used to model the tunneling current in MOS structures in the direct tunneling regime with reasonable accuracy [3,29,32]. However, in the context of compact modeling, the need for integration in (3), calls for the use of approximations that yield a closed-form expression for the tunneling current as a function of the applied voltage.

In the early studies on tunneling, one common approximation was to replace the Fermi–Dirac functions with the Fermi level for very low temperatures [26]. This kind of approximation provided insight into the physics of the process, but is clearly inappropriate for device modeling, because at the typical operating temperatures of electronic devices the full Fermi–Dirac statistics are required.

Another possible approximation is to neglect the effect of the finite availability of carriers for tunneling, given in (3) by the presence of the Fermi–Dirac functions, by assuming a degenerate inversion or accumulation layer in the semiconductor surface [24]. Under this assumption, Schuegraf et al. [19] proposed the following expression for the tunneling current in the direct tunneling regime:

$$J_{T}=\frac{q^{3}}{16\pi^{2}\hbar\phi_{b}}\frac{1}{\left[1-\sqrt{1-\frac{qV_{ox}}{\phi_{b}}}\right]^{2}}F_{ox}^{2}\times\exp\left\{-\frac{4}{3}\frac{\sqrt{2m_{ox}}\phi_{b}^{3/2}}{\hbar qF_{ox}}\left[1-\left(1-\frac{qV_{ox}}{\phi_{b}}\right)^{3/2}\right]\right\} \quad (13)$$

This can be used to model the tunneling current for moderate applied bias, where there is strong inversion or accumulation, but the field is not high enough to cause Fowler–Nordheim tunneling. However, for low voltages, where there is a limited availability of carriers for tunneling, this expression can overestimate the tunneling current by several orders of magnitude as shown in Fig. 8; besides,

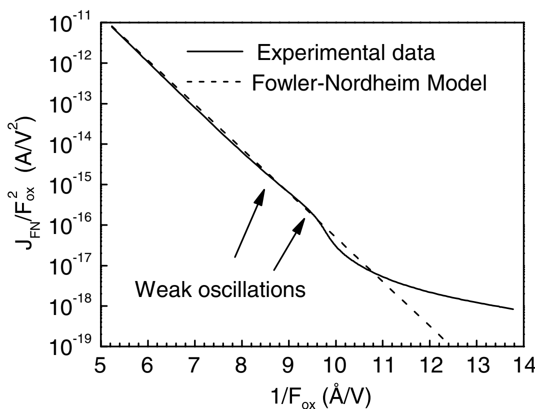

Fig. 7. Fowler–Nordheim plot of measured tunneling current (solid line) in a MOS structure ($t_{ox}=40.5$ Å), showing weak oscillations around the straight line (dashed line) that is predicted by the Fowler–Nordheim equation [10].

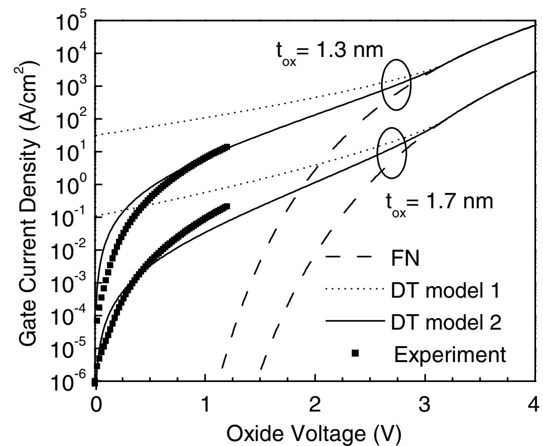

Fig. 8. Analytical models for the gate tunneling current and experimental data for 10 μm × 10 μm nMOS transistors with ultra-thin oxides at $V_{ds}=0$. The two direct tunneling current models correspond to (13) (DT model 1) and (14) (DT model 2) and the experimental data is from [15]. The calculations were made for $\phi_b=3.1$ eV and $m_{ox}=0.5m_e$.

in (13) the current does not go to zero for zero field, which is to be expected from basic physics. An alternative expression that overcomes this problem is [30]

$$J_{\mathrm{T}} = \frac{q^3}{16\pi^2\hbar\phi_{\mathrm{b}}}F_{\mathrm{ox}}^2 \times \exp\left\{-\frac{4}{3}\frac{\sqrt{2m_{\mathrm{ox}}}\phi_{\mathrm{b}}^{3/2}}{\hbar qF_{\mathrm{ox}}}\left[1-\left(1-\frac{qV_{\mathrm{ox}}}{\phi_{\mathrm{b}}}\right)^{3/2}\right]\right\} \quad (14)$$

but, as shown in Fig. 8, this model still has problems at low voltages for very thin oxides, where it cannot predict the correct trend for the current even if the effective mass in the oxide and the barrier height are used as fitting parameters. Accurate modeling of the direct tunneling current in the moderate inversion or accumulation regime requires calculating the relation between the surface potential and the applied voltage [33], so the gate current can not be expressed analytically in terms of the applied voltage.

Lee and Hu [34] derived a quasi-empirical model that takes into account the state availability. It will be presented later in this work, in the context of the different tunneling current components.

Another common approximation is to assume that the tunneling probability is constant for all energies, so the tunneling probability can be taken out of the integral in (3) [13,35]. Typically, the tunneling probability is calculated for the energy at the bottom of the conduction band.

2.3. Energy quantization in the inversion and accumulation layers

One of the basic assumptions in the derivation of the Fowler–Nordheim formula and in the calculations for direct tunneling shown in the previous section, is that the electrons in the semiconductor can be treated as a three-dimensional free Fermi gas. However, the electric field in the gate–channel region can create a band bending strong enough that confines the electrons to a narrow potential well close to the semiconductor surface, and their energy is then quantized in the direction normal to this surface, as shown in Fig. 9. This is often referred to as a 2-dimensional electron gas (2-DEG), because the energy of the electrons in one direction is quantized, but the energies in the other two directions can have any value [36].

One of the effects of this quantization is that the carrier concentration distribution in the channel is different from that obtained from the classical theory, which is based on the assumption of energy continuum [37]. The peak of the carrier concentration lies somewhere in the bulk of the semiconductor, instead of at the surface, so the semiconductor capacitance is overestimated whenever quantum effects are not taken into account [37]. The quantization effects in the substrate also results in an increase of the threshold voltage of MOS transistors. The effect is especially important for highly doped substrates semiconductor substrates [37,40], as shown in Fig. 10.

For the calculation of the gate tunneling current, the main consequence of quantization of electron energies in the accumulation/inversion layers is that the electrons in the semiconductor can no longer be modeled as a free electron gas, and the contribution of each of the quantized energy levels must be calculated separately. In this case, instead of a "transmission probability", a "carrier tunneling lifetime" is used for the calculation of the tunneling current from the quantized levels [31]. These levels are often referred to as quasi-bound states, because electrons can tunnel through the barrier, and thus are not completely bound inside the potential well.

Since the carrier distribution affects the potential, and the potential in turn affects the carrier distribution, in order to find the quantized energy levels it is necessary to solve simultaneously Schrödinger's equation for the wave function of the electrons:

$$\frac{\hbar^2}{2m_x^*}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi_i(x) + [E_i - \phi(x)]\psi_i(x) = 0 \quad (15)$$

and Poisson's equation for the potential:

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\phi(x) = -q\frac{\rho(x)}{\varepsilon_{\mathrm{s}}} \quad (16)$$

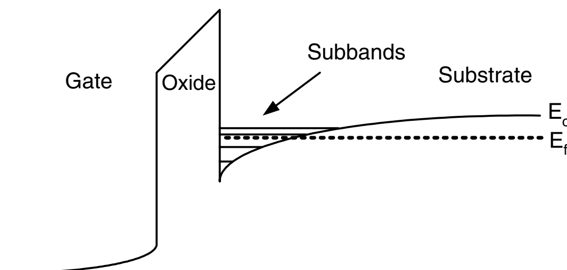

Fig. 9. Schematic band profile for a poly $Si–SiO_2–Si$ structure in inversion, showing the formation of sub-bands due to carrier confinement in a narrow potential well [39].

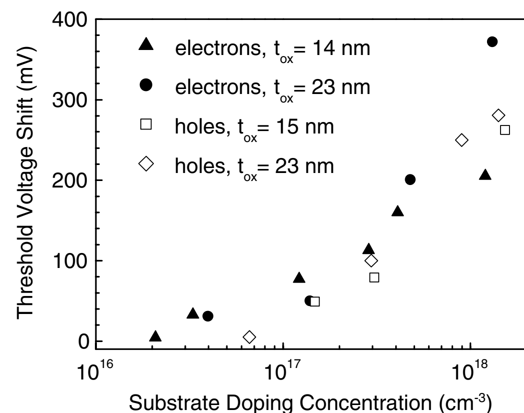

Fig. 10. Variation in the threshold voltage with respect to the classical theory caused by quantum-mechanical effects in the substrate for electron and hole inversion layers [40].

where ψ_i is the wave function for the ith energy state ($i = 1, 2, 3, \ldots$), ρ is the charge density in the semiconductor and ε_s is the electrical permittivity of the semiconductor.

Because of the band structure of the semiconductor, there are typically several different effective masses corresponding to different energy sub-bands, and Eq. (15) is expressed as

$$\frac{\hbar^2}{2m^*_{xj}}\frac{d^2}{dx^2}\psi_{ij}(x) + [E_{ij} - \phi(x)]\psi_{ij}(x) = 0 \tag{17}$$

where the index j corresponds to the different effective masses. For the case of silicon with a surface orientation of $\langle 100\rangle$, Eq. (15) has to be solved for $j = 1$ with $m^*_{x1} = m^*_l$ and $j = 2$ with $m^*_{x2} = m^*_t$, where m^*_l and m^*_t are the longitudinal and transversal electron mass, respectively [37].

Since the potential ϕ appears both in Schrödinger's and Poisson's equations, to find the exact distribution of the potential and the discrete energy states both equations must be solved in a fully consistent way. Moreover, the wave function ψ_{ij} appears indirectly in Poisson's equation through the charge density ρ. For example, when the electron energies are quantized, the charge density in strong inversion for an nMOS structure is given by [37]:

$$\rho(x) = -q\left[(N_A - N_D) + \sum_j \sum_i N_{ij}|\psi_{ij}(x)|^2 - p(x)\right] \tag{18}$$

where $(N_A - N_D)$ is the net doping concentration in the semiconductor, $p(x)$ is the hole concentration, which can be calculated classically, and N_{ij}, the number of electrons occupying the ijth energy level, is given by

$$N_{ij} = \frac{g_{vj}m^*_{dj}}{\pi\hbar^2}kT\ln\left[1 + \exp\left(\frac{E_{Fn} - E_{ij}}{kT}\right)\right] \tag{19}$$

where g_{vj} is the degeneracy of the jth valley, m_{dj} is the density of states effective mass and E_{Fn} is the electron quasi-Fermi level in the semiconductor. In $\langle 100\rangle$ silicon, for $j = 1$, $m^*_{d1} = m^*_t$ and for $j = 2$, $m^*_{d2} = (m^*_l m^*_t)^{1/2}$.

A common approximation which allows decoupling of Eqs. (16) and (17) is to assume that the potential ϕ in the semiconductor is independent of the exact carrier distribution and varies linearly with distance [37,41] as

$$\phi(x) = -qF_s x \tag{20}$$

where F_s is the electric field in the semiconductor surface. In this case, Schrödinger's wave equation takes the form:

$$\frac{d^2}{dx^2}\psi_{ij}(x) + \frac{2m^*_{xj}}{\hbar^2}[E_{ij} - qF_s x]\psi_{ij}(x) = 0 \tag{21}$$

This differential equation can be solved in terms of Airy functions [27], and if the barrier height at the semiconductor surface is assumed to be infinite (corresponding to a boundary condition of $\psi_{ij} = 0$ for $x = 0$), then the discrete energy levels are approximately given by

$$E_{ij} \approx \left(\frac{\hbar^2}{2m_{xj}}\right)^{1/3}\left[\frac{3}{2}\pi qF_s\left(i - \frac{1}{4}\right)\right]^{2/3} \tag{22}$$

As (22) is an approximate solution for "large" values of i, for $i = 1$, $i = 2$ and $i = 3$, the term $(i - 1/4)$ has to be replaced with 0.7587, 1.7540 and 2.7525, respectively [41].

Clearly, the assumption of an infinite potential barrier is not consistent with the presence of tunneling current, since an infinite barrier implies that there is zero probability of the wave function to penetrate into the barrier, corresponding to no tunneling current. If the actual barrier height is used as boundary condition, then the values of the discrete energies will be different, and as expected, the difference is more important for thinner oxides, as shown in Fig. 11 [42].

The linear potential approximation given by (20) is valid only in weak inversion [41]. For strong inversion or accumulation, either other approximations or the numerical solution of the wave and potential equations have to be used [41,43]. An additional complication that arises in the accumulation region is that near the surface of the semiconductor, there is a mixture of quantized energy levels corresponding to the bound electrons in the potential well, and unbound electrons whose wave functions extend into the bulk of the semiconductor, which also contribute to the surface potential [43]. This is not the case in inversion because the depletion layer separates the bound and unbound states.

Once the energy levels E_{ij} are known, for each of them a "quasi-bound state lifetime" can be calculated as [44]

$$\tau_{ij} = \frac{j\pi\hbar}{E_{ij}} \tag{23}$$

and the tunneling current will be the summation of the contribution from each of those quasi-bound states

$$J_T = q\sum_{i,j}\frac{N_{ij}P_{ij}}{\tau_{ij}} \tag{24}$$

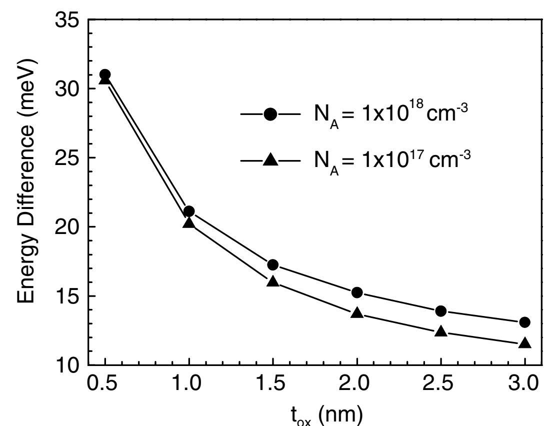

Fig. 11. Difference in the quasi-bound state energies calculated by assuming finite boundary conditions and infinite boundary conditions, versus oxide thickness [42].

where P_{ij} is the transmission probability associated with each energy level (calculated by any of the methods presented in the previous section) and N_{ij} is the carrier occupation for each sub-band, given by (19). The quasi-bound state lifetime given by (23) is derived using a semi-classical approach, where each electron is considered as a point-like charged particle bouncing inside the well [45].

It is also common to include the transmission coefficient as part of the lifetime (thus referred to as "tunneling lifetime" [44] or "decay lifetime" [39]), and then Eq. (24) becomes [39]

$$J_T = q \sum_{i,j} \frac{N_{ij}}{\tau_{ij}} \quad (25)$$

Unfortunately, in the literature the same symbol is used for both lifetimes.

A fully quantum-mechanical alternative to this semi-classical approach is to consider the potential well in the inversion/accumulation layer, and the barrier in the insulator, as a scattering center for electrons incident from the bulk. From the continuum of possible energies for the incident particle, at some particular energies, the wave equation will have a "resonant" peak inside the well, which corresponds to the quasi-bound states [28]. If the well is infinite, the energy is discrete and the state is bound, but since tunneling is allowed into the oxide, there is a "broadening" in the energy, which is related to the lifetime through the energy–time uncertainty principle. The complete solution of Schrödinger's equation in this case yields complex energies whose imaginary part Γ_{ij} [38] is known as the resonance width [23], and is a measure of the energy broadening of the quasi-bound state. This resonance width can be directly related to the tunneling lifetime by Heisenberg's uncertainty principle [38]:

$$\tau_{ij} = \frac{\hbar}{\Gamma_{ij}} \quad (26)$$

Note that this tunneling lifetime already has the information related to the tunneling probability, because the broadening of the energy levels is caused precisely by the tunneling through the barrier. The determination of Γ has to be performed numerically, and one of the methods proposed for this calculation is the so-called transverse-resonant method [38], which makes an analogy of the electron confined in the potential well to a waveguide with a varying refractive index.

It has been reported that the calculation of tunneling current considering energy quantization effects in the substrate, yields numerical results very similar to those obtained using the classical assumption of continuous energies in the substrate [39]. This has been explained by the presence of two compensating effects: the carrier distribution peak shifts away from the semiconductor surface, thus increasing the tunneling distance (lower tunneling probability), and the quasi-bound states have energies above the conduction band minimum, thus reducing the barrier height (higher tunneling probability) [39]. This is the case for bulk devices; for SOI devices, the tunneling current might be significantly affected by carrier confinement [11].

The steps outlined above for the calculation of the tunneling current from quasi-bound states are typically implemented numerically, often together with the self-consistent solution of Schrödinger and Poisson's equations [38,43,42]. For the purpose of compact modeling, the carrier quantization is taken into account by assuming a "barrier lowering" equal to the difference between the first energy level and the conduction band [46], or it is ignored altogether [29], on the basis of the self-compensating effects mentioned previously.

2.4. Tunneling current components

In addition to tunneling of electrons from the conduction band (ECB) at one side of the oxide barrier, into the conduction band at the other side, other tunneling mechanisms have been identified for modern MOS structures. These are electron tunneling from the valence band into the conduction band (often referred to as EVB), and hole tunneling into the valence band (HVB). The different mechanisms are shown schematically in a band diagram for a Si–SiO_2–Si structure in Fig. 12. Note that HVB tunneling is equivalent to electron tunneling from valence band to valence band, as opposed to EVB tunneling, where the electron goes from the valence band into the conduction band.

Additionally, in an MOS transistor, several different current components can be identified, as shown in Fig. 13. When the device operates in inversion, tunneling takes place between the gate and the channel (I_{gc}); current also flows between the gate and the bulk (I_{gb}) both in accumulation and inversion, and in all the operating regions there is tunneling in the region where the gate overlaps the source and drain (I_{gs} and I_{gd}). This last component is also known as edge direct tunneling (EDT), and it is of particular importance in short-channel devices, where the ratio

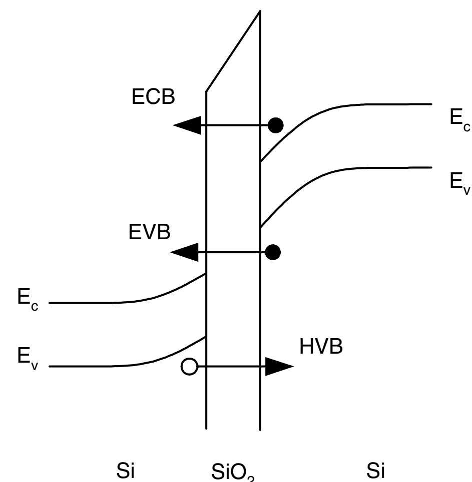

Fig. 12. Schematic representation of the tunneling components in a Si/SiO_2/Si structure [34].

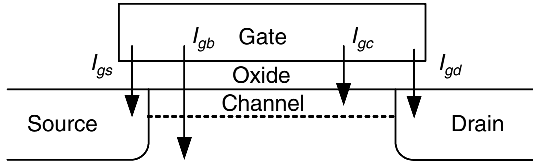

Fig. 13. Gate tunneling current components for a MOS transistor.

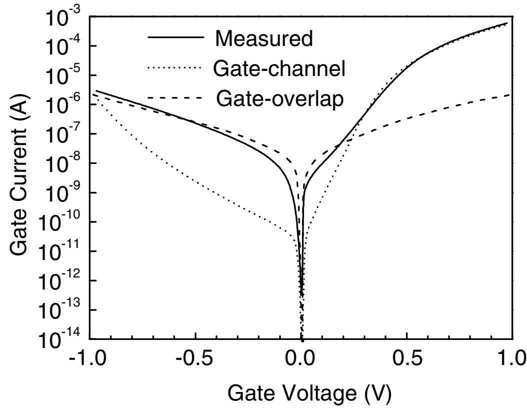

Fig. 14. Measured gate current for an n+ Si/SiO$_2$/p Si MOS capacitor with an n+ source, $t_{ox} = 1.6$ nm and $W/L = 562/62$ μm (solid line). The dashed and dotted lines show the calculated gate-overlap and gate–channel components, respectively [29]. The overlap current component is dominant at low positive and negative gate voltages (depletion–accumulation), even though the overlap area is much smaller than the channel area.

of the source–drain extensions to the channel length is higher [47]. Indeed, EDT current has been reported to be more significant than other leakage mechanisms such as gate-induced drain leakage (GIDL) and band-to-band tunneling (BTBT) for ultra-thin gate oxides, in the order of 1.2 nm [48]. Fig. 14 clearly shows how the overlap currents dominate the gate current in the accumulation and depletion regimes, even when the channel area is much larger than the overlap area.

Depending on the operating region, different tunneling mechanisms dominate the different components. Consider for example an nMOS in inversion: the gate-to-channel current (I_{gc}) is dominated by ECB, and there is also a smaller gate–bulk current I_{gb} caused by EVB at high voltages. In accumulation, there is no channel current, but only gate–bulk current, caused by EVB. Both in inversion and accumulation, EDT current is dominated by ECB. Table 1 summarizes the dominant mechanisms for each tunneling current component, both for nMOS and pMOS structures.

Table 1
Dominant current mechanisms for each tunneling current component [51]

Region of operation	Current component			
	I_{gc}	I_{gb}		I_{gs}, I_{gd}
	Inversion	$V_{ox} > 0$	$V_{ox} < 0$	All
NMOS	ECB	EVB	ECB	ECB
PMOS	HVB	ECB	EVB	HVB

Note that, for each tunneling mechanism, the probability of tunneling will depend on the height of the barrier, which is different for holes and electrons. It will also depend on the "availability" of carriers in the material from where the carriers are tunneling [49], and the availability of states in the receiving material [34], and these factors have to be taken into account when developing models for the tunneling current. The case of EVB is of particular interest because, due the band alignment, this current component is only important for relatively high voltages, because at low voltages, receiving states would be required in the energy gap of the semiconductor, as shown in Fig. 15.

Lee and Hu [34] developed a quasi-empirical model that takes into account the different "supplies" for the different tunneling mechanisms:

$$J = \frac{q^3}{16\pi^2\hbar\phi_b\varepsilon_{ox}} C(V_g, V_{ox}, t_{ox}, \phi_b) \times \exp\left\{-\frac{4}{3}\frac{\sqrt{2m_{ox}}\phi_b^{3/2}}{\hbar q|F_{ox}|}\left[1-\left(1-\frac{qV_{ox}}{\phi_b}\right)^{3/2}\right]\right\} \quad (27)$$

where ε_{ox} is the electric permittivity of the oxide, t_{ox} is the oxide thickness, V_{ox} is the voltage across the oxide and $C(V_g, V_{ox}, t_{ox}, \phi_b)$ is a "correction" function, developed by empirical fitting, given by

$$C(V_g, V_{ox}, t_{ox}, \phi_b) = \exp\left[\frac{20}{\phi_b}\left(\frac{|V_{ox}|-\phi_b}{\phi_{bo}}+1\right)^{\alpha}\left(1-\frac{|V_{ox}|}{\phi_b}\right)\right]\frac{V_g}{t_{ox}}N \quad (28)$$

where α is a fitting parameter and N, which represents the density of carriers in the injecting surface, is given by

$$N = \frac{\varepsilon_{ox}}{t_{ox}}\left\{n_{inv}V_t \ln\left[1+\exp\left(\frac{V_{ge}-V_{th}}{n_{inv}V_t}\right)\right] + n_{acc}V_t \ln\left[1+\exp\left(-\frac{V_g-V_{fb}}{n_{acc}V_t}\right)\right]\right\} \quad (29)$$

for ECB and HVB in inversion and accumulation, where V_{fb} is the flat-band voltage, V_{ge} is the gate voltage minus the gate depletion voltage (discussed in next section), V_g is the gate voltage, V_t is the thermal voltage (kT/q), V_{th} is the threshold voltage and n_{inv} and n_{acc} are fitting parameters. For EVB, the density of carriers is given by

$$N = \frac{\varepsilon_{ox}}{t_{ox}} n_{EVB}V_t \ln\left[1+\exp\left(\frac{|V_{ox}|-E_g/q}{n_{EVB}V_t}\right)\right] \quad (30)$$

where n_{EVB} is a fitting parameter and E_g is the semiconductor energy band gap.

This semi-empirical model is the basis for one of the circuit-level models that will be discussed later.

2.5. Channel current partition

So far we have implicitly been considering mostly MOS capacitor structures, where there is no drain-source region, or the drain and source are at the same potential.

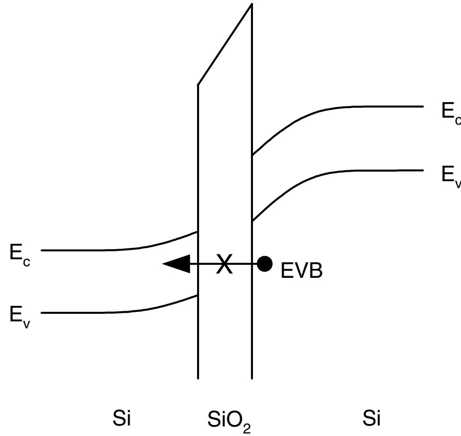

Fig. 15. Schematic band diagram showing that EVB tunneling for low voltages across the oxide would require receiving states inside the energy gap of the semiconductor [34].

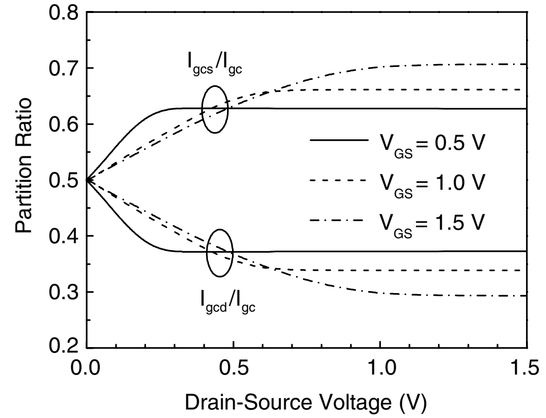

Fig. 16. Channel tunneling current partition ratio as a function of drain voltage for different gate–source voltages [60].

For MOS transistors with source and drain contacts, when a voltage is applied between drain and source, the tunneling current density becomes a function of the position y along the channel. The total gate current is then given by

$$I_g = W\int_0^L J_g(y)\mathrm{d}y \tag{31}$$

where W and L are the effective width and length of the channel. In (31), I_g and J_g can represent either gate–channel current or gate–bulk current, and in the case of the overlap regions, the integration would be performed along the source/drain extensions.

As was described in the previous section, in accumulation there is no channel current, and all the gate current goes to the bulk. In inversion however, in addition to the gate–bulk current there is a gate–channel current I_{gc}, part of which goes to the source and part to the drain. The question then arises as to what fraction of I_{gc} is transferred to the drain and what fraction to the source. It has been shown by solving the current continuity equation in the channel that the source and drain components are given by [46,50]:

$$I_{gcs} = W\int_0^L \left(1-\frac{y}{L}\right)J_g(y)\mathrm{d}y \tag{32}$$

and

$$I_{gcd} = W\int_0^L \frac{y}{L}J_g(y)\mathrm{d}y \tag{33}$$

These partition formulas have been verified by comparison with the numerical solution of the current continuity equation [50], and by using a channel segmentation model [46]. However, because of the difficulty of measuring the current partition ratio they have not been verified experimentally [35,46,51]. The partition ratios I_{gcs}/I_{gc} and I_{gcd}/I_{gc}, calculated using (32) and (33), are shown in Fig. 16 as a function of the drain–source bias. As expected, for zero drain voltage, the current is equally split between drain and source, and as the drain–source voltage increases, a bigger fraction of the current goes to the source terminal.

2.6. Polysilicon depletion

Another effect that must be considered when calculating the gate current is the polysilicon gate depletion. Being a semiconductor material, albeit highly doped, the polysilicon gate is subject to the formation of a very thin depletion layer, as opposed to an ideal conductor where the charge is distributed on an infinitely thin sheet at the surface.

The distribution of charge on this finite thickness region implies that there is a voltage drop on the depletion layer, which must be taken into account when determining the voltage across the oxide for the purpose of tunneling current calculation [19]. This voltage drop can be determined to various levels of accuracy, the simplest being a solution to the Poisson equation, under the depletion approximation. This yields an "effective" gate voltage given by [34]:

$$V_g = V_{fb} + \phi_s + \frac{q\varepsilon_s N_{gate} T_{ox}^2}{\varepsilon_{ox}^2}\left(\sqrt{1+\frac{2\varepsilon_{ox}^2(V_g - V_{fb} - \phi_s)}{q\varepsilon_s N_{gate} T_{ox}^2}} - 1\right) \tag{34}$$

where ϕ_s is the surface potential, ε_s and ε_{ox} are the dielectric constants of the semiconductor and insulator, respectively, V_{fb} is the flat-band voltage, N_{gate} is the doping concentration in the gate and T_{ox} is the oxide thickness.

As shown in Fig. 17, polysilicon depletion has to be taken into account when using *C–V* measurements to extract device parameters. Fig. 17 also shows the effect of the gate current in the measured *C–V* characteristics of thin-oxide MOSFETs.

When the Schrödinger and Poisson equations are solved self-consistently, the polysilicon gate must be included in the solution region, which should yield automatically the potential distribution in all the regions.

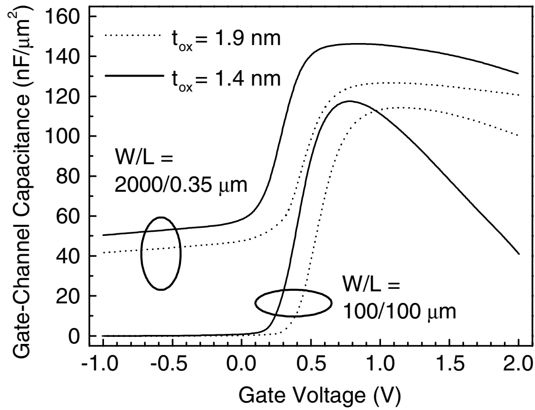

Fig. 17. Capacitance–Voltage measurements for thin-oxide nMOSFETS at $f = 1$ MHz [17]. For the thicker oxide (1.9 nm, dotted line) the capacitance reduction in inversion (high gate voltages) is caused mainly by polysilicon depletion. For the thinner oxide (1.4 nm, solid line), a sharp decrease in capacitance is observed for large-area devices; this can be attributed to the gate current. The non-zero accumulation capacitance is due to the source/drain overlap regions.

Similar to the inversion/accumulation layers in the bulk, quantization effects in the polysilicon depletion layer have been reported [52], but to the best of our knowledge, its impact on the calculation and modeling of the gate tunneling current has yet to be assessed.

2.7. Barrier lowering by image forces

The lowering of the potential barrier in the MOS structures caused by image forces (see Fig. 18) is often neglected in the calculation of the tunneling current, based on an argument given by Weinberg [9] in 1982: for large barriers (the case of Si–SiO_2) the image-force lowering of the barrier is very small, and this was supported by experimental evidence at the time.

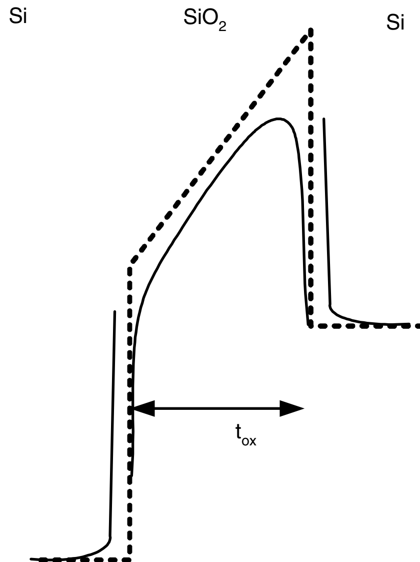

Fig. 18. Potential barrier without (dashed line) and with image forces (solid lines) at the Si–SiO_2 interface [54].

For very thin oxides, however, this might not be the case, and the barrier lowering can have an impact on the calculation of the tunneling current [53]. The use of an "effective" trapezoidal barrier, lowered with respect to the ideal one, has been proposed to account for image force effects [53]. Recent results, however, indicate that the tunneling current dependence on the field would not be modeled correctly by such method [54].

3. Gate current in compact MOS models

3.1. BSIM4

The Berkeley Short-channel IGFET Model (BSIM), one of the most widely used MOS models for circuit simulation, includes gate tunneling current starting from version 4.0.0 [55].

BSIM4 gate current model is based on the semi-empirical model of Lee and Hu [34], and it uses a common expression for all the tunneling current components and mechanisms [51]:

$$J_{\mathrm{T}} = \frac{q^3}{16\pi^2\hbar\phi_{\mathrm{b}}}\left(\frac{T_{\mathrm{oxref}}}{t_{\mathrm{ox}}P}\right)^{ntox}\frac{V_{\mathrm{aux}}V_{\mathrm{appl}}}{(t_{\mathrm{ox}}p)^2} \times \exp\left[-\frac{4}{3}\frac{\sqrt{2m^*_{\mathrm{ox}}}\phi_{\mathrm{b}}^{3/2}}{\hbar q}(\alpha-\beta|V_{\mathrm{ox}}|)(1+\gamma|V_{\mathrm{ox}}|)t_{\mathrm{ox}}p\right] \tag{35}$$

where T_{oxref} is a reference oxide thickness at which the parameters are extracted, ntox is a fitting parameter, V_{appl} is the applied voltage (which has a different meaning depending on the current component), and the function V_{aux} and the parameters α, β, γ and p depend on the tunneling mechanism (ECB, EVB or HVB), the region of operation (inversion, accumulation or depletion) and the current component (I_{gb}, I_{gc}, I_{gs} or I_{gd}). Table 2 shows the parameter values, as well as the definition of the auxiliary function V_{aux} and the applied voltage V_{appl}, for each operating region and current component.

Note the similarity between Eq. (35) and the classical Fowler–Nordheim tunneling Eq. (1). Here, the auxiliary function is used as a means to correct the inaccuracies caused by approximations such as the WKB, and also to account for the different "availability" of carriers for tunneling and of receiving states for each tunneling mechanism, as was discussed in Section 2.4.

Notice that in the expression for I_{gc} in Table 2 the "effective" gate voltage, which takes into account the depletion in the polysilicon region, is used instead of the applied gate voltage.

Eq. (35) provides the current densities. The currents are given by

$$I_{\mathrm{gb}} = W \cdot L \cdot J_{\mathrm{gb}} \tag{36}$$

Table 2
BSIM4 auxiliary function V_{aux} and parameters α, β and γ for gate tunneling current modeling

Region of operation	Current component	V_{aux}, V_g	α, β, γ, p
Acc.	J_{gb}	$V_{aux} = nigbacc \cdot V_t \ln\left[1+\exp\left(-\frac{V_{gb}-V_{fbzb}}{nigbacc*V_t}\right)\right]$ $V_{appl} = V_{gb}$	α = AIGBACC, β = BIGBACC, γ = CIGBACC, $p = 1$
Inv. and Depl.	J_{gb}	$V_{aux} = nigbinv \cdot V_t \ln\left[1+\exp\left(-\frac{V_{oxdepinv}-eigbinv}{nigbinv*V_t}\right)\right]$ $V_{appl} = V_{gb}$	α = AIGBINV, β = BIGBINV, γ = CIGBINV4, $p = 1$
	J_{gc}	$V_{aux} = nigc \cdot V_t \ln\left[1+\exp\left(-\frac{V_{gse}-V_{th0}}{nigc*V_t}\right)\right]$ $V_{appl} = V_{gse}$	α = AIGC, β = BIGC, γ = BIGC, $p = 1$
All	J_{gs}	$V_{aux} = \lvert V_{gs} - V_{fbsd}\rvert$ $V_{appl} = V_{gs}$	α = AIGSD, β = BIGSD, γ = CIGSD, p = POXEDGE
All	J_{gd}	$V_{aux} = \lvert V_{gd} - V_{fbsd}\rvert$ $V_{appl} = V_{gd}$	α = AIGSD, β = BIGSD, γ = CIGSD, p = POXEDGE

Here, $V_t = kT/q$ is the thermal voltage, V_{gb} is the gate–substrate voltage, V_{fbzb} is the flat-band voltage calculated from zero-bias threshold voltage, $V_{oxdepinv}$ is the voltage drop in the oxide in inversion and depletion, V_{gse} is the effective gate voltage taking into account the polysilicon depletion, V_{gs}, V_{ds} and V_{gd} are the gate–source, drain–source and gate–drain voltages, V_{th0} is the threshold voltage at $V_{bs} = 0$, V_{fbsd} is the flat-band voltage of the gate and source/drain diffusion areas, POXEDGE is a factor for the possible difference of the gate oxide thickness in the source/drain overlap regions, and *nigbacc*, *nigbinv*, *eigbinv*, *nigc*, AIGBACC, BIGBACC, CIGBACC, AIGBINV, BIGBINV, CIGBINV, AIGC, BIGC, CIGC, AIGSD, BIGSD and CIGSD are parameters that characterize the tunneling process in the different operating regions [55,51].

$$I_{gc} = W \cdot L \cdot J_{gc}, \tag{37}$$

$$I_{gs} = W \cdot \Delta L \cdot J_{gs}, \tag{38}$$

and

$$I_{gd} = W \cdot \Delta L \cdot J_{gd} \tag{39}$$

where ΔL is the length of the source/drain overlap.

One last consideration on the BSIM4 model is the partition of the gate–channel current I_{gc0} into the components I_{gcs} and I_{gcd}. Each of these components is given by

$$I_{gcs} = I_{gc0}\frac{\mathrm{PIGCD} \cdot V_{dseff} + \exp(-\mathrm{PIGCD} \cdot V_{dseff}) - 1}{\mathrm{PIGCD}^2 \cdot V_{dseff}^2} \tag{40}$$

and

$$I_{gcd} = I_{gc0}\frac{1 - [(\mathrm{PIGCD} \cdot V_{dseff} + 1) \cdot \exp(-\mathrm{PIGCD} \cdot V_{dseff})]}{\mathrm{PIGCD}^2 \cdot V_{dseff}^2} \tag{41}$$

where I_{gc0} is the gate–channel current at zero V_{ds}, PIGCD is a model parameter and V_{dseff} is an effective drain–source voltage, defined to ensure a smooth transition from the triode region to the saturation region. These equations are derived from the current continuity equation along the channel under the assumption that the gate current is small compared to the drain current [51].

3.2. Philips MOS Model 11

Philips MOS Model 11 is based on the explicit formulation of the surface potential, defined as the electrostatic potential at the gate oxide/substrate interface with respect to the neutral bulk [56]. This approach is different from the so-called threshold voltage based models like BSIM3 or Philips MOS 9 model [57], that use separate expressions for the drain current in the weak-inversion and strong inversion regions, and the moderate inversion region is modeled through the use of smoothing functions. In Philips MOS Model 11 model, it is assumed that the tunneling current is a small perturbation, and thus the surface potential is not affected by it.

The dominant tunneling mechanisms considered in MOS Model 11 for each region of operation are presented in Table 3. Note that MOS Model 11 does not take into account electron valence band (EVB) tunneling, and therefore, the gate-to-bulk current in inversion is not considered.

For all the tunneling components and regions of operation the tunneling probability is given by

$$P(V_{ox}, \chi_B, B) = \begin{cases} \exp\left[-B\dfrac{1-\left(1-\frac{V_{ox}}{\chi_B}\right)^{3/2}}{V_{ox}}\right] & V_{ox} < \chi_B \\ \exp(-B/V_{ox}) & V_{ox} \geqslant \chi_B \end{cases} \tag{42}$$

where V_{ox}, χ_B, and B depend on the tunneling mechanism and region of operation. The expression for $V_{ox} > \chi_B$ corresponds to Fowler–Nordheim tunneling (1), and the expression for $V_{ox} < \chi_B$ is essentially the WKB direct tunneling probability with a triangular barrier for tunneling from a single energy level.

In inversion, the gate–channel current density, considering only the gate–channel component and neglecting the gate–bulk current caused by EVB, is proportional to the

Table 3
Dominant current components in MOS Model 11 [56]

	Current component		
	I_{gc} Inversion	I_{gb} Accumulation	I_{gs}, I_{gd} All
NMOS	ECB	ECB	ECB
PMOS	HVB	ECB	HVB

tunneling probability, the oxide voltage V_{ox} and the inversion layer charge Q_{inv} and is

$$J_G \propto -V_{ox} \cdot Q_{inv} \cdot P(V_{ox}, \chi_{Beff}, B_{inv}) \quad (43)$$

where the proportionality constant is a fitting parameter and the parameters χ_{Beff} and B_{eff} are calculated taking into account quantization in the inversion layer, by calculating a "barrier lowering" equal to the difference between the conduction band and the first discrete energy level.

The total gate channel current is found by integrating (43) along the channel and the partition of the gate–channel current is similar to that discussed in Section 2.5. The respective integrals are solved analytically by linearizing the exponent in the tunneling probability.

In accumulation, the gate current density is given by

$$J_G \propto -V_{ox} \cdot Q_{acc} \cdot P(-V_{ox}, \chi_{Bacc}, B_{acc}) \quad (44)$$

where Q_{acc} is the accumulation layer charge, and χ_{Bacc} and B_{acc} are the barrier height in accumulation and a fitting parameter, respectively. In accumulation, quantization effects are not considered.

The tunneling current in the source/drain overlap regions is given by

$$I_{Gov} \propto \mp V_{ov} \cdot Q_{ov} \cdot P(\pm V_{ov}, \chi_{Binv}, B_{inv}) \quad (45)$$

where V_{ov} and Q_{ov} are the voltage and the total charge density in the source/drain overlaps and the signs depend on whether the gate–source voltage is greater or lower than an "effective flat-band voltage" that is defined for these regions.

The EKV compact model [58] uses an approach similar to MOS Model 11 for gate current modeling, with adaptations for the EKV charge based formalism [59].

3.3. SP model

SP is a surface potential-based compact MOS model [60], and its gate current model is based in (3). In order to avoid the integral, the tunneling probability and the supply function are assumed to be independent of the energy, so the tunneling current density is given by

$$J_G(y) \approx J_0 \cdot P \cdot S \quad (46)$$

where P is the tunneling probability, S is the supply function and

$$J_0 = \frac{q m_s^* k^2 T^2}{2\pi^2 \hbar^3} \quad (47)$$

The tunneling probability is derived using the WKB approximation (9), assuming a parabolic dispersion relation in the oxide as in (10) and a triangular potential barrier, which results in

$$P = \exp\left(-\frac{4}{3}\frac{\sqrt{2m_{ox}^*\phi_b}}{q\hbar} t_{ox} \frac{\phi_b}{V_{ox}}\left[1-\left(1-\frac{qV_{ox}}{\phi_b}\right)^{3/2}\right]\right) \quad (48)$$

In order to simplify the evaluation of P, the exponent of (48) is approximated using

$$f\left(\frac{\phi_b}{V_{ox}}\right) \equiv \frac{\phi_b}{V_{ox}}\left[1-\left(1-\frac{V_{ox}}{\phi_b}\right)^{3/2}\right] \approx G_1 + G_2\frac{V_{ox}}{\phi_b} + G_3\left(\frac{V_{ox}}{\phi_b}\right)^2 \quad (49)$$

where the coefficients G_1, G_2 and G_3, which have a theoretical value given by the Taylor series expansion of (49), are used as fitting parameters to account for the errors of the WKB approximation, the assumption of a single energy level and uncertainties in the values of the effective masses and barrier height.

The supply function is equivalent to (4), but is rewritten in terms of the surface potential ϕ_s[3] as

$$S(\phi_s) = \ln\left\{\frac{1+\exp[(\phi_s - \phi_n - \phi_b - E_t/q)/V_t]}{1+\exp[(\phi_s - V_{gb} - \phi_b - E_t/q)/V_t]}\right\} \quad (50)$$

where ϕ_n is the quasi-Fermi level splitting, ϕ_b is the electron quasi-Fermi level in the bulk and E_t is the single energy level (measured from the conduction band edge) from which the carriers are assumed to tunnel in (46). In order to account for possible differences in the tunneling parameters at the polysilicon–insulator and silicon–insulator interfaces, ψ_t has different values for $V_{ox} > 0$, which corresponds to electron tunneling from bulk to gate, and for $V_{ox} < 0$, corresponding to electron tunneling from gate to bulk. To provide a smooth transition around $V_{ox} = 0$, the expression

$$E_t = q\frac{1}{2}\left[\sqrt{(V_{ox}+G_0V_t)^2 + 0.01} - (V_{ox}+G_0V_t)\right] \quad (51)$$

is used, where G_0 is an additional parameter.

The total gate–channel current is found by integrating (46) along the channel. Instead of performing this integration numerically, an approximate closed-form expression for I_{gc} is derived by linearizing the supply function (50) and the tunneling probability exponent (49) around the point y_m in the channel where the surface potential is equal to

$$\phi_m = \frac{\phi_{sd} + \phi_{ss}}{2} \quad (52)$$

where ϕ_{ss} and ϕ_{sd} are the surface potentials in the drain and source sides of the channel.

The overlap currents are also modeled in SP. The position dependence is neglected, and the currents are given directly by

[3] In what follows the symbol ϕ_s will be used for the surface potential; in all of the previous discussions, the symbol ϕ was used for the potential energy; the context should clearly indicate the meaning of each symbol.

$$I_{\mathrm{GSOV}} = WL_{\mathrm{ov}}J_0F_{\mathrm{s}}(\phi_{\mathrm{sov}}) \times \exp\left[-\frac{4}{3}\frac{\sqrt{2m^*_{\mathrm{ox}}\phi_{\mathrm{b}}}}{q\hbar}t_{\mathrm{ox}}f\left(\frac{V_{\mathrm{gs}}-\phi_{\mathrm{sov}}}{\phi_{\mathrm{b}}}\right)\right] \quad (53)$$

and

$$I_{\mathrm{GDOV}} = WL_{\mathrm{ov}}J_0F_{\mathrm{s}}(\phi_{\mathrm{dov}}) \times \exp\left[-\frac{4}{3}\frac{\sqrt{2m^*_{\mathrm{ox}}\phi_{\mathrm{b}}}}{q\hbar}t_{\mathrm{ox}}f\left(\frac{V_{\mathrm{gd}}-\phi_{\mathrm{dov}}}{\phi_{\mathrm{b}}}\right)\right] \quad (54)$$

where V_{gs} and V_{gd} are the gate–source and drain–source voltages, L_{ov} is the overlap length, ϕ_{sov} and ϕ_{dov} are the surface potentials in the source and drain overlaps.

The gate–channel current partition is given by (32) and (33), and the integration is performed analytically by using the same linearization discussed above.

3.4. HiSIM

The Hiroshima University STARC IGFET Model (HiSIM) is a compact model based on the drift-diffusion approximation for the drain current, and it describes analytically all the device characteristics by the surface potential at the source and drain sides of the MOSFET's channel, ϕ_{S0} and ϕ_{SL}, respectively [61].

The gate current model in HiSIM, which considers only ECB tunneling and is based on band-to-band tunneling theory, is given by [61]

$$I_{\mathrm{G}} = q\cdot \mathrm{GLEAK1}\cdot W\cdot L\cdot\frac{F^2}{\sqrt{E_{\mathrm{g}}}}\exp\left(-\mathrm{GLEAK2}\frac{E_{\mathrm{g}}^{3/2}}{F}\right) \quad (55)$$

where

$$F = \frac{V'_{\mathrm{G}}-\phi_{\mathrm{s}}}{T_{\mathrm{ox}}} \quad (56)$$

$$\phi_{\mathrm{s}} = \frac{\phi_{\mathrm{S0}}-\phi_{\mathrm{SL}}}{\mathrm{GLEAK3}} \quad (57)$$

$$V'_{\mathrm{G}} = V_{\mathrm{gs}} - V_{\mathrm{fb}} + \Delta V_{\mathrm{th}} \quad (58)$$

and E_{g} is energy gap of the bulk semiconductor, ΔV_{th} is the threshold voltage shift with respect to the long-channel V_{t} due to short-channel effects and GLEAK1, GLEAK2 and GLEAK3 are model parameters.

The gate current partition in HiSIM is based in (32) and (33). The tunneling current density is assumed to vary linearly with the position along the channel, which allows for analytical expressions to be obtained for the drain and source components of the gate current in terms of ϕ_{S0} and ϕ_{SL} [62].

3.5. Comparison of gate current compact models

Table 4 summarizes some key differences between BSIM4, MOS Model 11, SP and HiSIM as far as tunneling current modeling is concerned.

4. Impact of gate tunneling current

4.1. Digital and analog circuit performance

The possibility to use MOS devices with gate oxide thickness as low as 1.5 nm was demonstrated already in 1994 [4]. Despite the high gate tunneling current density caused by such a thin oxide, for very short channel devices, an acceptable performance is obtained, since the gate current scaling is directly proportional to the channel length, while the drain current scales inversely with channel length.

It has also been argued that for very high performance applications, gate current densities of up to 1000 A/cm^2 (corresponding to oxide thicknesses in the range of 1–1.5 nm) can be tolerated, assuming that heat removal and reliability issues are not of concern [63]. Fig. 19 shows the ITRS 2004 prediction for the gate current and gate current limit for high-performance logic; beyond 2007, the required oxide thickness of 0.9 nm cannot meet the required power dissipation limits if gate dielectrics based on silicon dioxide and silicon nitride are used.

For applications such as memory, low-power and analog circuits, requirements of very low gate currents or the need for large-geometry devices might set the minimum gate oxide thickness at a higher level, in the range of 1.8–2.6 nm [63,64]. According to the 2004 ITRS, the power dissipation caused by gate current will reach its limit around 2006 for both low-operating power and low-standby power, with equivalent oxide thicknesses of 1.3 nm and 1.9 nm, respectively [2].

Some studies have been conducted on the influence of tunneling current in different circuit building blocks, and will be discussed next.

Probably the first circuit application that comes to mind when considering current leakage through the gate is the sample-and-hold cell, depicted in Fig. 20. For this circuit, when the clock signal goes low, corresponding to the "hold" phase, the voltage in the capacitor should remain

Table 4
Comparison of gate current compact models

	BSIM4	MOS Model 11	SP	HiSIM
Tunneling mechanisms	ECB, HVB, EVB	ECB, HVB	ECB	ECB
Considers quantization in acc./inv. layer for tunneling current calculation	No	Yes	No	No
Number of parameters for gate tunneling modeling	22	7	4	3
Overlap current modeling	Yes	Yes	Yes	No

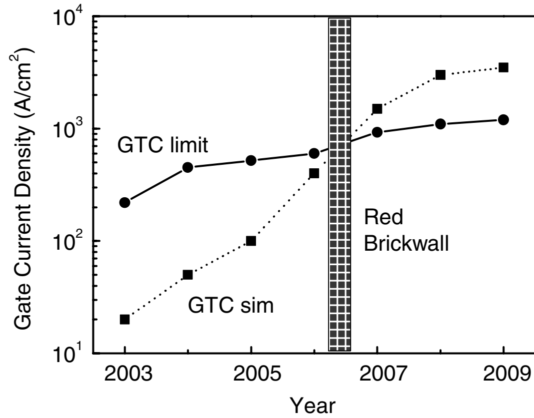

Fig. 19. ITRS gate leakage current limit for high-performance logic (HP) [2]. The gate current limit was calculated from power dissipation considerations.

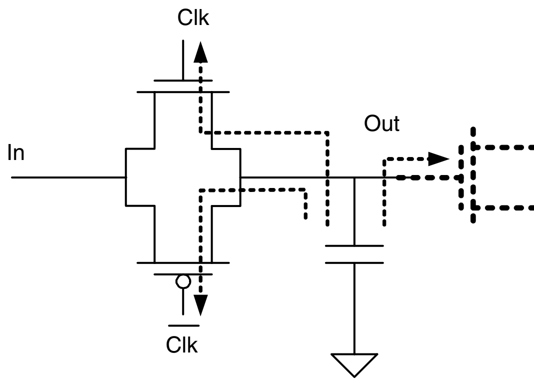

Fig. 20. CMOS sample and hold circuit, showing the possible leakage paths due to the gate current.

constant. In the presence of gate current, this is not the case, since an important amount of current flows through the source/drain overlap extensions of the sampling transistors [65] or through the gate of the transistor connected to the output to read out the signal [66]. If this circuit were to be used, for example, in an A/D converter, the inability to hold the charge for a certain time period implies that the sampling frequency has to be considerably increased [65].

Another circuit of special interest is the basic building block of CMOS logic: the CMOS inverter. Some of the key features of complementary MOS technology, such as the extremely low standby power consumption, the very high fan in/fan out and the good noise immunity might be affected by the gate current. The static component of power dissipation, which in thick-oxide technology is very low compared to the dynamic power dissipation, increases exponentially with the decrease in the oxide thickness, by as much as one order of magnitude per 0.3 nm of oxide thickness [64].

The voltage swing can also be affected by tunneling current. The gate current flowing in one stage, which is provided by the previous stage, causes a voltage drop in the channel of the devices along this current path, as illustrated in Fig. 21. This in turn implies that the logic low and high voltages start deviating appreciably from the supply source values, thus degrading the noise margin. Because of the higher tunneling current in nMOS devices compared to pMOS, the "high" logic level is more affected than the "low" level [67]. Simulation studies indicate that this effect might not be very significant for oxide thickness down to 1.1 nm [67]. However, the DC coupling illustrated in Fig. 21 complicates the estimation of power consumption and voltage swing levels in multiple-stage circuits, because each stage cannot be treated independently.

The switching performance of logic circuits can also be affected by the gate current, especially in dynamic logic circuits, where the circuit operation depends on the ability of the parasitic capacitors to hold charge for certain time. It has been shown from simulations that for very thin oxides, gate current can cause glitches or even erroneous logic levels in dynamic CMOS circuits [47].

Some results have been published regarding the influence of tunneling current on the maximum operating frequency [46,68], power and current gain [68,69], input impedance [46] and device matching [46,69], but a comprehensive study of the impact of tunneling current in the performance of MOS RF and analog circuits is still lacking.

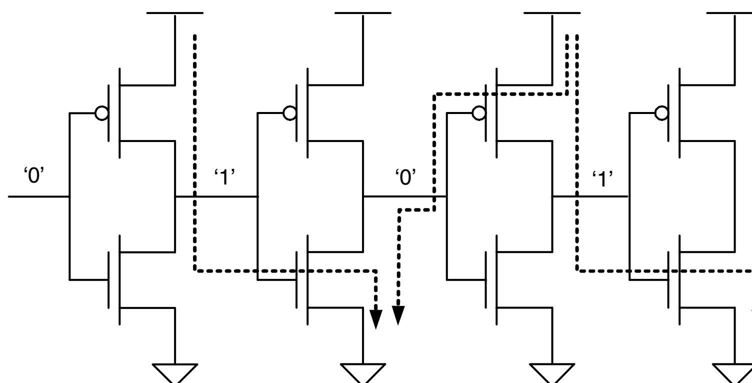

Fig. 21. CMOS inverter chain, showing the flow of DC current between stages caused by gate tunneling current [67].

The effect of gate current on the noise performance of MOS devices has received some more attention [70–74], and because of the importance of this topic it will be discussed in a separate section.

4.2. Noise

The presence of noise, the fluctuations in currents or voltages in electronic circuits, is a key issue in the design of digital, analog and RF applications. In logic circuits, it can cause bit errors, and in analog and RF systems it sets the lower limit of the dynamic range.

Since the characterization and modeling of the noise characteristics of MOS devices with large gate tunneling currents is still an active subject of research, this discussion will focus on the additional noise sources caused by gate current, and the physical mechanisms that have been proposed to explain them.

In MOS devices without gate current, several noise sources have been identified. At high frequencies, the dominant noise source is thermal noise in the channel [75], which in turn causes an induced gate noise, through the capacitive coupling between the channel and the gate. At low frequencies, noise in the drain current with a frequency spectrum proportional to $1/f^{\gamma}$ is observed, with γ typically between 0.7 and 1.3 [76]. Other noise sources include the thermal noise caused by parasitic resistances and avalanche noise for high drain–source voltages [72]. While $1/f$ noise is dominant at low frequencies, it is well known that it can affect the performance of high frequency circuits such as voltage-controlled-oscillators, due to frequency up-conversion [78].

Gate tunneling current, being generated by discrete carriers randomly crossing a barrier, is expected to have a shot noise component with a frequency-independent spectral density given by [79]

$$S_{i_g} = 2 \cdot q \cdot I_G \tag{59}$$

This white-like noise has been observed for ultra-thin oxide transistors at high frequencies, and Eq. (14) gives a good description of the dependence of this noise on gate current [71],[72].

In addition to this shot noise, low-frequency noise with an approximate $1/f$ dependence has also been reported for MOS capacitors [80] and in the gate current of MOS transistors [71,72].

Scholten et al. suggested that part of this low-frequency noise in the gate current is actually induced by the low-frequency noise in the drain current, since the tunneling current creates a DC coupling between the channel and the gate. However, this is not enough to account for all the low-frequency noise in the gate current [72].

The excess low-frequency noise in the gate current is often attributed to trap-related mechanisms, where charge traps in the oxide cause fluctuations in the gate current. However, there is no consensus on the detailed microscopic mechanism through which this trap-assisted tunneling causes the $1/f$ noise. Key in the determination of such mechanism is the dependence of low-frequency noise on the bias conditions. Alers et al. [80] found an exponential dependence of the low-frequency noise with the applied voltage, where the noise increases with decreasing gate voltage; this led them to propose that the low-frequency noise is caused by trap-assisted tunneling, where traps in the oxide, close to the emitting electrode, make tunneling a two-step process. This model was able to explain the dependence of LF noise on voltage observed by them. Other authors, however, [71], have not found such dependence on applied voltage, but a quadratic dependence on gate current instead, which is common to other $1/f$ noise mechanisms [77].

Lee and Bosman [73,74] proposed a model where the $1/f$ noise in the gate current is caused by the local fluctuations of the tunneling probability, caused in turn by the presence of fluctuating traps. This model also yields a quadratic dependence of LF noise on gate current; however, their experimental results for nitrided SiO_2 MOSFETs show a deviation from this trend at low currents. This was attributed to an enhancement of the shot noise at low currents (voltages), caused by generation-recombination processes in "fast" traps.

As for the implications of the enhancement of noise caused by gate current, it has been shown that, similar to the case of MESFETs with gate leakage current, the effect of gate current on the noise parameters of MOSFETs is more important at relatively low frequencies [70,74,81–83].

Gate current noise can also affect the observed drain current noise. At high frequencies, it has been shown that the gate shot noise results in a drain shot noise component which is fully correlated to the gate noise [82]. The same effect has also been observed at low frequencies: when the gate voltage is high enough to cause a significant gate current to flow, the low-frequency drain current noise increases beyond what can be explained by the conventional low-frequency noise theories, and this enhanced drain current noise is highly correlated to the gate current noise [71].

5. Conclusions

The basic physics of gate tunneling current modeling were presented. It was shown how the relatively simple Fowler–Nordheim model for tunneling fails to take into account several effects present in state-of-the-art MOSFET technology, such as direct tunneling or valence band tunneling. Also, the challenges for compact modeling of gate tunneling current with high accuracy and physical meaning were highlighted.

The modeling approach to gate tunneling used in several the industry-standard compact MOS models was presented, as well as some of their limitations such as numerous simplifications or non-physical modeling which result in a large number of parameters or inaccurate modeling.

The potential impact of tunneling current on circuit performance was discussed, both for logic and analog/RF circuits. Even though most applications are affected in one way or another, this should not create a barrier for the continued down-scaling of MOS devices, but will certainly create new challenges for device, circuit and system designers.

The additional noise due to the presence of tunneling current, and the physical mechanisms that have been proposed to explain it were presented at both high and low frequencies. The importance of noise in analog and RF circuits, calls for attention to the study of these issues, as well as to the development of compact models suitable for use in circuit simulators.

References

[1] Dennard RH, Gaensslen FH, Yu H-N, Rideout VL, Bassous E, Leblanc AR. Design of ion-implanted MOSFET's with very small physical dimensions. IEEE J Solid-State Circ 1974;SC-9(5):256–68.

[2] International technology roadmap for semiconductors, 2004 Update. Available from: <http://public.itrs.net>.

[3] Majkusiak B. Gate tunnel current in an MOS transistor. IEEE Trans Electron Dev 1990;37(4):1087–92.

[4] Momose HS, Ono M, Yoshitomi T, Ohguro T, Nakamura S, Saito M, et al. Tunneling gate oxide approach to ultra-high current drive in small geometry MOSFET's. IEDM Tech Dig 1994:593–6.

[5] Kish LB. End of Moore's law: thermal (noise) death of integration in micro and nano electronics. Phys Lett A 2002;305:144–9.

[6] Naseh S, Deen MJ, Marinov O. Effects of hot-carrier stress on the performance of the LC-tank CMOS oscillators. IEEE Trans Electron Dev 2003;50(5):1334–9.

[7] Iwai H. CMOS downsizing toward sub-10 nm. Solid-State Electron 2004;48:497–503.

[8] Lezlinger M, Snow EH. Fowler–Nordheim tunneling into thermally grown SiO_2. J Appl Phys 1969;40(1):278–83.

[9] Weinberg ZA. On tunneling in metal–oxide–silicon structures. J Appl Phys 1982;53(7):5052–6.

[10] Maserjian J. Tunneling in thin MOS structures. J Vac Sci Technol 1974;11(6):996–1003.

[11] Chang L, Yang KJ, Yeo Y-C, Polishchuk I, King T-J, Hu C. Direct-tunneling gate leakage current in double-gate and ultrathin body MOSFETs. IEEE Trans Electron Dev 2002;49(12):2288–95.

[12] Govoreanu B, Blomme P, Henson K, Van Houdt J, De Meyer K. An effective model for analyzing tunneling gate leakage currents through ultrathin oxides and high-*k* gate stacks from Si inversion layers. Solid-State Electron 2004;48:617–25.

[13] Vogel EM, Ahmed KZ, Hornung B, Henson WK, McLarty PK, Lucovsky G, et al. Modeled tunnel currents for high dielectric constant dielectrics. IEEE Trans Electron Dev 1998;45(6):1350–5.

[14] Mukhopadhyay S, Neau C, Cakici RT, Agarwal A, Kim CH, Roy K. Gate leakage reduction for scaled devices using transistor stacking. IEEE Trans VLSI Syst 2003;11(4):716–30.

[15] Marin M, Deen MJ, de Murcia M, Llinares P, Vildeuil J-C. A new method for the channel length extraction in MOSFETs with sub-2-nm gate oxide. IEEE Electron Dev Lett 2004;25(4):202–4.

[16] Clerc R, De Salvo B, Ghibaudo G, Reimbold G, Pananakakis G. Electrical characterization and modeling of MOS structures with an ultra-thin oxide. Solid-State Electron 2002;46(Mar):407–16.

[17] Ahmed K, Ibok E, Yeap GC-F, Xiang Q, Ogle B, Wortman JJ, et al. Impact of tunnel currents and channel resistance on the characterization of channel inversion layer charge and polysilicon-gate depletion of sub-20-Å gate oxide MOSFETs. IEEE Trans Electron Dev 1999;46(8):1650–5.

[18] Fowler RH, Nordheim L. Electron emission in intense electric fields. Proc Roy Soc A 1928;119(781):173–81.

[19] Schuegraf K, King CC, Hu C. Ultra-thin silicon dioxide leakage current and scaling limit. Dig 1992 Symp VLSI 1992(June):18–9.

[20] Harrison WA. Tunneling from an independent-particle point of view. Phys Rev 1961;123(1):85–9.

[21] Price PJ, Radcliffe JM. Esaki tunneling. IBM J Res Dev 1959;3:364–71.

[22] Stratton R. Volt–current characteristics for tunneling through insulating films. J Phys Chem Solids 1962;23:1177–90.

[23] Merzbacher E. Quantum mechanics. New York: John Wiley & Sons; 1962.

[24] Depas M, Vermeire B, Mertens PW, van Meirhaeghe RL, Heyns MM. Determination of tunneling parameters in ultra-thin oxide layer poly-Si/SiO_2/Si structures. Solid-State Electron 1995;38(8):1465–71.

[25] Maserjian J, Petersson GP. Tunneling through thin MOS structures: dependence on energy ($E - \kappa$). Appl Phys Lett 1974;25(1):50–2.

[26] Gundlach KH. Zur Berechnung Des Tunnelstroms Durch Eine Trapezförmige Potentialstufe. Solid-State Electron 1966;9:949–57.

[27] Abramowitz M, Stegun IA, editors. Handbook of mathematical functions, with formulas, graphs, and mathematical tables. New York: Dover Publications; 1965.

[28] Shih W-K, Wang EX, Jallepalli S, Leon F, Maziar CM, Taschjr AF. Modeling gate leakage current in nMOS structures due to tunneling through an ultra-thin oxide. Solid-State Electron 1998;42(6):997–1006.

[29] Cai J, Sah C-T. Gate tunneling currents in ultrathin oxide metal–oxide–silicon transistors. J Appl Phys 2001;89(4):2272–85.

[30] Schuegraf KF, Hu C. Hole injection SiO_2 breakdown model for very low voltage lifetime extrapolation. IEEE Trans Electron Dev 1994;41(5):761–7.

[31] Register LF, Rosenbaum E, Yang K. Analytic model for direct tunneling current in polycrystalline silicon-gate metal–oxide–semiconductor devices. Appl Phys Lett 1999;74(3).

[32] Majkusiak B, Jakubowski A. On electron tunneling in the metal–insulator–semiconductor systems including various electron effective masses. J Appl Phys 1985;58(8):3141–4.

[33] Depas M, Vermeire B, Meirhaeghe RL, Laflère WH, Cardon F. Electrical characteristics of Al/SiO_2/*n*-Si tunnel diodes with an oxide layer grown by rapid thermal oxidation. Solid-State Electron 1994;37(3):433–41.

[34] Lee W-C, Hu C. Modeling CMOS tunneling currents through ultrathin gate oxide due to conduction- and valence-band electron and hole tunneling. IEEE Trans Electron Dev 2001;48(7):1366–73.

[35] Gu X, Gildenblat G, Workmann GO, Veeraraghavan S, Shapira S, Stiles K. A surface potential-based compact model of n-MOSFET gate-tunneling current. IEEE Trans Electron Dev 2004;51(1):127–35.

[36] Ando T, Fowler AB, Stern F. Electronic properties of two-dimensional systems. Rev Mod Phys 1982;54(2):437–672.

[37] Janik T, Majkusiak B. Analysis of the MOS transistor based on the self-consistent solution to the Schrödinger and Poisson equations and on the local mobility model. IEEE Trans Electron Dev 1998;45(6):1263–71.

[38] Lo S-H, Buchanan DA, Taur Y, Wang W. Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET's. IEEE Electron Dev Lett 1997;18(5):209–11.

[39] Rana F, Tiwari S, Buchanan DA. Self-consistent modeling of accumulation layers and tunneling currents through very thin oxides. Appl Phys Lett 1996;69(8):1104–6.

[40] Chindalore G, Hareland SA, Jallepalli S, Tasch Jr AF, Maziar CM, Chia VKF, et al. Experimental determination of threshold voltage shifts due to quantum mechanical effects in MOS electron and hole inversion layers. IEEE Electron Dev Lett 1997;18(5):206–8.

[41] Stern F. Self-Consistent results for n-type Si inversion layers. Phys Rev B 1972;5(12):4891–9.

[42] Liu X, Kang J, Han R. Direct tunneling current model for MOS devices with ultra-thin gate oxide including quantization effect and polysilicon depletion effect. Solid State Commun 2003;125:219–23.

[43] Suñé J, Olivo P, Riccó B. Self-consistent solution of the Poisson and Schrödinger equations in accumulated semiconductor–insulator interfaces. J Appl Phys 1991;70(1):337–45.

[44] Yang N, Henson WK, Hauser JR, Wortman JJ. Modeling study of ultrathin gate oxides using direct tunneling current and capacitance–voltage measurements in MOS devices. IEEE Trans Electron Dev 1999;46(7):1464–71.

[45] Dalla Serra A, Abramo A, Palestri P, Selmi L, Widdershoven F. Closed- and open-boundary models for gate-current calculation in n-MOSFETs. IEEE Trans Electron Dev 2001;48(8):1811–5.

[46] van Langevelde R, Scholten AJ, Duffy R, Cubaynes FN, Knitel MJ, Klaasen DBM. Gate current: modeling, ΔL extraction and impact on RF performance. IEDM Tech Dig 2001:289–92.

[47] Choi C-H, Nam K-Y, Yu Z, Dutton RW. Impact of gate direct tunneling current on circuit performance: a simulation study. IEEE Trans Electron Dev 2001;48(12):2823–9.

[48] Yang KN, Huang HT, Chen MJ, Lin YM, Yu MC, Jang SM, et al. Characterization and modeling of edge direct tunneling (EDT) leakage in ultrathin gate oxide MOSFETs. IEEE Trans Electron Dev 2001;48(6):1159–64.

[49] Shi Y, Ma TP, Prasad S, Dhanda S. Polarity dependent gate tunneling currents in dual-gate CMOSFET's. IEEE Trans Electron Dev 1998;45(11):2355–60.

[50] Shih WK, Rios R, Packan P, Mistry K, Abbott T. A general partition scheme for gate leakage current suitable for MOSFET compact models. IEDM Tech Dig 2001:293–6.

[51] Cao KM, Lee W-C, Liu W, Jin X, Su P, Fung SKH, et al. BSIM4 gate leakage model including source–drain partition. IEDM Tech Dig 2000:815–8.

[52] Spinelli AS, Pacelli A, Lacaita AL. The impact of polysilicon quantization on ultra-thin oxide MOSFET characteristics. First European Workshop on Ultimate Integration of Silicon, January 2000.

[53] Schenk A, Heiser G. Modeling and simulation of tunneling through ultra-thin gate dielectrics. J Appl Phys 1997;81(12):7900–8.

[54] Goldman EI, Kukharskaya NF, Zhdan AG. The effect of imaging forces in ultra thin gate insulator on the tunneling current and its oscillations at the region of transition from the direct tunneling to the Fowler–Nordheim tunneling. Solid-State Electron 2004;48(5):831–6.

[55] BSIM 4.4.0 MOSFET Model User's Manual, Department of Electrical Engineering and Computer Sciences, University of California, March 2004. Available from: <http://www-device.eecs.berkeley.edu/~bsim3/bsim4.html>.

[56] The MOS model, level 1101, Philips Semiconductors. Available from: <http://www.semiconductors.philips.com/Philips_Models/mos_models/model11/index.html>.

[57] The MOS model, level 903, Philips Semiconductors. Available from: <http://www.semiconductors.philips.com/Philips_Models/mos_models/model9/index.html>.

[58] Enz C, Krummenacher F, Vittoz E. An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications. Analog Integr Circ Signal Process 1995:83–114.

[59] Krummenacher F. private communication.

[60] Gildenblat G, Wang H, Chen T-L, Gu X, Cai X. SP: an advanced surface-potential-based compact MOSFET model. IEEE J Solid-State Circ 2004;39(9):1394–406.

[61] HiSIM1.1.1 User's Manual, Semiconductor Technology Academic Research Center (STARC). Available from: <http://www.starc.or.jp/kaihatu/pdgr/hisim/hisim.html>.

[62] Ngo Q, Navarro D, Mizoguchi T, Hosakawa S, Ueno H, Miura-Mattausch M, et al. Gate current partitioning in MOSFET models for circuit simulation. Tech Proc 2003 Nanotechnology Conference 2003;2(Feb):322–5.

[63] Frank DJ, Dennard RH, Nowak E, Solomon PM, Taur Y, Philip Wong H-S. Device scaling limits of Si MOSFETs and their application dependencies. Proc IEEE 2001;89(3):259–88.

[64] Henson WK, Yang N, Kubicek S, Vogel EM, Wortman JJ, De Meyer K, et al. Analysis of leakage currents and impact on off-state power consumption for CMOS technology in the 100-nm regime. IEEE Trans Electron Dev 2000;47(7):1393–400.

[65] Bhat N, Chandrabhan C. Analog CMOS performance degradation due to edge direct tunneling (EDT) current in sub-100 nm technology. J Semicond Tech Sci 2003;3(3):139–44.

[66] Chatterjee A, Mosher D, Sridhar S, Nandakumar M, Aur S, Chen Z, et al. Analog integration in 0.35 μm Cu metal pitch, 0.1 μm gate length, low-power digital CMOS technology. IEDM Tech Dig 2001:211–4.

[67] Marras A, De Munari I, Vescovi D, Ciampolini P. Performance evaluation of ultra-thin gate-oxide CMOS circuits. Solid-State Electron 2004;48:551–9.

[68] Momose HS, Morifuji E, Yoshitomi T, Ohguro T, Saito M, Iwai H. Cutoff frequency and propagation delay time of 1.5-nm gate oxide CMOS. IEEE Trans Electron Dev 2001;48(6):1165–74.

[69] Annema AJ, Nauta B, van Langevelde R, Tuinhout H. Analog circuits in ultra-deep-submicron CMOS. IEEE J Solid-State Circ 2005;40(1):132–43.

[70] Deen MJ, Ranuárez JC, Chen C-H. Effect of the gate tunneling current on the high-frequency noise of MOSFETs. In: Workshop on compact modeling, Anaheim, CA, 8–12 May 2005. p. 35–9.

[71] Valenza M, Hoffmann A, Sodini D, Laigle A, Martinez F, Rigaud D. Overview of the impact of downscaling technology on $1/f$ noise in p-MOSFETs to 90 nm. IEE Proc G—Circ Dev Syst 2004;151(2):102–10.

[72] Scholten AJ, Tiemeijer LF, van Langevelde R, Havens RJ, Zegers-van Duijnhoven ATA, Venezia VC. Noise modeling for RF CMOS circuit simulation. IEEE Trans Electron Dev 2003;50(3):618–32.

[73] Lee J, Bosman G, Green KR, Ladwig D. Noise model of gate-leakage current in ultrathin oxide MOSFETs. IEEE Trans Electron Dev 2003;50(12):2499–506.

[74] Lee J, Bosman G. Comprehensive noise performance of ultrathin oxide MOSFETs at low frequencies. Solid-State Electron 2004;48:61–71.

[75] Chen CH, Deen MJ. Channel noise modeling of deep sub-micron MOSFETs. IEEE Trans Electron Dev 2002;49(8):1484–7.

[76] Simoen E, Claeys C. On the flicker noise in submicron silicon MOSFETs'. Solid-State Electron 1999;43:865–82.

[77] Deen MJ, Pascal F. Review of low-frequency noise behaviour of polysilicon emitter bipolar junction transistors. IEE Proc G—Circ Dev Syst 2004;151(2):125–37.

[78] Hajimiri A, Lee TH. A general theory of phase noise in electrical oscillators. IEEE J Solid-State Circ 1998;33(2):179–94.

[79] van der Ziel A. Noise in solid state devices and circuits. New York: Wiley; 1986.

[80] Alers GB, Krisch KS, Monroe D, Weir BE, Chang AM. Tunneling current noise in thin gate oxides. Appl Phys Lett 1996;69(19):2885–7.

[81] Fiegna C. Analysis of gate shot noise in MOSFETs with ultrathin gate oxides. IEEE Electron Dev Lett 2003;24(2):108–10.

[82] Paillopncy G, Iñiguez B, Dambrine G, Danneville F. Influence of a tunneling gate current on the noise performance of SOI MOSFETs. In: Proceedings of the 2004 IEEE international SOI conference, October 2004. p. 55–7.

[83] Vickes H-O, Ferndahl M, Masud A, Zirath H. The influence of the gate leakage current and the gate resistance on the noise and gain performances of 90-nm CMOS for micro- and millimeter-wave frequencies. IEEE MTT-S Int Microwave Symp Dig 2004;2(June):969–72.