

Quantum effects in the simulation of conventional devices

A. Abramo^a, C. Fiegna^b, P. Casarini^c

^aDIEGM, University of Udine
via delle Scienze 208, I-33100 Udine, Italy

^bDept. of Engineering, University of Ferrara
via Saragat 1, I-44100 Ferrara, Italy

^cINFN, University of Modena
via Campi 213/A, I-41100 Modena, Italy

Abstract

It is widely known that a fundamental role in the evolution of modern solid-state devices is played by scaling theories. The constant increase of the circuit complexity, the reduction of their dimensions and power consumption, in fact, is made possible mainly due to device shrinking. Of course, this progress wouldn't have happened without the parallel evolution of semiconductor technologies, which, in turn, probably wouldn't have progressed this much if the performance limits of MOS transistors had been reached sooner. Therefore, it is important to understand and try to predict these limits, possibly to avoid them circumventing their origin, ultimately to delay as much as possible the need of a different technology. To this purpose, from the theoretical side it is important to identify the correct physical frame in which investigations have to be performed, with the aim of bridging the gap between experiments and models, and, in essence, to be confident on the prediction ability of the simulation tools. In this paper we focus our attention on the modeling of quantum effects in MOS transistors, presenting some recent applications concerning quantum effects in MOSFETs.

1. Introduction

Scaling strategies fix strict rules to properly reduce the device dimensions with an overall increase in device performance [1, 2]. In particular, high doping concentrations ($\approx 10^{18}\text{cm}^{-3}$) and thin gate oxides ($\approx 3\text{nm}$) are required to limit 2D short-channel effects (SCE) in ultra-short (US-) MOSFETs with gate length (L_G) below $0.1\ \mu\text{m}$ [3].

This results in a sensible quantization at the Si-SiO₂ interface of the MOS system, whose main consequence is a relevant reduction of low-field mobility with respect to bulk silicon [4]. In particular, higher channel doping concentrations produce stronger energy quantization that, in turn, causes quantum induced threshold voltage shift (Q-DV) [5, 6]. In addition, the displacement of the charge distribution from the interface causes a gate capacitance reduction with respect to $C_{OX} = \epsilon_{OX}/t_{OX}$ [7, 8], that becomes more relevant as t_{OX} is scaled down and the channel doping is decreased.

This experimental evidence is sufficient to state the limitation of classical models, where the charge density peaks at the Si-SiO₂ interface, and suggests that quantum physics shall be in some way included inside charge transport models to describe these effects on a quantitative basis.

2. The modeling framework

A possible way, as usual, is to approach the problem in the most fundamental way, i.e. solving the whole transport on a quantum mechanical basis. Several treatment of the quantum transport problem have been proposed in the literature, based on different formalisms, such as density matrix [9], Green's functions [10, 11], and Wigner function [12, 13]. In our opinion, however, the fully-quantum analysis of carrier transport is still computationally too intensive to be used for the simulation of

real devices. For this reason, *decoupled* approaches have been proposed, to include quantum features inside the standard tools for devices transport simulation. In practice, this often means to solve some form of the steady-state Schrödinger equation to find the eigenvalues of the problem, and to spend this information inside a traditional simulator. This has been proposed in different ways [14, 15, 16, 6, 17, 18, 19].

However, since most of the methods have to deal with time-independent Schrödinger equation, whose solution can be a non trivial task especially in 2D, a small, thus limited, overview of some possible approaches to its solution may be of some help.

In principle, the time-independent Schrödinger equation describes the coherent motion of an electron subject to all internal potentials: that induced by the crystal to whom the electron belongs (U_C), that due to the effect of external forces on the internal potential (U_E), and that induced by perturbations (i.e. scattering processes) of different nature (U_S):

$$-\frac{\hbar^2}{2m}\nabla^2\Psi(\mathbf{r}) + [U_E(\mathbf{r}) + U_C(\mathbf{r}) + U_S(\mathbf{r})]\Psi(\mathbf{r}) = \epsilon\Psi(\mathbf{r}) \quad (1)$$

where no time dependence has been assumed also in the scattering potential. Since for simplicity we limit ourselves to the case of the coherent motion of particles, the scattering potential is assumed to be zero.

Usually, the effect of the crystal goes inside the so called *band structure*, computed at different level of approximation. This results in a new equation with the same form as the original one:

$$\epsilon(-i\nabla)\Psi(\mathbf{r}) + U_E(\mathbf{r})\Psi(\mathbf{r}) = \epsilon\Psi(\mathbf{r}) \quad (2)$$

where the rapidly varying internal potential $U_C(\mathbf{r})$ has formally disappeared but its contribution has been actually taken into account by the differential operator $\epsilon(-i\nabla)$, where this notation means that the $\epsilon(k_x, k_y, k_z)$ function (i.e. the band structure) reads $\epsilon(-i\frac{\partial}{\partial x}, -i\frac{\partial}{\partial y}, -i\frac{\partial}{\partial z})$ after the usual substitution of quantum mechanics. This equation is called *envelope function equation*, meaning that its eigenstates, i.e. its solutions, don't oscillate on the atomic scale of the crystal, but describe a sort of envelope behavior of the electrons, as if they were subject to the effect of external forces only, which is a much more slowly varying potential inside the crystal. Although in this form, the equation is still not easy to solve, due to the complexity of the differential operator. However, in principle it can be solved numerically in the momentum representation [14, 20].

To further simplify the description, isotropic parabolic mass approximation is often assumed. In this case the differential operator goes back to its original form, but the free electron mass is replaced with the *effective mass*, representing the electron around the energy minimum of the crystal band structure:

$$-\frac{\hbar^2}{2m^*}\nabla^2\Psi(\mathbf{r}) + U_E(\mathbf{r})\Psi(\mathbf{r}) = \epsilon\Psi(\mathbf{r}) \quad (3)$$

This approximation is widely used and can be considered reasonably accurate when dealing with semiconductors with low anisotropy, as it is the case of conduction band electrons in silicon, and/or when a rather low energy regime can be assumed. Corrections to account for the non-parabolicity of the bands can be easily included through first order perturbation theory [14, 21], extending the validity of the model to slightly higher energies.

From the numerical point of view, the *natural* way to solve the Schrödinger equation is to discretize it in the \mathbf{r} representation, i.e. to replace the differential operator with its finite differences equivalent:

$$\nabla^2 f_{ij} \rightarrow \frac{f_{i-1} - 2f_i + f_{i+1}}{(\Delta x)^2} + \frac{f_{j-1} - 2f_j + f_{j+1}}{(\Delta y)^2} \quad (4)$$

where (i, j) are the index of a generic node of the 2D space grid, to substitute the potential $U_E(\mathbf{r})$ with its discrete (say piecewise linear) approximation U_E^{ij} and solving the resulting system of linear equations. In the 1D case and for uniform real space discretization, many solutions of a tridiagonal system of equations must be performed, the second derivative connecting each grid point to its two nearest neighbors. In the 2D case, instead, the node connectivity is higher, thus the resulting matrix is no longer tridiagonal. Given the number of grid nodes needed to describe in some detail a MOS structure, the matrix can be of some size if stored in a dense form. However, since the node connectivity is usually limited, the use of the sparse representation is certainly of some advantage.

Similar results can be obtained in the frame of the finite elements formalism [22, 23]. In this case a variational approach is used, where the energy related with the linear functional associated with the

differential operator gets minimized in presence of a family of trial (for example piecewise linear) solutions $\bar{\Psi}$:

$$\langle \bar{\Psi} | \mathcal{H} | \Psi \rangle = \epsilon \langle \bar{\Psi} | \Psi \rangle \quad (5)$$

The method gives rise to a linear system of equations equivalent to that obtained through finite differences, a part from the more sound mathematical basis of the finite element approach, whose level of accuracy can also be increased by using higher order *shape functions* to better approximate the general solution [22, 23].

All methods to solve the Schrödinger equation in its \mathbf{r} representation, unfortunately, suffer for a serious intrinsic limitation: since wave functions related to subsequent eigenvalues increase by one the number of their space oscillations, the number of grid points needed to describe such oscillations at the energies typical of MOS operation are too many to be handled with standard computing resources. This is particularly true in 2D systems, where the number of points follows a quadratic law of increase. For example, let us take a thermal electron inside the source (for simplicity assuming $E = 0\text{eV}$ there) traveling ballistically towards the drain of a $L = 0.1\mu\text{m}$ MOS biased at $V_{DS} = 1.5\text{V}$. From the expression of the electron energy (in the parabolic mass approximation) $E = \hbar^2 k_x^2 / (2m_x^*)$ it can be easily derived that, assuming $m_x^* = 0.32m_0$, the number of oscillations the electron wave function performs from source to drain is $(k_x \cdot L) \approx 355$, which requires a number of points in the longitudinal direction twice as bigger, thus a too big number of grid points to describe the MOS in 2D.

To overcome the problem, the Schrödinger equation can be solved in the \mathbf{k} representation. A proper basis of the Hilbert space is chosen, and the eigenfunctions are determined in terms of their spectral components. For example, in the case of open boundary systems the general solution of the Schrödinger equation can be expressed as the weighted superposition of planes waves:

$$\Psi(\mathbf{r}) = \sum_{n=1}^N A_n e^{i\mathbf{k}_n \cdot \mathbf{r}} \quad (6)$$

while in the case of closed systems we can write:

$$\Psi(\mathbf{r}) = \sum_{m=1}^{N_x} \sum_{n=1}^{N_y} A_{mn} \sin\left(\frac{m\pi}{L_x}\right) \sin\left(\frac{n\pi}{L_y}\right) \quad (7)$$

In both cases the A 's are the unknown to be determined. The limit of the method stands in the number of \mathbf{k} components to be included to accurately describe the solution in presence of rapidly varying potential profiles, such as in the case of tunneling through the discontinuous Si-SiO₂-Si barrier. From the numerical point of view, this method requires to perform N or $N_x \cdot N_y$ Fast Fourier Transforms in the 1D or 2D case, respectively, which may result in a sensible computational work depending on the number of eigenstates that are needed for the proper description of the physical phenomena. In addition, no advantage is obtained from the sparse matrix storage, since the evaluation of the matrix elements $U_{kk'} = \langle \Psi | U_E | \Psi \rangle \forall k, k'$ is required, resulting in a dense matrix.

All the previous methods assume that no simplifications are made on the potential profile. A last solution method, again in real space, relies instead on a simplification than can be applied to the potential profile. In general, the case of 2D open boundary systems cannot be solved by separation of variables, due to the non separable nature both of the potential profile and of the boundary conditions. However, the system can be discretized into strips, and in each of them, say the n^{th} , the potential can be considered constant along one direction, say y :

$$U_E(x, y) = U_0^n + U^n(x) \quad (8)$$

The approximation, in essence, substitutes the y dependence of the potential with a piecewise constant form, which can be acceptable if the number of strips is sufficiently high where the variation of the potential in the y direction is large. Under this assumption, the general solution inside the l^{th} strip can be written as:

$$\Psi^l(x, y^l) = \sum_{n=1}^N \xi_n^l(x) \left[A_n^l e^{ik_y^l y^l} + D_n^l e^{-ik_y^l y^l} \right] \quad (9)$$

where N is the number of bound states ξ_n^l in the x direction inside the l^{th} strip, k_y^l is the y wave vector component inside the l^{th} strip that relates to the total energy of the eigenstate through the

relation $E - U_0^n = E_x + \frac{\hbar^2(k'_y)^2}{2m_y^*}$, m_x^* and m_y^* the effective masses in the two directions, and where we have assumed open boundary conditions injecting in the y direction only. Then, the global eigenfunctions, i.e. the A_n^l and B_n^l complex constants, can be obtained imposing the y continuity of the $\Psi^l(x, y^l)$ and of their first derivative in N points along the x boundary of each strip, and imposing the global boundary conditions.

3. Applications to conventional MOSFET devices

So far the more relevant applications concerning the quantum simulation of MOS devices have consisted in 1D studies. In fact, under the assumption of negligible electric field in the longitudinal direction, translational symmetry can be assumed along the directions parallel to the Si-SiO₂ interface, electrons being quantized in the normal direction only. Under these assumption, separation of variables leads to a 1D Schrödinger equation [24, 25]. As application of this methodology, we simulated one-dimensional MOS capacitors representative of both bulk and SOI technologies.

The issues of Q-DV, low-field electron effective mobility, and gate-to-channel capacitance have been investigated by a one-dimensional approach. Therefore, the results obtained are valid for low applied drain-to-source voltages, i.e. in the linear MOSFET regime, or, more in general, when two-dimensional effects can be assumed to be small. Despite to this limitation, the results of this approach are anyway valid as a comparison between the different doping solutions analyzed. In fact, since the concept of *short MOSFET* (i.e. suffering of relevant 2D short channel effects) is relative to the adopted technology, even a 0.1 micron MOSFET fabricated with a 50 nm technology and operating at low voltages [3] can be considered as a *long channel* device. Therefore, for such a device 1D results are relevant.

The simulations were carried out by self-consistently solving the Poisson and Schrödinger equations in a Gummel-like iterative procedure [26]. The envelope function equation is solved to determine the eigenvalues and eigenvectors of the system. In this solution, the six-fold ellipsoidal symmetry is assumed for the silicon, with the usual values for longitudinal and transverse masses (0.19 and 0.915 in free electron mass units, respectively), together with a parabolic energy vs. k-vector dispersion relationship, adequate for the case of n-channel MOS structures [21]. A zero wave function boundary condition is forced at the quantum system boundaries, i.e. in the substrate and at the oxide interface. Therefore, wave function penetration into the oxide is neglected.

4. Simulated devices and results

We considered bulk MOS structures featuring an ideal doping profile, as could be obtained by epitaxial growth. A low-doped EPI (with doping concentration N_{EPI} and thickness t_{EPI}) covers a highly doped region (doping concentration N_A) acting as a background plane to suppress punch-through. The purpose of the surface layer is to improve low-field mobility by reducing the transverse electric field (TEF) (i.e. surface roughness scattering) and the effect of ionized impurities. Since we are investigating a possible structure for MOS devices with $L_G \leq 0.1 \mu\text{m}$, we assumed very thin oxide (typically $t_{OX}=3\text{ nm}$) and epitaxial-layer thicknesses (t_{SI} as thin as 10 nm).

SOI devices are also investigated since they are proposed as an alternative to bulk MOSFETs for the fabrication of devices with gate length around $0.1\mu\text{m}$ and below. In particular, the double-gate MOS structure (DGM) has been proposed due to its superior control of short channel effects, the lower sub-threshold leakage, and the higher currents compared to a single-gate SOI MOS (SGM) of same area [27, 28, 29]. In our simulations we considered very thin silicon film MOS SOI devices (t_{SI} down to 2.5 nm) in order to study the impact of size quantization on the main device characteristics.

Fig. 1 reports the diagram of the conduction band across the MOS structure and the first 20 quantized energy levels self-consistently computed inside a uniformly doped MOSFET with doping concentration $N_A=1 \times 10^{17} \text{ cm}^{-3}$ and inversion electron sheet concentration $N_{INV}=6 \times 10^{12} \text{ cm}^{-2}$. Comparison with a single-gate SOI device, featuring $t_{SI}=10 \text{ nm}$ and biased for the same inversion layer concentration, shows that in the thin SOI case relevant size quantization occurs, as the energy levels are more separated from each other. As a consequence, some differences between the two structures may be expected in terms of threshold voltage and carrier mobility.

The impact of the coupling between Schrödinger and Poisson equations in the presence of very strong quantization is shown in Fig. 2. When very thin double-gate structures are considered, the quantum charge distribution has a very strong impact on the electrostatic potential distribution, which is very different from the one computed classically.

In the following subsections we will discuss the effect of such evidence on threshold voltage, effective mobility and gate capacitance results.

4.1. Threshold voltage results

We have simulated the EPI structure with t_{EPI} down to 10 nm and t_{OX} down to 3 nm, that was investigated in [30] for application to MOSFETs with L_G down to 50 nm. Its threshold voltage (V_{TH}), computed as the linear extrapolation of the simulated quantum inversion charge Q_S vs. voltage characteristic, has been studied. In particular, we focused on the threshold voltage shift with respect to the classical solution (i.e. the Q-DV) for the EPI structure and for the more conventional highly doped (UHD) structure obtained setting $t_{EPI}=0$. Due to the low surface doping, a much lower Q-DV could be expected in the EPI case compared to UHD. Instead, similar Q-DV is obtained for UHD and EPI with small t_{EPI} . In fact, for $t_{EPI}=10$ nm and t_{OX} ranging from 3 to 10 nm, we obtained only a 20% reduction of Q-DV with respect to UHD, showing that this effect does not simply depend on the doping at the interface, but rather, a long-range dependence on doping holds. This point is emphasized by Fig. 3-a, where Q-DV is reported as a function of t_{EPI} : a relatively thick EPI is needed to eliminate the quantum effects originated by the presence of the high-doped background plane. Fig. 3-b, instead, shows how V_{TH} can be controlled by properly selecting t_{EPI} : while the UHD MOS presents a threshold voltage that is too large for low supply voltage applications ($V_{DD} \approx 1-1.5$ V) the introduction of the low-doped layer provides a further degree of freedom, additional to the doping of the background-plane, for designing the threshold voltage according to circuit applications. The reported results also clearly show that simple Q-DV models based on average channel doping values cannot be applied to the highly non-uniform cases needed for US-MOSFETs.

The case of single- and double-gate SOI structures is reported in Fig. 4, where the threshold voltage is reported as a function of silicon thickness for single- and double-gate SOI structures featuring uniform doping concentration (10^{17} cm^{-3}) and gate oxide thickness $t_{OX}=3$ nm. By comparing the threshold voltages calculated classically and quantum-mechanically for the SGM case, we notice that for large t_{SI} the threshold voltage is coincident with that of the bulk MOSFET for the same doping and that quantum calculations provide larger threshold compared to the results of classical simulations, as in the usual bulk MOSFET case [5]. As t_{SI} is reduced to reach the fully depleted case, V_{TH} decreases due to the reduction of fixed charge. In the quantum-mechanical case, the threshold voltage starts to increase below a critical thickness (≈ 10 nm) due to the reduction of the energy density of states that is a consequence of increasing size quantization (see also Fig. 1). A similar qualitative picture is found in the DGM case, where a shift of the V_{TH} characteristic towards larger thicknesses occurs due to the presence of the two gates, both contributing to the silicon film depletion.

4.2. Effective mobility results

We can address channel electron effective mobility starting from Fig. 5, showing the profiles of the quantum electron density in structures that feature $N_A=1 \times 10^{18} \text{ cm}^{-3}$, $N_{EPI}=1 \times 10^{15} \text{ cm}^{-3}$, $t_{OX}=3$ nm, and different t_{EPI} . The comparison has been performed for a given inversion sheet density above threshold ($N_S=3 \times 10^{12} \text{ cm}^{-2}$). For increasing t_{EPI} , charge confinement reduces due to the reduction of transverse electric field.

In fact, the introduction of low-doped EPI effectively decreases the effective electric field (EEF), reported in Fig. 6 as a function of t_{EPI} for a given above-threshold inversion sheet density. Here, the EEF has been computed as the mean value of the electric field component normal to the interface averaged over the carrier density. Consequently, no surprise if improved effective mobility (μ_{eff}) is obtained for larger t_{EPI} extracting the values from experimental mobility vs. effective field curves [4], as it is again shown in Fig. 6.

The μ_{eff} calculated in the relaxation time approximation including the effects of inter- and intra-sub band phonon scattering and surface roughness, and consistently with the sub bands and wavefunctions computed solving the Schrödinger and Poisson equations, are given in Fig. 7, confirming that a decrease of effective field due to the reduction of the depletion charge produces an improved channel mobility in the epitaxial MOSFETs. It should be mentioned that UHD devices are expected to present even larger disadvantages for low inversion charge densities [4], due to the reduced screening of the Coulomb doping scatterer.

Similar calculations have been performed for very thin single- and double-gate SOI MOSFETs. Fig. 8 reports the results obtained for SGM structures with uniform doping concentration of 10^{15} cm^{-3} . The phonon limited mobility is a non-trivial function of the silicon thickness, since two competing mechanisms coexist: a) the reduction of the silicon thickness increases the quantization and decreases the available density of states, leading to a mobility enhancement; b) for decreasing t_{SI} , instead, a stronger surface confinement of the wave functions occurs, that leads to an increase of the overlap factors and, therefore, of the scattering rates and, consequently, to a decrease of the mobility.

relation $E - U_0^n = E_x + \frac{\hbar^2(k_y^l)^2}{2m_y^*}$, m_x^* and m_y^* the effective masses in the two directions, and where we have assumed open boundary conditions injecting in the y direction only. Then, the global eigenfunctions, i.e. the A_n^l and B_n^l complex constants, can be obtained imposing the y continuity of the $\Psi^l(x, y^l)$ and of their first derivative in N points along the x boundary of each strip, and imposing the global boundary conditions.

3. Applications to conventional MOSFET devices

So far the more relevant applications concerning the quantum simulation of MOS devices have consisted in 1D studies. In fact, under the assumption of negligible electric field in the longitudinal direction, translational symmetry can be assumed along the directions parallel to the Si-SiO₂ interface, electrons being quantized in the normal direction only. Under these assumption, separation of variables leads to a 1D Schrödinger equation [24, 25]. As application of this methodology, we simulated one-dimensional MOS capacitors representative of both bulk and SOI technologies.

The issues of Q-DV, low-field electron effective mobility, and gate-to-channel capacitance have been investigated by a one-dimensional approach. Therefore, the results obtained are valid for low applied drain-to-source voltages, i.e. in the linear MOSFET regime, or, more in general, when two-dimensional effects can be assumed to be small. Despite to this limitation, the results of this approach are anyway valid as a comparison between the different doping solutions analyzed. In fact, since the concept of *short MOSFET* (i.e. suffering of relevant 2D short channel effects) is relative to the adopted technology, even a 0.1 micron MOSFET fabricated with a 50 nm technology and operating at low voltages [3] can be considered as a *long channel* device. Therefore, for such a device 1D results are relevant.

The simulations were carried out by self-consistently solving the Poisson and Schrödinger equations in a Gummel-like iterative procedure [26]. The envelope function equation is solved to determine the eigenvalues and eigenvectors of the system. In this solution, the six-fold ellipsoidal symmetry is assumed for the silicon, with the usual values for longitudinal and transverse masses (0.19 and 0.915 in free electron mass units, respectively), together with a parabolic energy vs. k-vector dispersion relationship, adequate for the case of n-channel MOS structures [21]. A zero wave function boundary condition is forced at the quantum system boundaries, i.e. in the substrate and at the oxide interface. Therefore, wave function penetration into the oxide is neglected.

4. Simulated devices and results

We considered bulk MOS structures featuring an ideal doping profile, as could be obtained by epitaxial growth. A low-doped EPI (with doping concentration N_{EPI} and thickness t_{EPI}) covers a highly doped region (doping concentration N_A) acting as a background plane to suppress punch-through. The purpose of the surface layer is to improve low-field mobility by reducing the transverse electric field (TEF) (i.e. surface roughness scattering) and the effect of ionized impurities. Since we are investigating a possible structure for MOS devices with $L_G \leq 0.1 \mu\text{m}$, we assumed very thin oxide (typically $t_{OX}=3 \text{ nm}$) and epitaxial-layer thicknesses (t_{SI} as thin as 10 nm).

SOI devices are also investigated since they are proposed as an alternative to bulk MOSFETs for the fabrication of devices with gate length around $0.1 \mu\text{m}$ and below. In particular, the double-gate MOS structure (DGM) has been proposed due to its superior control of short channel effects, the lower sub-threshold leakage, and the higher currents compared to a single-gate SOI MOS (SGM) of same area [27, 28, 29]. In our simulations we considered very thin silicon film MOS SOI devices (t_{SI} down to 2.5 nm) in order to study the impact of size quantization on the main device characteristics.

Fig. 1 reports the diagram of the conduction band across the MOS structure and the first 20 quantized energy levels self-consistently computed inside a uniformly doped MOSFET with doping concentration $N_A=1 \times 10^{17} \text{ cm}^{-3}$ and inversion electron sheet concentration $N_{INV}=6 \times 10^{12} \text{ cm}^{-2}$. Comparison with a single-gate SOI device, featuring $t_{SI}=10 \text{ nm}$ and biased for the same inversion layer concentration, shows that in the thin SOI case relevant size quantization occurs, as the energy levels are more separated from each other. As a consequence, some differences between the two structures may be expected in terms of threshold voltage and carrier mobility.

The impact of the coupling between Schrödinger and Poisson equations in the presence of very strong quantization is shown in Fig. 2. When very thin double-gate structures are considered, the quantum charge distribution has a very strong impact on the electrostatic potential distribution, which is very different from the one computed classically.

In the following subsections we will discuss the effect of such evidence on threshold voltage, effective mobility and gate capacitance results.

4.1. Threshold voltage results

We have simulated the EPI structure with t_{EPI} down to 10 nm and t_{OX} down to 3 nm, that was investigated in [30] for application to MOSFETs with L_G down to 50 nm. Its threshold voltage (V_{TH}), computed as the linear extrapolation of the simulated quantum inversion charge Q_S vs. voltage characteristic, has been studied. In particular, we focused on the threshold voltage shift with respect to the classical solution (i.e. the Q-DV) for the EPI structure and for the more conventional highly doped (UHD) structure obtained setting $t_{EPI}=0$. Due to the low surface doping, a much lower Q-DV could be expected in the EPI case compared to UHD. Instead, similar Q-DV is obtained for UHD and EPI with small t_{EPI} . In fact, for $t_{EPI}=10$ nm and t_{OX} ranging from 3 to 10 nm, we obtained only a 20% reduction of Q-DV with respect to UHD, showing that this effect does not simply depend on the doping at the interface, but rather, a long-range dependence on doping holds. This point is emphasized by Fig. 3-a, where Q-DV is reported as a function of t_{EPI} : a relatively thick EPI is needed to eliminate the quantum effects originated by the presence of the high-doped background plane. Fig. 3-b, instead, shows how V_{TH} can be controlled by properly selecting t_{EPI} : while the UHD MOS presents a threshold voltage that is too large for low supply voltage applications ($V_{DD} \approx 1-1.5$ V) the introduction of the low-doped layer provides a further degree of freedom, additional to the doping of the background-plane, for designing the threshold voltage according to circuit applications. The reported results also clearly show that simple Q-DV models based on average channel doping values cannot be applied to the highly non-uniform cases needed for US-MOSFETs.

The case of single- and double-gate SOI structures is reported in Fig. 4, where the threshold voltage is reported as a function of silicon thickness for single- and double-gate SOI structures featuring uniform doping concentration (10^{17}cm^{-3}) and gate oxide thickness $t_{OX}=3$ nm. By comparing the threshold voltages calculated classically and quantum-mechanically for the SGM case, we notice that for large t_{SI} the threshold voltage is coincident with that of the bulk MOSFET for the same doping and that quantum calculations provide larger threshold compared to the results of classical simulations, as in the usual bulk MOSFET case [5]. As t_{SI} is reduced to reach the fully depleted case, V_{TH} decreases due to the reduction of fixed charge. In the quantum-mechanical case, the threshold voltage starts to increase below a critical thickness (≈ 10 nm) due to the reduction of the energy density of states that is a consequence of increasing size quantization (see also Fig. 1). A similar qualitative picture is found in the DGM case, where a shift of the V_{TH} characteristic towards larger thicknesses occurs due to the presence of the two gates, both contributing to the silicon film depletion.

4.2. Effective mobility results

We can address channel electron effective mobility starting from Fig. 5, showing the profiles of the quantum electron density in structures that feature $N_A=1 \times 10^{18} \text{cm}^{-3}$, $N_{EPI}=1 \times 10^{15} \text{cm}^{-3}$, $t_{OX}=3$ nm, and different t_{EPI} . The comparison has been performed for a given inversion sheet density above threshold ($N_S=3 \times 10^{12} \text{cm}^{-2}$). For increasing t_{EPI} , charge confinement reduces due to the reduction of transverse electric field.

In fact, the introduction of low-doped EPI effectively decreases the effective electric field (EEF), reported in Fig. 6 as a function of t_{EPI} for a given above-threshold inversion sheet density. Here, the EEF has been computed as the mean value of the electric field component normal to the interface averaged over the carrier density. Consequently, no surprise if improved effective mobility (μ_{eff}) is obtained for larger t_{EPI} extracting the values from experimental mobility vs. effective field curves [4], as it is again shown in Fig. 6.

The μ_{eff} calculated in the relaxation time approximation including the effects of inter- and intra-sub band phonon scattering and surface roughness, and consistently with the sub bands and wavefunctions computed solving the Schrödinger and Poisson equations, are given in Fig. 7, confirming that a decrease of effective field due to the reduction of the depletion charge produces an improved channel mobility in the epitaxial MOSFETs. It should be mentioned that UHD devices are expected to present even larger disadvantages for low inversion charge densities [4], due to the reduced screening of the Coulomb doping scatterer.

Similar calculations have been performed for very thin single- and double-gate SOI MOSFETs. Fig. 8 reports the results obtained for SGM structures with uniform doping concentration of 10^{16}cm^{-3} . The phonon limited mobility is a non-trivial function of the silicon thickness, since two competing mechanisms coexist: a) the reduction of the silicon thickness increases the quantization and decreases the available density of states, leading to a mobility enhancement; b) for decreasing t_{SI} , instead, a stronger surface confinement of the wave functions occurs, that leads to an increase of the overlap factors and, therefore, of the scattering rates and, consequently, to a decrease of the mobility.

Therefore, in these calculations we did not include the effects of surface roughness, since the usual modeling approach is not adequate to account for the interaction of the wave-function with both the upper and lower interfaces.

4.3. Gate capacitance results

Another important issue relates to the total MOS capacitance C_{TOT} , i.e. the series of the oxide capacitance $C_{OX}=\epsilon_{OX}/t_{OX}$ and of the inversion layer capacitance $C_{INV}=dQ_S/d\varphi_S$, φ_S being the Si surface potential. As mentioned, C_{TOT} in quantized inversion layers reduces with respect to C_{OX} as t_{OX} is decreased, due to the larger effective oxide thickness related to the displacement of the inversion charge from the device interface. As shown in [7] for uniform channels, such a reduction becomes more relevant for lower doping. Consequently, compared to the UHD case a lower capacitance should be expected for the EPI one. Fig. 9-a reports $C_{TOT}=dQ_S/dV_G$ as a function of t_{OX} , for $t_{EPI}=10\text{nm}$ and 0nm . As can be seen, the capacitance degradation increases at thinner oxides, but no additional degradation occurs if thin low-doped EPI is introduced. As reported in Fig. 9-b, the simulated results indicate that no serious degradation shall be expected by the application to UHDs, such as those described in [30], of EPI layers up to 30-40nm.

Results obtained by classical simulation accounting for Fermi-Dirac statistics are also reported. As can be seen degenerate statistics can only partially account for the capacitance reduction occurring at thin gate oxides, as already reported in [8]. Furthermore, classical simulations underestimate the capacitance dependence on t_{EPI} .

Gate capacitance calculations have been performed for SOI structures too, in order to compare SGM and DGM. Such a comparison is important because in the case of ultra-thin DGM devices (t_{SJ} comparable with the displacement of electrons from the oxide interface) biased at low gate voltages, the electron concentration peaks in the middle of the silicon layer. This effect is due to the interaction of the two gate electrodes, and may be the sign of a larger inversion sheet density for the DGM compared to the SGM with same t_{SJ} and biased at the same gate drive. In particular, in presence of substantial volume inversion, a larger inversion charge and a larger derivative of the same quantity with respect to the gate voltage (dQ_S/dV_G , $Q_S=q\cdot N_S$ being the inversion charge sheet density) could be expected in the DGM case. Simulation results show that even when t_{SJ} is comparable with the displacement of the charge peak from the interface, the increase of DGM inversion charge and capacitance occurs only for bias points close to the threshold voltage, and almost negligibly, while above threshold the two structures show coincident behaviors. This is shown in Fig. 10, reporting the comparison between the dQ_S/dV_G of a DGM and of a SGM with $t_{SJ}=5\text{nm}$ and the same effective width (i.e. the charge per unit width of the SGM is multiplied by a factor of two). From this result we may not expect large improvements in the current and transconductance as a direct consequence of the increase of the inversion charge due to the interaction between the two gates of the DGM structure.

5. Conclusions

In this paper some of the more relevant quantum effects taking place in bulk and SOI MOSFETs have been investigated through 1D simulations. Few methods for the solution of Schrödinger equation have been reviewed to the purpose of orienting in the choice of an accurate and efficient algorithm. Simulation results concerning the evidence of quantum effects in MOSFETs, such as quantum threshold voltage shift, capacitance degradation, and mobility issues, have been presented, also comparing standard MOS structures with some of the new promising ones.

References

- [1] R. Dennard *et al.*, IEEE J. Solid-State Circuits **9**, 256 (1974).
- [2] G. Baccarani, M. R. Wordeman, and R. H. Dennard, IEEE Trans. Electron Devices **31**, 452 (1984).
- [3] M. Ono *et al.*, IEEE Trans. Electron Devices **42**, 1822 (1995).
- [4] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, IEEE Trans. Electron Devices **41**, 2357 (1994).
- [5] M. van Dort *et al.*, IEEE Trans. Electron Devices **39**, 932 (1992).
- [6] S. Harelund *et al.*, IEEE Trans. Electron Devices **43**, 90 (1996).
- [7] S. Takagi and A. Toriumi, IEEE Trans. Electron Devices **42**, 2125 (1995).
- [8] K. Krisch, J. Bude, and L. Manchanda, IEEE Electron Device Lett. **11**, 521 (1997).

- [9] T. Kuhn and F. Rossi, *Phys. Rev. B* **46**, 7496 (1992).
 [10] R. Lake, G. Klimeck, R. Bowen, and D. Javanovic, *J. Appl. Phys.* **81**, 7845 (1997).
 [11] D. Vasiljeska and D. Ferry, *IEEE Trans. Electron Devices* **44**, 577 (1997).
 [12] W. Frenseley, *Rev. Mod. Phys.* **62**, 745 (1990).
 [13] P. Bordone *et al.*, *Phys. Stat. Sol. B* **204**, 303 (1997).
 [14] M. Fischetti and S. Laux, *Phys. Rev. B* **48**, 2244 (1993).
 [15] M. van Dort, P. Woerlee, and A. Walker, *Solid State Electron.* **37**, 411 (1994).
 [16] A. Spinelli, A. Benvenuti, and A. Pacelli, *IEDM Tech. Dig.* (1996), p. 399.
 [17] S.-H. Lo, D. Buchanan, Y. Taur, and W. Wang, *IEEE Electron Device Lett.* **18**, 209 (1997).
 [18] J. Lopez-Villanueva *et al.*, *IEEE Trans. Electron Devices* **44**, 1915 (1997).
 [19] C. Bowen *et al.*, *IEDM Tech. Dig.* (1997), p. 869.
 [20] M. Fischetti, S. Laux, and D. DiMaria, *Appl. Surf. Sci.* **32**, 578 (1989).
 [21] S. Jallepalli *et al.*, *IEEE Trans. Electron Devices* **44**, 297 (1997).
 [22] G. Strang and G. Fix, *An analysis of the finite element method* (Prentice-Hall, Englewood Cliffs, N.J., 1973).
 [23] O. Zienkiewicz, *The finite element method* (McGraw-Hill, London, 1977).
 [24] F. Stern and W. Howard, *Phys. Rev.* **163**, 816 (1967).
 [25] F. Stern, *Phys. Rev. B* **5**, 4891 (1972).
 [26] A. Abramo, J. Bude, F. Venturi, and M. Pinto, *IEEE Electron Device Lett.* **17**, 59 (1996).
 [27] F. Balestra *et al.*, *IEEE Electron Device Lett.* **8**, 410 (1987).
 [28] J. Colinge *et al.*, *IEDM Tech. Dig.* (1990), p. 595.
 [29] D. Frank, S. Laux, and M. Fischetti, *IEDM Tech. Dig.* (1992), p. 553.
 [30] C. Fiegna *et al.*, *IEEE Trans. Electron Devices* **41**, 941 (1994).

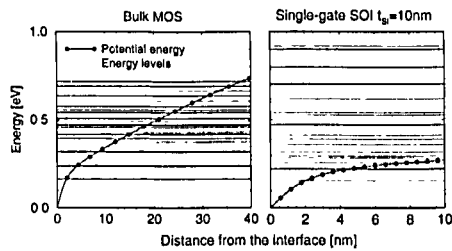


Figure 1: Potential energy and quantized energy levels for a bulk MOS (left) and a single-gate SOI MOS (right) structures with $t_{SI}=10$ nm. Uniform doping 10^{17} cm^{-3} , $t_{OX}=3$ nm, and $N_{INV}=6 \times 10^{12} \text{ cm}^{-2}$ in both devices.

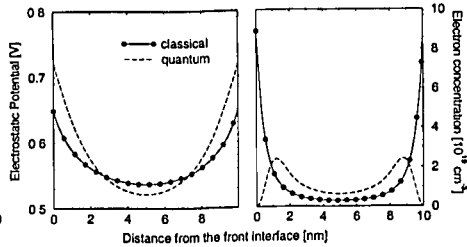


Figure 2: Electrostatic potential (left) and electron concentration (right) calculated classically and quantum mechanically in a double-gate SOI with $t_{SI}=10$ nm, $t_{OX}=3$ nm.

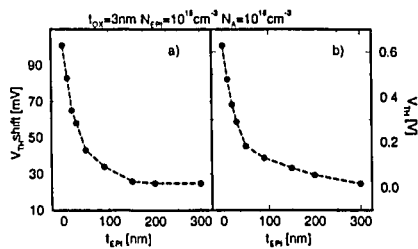


Figure 3: a) Quantum induced V_{TH} shift vs. EPI thickness; b) V_{TH} vs. EPI thickness. $t_{OX}=3$ nm, $N_{EPI}=10^{16} \text{ cm}^{-3}$, $N_A=10^{18} \text{ cm}^{-3}$.

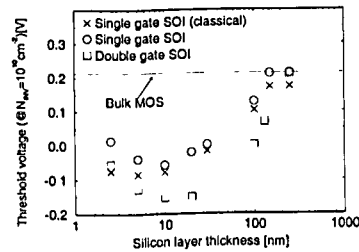


Figure 4: Threshold voltage vs Silicon layer thickness for single- and double-gate SOI MOS structures.

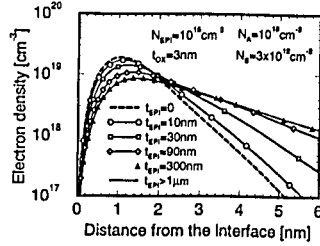


Figure 5: Electron density vs. position for different EPI thicknesses. Symbols are for visualization only and do not correspond to actual discretization points. The bias V_G was set in order to obtain the same charge sheet density for all structures.

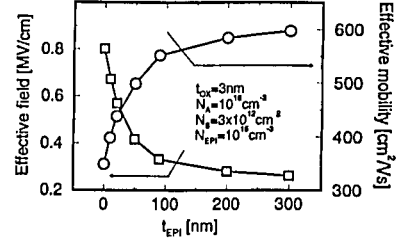


Figure 6: Dependence on EPI thickness of calculated effective electric field (squares, left) and corresponding effective mobility extracted from the experimental mobility curve of [4] (circles, right). The bias V_G was set in order to obtain the same charge sheet density for all structures.

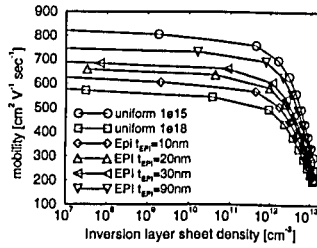


Figure 7: Mobility as a function of inversion layer sheet density for different epitaxial layer thickness.

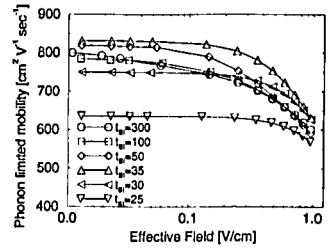


Figure 8: Phonon-limited mobility as a function of effective field for single-gate SOI MOS structures and different silicon layer thicknesses.

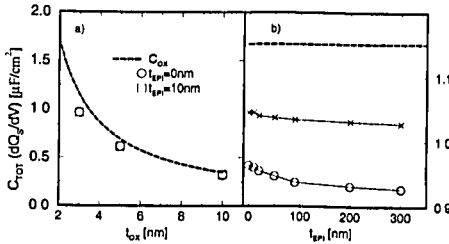


Figure 9: a) Total capacitance vs. oxide thickness for $t_{EPI}=0, 10$ nm. Symbols: simulated results. Dashed-line: C_{OX} . b) Total capacitance vs. t_{EPI} for $t_{OX}=3$ nm. Circles: simulated results; crosses: results of classical simulation. Dashed-line: C_{OX} . All simulated structures feature $N_{EPI}=10^{16}cm^{-3}$, $N_A=10^{18}cm^{-3}$. The bias V_G was always set in order to obtain the same charge sheet density ($N_S=3 \times 10^{12}cm^{-2}$) for all structures.

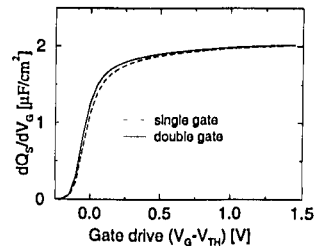


Figure 10: Comparison of dQ_S/dV_G for a double-gate (solid) and a single-gate MOSFET with $t_{SJ}=5$ nm and the same effective width (i.e. the charge per unit width of the SGM is multiplied by a factor of two).