Journal of Computational Electronics 1: 453–465, 2002
© 2003 Kluwer Academic Publishers. *Manufactured in The Netherlands*.

The Role of Quantization Effects on the Operation of 50 nm MOSFETs, 250 nm FIBMOS Devices and Narrow-Width SOI Device Structures

D. VASILESKA, I. KNEZEVIC, R. AKIS, S. AHMED AND D.K. FERRY
Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287-5706, USA

Abstract. We investigate the role of the quantum-mechanical space-quantization effects on the operation of a 50 nm MOSFET device, an asymmetric 250 nm FIBMOS device and a narrow-width SOI device structure. We find that space-quantization effects give rise to larger average displacement of the carriers from the interface proper and lower sheet electron density in both the regular and the asymmetric MOSFET device structures. The effect is even more pronounced in the narrow-width SOI device due to the presence of a two-dimensional confinement (both vertical and along the width direction). The reduction in the sheet electron density, in turn, gives rise to shift in the devices threshold voltage, on the order of 100–200 mV, depending upon the device structure being investigated. This leads to 20–40% decrease of the device on-state current which depends upon the gate bias. Hence, to properly describe the operation of future ultra-small devices it is mandatory to incorporate quantum-mechanical space quantization effects into existing classical device simulators (drift-diffusion, hydrodynamics or Monte Carlo particle-based simulators) since first-principle quantum-mechanical calculations (direct solution of the many-body Schrödinger equation, Green's functions method, etc.) are still limited to one-dimensional structures and rely on a number of approximations.

Keywords: MOSFETs, asymmetric device structures, narrow-width SOI devices, space-quantization effect, threshold voltage degradation, on-state current degradation

1. Introduction

The end of the CMOS scaling is foreseeable, although the specific combination of factors that will end scaling is not known. As industry moves through the two to three year technology generation cycles, the complexity of the processes is increasing and the demands on fabrication equipment performance are more strenuous. For example, in order to fabricate a 25 nm device, lithography tools are needed that operate beyond optical lithography limits. The performance of the scaled device in the 25 nm regime is itself problematical. The gate oxide has to be aggressively scaled down to enhance device performance, but, as it approaches 1 nm, tunneling through the gate oxide creates unacceptably large off-state currents, dramatically increasing quiescent power consumption, and rendering the device impractical for analog applications due to unacceptable noise level. Source to drain tunneling is another mechanism that contributes to the off-state current. Another issue that will pose serious problems is that the dopant distribution in future ultra-small devices can no longer be regarded as statistically uniform. Namely, statistical uniformity requires a large number of dopants, whereas in a 25 nm MOSFET device one has less than one hundred dopant atoms in the junction region, so the number and location of each dopant atom will play an important role in determining the device behavior. The challenge of precisely placing small numbers of dopants may represent a brick wall and could end conventional MOSFET scaling. These effects have inspired a number of recent advances in the computer simulation of small semiconductor devices. For instance, tunneling through ultrathin oxides was studied by Städele, Tuttle and Hess 2001, Städele *et al.* (2000), and Demkov, Zhang and Loechelt (2001). Source-to-drain tunneling was investigated by Watling, Brown and Asenov (2002) using the

density gradient approach, whereas Vasileska, Gross and Ferry (1998) and Asenov *et al.* (2001) analyzed the role of random dopant distribution on the device behavior.

However, there is an additional important feature, which needs to be accounted for when simulating small semiconductor devices, and which will be our primary focus in the remainder of this paper: the quantum-mechanical space-quantization effect. This effect arises in state-of-the-art devices due to the very high substrate doping used to prevent the punchthrough effect. One consequence of the space-quantization effect is higher average displacement of the carriers from the interface proper, which leads to effective oxide thickness increase and total gate capacitance degradation. Additional degradation of the total gate capacitance arises from the poly-depletion effects that lead to additional capacitance component in series with the oxide and the inversion layer capacitance (Vasileska, Schroder and Ferry 1997). The total gate-capacitance degradation due to these two effects is clearly seen in Fig. 1, where we show the total gate capacitance for different technology generations (as a function of the oxide thickness). We use Fermi-Dirac (F-D) and Maxwell-Boltzmann (M-B) statistics for the case of classical charge description and F-D statistics for the case of quantum-mechanical charge description. From the results shown in Fig. 1, it follows that the quantum-mechanical charge description leads to approximately 20% degradation of the total gate capacitance when compared to the classical charge description model for the end of the roadmap devices with oxide thickness between 1 and 1.5 nm. Further capacitance degradation, of up to 40% for the smallest oxide thickness, is due to the additional poly-depletion effect. The total gate capacitance

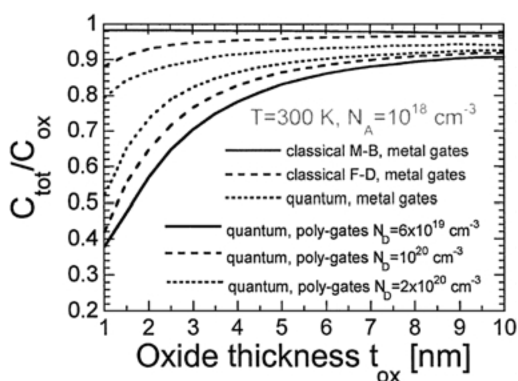

Figure 1. Total gate capacitance as a function of the oxide thickness.

degradation, on the other hand, when combined with the quantum-mechanical band-gap widening effect due to the reduced density of states, gives rise to a reduction of the sheet electron density. This, in turn, increases the threshold voltage and, at the same time, degrades the device transconductance. Hence, to properly describe the operation of future ultra-small devices, it becomes mandatory to incorporate quantum-mechanical and poly-depletion effects into the device simulators.

In this paper, we will describe the use of *an effective potential approach*, introduced by Ferry (2000), to account for the quantum-mechanical space quantization effect in the simulation of small devices. We have incorporated this approach into our Monte Carlo particle-based simulator to investigate the role of the quantum-mechanical space-quantization effects on the operation of a conventional 50 nm channel length MOSFET (Formicone *et al.* 2002) device and a 250 nm channel length FIBMOS device structure (Knezevic, Vasileska and Ferry 2002, Knezevic *et al.* 2002a, b). In addition, we have utilized both the effective potential approach and the transfer matrix approach in the investigations of the operation of a narrow-width SOI device structure.

The paper is organized as follows. In Section 2, we describe the use of the effective potential approach in treatment of quantum effects. A brief description of the Monte Carlo device simulator is given in Section 3. In Section 4 we present the simulation results for the 50 nm MOSFET device. The usefulness of FIBMOS device structures and the role played by quantum-mechanical space-quantization effects in its operation is discussed in detail in Section 5. Simulation results for the SOI device, obtained via both the effective potential and the transfer matrix approach, are presented in Section 6. We end the paper with conclusions regarding this work, and present some future directions of research.

2. Quantum-Mechanical Space Quantization Effect: The Effective Potential Approach

One way of incorporating quantum-mechanical space-quantization effects is to use 1D or 2D Schrödinger equation solvers coupled with existing particle-based simulator (Assad *et al.* 2000). This is not practical, however, because the repeated solution of the Schrödinger equation adds in to the required CPU time. There have been several interesting proposals in the past that avoid

this problem. All of them are based on the idea of quantum potentials that derive from the hydrodynamic formulation of quantum mechanics first introduced by de Broglie (1926, 1927) and Madelung (1926) and later developed by Bohm (1952a, b). In this picture, the wave function is written in complex form in terms of its amplitude and phase, and, when substituted back into the Schrödinger equation leads to coupled equations of motion for the density and phase, of the form

$$\frac{\partial \rho(\mathbf{r},t)}{\partial t} + \nabla \cdot \left(\rho \frac{1}{m} \nabla S \right) = 0, \quad (1)$$

$$-\frac{\partial S(\mathbf{r},t)}{\partial t} = \frac{1}{2m}(\nabla S)^2 + V(\mathbf{r},t) + Q(\rho,\mathbf{r},t), \quad (2)$$

where the probability density is $\rho(\mathbf{r},t) = R(\mathbf{r},t)^2$. With identification of the velocity as $\mathbf{v} = \nabla S/m$, and the flux as $\mathbf{j} = \rho \mathbf{v}$, Eq. (1) is the continuity equation. Hence, Eqs. (1) and (2), arising from this so called Madelung transformation to the Schrödinger equation, have the form of classical hydrodynamic equations with the addition of an extra potential, often referred to as the *quantum* or *Bohm* potential, written as

$$Q = -\frac{\hbar^2}{2mR}\nabla^2 R \rightarrow -\frac{\hbar^2}{2m\sqrt{n}}\frac{\partial^2 \sqrt{n}}{\partial x^2}, \quad (3)$$

where the square root of the density n is proportional to the magnitude of the wave function R. The Bohm potential essentially represents a field through which the particle interacts with itself. It has been used, for example, in the study of wave packet tunneling through barriers (Dewdney and Hiley 1982), where the effect of the quantum potential is seen to lower or smooth barriers, and hence allow particles to leak through. A standard way to include quantum effects into classical simulation tools is to add the above-described quantum potential to the mean-field potential computed from solving Poisson's equation. Such potential corrections have been employed mostly in the context of fluid approximations leading to the so-called *quantum-hydrodynamic* (QHD) equations (Gardner 1994).

In recent years, sophisticated Monte Carlo particle-based solvers have been developed by various research groups, based on analytic or a full-band structure (e.g., Duncan, Ravaioli and Jakumeit (1998), Takeda, Mori and Hamaguchi (2002), and Saraniti *et al.* (2002)). Particle-based solvers are the only way to capture the velocity overshoot effect, which is prominent in small devices due to high electric fields, and thus they give a faithful description of the on-state regime (Knezevic, Vasileska and Ferry 2002). However, they are often inadequate for the description of subthreshold regime and leakage, but there has been recent work regarding these issues (e.g., Kan, Narayan and Pei (2002)). In analogy to the smoothed potential representations discussed for the QHD model above, it is desirable to define a smooth quantum potential for use in quantum particle based simulation. The actual quantum potential (3) may be included in particle simulators (Tsuchiya, Winstead and Ravaioli 2001), but since it depends on the derivatives of the density, this can lead to the amplification of noise characteristic for particle simulators. Ferry (2000), based on the idea of Feynman and Kleinert (1986), suggested an *effective potential* that is derived from a wave packet description of particle motion, where the extent of the wave packet is defined from the range of wavevectors established by the thermalized distribution function (specified via the carrier temperature). The effective potential seen by electrons is then represented as a convolution of the Hartree potential obtained from solving Poisson's equation and a Gaussian function. In other words,

$$V_{eff}(x) = \frac{1}{\sqrt{2\pi}a_0}\int_{-\infty}^{\infty} V(x')\exp\left(-\frac{(x-x')^2}{2a_0^2}\right)dx', \quad (4)$$

where $V(x')$ is the actual potential, and a_0 is the spacial spread of the wavepacket. The effective potential accounts for the *size of the electron* and its associated wavepacket, which feels the presence of barriers, etc. at a distance. From this Ansatz, the actual particle is treated as point-like in the presence of the effective potential associated with its wave-like nature, leading back to a classical particle simulation scheme. The additional simulation time needed to implement the aforementioned scheme is less than 10%.

When implementing the effective potential approach due to Ferry (2000), one of the questions that one would naturally ask is related to the actual value of the Gaussian smoothing parameter a_0. Also, there has been a debate as to whether the smoothing parameter will depend upon the shape of the confining potential or the substrate doping density. To verify the applicability of the approach and investigate the proper values of the Gaussian smoothing parameter, we have simulated a simple MOS capacitor structure in which

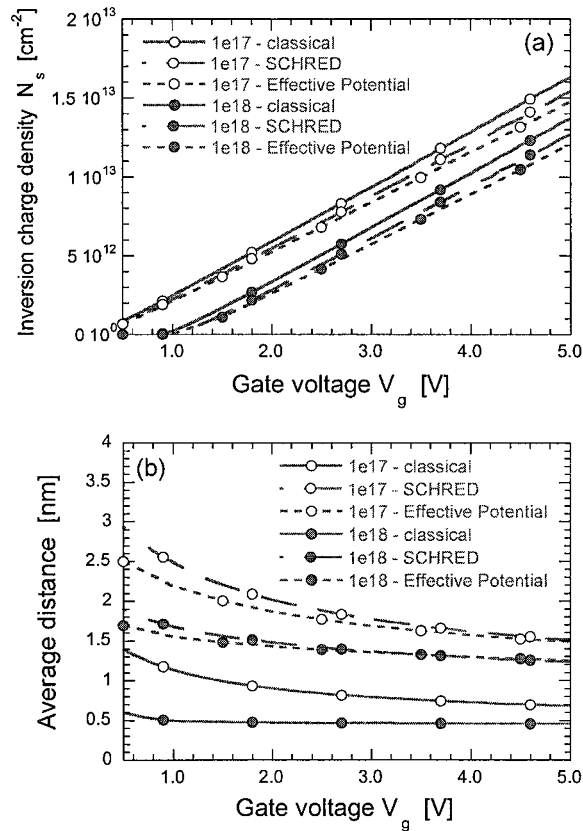

Figure 2. (a) Sheet electron density and (b) average displacement of the carriers from the semiconductor/oxide interface, as a function of the gate voltage. For each doping density, we use classical charge description, fully quantum-mechanical charge description via SCHRED (see http://punch. ecn.purdue.edu), and Gaussian smoothing of the Hartree potential (effective potential approach).

the band-bending leads to triangular confinement. The oxide thickness of the MOS capacitors being simulated is 6 nm, and the substrate doping equals 10^{17} and 10^{18} cm^{-3}, respectively. The results of these simulations are shown in Fig. 2(a) and (b), where we plot the gate-voltage dependence of the inversion charge density, and the average distance of the carriers from the interface proper (Akis *et al.* 2001). From the results shown in Fig. 2, it is obvious that the use of a single Gaussian smoothing parameter $a_0 = 0.5$ nm along the growth direction can quite accurately describe the reduction in the inversion layer electron density due to quantum-mechanical space-quantization effect. It also leads to accurate description of the displacement of the carriers away from the interface proper which, in turn, gives rise to quantum capacitance in series with the oxide capacitance.

3. Description of the Monte Carlo Particle-Based Simulator

The Monte Carlo model, used in the transport portion of the is based on the usual Si band-structure for three-dimensional electrons in a set of non-parabolic Δ-valleys with energy-dependent effective masses. The six conduction band valleys are included through three pairs: valley pair 1 pointing in the (100) direction, valley pair 2 in the (010) direction, and valley pair 3 in the (001) direction. The explicit inclusion of the longitudinal and transverse masses is important and this is done in the program using the Herring-Vogt transformation (Herring and Vogt 1956). Intravalley scattering is limited to acoustic phonons. For the intervalley scattering, we include both g- and f-phonon processes. It is important to note that, by group symmetry considerations, the zeroth-order low-energy f- and g-phonon processes are forbidden. Nevertheless, three zeroth-order f-phonons and three zeroth-order g-phonons with various energies are usually assumed (Jacoboni and Reggiani 1983). We have taken into account this selection rule and have considered two high-energy f- and g-phonons and two low-energy f- and g-phonons. The high-energy phonon scattering processes are included via the usual zeroth-order interaction term, and the two low-energy phonons are treated via a first-order process (Ferry 1976). The first-order process is not really important for low-energy electrons but gives a significant contribution for high-energy electrons. The low-energy phonons are important in achieving a smooth velocity saturation curve, especially at low temperatures. The phonon energies and coupling constants in our model are determined so that the experimental temperature-dependent mobility and velocity-field characteristics are consistently recovered (Gross, Vasileska and Ferry 2000). At present, impact ionization, surface-roughness and Coulomb scattering are not included in the model. They are omitted as they tend to mask the role of the space-quantization effects on the overall device performance.

In solving Poisson's equation, the Ensemble Monte Carlo (EMC) simulation is used to obtain charge distribution in the device. We use the Incomplete Lower-Upper decomposition method for the solution of the 2D Poisson equation. After solving for the potential, the electric fields are calculated and used in the free-flight portion of the Monte Carlo transport kernel. However, the charges obtained from the EMC simulation are usually distributed within the continuous mesh cell instead

of on the discrete grid points. The particle mesh method (PM) is used to perform the switch between the continuum in a cell and discrete grid points at the corners of the cell. The procedures of PM coupling is outlined below:

- Assign the charges from within the continuous mesh cell onto the discrete grid points.
- Solve Poisson equation for potentials at those points.
- Calculate electric fields at those points from the potential profiles and obtain the electric fields for particles within the cell.

The charge assignment to each mesh-point depends on the particular scheme that is used. A proper scheme must ensure proper coupling between the charged particles and the Coulomb forces acting on the particles. Therefore, the charge assignment scheme must maintain zero self-forces and a good spatial accuracy of the forces. To achieve this, two major methods have been implemented in the present version of the code: the Nearest-Grid-Point (NGP) scheme and the Cloud-in-Cell (CIC) scheme. The cloud-in-cell scheme (CIC) produces a smoother force interpolation, but introduces self-forces on non-uniform meshes. These issues have been dealt with extensively by Hockney and Eastwood (1981), and quite recently by Laux (1996).

The device current is determined by using two different, but consistent methods. First, by keeping track of the charges entering and exiting each terminal, the net number of charges over a period of the simulation can be used to calculate the terminal current. The method is quite noisy, due to the discrete nature of the electrons. In the second method, the sum of the electron velocities in a portion of the device is used to calculate the current. For this purpose, the device is divided into several sections along the x-axis (along the semiconductor/oxide interface). The number of electrons, and their corresponding velocities, are added up for each section after each free-flight time step. The total x-velocity in each section is then averaged over several time steps to determine the current for that section. The total device current is determined from the average of several sections, which gives a much smoother result than that based on counting the terminal charges. By breaking the device into sections, individual section currents can be compared to verify that the currents are uniform. In addition, sections near the source and drain regions may have a high y-component (vertical) in their velocity and should be excluded from the current calculations. Finally, by using several sections in the channel, the average energy and velocity of electrons along the channel is checked to ensure proper physical characteristics.

4. MOSFET Simulations

The first device structure we simulate is a 50 nm MOSFET that is schematically shown in Fig. 3, with channel doping of 10^{18} cm^{-3}, source/drain doping of 10^{19} cm^{-3} and junction depth of 36 nm. The oxide thickness is 2 nm and the width of the device being simulated is 0.8 μm. A non-uniform tensor-product grid is used with 0.5 nm mesh size normal to the interface and 1.0 nm parallel to the interface in the active channel region. The Monte Carlo procedure discussed in Section 2 uses analytic, non-parabolic bands and particles are simulated throughout the whole device (rather than just in the channel). In the actual simulation, ~30,000 particles are included, although most of the particles reside in the source and drain regions. Values of the Gaussian smoothing parameter a_0 are taken to be 0.5 nm normal to the interface (in agreement with the simulations of the MOS capacitor structure discussed in the introduction part of this article) and 1.0 nm along the channel.

In Fig. 4, we show the conduction band edge, found in the device simulation, for the bias conditions $V_G = V_D = 1$ V. From the results shown in the bottom panels, one can clearly see that the introduction of the effective potential shifts the conduction band edge upwards, thus accounting for the so-called band-gap widening effect due to the quantum mechanical quantization in the triangular potential well near the Si-SiO_2 interface. As mentioned, the upward shift of the conduction band edge leads to a reduction of the carrier density at the interface proper. This is shown in Fig. 5, where we plot

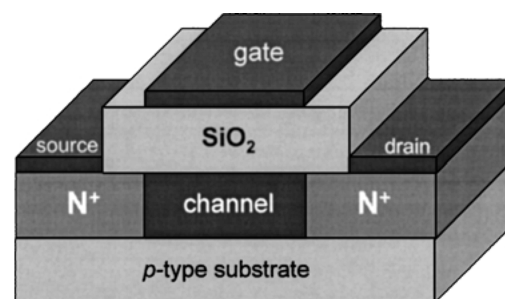

Figure 3. Schematics of the 50 nm MOSFET device structure being simulated.

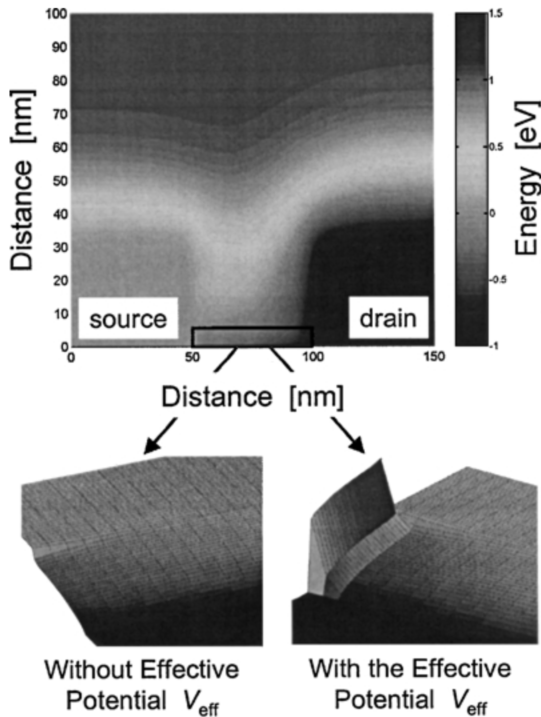

Figure 4. Conduction band edge (top panel) for applied bias $V_G = V_D = 1$ V. In the lower right (left) panel we plot the conduction band edge near the Si/SiO_2 interface when the effective potential is included (omitted) in the simulations.

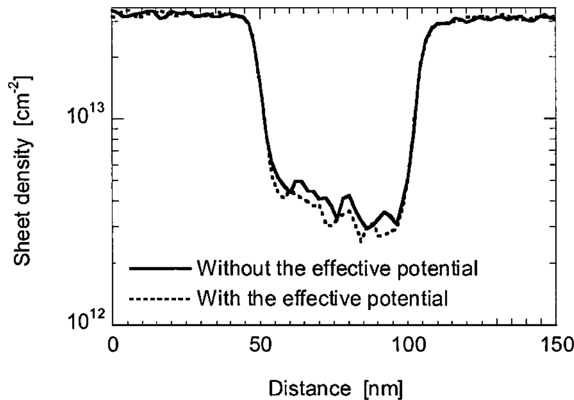

Figure 5. Integrated electron density along the channel for $V_G = V_D = 1$ V.

the integrated (sheet) electron density along the channel for the same applied bias. Also, the electron density is moved away from the interface, which is more clearly seen from the results shown in Fig. 6 for two different bias conditions. We note that, for a drain voltage of 0.4 V, the channel is continuous from the source to the drain. However, with a drain bias of 1 V, the channel is clearly pinched off due to saturation effects. It is also noticeable that the peak of the density is lower in the effective potential case, due to the quantization energy (and higher threshold voltage). Moreover, the carriers are more localized away from the interface. A quantitative description of the average displacement of the electrons from the interface is shown in Fig. 7 for the bias conditions of Fig. 6.

The quantization of charge in the inversion layer produces an expected increase of the threshold voltage in the channel. In Fig. 8, we plot the linear drain current as a function of the gate voltage, for a drain voltage of 0.1 V. It is clear that the turn-on of current is shifted by about 80 mV. The actual threshold voltage is quite difficult to determine in a Monte Carlo simulation. Here, we average the charge over the entire channel in order to estimate the inversion charge at any one point. This is done with low drain bias to keep the channel as homogeneous as possible.

In Fig. 9, we plot the output characteristics for the simulated device discussed here. The reduction in output current that is found is largely due to the increase in the threshold voltage in the device. Consequently, the quantization mainly affects this threshold voltage, which is a property of the confinement *normal* to the interface, whereas the transport is largely parallel to the interface.

5. FIBMOS Simulations

In the design of ultra-small devices, one has to deal with two contradictory requirements: (1) the reduction of short-channel effects, that necessitate the use of high substrate doping densities, and (2) hot-carrier reliability increase and substrate current reduction, that require a reduction of the in-plane electric fields and, therefore, smaller substrate doping density. A device structure that satisfies both requirements must have asymmetric doping profiles, and there has been a vast amount of theoretical and experimental effort to predict optimal device structures that can operate reliably at higher drain voltages and do not exhibit pronounced short-channel effects. Some representative devices proposed include lightly-doped drain (LDD) devices, gate overlapped LDD structure (GOLD), halo source GOLD drain (HS-GOLD) (Buti *et al.* 1991) and graded-channel MOS (GCMOS) devices (Ma *et al.* 1997). The last device structure overcomes the incompatible requirements of short-channel effects and hot-carrier resistance. As

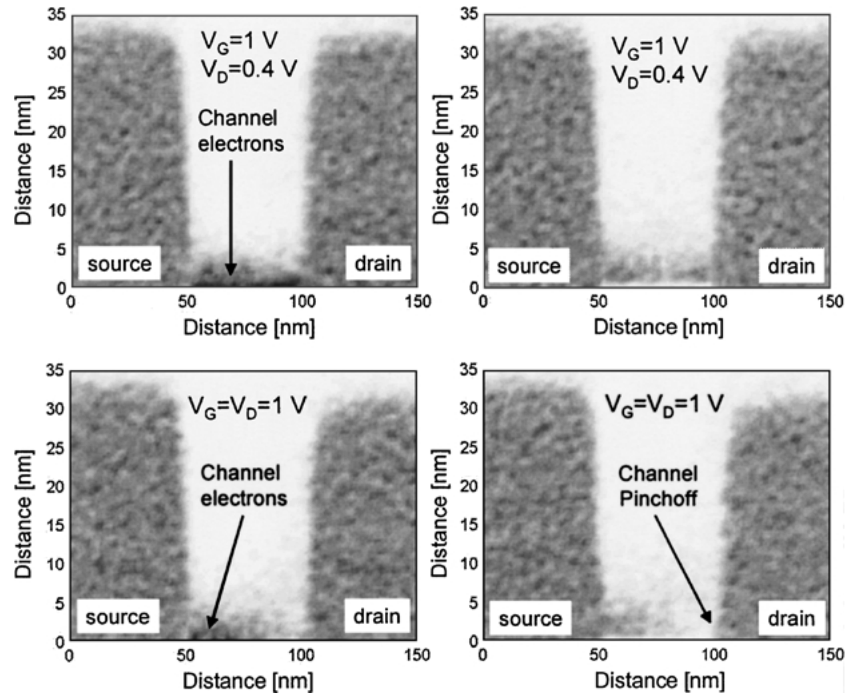

Figure 6. The electron density for two different bias levels. The left-hand panels are without the effective potential, while the right-hand panels include the effective potential.

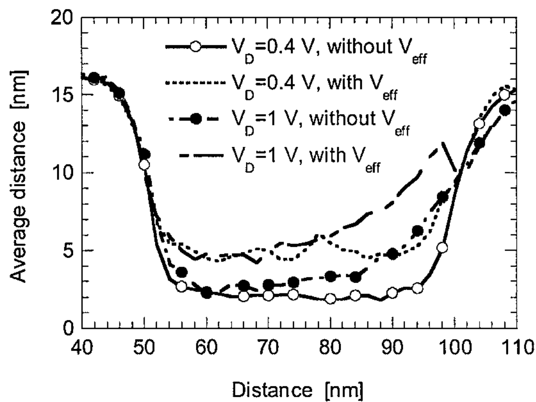

Figure 7. The "center of charge" for the inversion electrons along the channel for the bias conditions of Fig. 6.

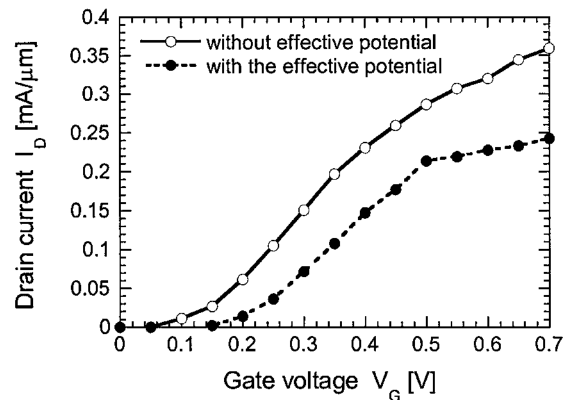

Figure 8. Device transfer characteristics. The inclusion of the effective potential provides a threshold shift due to the quantum-mechanical space-quantization effect.

such, it is a rather promising way, in terms of reliability and performance, for downward scaling without the need to further reducing the power supply. Further improvements in this device structure have been made with the use of focused-ion-beam (FIB) implantation (Vignaud *et al.* 1992), that provides point by point control of the lateral dopant density at the source end of the channel, resulting in the so-called FIBMOS device structure.

A schematic of this last type of device structure being investigated in the present study is shown in Fig. 10. The substrate doping equals 10^{16} cm^{-3}, source/drain doping is 10^{19} cm^{-3}, junction depth is 36 nm and bulk depth equals 400 nm. The oxide thickness is 5 nm and the width of the device being simulated is 1.4 μm. The width of the FIB region is 70 nm, the depth equals the junction depth, and the doping is 1.6×10^{18} cm^{-3}.

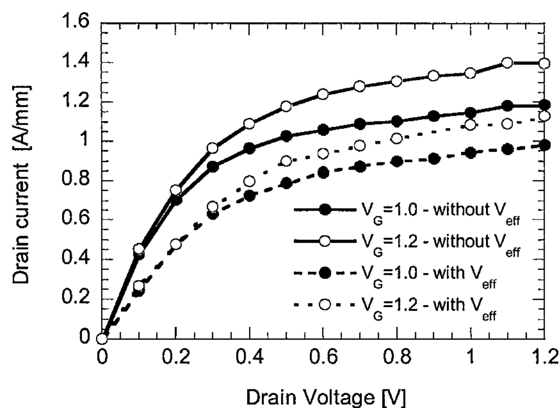

Figure 9. The output characteristics for the device discussed here. The decrease in output current with the inclusion of the effective potential is largely due to the threshold voltage increase.

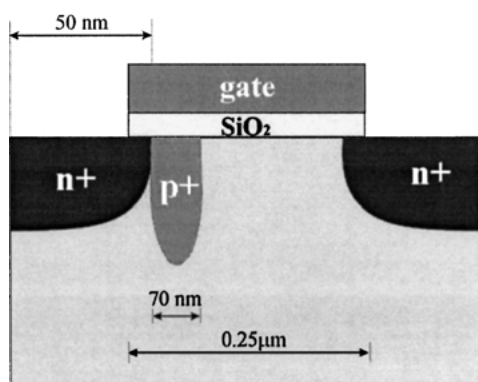

Figure 10. Schematic description of the asymmetric focused ion beam (FIBMOS) device structure being simulated.

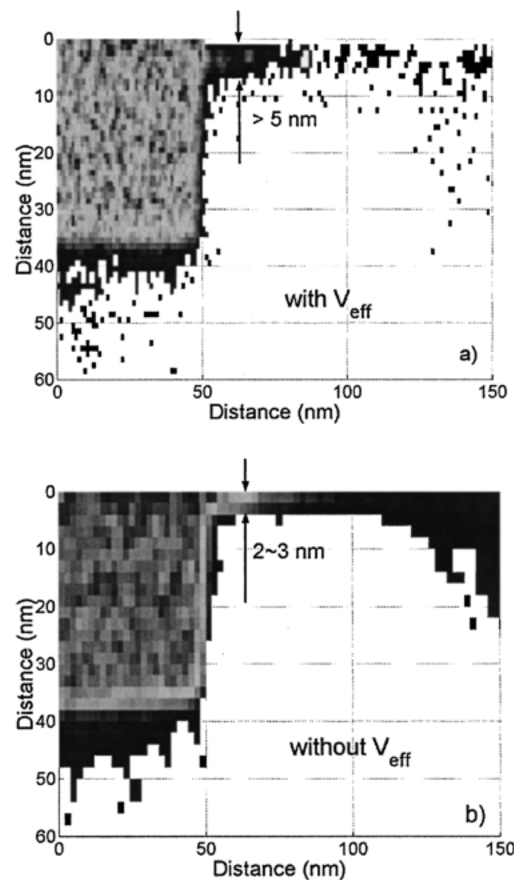

Figure 11. Electron density near the source end of the channel, for $V_G = V_D = 1$ V. Top panel corresponds to the case when V_{eff} is included in the model and the bottom panel corresponds to the case without V_{eff}.

When simulating this device we used the same mesh size and the same smoothing parameters as for the case of the MOSFET device. In here as well, we find upward shift of the conduction bands that accounts for the band-gap widening effect, that leads to a reduction of the carrier density at the interface proper. Also, the electron density is moved away from the interface because of the additional perpendicular electric field in the vicinity of the Si-SiO_2 interface. This later observation is more clearly seen in the results shown in Fig. 11, where we show the electron density at the source end of the channel and the average displacement of the carriers from the semiconductor/oxide interface for the bias conditions $V_G = V_D = 1$ V. For example, for the case when we include V_{eff} into the model, there is a spread of about 5 nm of the electron density in the vicinity of the p^+-implant (top panel of Fig. 11). On the other hand, when V_{eff} is omitted from the model, the electron density spread in the vicinity of the p^+-implant reduces to ~2–3 nm (bottom panel of Fig. 11), which suggests closer confinement of the carriers from the interface proper. As shown in Fig. 12, the upward shift of the conduction band edge leads to approximately a factor of two decrease in the sheet electron density which is consistent with the quantum-mechanical simulations for the MOSFET device shown in Fig. 7.

The above-described quantization of charge in the inversion layer produces an expected increase of the threshold voltage in the channel. This observation is more clearly seen from the results shown in Fig. 13, where we plot the linear drain current I_D as a function of the gate voltage V_G, for a drain voltage $V_D = 0.4$ V. If we take as a criterion for determining the threshold voltage being equal the gate voltage for which the drain current equals 10 μA/μm, we obtain a shift of

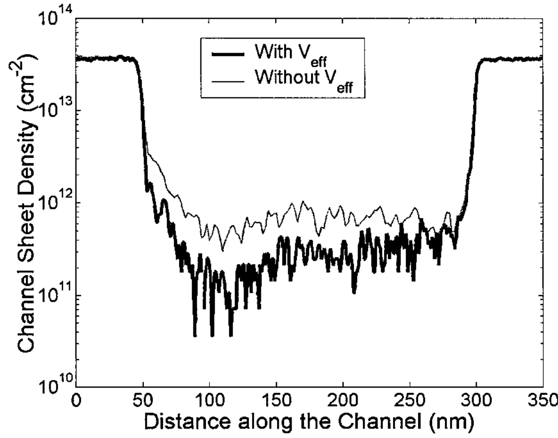

Figure 12. Sheet carrier density for $V_G = V_D = 1$ V, with and without V_{eff}.

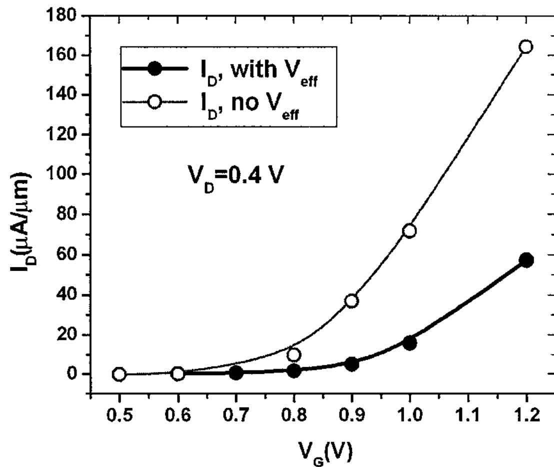

Figure 13. Transfer characteristics of a FIBMOS device. We use $V_D = 0.4$ V in these simulations to reduce the statistical noise in the current estimation.

approximately 200 mV when space-quantization effects are included in the model. This result, taking into account the very high doping density of the p^+-implant, is consistent with some previous observations (Vasileska and Ferry 1998).

The device output characteristics are shown in Fig. 14 for V_G equal to 1 V and 1.2 V, with and without V_{eff}. There are three noteworthy features in this figure: with V_{eff}, the drive current is reduced, the threshold voltage (V_{th}) is increased, and the transconductance is degraded. The inclusion of V_{eff} reduces the drain current I_D, due to two factors: the reduced average carrier velocity and the reduced sheet density, the latter being predominant. Also, since the slope of the $I_D - V_D$ curve

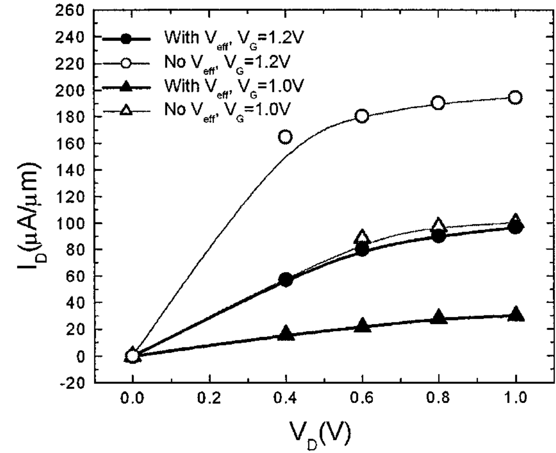

Figure 14. Output characteristics of a FIBMOS device for gate voltage $V_G = 1$ V and $V_G = 1.2$ V. Note the larger degradation of the on-state current for higher values of the gate voltage V_G.

in the linear region is proportional to $V_G - V_{th}$, we see that for a given V_G the inclusion of V_{eff} increases V_{th}, in agreement with the reduction in sheet density, as presented in Fig. 12. Finally, if for a given V_D we analyze the current increase between V_G equal to 1 V and 1.2 V, which is (roughly) proportional to transconductance, it is clear that the transconductance is lower with V_{eff} included, as expected according to the charge set-back depicted in Fig. 11.

In Fig. 15(a) and (b), we show the average carrier energy and the average velocity of the carriers along the channel. We find that the average carrier energy peaks at the p^+-implant region where the high doping density leads to very large electric fields. At the drain end of the channel, the low in-plane fields and significant phonon scattering give rise to a reduction of the average electron energy. These trends are observable regardless of whether we included or omitted the effective potential in the theoretical model. However, there is one drastic difference when comparing the two cases. The upward shift of the conduction band, which mimics the band-gap widening effect, leads to increase of the average electron energy of approximately 100 to 150 meV when the effective potential is included in the model. We observe the same trends in the average carrier velocity, shown in Fig. 15(b), as for the average electron energy. Namely, velocity overshoot effect is quite evident in the right portion of the p^+-implant region due to the large in-plane electric fields that give rise to fast acceleration of the carriers. Once the carriers enter the low-field region (the middle and drain end of the channel), the average electron velocity initially

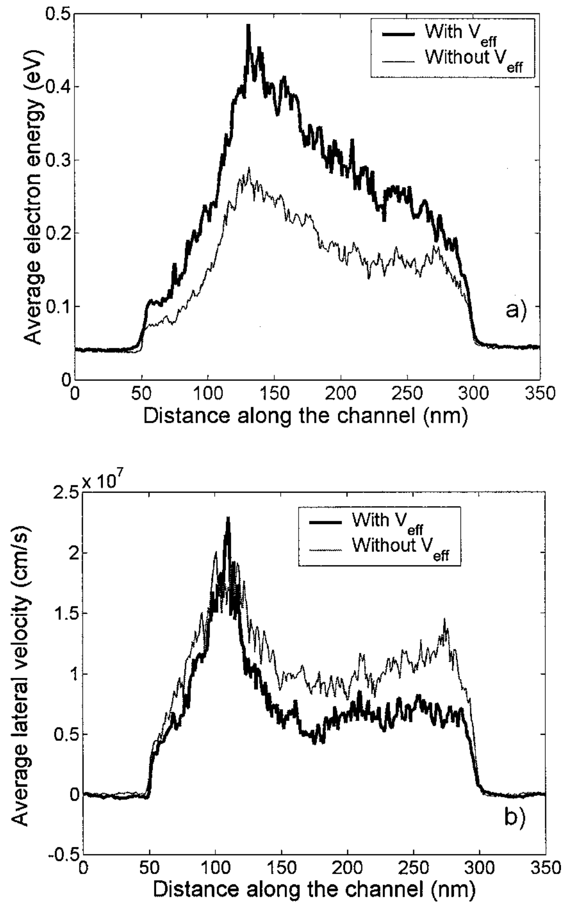

Figure 15. (a) Average electron energy and (b) average electron velocity for $V_G = V_D = 1$ V.

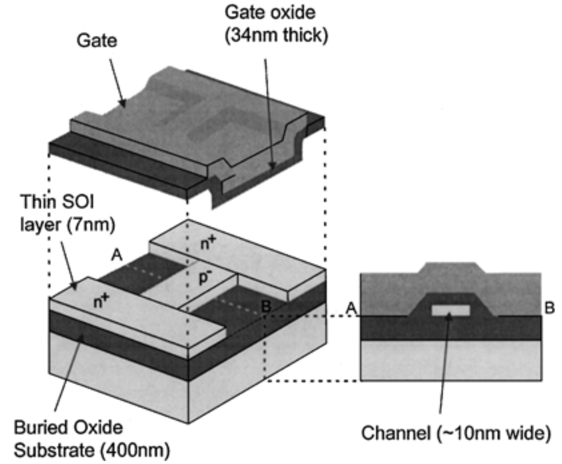

Figure 16. Schematic description of the SOI device structure.

decreases and then exhibits a very slight increase. Also noticeable is the fact that the inclusion of V_{eff} does not influence the magnitude of the velocity overshoot, but has significant influence on the velocity at the drain-end of the channel. There, a reduction in the velocity is observed because of the in-plane electric field reduction due to the lateral smearing of the potential energy profile. The above effects are consistent with the drain current reduction discussed in conjunction with Fig. 14.

6. Simulations of Narrow-Width SOI Device Structure

The last device structure we simulate is a narrow channel silicon-on-insulator (SOI) device (see Fig. 16) that consist of a thick silicon substrate, on top of which 400 nm of buried oxide is grown. The thickness of the SOI layer is 7 nm, with p^- region width between 5 and 15 nm. On top of the SOI layer sits the gate-oxide layer, the thickness of which is 34 nm.

When modeling this device structure we employ two different methods:

- The effective potential approach in conjunction with the self-consistent solution of the 2D Schrödinger equation, to calculate the line electron density.
- Use of the classical solution of the electrostatic confinement combined with the Landauer approach to calculate the on-state current as a function of the gate voltage.

The description of these two methods and the discussion of the simulation results obtained is given below.

6.1. *Effective Potential Approach Versus Self-Consistent Solution of the 2D Schrödinger Equation*

When solving the 2D Schrödinger equation, we have taken into account the pronounced mass anisotropy in the Si material system and the multi-valley nature of the lowest conduction bands. Namely, the six conduction band valleys in Si are included through a standard three-valley model. Valley pair 1 points along the (100) direction having $m_x = m_l = 0.91\ m_0$ and $m_y = m_z = m_t = 0.19\ m_0$. Valley pair 2 points towards the (010) direction and has $m_x = m_z = m_t$ and $m_y = m_l$, and valley pair 3 points in the (001) direction, having $m_x = m_y = m_t$ and $m_z = m_l$. As a result of the above, at each iteration step, we solve the 2D

Schrödinger equation, of the form

$$\left[-\frac{\hbar^2}{2m_y^\nu}\frac{\partial^2}{\partial y^2}-\frac{\hbar^2}{2m_z^\nu}\frac{\partial^2}{\partial z^2}+V(y,z)\right]\psi_j^\nu(y,z) = E_j^\nu\psi_j^\nu(y,z) \tag{5}$$

three times, i.e. for each equivalent valley pair ν. Once the energy eigenstates and the corresponding eigenfunctions are known, the electron density is found by using

$$n(y,z) = 2\sum_{\nu=1}^{3}\sum_j N_j^\nu\left|\psi_j^\nu(y,z)\right|^2 \tag{6}$$

where the factor of 2 accounts for valley degeneracy, the double sum represents summation over all energy eigenstates (index j) belonging to each of the three valley pairs (index ν) and the line charge density is given by

$$N_j^\nu = \frac{1}{\pi\hbar}\sqrt{2m_x^\nu k_B T}\cdot F_{-1/2}\left(\frac{E_F-E_j^\nu}{k_B T}\right), \tag{7}$$

where T is the temperature and k_B is the Boltzmann constant. In the actual evaluation of the Fermi-Dirac integral of order $-1/2$, which appears in Eq. (7), we use the approximate expression given in Abramo *et al.* (2000).

The calculated gate-voltage dependence of the line density, for the test device structure with homogeneous confinement along the x-axis, is shown in Fig. 17. For each wire width (7, 10 and 15 nm) we use both the effective potential approach discussed in the introduction part of this article, and the self-consistent solution of the

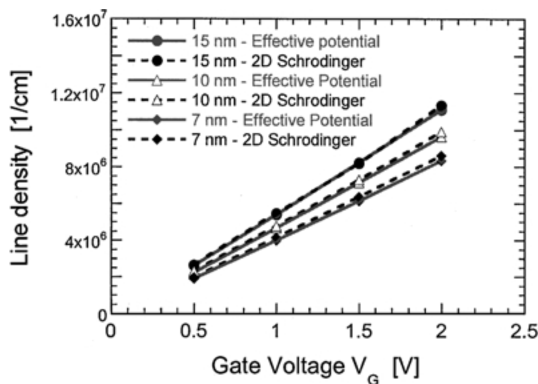

Figure 17. Variation of the line charge density for a quantum wire that represents the channel region of the SOI device structure from Fig. 16. The wire width equals 7, 10 and 15 nm.

2D Schrödinger-3D Poisson problem. Excellent agreement is observed between the two approaches when using the theoretical value for the Gaussian smoothing parameter of 0.64 nm. This result suggests that the effective potential approach can be successfully used even for complicated confining potentials. Also note that slightly lower values of the Gaussian smoothing parameter were needed when simulating triangular potential well confinements that occur in conventional MOSFETs and FIBMOS devices.

6.2. *Classical Solution Combined with Landauer's Formalism*

Another approach to modeling the SOI device is to think of it as quantum wire. This viewpoint is particularly useful in the ballistic regime, where most of the scattering comes from the device boundaries. As channel lengths become shorter, it becomes increasingly more appropriate to model devices in this manner. A quantum wire is essentially a waveguide for propagating electron waves. Depending on the electron density and the width of the wire, only a certain number of quantized modes are allowed to propagate. The amount of current that is passed by the device then depends on the transmission probability of these modes. According to the formalism originally developed by Landauer (1970) and extended by Büttiker (1988), the source-drain current, I_D, can be expressed as the integral

$$I_D = \frac{2e}{h}\int (f(E)-f(E-eV_{DS})) \times \sum_{nm}|t_{nm}(E,V_{DS},V_G)|^2 dE, \tag{8}$$

where V_{DS} is the voltage drop from source to drain, $f(E)$ is the Fermi function for energy E, and t_{nm} is the transmission amplitude going from mode n to mode m, and the summation is over all propagating modes. Thus, obtaining the current reduces to computing these quantum mechanical transmission amplitudes. There are a number of different ways for doing this, but a method that we have used with great success is that of Usuki *et al.* (1995), who developed an approach based on numerically stabilized variant of the transfer matrix method. To begin, the Schrödinger equation is mapped onto a finite-difference mesh, on a square lattice of lattice constant a. Since the wires are of finite width, extending a given number (M) of lattice sites across, one can work in terms of slices, where ψ_j is an

M-dimensional vector containing the wave function amplitudes of the jth slice. The discretized Schrödinger equation, keeping terms up to first order in the approximation of the derivative, has the form:

$$(E - \mathbf{H}_j)\psi_j + \mathbf{H}_{j,j-1}\psi_{j-1} + \mathbf{H}_{j,j+1}\psi_{j+1} = 0. \quad (9)$$

In the above equation, the $\mathbf{H}_j$ matrices represent Hamiltonians for individual slices, and the matrices $\mathbf{H}_{j,j-1}$ and $\mathbf{H}_{j,j+1}$ give the inter-slice coupling. By approximating the derivative by a finite difference, the kinetic energy terms of Schrödinger's equation get mapped onto a tight-binding model with $t = -\hbar^2/2m * a^2$ representing nearest neighbor hopping. The potential V at site (i, j) simply adds to the on-site energies, which appear along the diagonal of the $\mathbf{H}_j$ matrices. Transfer matrices, that allow translation from source to drain, can be derived using Eq. (9). These allow for the modes of the wire to be determined. Setting the boundary condition that these modes be occupied at the source end with unit amplitude, one obtains the transmission amplitudes that enter Eq. (8) by translating across the system. By using some clever matrix manipulations, Usuki overcame the problems created by the exponentially growing and decaying contributions of evanescent modes. Rather than just multiplying transfer matrices, the translation scheme is turned into an iterative procedure, which provides numerical stability by not allowing the evanescent contributions to diverge.

Simulation results for the device transfer characteristics, that utilize the above described approach, are shown in Fig. 18. The channel width is equal to 5, 7, 10 and 13 nm. We observe a drastic decrease of the drain

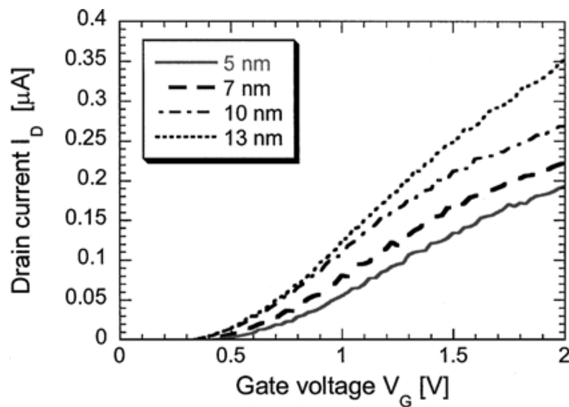

Figure 18. Transfer characteristics of the SOI device with channel width equal to 5, 7, 10 and 13 nm. The drain voltage V_D equals 10 mV.

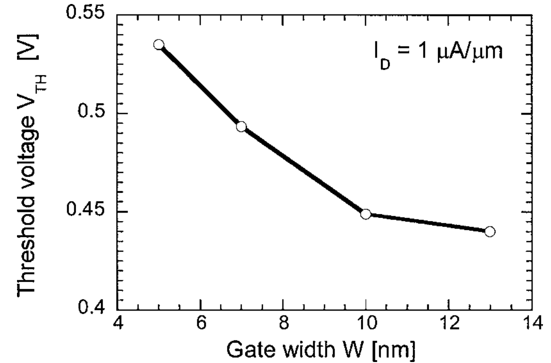

Figure 19. Threshold voltage as a function of the gate width.

current with decreasing channel width due to the lateral space-quantization effect. This observation is seen more clearly from the results shown in Fig. 19, where we plot the device threshold voltage as a function of the channel width. Note that the threshold voltage equals the gate voltage for which the current equals 1 μA/μm.

7. Conclusions

In conclusion, we have investigated the role played by quantum-mechanical space-quantization effects on the operation of a 50 nm MOSFET device, a 250 nm FIBMOS device and a narrow-width SOI device structure. In all three cases we find significant shift in the threshold voltage. For the MOSFET device the threshold voltage shift is on the order of 80 mV, whereas for the FIBMOS device we find threshold voltage shift of about 200 mV, mainly due to the very high doping of the FIB implant and its importance on the overall device performance. Also, an increase in the average electron energy is observed, which is consistent with the decrease in the average drift velocity at the drain end of the channel. The shift in the threshold voltage gives rise to on-state current reduction between 20% and 40%, depending upon the gate bias. For the narrow-width SOI device structure significant dependence of the threshold voltage is observed upon the channel width. The results presented in this paper for all the structures being considered suggest that it is crucial to include quantization effects into existing device simulators when trying to predict the performance of state-of-the-art devices that exhibit very high substrate doping densities. However, further investigations are needed to verify the appropriateness of the approach by comparing it with similar approaches discussed in the introduction part of this article.

Acknowledgments

The authors greatly appreciate the discussions with Prof. Stephen M. Goodnick and Massimo Fischetti during the preparation of this manuscript. This work is supported in part by the Office of Naval Research under Contract No. N000149910318 and the National Science Foundation under NSF-CAREER ECS-9875051.

References

Abramo A., Cardin A., Selmi L., and Sangiorgi E. 2000. IEEE Trans. Electron Dev. 47: 1858.
Akis R., Milicic S., Ferry D.K., and Vasileska D. 2001. In Proceedings of the 4th International Conference on Modeling and Simulation of Microsystems. Computational Publications, Boston, p. 550.
Asenov A., Slavcheva G., Brown A.R., Davies J.H., Saini S. 2001. IEEE Trans. Electron Dev. 48: 722.
Assad F., Ren Z., Vasileska D., Datta S., and Lundstrom M. 2000. IEEE Trans. Electron Dev. 44: 232.
Bohm D. 1952a. Phys. Rev. 85: 166.
Bohm D. 1952b. Phys. Rev. 85: 180.
Buti T.N., Ogura S., Rovedo N., and Tobimatsu K. 1991. IEEE Trans. Electron Devices 38: 1757.
Buttiker M. 1988. IBM J. Res. Develop. 32: 63.
de Broglie L. 1926. C. R. Acad. Sci. Paris 183: 447.
de Broglie L. 1927. C. R. Acad. Sci. Paris 184: 273.
Demkov A.A., Zhang X.D., Loechelt H. 2001. VLSI Design 13: 135.
Dewdney C. and Hiley B.J. 1982. Found. Phys. 12: 27.
Duncan A., Ravaioli U., and Jakumeit J. 1998. IEEE Trans. Electron Dev. 45: 867.
Ferry D.K. 1976. Phys. Rev. B 14: 1605.
Ferry D.K. 2000. Superlattices and Microstructures 27: 61.
Feynman P. and Kleinert H. 1986. Phys. Rev. A 34: 5080.
Formicone G.F., Saraniti M., Vasileska D.Z., and Ferry D.K., IEEE Trans. Electron Dev. 49: 125.
Gardner C.L. 1994. SIAM J. Appl. Math. 54: 409.
Gross W.J., Vasileska D., and Ferry D.K. 2000. VLSI Design 10: 437.
Herring C. and Vogt E. 1956. Phys. Rev. 101: 944.
Hockney R.W. and Eastwood J.W. 1981. Computer Simulation Using Particles. McGraw-Hill, Maidenhead.
Jacoboni C. and Reggiani L. 1983. Rev. Modern Phys. 55: 645.
Kan E.C., Narayan V., and Pei G. 2002. Journal of Computational Electronics, in press.
Knezevic I., Vasileska D., Akis R., Kang J., He X., and Schroder D. K. 2002a. Physica B 314: 386.
Knezevic I., Vasileska D., and Ferry D.K. 2002. IEEE Trans. Electron. Dev. 49: 1019, in press.
Knezevic I., Vasileska D., He X., Schroder D.K., and Ferry D.K. 2002b. Journal of Computational Electronics, in press.
Landauer R. 1957, 1970. IBM J. Res. Develop 1: 233; Philos. Mag. 21: 863.
Laux S.E. 1996. IEEE Trans. CAD Integr. Circ. Syst. 15: 1266.
Ma J., Liang H.-B., Pryor R.A., Cheng S., Kaneshiro M.H., Kyono C.S., and Papworth K. 1997. IEEE Trans. Very Large Scale Integration (VLSI) Systems 5: 352.
Madelung E. 1926. Z. Phys. 40: 322.
Saraniti M., Tang J., Goodnick S., and Wigger S. 2002. Journal of Computational Electronics, in press.
Städele M. Tuttle B.R., and Hess K. 2001. J. Appl. Phys. 89: 348.
Städele M. *et al*. 2000. Superlatt. Microst. 27: 405.
Takeda H., Mori N., and Hamaguchi C. 2002. Journal of Computational Electronic, in press.
Tsuchiya H., Winstead B., and Ravaioli U. 2001. VLSI Design 13: 335.
Vasileska D. and Ferry D.K. 1998. In Technical Proceedings of the First International Conference on Modeling and Simulation of Microsystems, Semiconductors, Sensors and Actuators. Computational Publications, Boston, p. 408.
Vasileska D., Gross W.J., and Ferry D.K. 1998. In Extended Abstracts of 1998 Sixth International Workshop on Computational Electronics (IWCE-6), Osaka, Japan, 19–21 Oct. 1998, pp. 259–262.
Vasileska D., Schroder D.K., and Ferry D.K. 1997. IEEE Trans. Electron Dev. 44: 584.
Vignaud D., Etchin S., Liao K.S., Musil C.R., Antoniadis D.A., and Melngailis J. 1992. Appl. Phys. Lett. 60: 2267.
Watling J.R., Brown A.R., and Asenov A. 2002. Journal of Computational Electronics, in press.
Usuki T., Saito M., Takatsu M., Kiehl R.A., and Yokoyama N., Phys. Rev. B 52: 8244.