nanoMOS 2.0:
A Two -Dimensional Simulator for Quantum Transport in Double-Gate MOSFETs

Zhibin Ren, Ramesh Venugopal, Sebastien Goasguen,
Supriyo Datta, and Mark S. Lundstrom
Purdue University
1285 Electrical Engineering Building
West Lafayette, IN 47907-1285

ABSTRACT

A program to numerically simulate quantum transport in double gate MOSFETs is described. The program uses a Green's function approach and a simple treatment of scattering based on the idea of so-called Büttiker probes. The double gate device geometry permits an efficient mode space approach that dramatically lowers the computational burden and permits use as a design tool. Also implemented for comparison are a ballistic solution of the Boltzmann Transport Equation and the drift-diffusion approaches. The program is described and some examples of the use of nanoMOS for 10nm double gate MOSFETs are presented.

I. INTRODUCION

As MOSFET channel lengths rapidly shrink below 100nm, current research focuses on understanding device physics and ultimate scaling limits as well as the practical issues that need to be addressed to achieve channel lengths at the 10nm scale [Fra98, Kno02, Ren00, Ren01]. Computational studies can help address these issues, but the methods commonly used in computer-aided design tools do not include (or do so only phenomenologically) important quantum mechanical effects. What is needed is a full quantum transport model – one that describes quantum phenomena (especially confinement and tunneling) as well as phase randomizing scattering. Computational efficiency is necessary to permit use for engineering design and to explore numerous design options. It would be useful for such a program to be based on an approach that can be extended to non-conventional devices, such as carbon nanotube FETs [Bac01, Der01, McE02] and other molecular transistors. In this paper we describe such a computer simulation tool, nanoMOS2.0.

The program, nanoMOS, uses the non-equilibrium Green's function (NEGF) approach, which provides a rigorous description of quantum transport and interactions that randomize phase [Lak92]. The method has been developed into a practical tool for simulating one-dimensional resonant tunneling diodes [Jpl02], and more recently, it has been extended to two-dimensional bulk MOSFETs [Jov00, Svi00, Svi01]. It is also now widely used to explore conduction at the mesoscopic [Dat97] and molecular scales [Tia98]. In comparison to phenomenological quantum approaches [Raf98, Shi00], the Green's function method offers the advantage of rigor, but the cost can be a heavy computational burden. The simple geometry of the fully-depleted SOI MOSFET (single or double gate) reduces this computational burden. It also permits the use of a mode space approach that, while not suitable for bulk MOSFETs, dramatically lowers the computational time for ultra-thin-body SOI MOSFETs [Ven02].

For ballistic transport, the NEGF formalism is equivalent to solving the Schrödinger equation, but the NEGF formalism is readily extendible to two-and three-dimensions or to the use of molecular orbitals. It also provides a clear prescription for treating large contacts attached to an intrinsic device. The NEGF formalism offers a sound conceptual basis upon which additional physics (such as scattering) can be included as needed. See [Dat00, Pau02] for a tutorial introduction to the NEGF method and [Dam02] for an example of its application to conduction in molecules.

This paper describes nanoMOS2.0, a simulation program that uses the NEGF approach to simulate fully-depleted, nanoscale, SOI MOSFETs. It begins in Sec. II with a description of the methods focusing on the mode space approach. Section III discusses some questions relating to the numerical implementation. In Sec. IV, we describe a phenomenological treatment of scattering, and in Sec. V some results are presented.

II. APPROACH

In this section, we describe the approach used in nanoMOS from a wavefunction perspective, because ultra-thin body DG MOSFETs can be treated by a 1D approach, and for 1D quantum transport with the simple scattering model currently implemented in nanoMOS, the wavefunction and NEGF approaches are equivalent. We use a wavefunction description because it is more familiar to readers of this journal, but in the appendix, we translate the key equations into the

language of the NEGF formalism. Readers are referred to a tutorial introduction [Dat00] or to standard references [Kel65, Lan76] for a more extensive discussion of the NEGF approach.

Figure 1 sketches the double gate MOSFET geometry that we assume; we seek a solution for the intrinsic device only. Because we assume that the device is wide in the y-direction, the wavefunction can be written as $\psi(x,y,z) = \Phi(x,z)e^{ik_y y}/\sqrt{W}$, where $\Phi(x,z)$ is obtained from

$$-\frac{\hbar^2}{2m_x^*}\frac{\partial^2\Phi(x,z)}{\partial x^2} - \frac{\hbar^2}{2m_z^*}\frac{\partial^2\Phi(x,z)}{\partial z^2} + E_C(x,z)\Phi(x,z) = \left(E - E_{k_y}\right)\Phi(x,z) = E_\ell\Phi(x,z), \tag{1}$$

where E_ℓ is the longitudinal energy. One could directly discretize eq. (1) in real space, which is often necessary, but it imposes a heavy computational burden [Ven02]. For thin body, SOI MOSFETs, however, quantum confinement in the z-direction introduces subbands, and for a thin body, only a few subbands are occupied. Accordingly, we expand the wavefunction in an orthonormal basis as [Ven02],

$$\Phi(x,z) = \sum_m \delta(x - x')\Psi_m(x',z)\phi_m(x'), \tag{2}$$

where $\Psi_m(x,z)$ are the eigenfunctions (modes) associated with confinement in the z-direction as defined in Fig. 1. These eigenfunctions and the associated eigenenergies are obtained by solving a one-dimensional wave equation in the z-direction at each vertical slice at x,

$$-\frac{\hbar^2}{2m_z^*}\frac{\partial^2\Psi_m(x,z)}{\partial z^2} + E_C(x,z)\Psi_m(x,z) = E_m(x)\Psi_m(x,z) \tag{3}$$

The energy, $E_m(x)$, represents the bottom of subband, m, which varies with position, x, along the channel. The envelope wavefunctions are assumed to be zero at the oxide/Si interfaces if electron penetration into the oxide regions is neglected (otherwise, the zero boundary is extended to the gate contact/oxide interfaces).

By using the orthonormal basis of eq. (2), we can transform eq. (1) to a mode-space basis. By retaining only a few occupied modes, the computational burden can be significantly reduced compared to a direct discretization in a two-dimensional real space [Dam00]. The geometry of the double gate MOSFET offers yet another simplification. If we assume that the shape of the confined mode does not change along *x*, i.e. $\partial\Psi_m(x,z)/\partial x \approx 0$
then eq. (2) becomes [Ren01a, Ven02]

$$-\frac{\hbar^2}{2m_x^*}\frac{\partial^2\phi_m(x)}{\partial x^2} + E_m(x)\,\phi_m(x) = \left(E - E_{k_y}\right)\phi_m(x) = E_\ell\phi_m(x)\,, \tag{4}$$

where $\phi_m(x)$ is the expansion coefficient of $\Phi(x,z)$ with respect to the mode-space eigenvector, $\Psi_m(x,z)$.

Equation (4) is the decoupled mode space transformation of eq. (1); the key assumption is that the shape of the mode does not change rapidly along *x*. Although the potential, $E_C(x,z)$, varies from source to drain, the largest variation is with position, *x*. For a thin body, the shape of the mode (i.e. its *z*-dependence) changes slowly along the channel. A careful comparison of the decoupled approach with a direct, real-space discretization shows that the decoupled approach produces results that are essentially identical to the exact solution for body thicknesses of up to at least 5 nm [Ven02]. There are, however, conditions for which this the decoupled approach fails [Ven02]. One example is the transition from the flared out contact to the ultra-thin channel (which is not treated in nanoMOS), and another is the bulk MOSFET.

Equation (4) is a 1D wave equation that greatly reduces the size of the 2D problem. In the ballistic limit, each mode (subband) can be treated separately. Each transverse mode is also independent, because we assumed there is no potential variation in the *y*-direction. In this 1D equation, instead of the bottom of the conduction band, $E_C(x,z)$, we have the bottom of the relevant subband, $E_m(x)$, which is determined by a solution to eq. (3) in the confinement direction. (Accordingly, when we plot energy band diagrams, we will plot $E_m(x)$ rather than the bottom of the conduction band, $E_C(x,z)$). We also see from eq. (4) that the relevant energy for the 1D problem is not the total energy, *E*, but the longitudinal energy, E_l.

Because the decoupled mode space solution reduces the problem to a set of 1D problems, one for each subband and transverse energy, the solution procedure is much like that of a true 1D problem. Figure 2 is a sketch of the subband energy versus position for one subband. Semi-infinite contacts are attached to the device at the source and drain ends. Because the potential in the contacts is assumed to be uniform, the solutions in the semi-infinite contacts are plane waves. If a unit amplitude wave is injected from the left (source) contact, then some portion reflects from the device and some transmits across and exits the perfectly absorbing right (drain) contact,

$$\phi_m(x) = \frac{1}{\sqrt{L}} e^{ik_x x} + \frac{r_m}{\sqrt{L}} e^{-ik_x x} \qquad x < 0 \tag{5a}$$

and

$$\phi_m(x) = t_m e^{ik_x x}, \qquad x > L \tag{5b}$$

where L is an arbitrary normalization length.

By solving eq. (4) subject to the boundary conditions, eqns. (5a) and (5b), we find the wavefunction due to the injection of a unit amplitude wave from the source. The corresponding electron density vs. position for confined mode, m, is obtained by summing the contribution for each transverse mode,

$$n_1^m(x) = \frac{1}{W} \sum_{k_y} \left\{ \frac{1}{L} \sum_{k_x} \left| \phi_m(x, k_x) \right|^2 f(E_F - E) \right\}, \tag{6}$$

where k_x refers to the x-component of the wavevector of an electron with energy, E, *in the source contact* and the subscript, 1, refers to injection from the first (left or source) contact. The probability that the state at energy, E, in the contact is occupied is give by a Fermi function because we assume that scattering maintains thermal equilibrium in the contacts.

Since the width in the transverse direction, W, is assumed to be large, we can convert the sum over transverse states to an integral over the transverse energy. We can also convert the sum over the injected k_x states to an integral over longitudinal energy. The integral over transverse energy can be done analytically and the final result expressed as [Ren01a]

$$n_1^m(x) = \int_{-\infty}^{\infty} n_1^m(x, E_\ell) dE_\ell \,, \tag{7a}$$

where we let the top of the band approach infinity and

$$n_1^m(x, E_\ell) = \frac{1}{\hbar a}\sqrt{\frac{2m_y^* k_B T}{\pi}} \mathcal{F}_{-1/2}(E_F - E_\ell) \frac{A_1^m(E_\ell, x)}{2\pi} = F(E_F - E_\ell) \frac{A_1^m(E_\ell, x)}{2\pi}, \tag{7b}$$

with

$$\frac{A_1^m(x, E_\ell)}{2\pi} \equiv \left[\frac{a}{\pi}\frac{dk_x}{dE}\left|\phi_m(x, k_x)\right|^2\right], \tag{8}$$

being the local density of states due to injection from the source (includes spin degeneracy). In eq. (7b) $\mathcal{F}_{-1/2}$ is the Fermi-Dirac integral of order –1/2 [Bla82] and F is the Fermi function integrated over the transverse energy. To find the total electron density within the device, we sum the contributions from each subband and valley and include the contribution due to injection from the drain,

$$n(x) = n_1(x) + n_2(x) = \sum_m \int_0^{\infty} \left[F(E_F - E_\ell) \frac{A_1^m(x, E_\ell)}{2\pi} + F(E_F - qV_D - E_\ell) \frac{A_2^m(x, E_\ell)}{2\pi} \right] dE_\ell \,, \tag{9}$$

where $A_2^m / 2\pi$ is the local density of states due to injection from the drain, which is computed from the wavefunction with boundary conditions analogous to eqs. (5a) and (5b).

To obtain the current due to injection from the source, we evaluate

$$I_m = \frac{q}{W}\sum_{k_y}\frac{1}{L}\sum_{k_x} T_m^{1-2}(E)\upsilon_x(E)\left[f(E_F - E) - f(E_F - E - qV_D)\right], \tag{10}$$

where T_m^{1-2} is the current transmission coefficient from contact 1 to contact 2 for electrons in subband, m. The sums can be converted to integrals and the integral over transverse energy performed analytically to find

$$I_m = \int_0^\infty I_m(E_\ell)dE_\ell \ , \tag{11a}$$

where

$$I_m(E_\ell) = \frac{2q}{h}T_m^{1-2}(E_\ell)\left[F(E_F - E_\ell) - F(E_F - E_\ell - qV_D)\right], \tag{11b}$$

and F is the Fermi function integrated over transverse energy. The total current is obtained by summing the contributions from each subband and valley.

Finally, the current transmission coefficient for mode, $m, T_m^{1-2}(E_\ell)$ is

$$T_m^{1-2}(E_\ell) = \frac{I_{trans}}{I_{inc}} = 1 - \left|r_m\right|^2, \tag{12}$$

which, from eq. (5a) can be expressed in terms of the computed wavefunction as

$$T_m^{1-2}(E_\ell) = 1 - \left|\phi_m(0) - 1\right|^2. \tag{13}$$

III. Numerical Solution

A. *Solving for the wavefunction*

To evaluate the expression for the electron density and current in the previous section, the wavefunction within the device must be known. In the confinement direction, eq. (3) is a standard eigenvalue problem solved by finite differences. In the longitudinal direction, we discretize eq. (4) on a finite difference grid imposing the boundary conditions, eqs. (5a) and (5b), to find

$$[E\mathbf{I} - \mathbf{H} - \Sigma_1 - \Sigma_2]\phi = i\gamma_1 , \tag{14}$$

where **H** is the N x N discretized Hamiltonian operator,

$$H_{i,j} = -t_0\delta_{i-1,j} + 2t_0\delta_{i,j} - t_0\delta_{i+1,j} , \tag{15}$$

with an on-site energy of

$$t_0 = \frac{\hbar^2}{2m_x^* a^2} , \tag{16}$$

for a finite difference grid with node spacing, *a*. The N x N "self-energy" matrices, Σ_1 and Σ_2 account for the open boundary conditions [eqns. (5a) and (5b)] and are

$$\Sigma_1(i,j) = -t_0 e^{ik_x x}\delta_{1,j} \tag{17a}$$

$$\Sigma_2(i,j) = -t_0 e^{ik_x x}\delta_{N,j} . \tag{17b}$$

The N x 1 vector, γ_1 is a source term accounting for injection from the left contact (source). It has only one non-zero component, the first

$$\gamma_1(1) = i\left[\Sigma_1(1,1) - \Sigma_1^*(1,1)\right] = 2t_0 \sin k_x a = \hbar\upsilon\left(k_x\right)/a. \qquad (18)$$

For injection from the drain, we use boundary conditions analogous to eqs. (5a) and (5b), and the corresponding γ_2 vector has a non-zero component in position, *N*.

The solution to eqn. (14) gives the value of the wavefunction, ϕ, at each of the *N* finite difference nodes. The formal solution is

$$\phi = i\mathbf{G}\gamma, \qquad (19)$$

where

$$\mathbf{G} = \left[E\mathbf{I} - \mathbf{H} - \Sigma_1 - \Sigma_2\right]^{-1} \qquad (20)$$

is the retarded Green's function in a discrete basis. Because the matrix in eq. (14) is tridiagonal, it can be efficiently solved by Gaussian elimination; the carrier density and current are then evaluated from the computed wavefunctions. Alternatively, as shown in the appendix, we can express all of the results in terms of the *N* x *N* retarded Green's function. For this simple, 1D problem, the Green's function approach may appear to be a complicated way to solve a simple problem. The advantage of the NEGF formalism become apparent when extensions to two and three-dimensions are contemplated, when an atomic basis is essential, or when a rigorous treatment of scattering is necessary.

B. *Poisson's equation*

The program, nanoMOS, computes a self-consistent solution to a quantum transport equation and Poisson's equation,

$$\nabla^2 V(x,z) = \frac{q}{\varepsilon}\left[n(x,z) - \left(N_D^+ - N_A^-\right)\right]. \tag{21}$$

Equation (7) gives the electron density per unit area within the device; to convert it to a density per unit volume, it is distributed in the *z*-direction according to the computed eigenfunction in the confinement direction. Direct use of eq. (21) leads to slow convergence. Instead, we solve a nonlinear Poisson equation. The electron density evaluated from the wavefunction can be related to a quasi-Fermi level by

$$n(x,z) \equiv N_{3D} F_{1/2}\left[F_n(x,z) - qV(x,z)\right], \tag{22}$$

where N_{3D} is the three-dimensional effective density-of-states, and the function, $F_{1/2}$, is the Fermi-Dirac integral of order 1/2. We regard eq. (22) as a mathematical change of variables to a quantity, F_n, whose physical significance under off-equilibrium conditions is unclear, but unimportant. When eq. (22) is inserted into eq. (21), a nonlinear Poisson equation results. After computing $n(x, z)$ from the quantum transport model, a corresponding $F_n(x, z)$ is computed from eq. (22). This "quasi-Fermi level" is used in a nonlinear Poisson equation that is obtained by inserting eq. (22) in eq. (21) and solved by Newton's method. The advantage of this approach is that it builds a negative feedback into the iterative process. If the potential increases (conduction band decreases) during the Poission solution, the subsequent transport solution will increase $n(x, z)$ as carriers flow to regions of lower energy. This coupling is built into the Poisson equation when eq. (22) is used. The approach has proven effective in previous quantum and semiclassical transport simulations [Cah87, Ven89] and proved similarly effective here.

C. Boundary conditions

Boundary conditions must be specified for both the transport equation and for Poisson's equation. For Poisson's equation, the boundary conditions at the gate electrodes (Dirichlet) and at the oxide/air interfaces (Neumann) are standard. Boundary conditions at the interface between the S/D extensions and the flared out contacts are also needed. A more complete solution would include the flared out region to explore the resistive drops that may occurs at the wide/narrow transition. In this work, we seek simple, upper limit boundary conditions that define our "ideal" contacts.

External to the intrinsic device being simulated, we assume large source and drain contacts where scattering maintains thermal equilibrium. We solve the wave equation assuming a unit amplitude injected wave, then weight by the Fermi function of the appropriate contact. The Fermi levels in the expressions presented in Sec. II refer to the Fermi levels in the equilibrium source and drain contacts that are external to the region being simulated. The self-energy matrices are expressed in terms of the wavevector, k_x, in the contact from which the electrons are injected. For a uniform contact with simple bands,

$$E(k) = \frac{\hbar^2 k^2}{2m^*}, \tag{23a}$$

but the discrete grid modifies this relation to

$$E(k) = 2t_0(1 - \cos ka), \tag{23b}$$

where a is the grid spacing. The two expressions are equal when $E << 4t_0$. Care must be taken to ensure that the maximum energy used is well below the top of the band so that eq. (23b) approximates eq. (23a). It should also be recognized that the maximum energy that can be safely used depends on the grid spacing, a.

For ballistic transport, the boundary conditions on Poisson's equation must be carefully specified. Figure 3 is a sketch of a generic subband profile from the source to the drain. Under

low gate bias and high drain bias, the source-to-channel barrier is high. Therefore most electrons injected from the source (empty circles in Fig. 3) reflect from the source to channel potential barrier, and both positive and negative velocity states are occupied in the n^+ source extension as a result of source injection. Under high gate and drain bias, however, fewer electrons are reflected from the barrier, and the electron density in the n^+ source extension decreases (there is very little contribution to the overall electron concentration at the source end due to drain side injection as the drain voltage is high). If the potential at the source contact is fixed, space-charge neutrality cannot be maintained. (Note that the location is the Fermi level is fixed by the large, thermal equilibrium source reservoir.) To maintain space-charge neutrality in the source/drain (S/D) extensions, we allow the contact potential to float by specifying

$$\frac{dV(x,z)}{dx} = 0 \tag{24}$$

at the idealized source and drain contacts. This boundary condition causes the potential at the source end to float down, thus allowing more carriers to enter into the device from the source reservoir. This ensures that space-charge neutrality is maintained at the source end irrespective of the biasing condition. The floating boundary condition eliminates the need to resolve the potential at the point where the thin body couples to the large S/D. It represents an ideal upper limit contact. In practice, the transition from the flared out S/D to the ultra-thin intrinsic device could introduce quantum mechanical parasitic resistances that are not treated within nanoMOS. The validity of this approach is supported by the results that will be shown in Sec. V.

D. *Solution Procedure*

The solution procedure is summarized in Fig. 4; it consists of the following steps:

1) An initial guess for the 2D electrostatic potential is defined. (We use a drift-diffusion simulation as our initial guess.)

2) For each position, x, along the channel, eq. (3) is solved to find the eigenfunctions and eigenenergies versus position. A grid spacing of a = 3Å in the x-direction is typical.

3) A grid in longitudinal energy is defined (with a typical grid spacing of 0.5 meV) and eq. (4) is solved to find the wavefunction due to injection of a unit amplitude wave at longitudinal energy, E_ℓ, from both the source and drain contacts. The electron density for each mode is evaluated from eq. (7b). This step is repeated for each longitudinal energy in the grid and for each occupied confined mode (as set by the user). The total carrier density is obtained by summing the contributions from each longitudinal energy, confined mode, and valley as in eq. (9).

4) The nonlinear Poisson equation is solved to update the electrostatic potential. The maximum change in potential is compared to the convergence criteria, and the process continues until convergence is achieved. The minimum change in potential that can be achieved is essentially the spacing in the energy grid.

5) After convergence is achieved, the transmission coefficient and current at each energy are evaluated from eq. (13). The contributions from each energy are summed to find the total current from eq. (11). The process is repeated for each mode, and the results summed to compute the total current.

Although we have described the solution process from a wavefunction perspective, the actual equations solved in nanoMOS are expressed in the Green's function formalism as described in the appendix.

IV. TREATMENT OF SCATTERING

The NEGF formalism provides a prescription for including scattering, but scattering greatly increases the computational burden because it couples the longitudinal and transverse energies. Instead of treating transverse modes as independent and integrating over them analytically, we need a grid in transverse energy too. And instead of treating each longitudinal energy independently, they must be coupled. Scattering has been treated rigorously in NEGF simulations of MOSFETs [Svi02, Ven02a], but the resulting computational burden limits the use of such simulations. The advantage of the NEGF approach lies in its ability to treat quantum

transport using a atomic level Hamiltonian, if necessary, but we also need a simple way to capture the main effects of scattering.

We treat scattering in a way that is analogous to the well-known relaxation time approximation to the collision operator for the Boltzmann equation,

$$\hat{C}f = -\frac{(f - f_0)}{\tau} = -\frac{f}{\tau} + \frac{f_0}{\tau}, \tag{25}$$

which describes carriers out-scattering from a state at a rate, $1/\tau$, and in-scattering from a thermal equilibrium distribution. (Under non-equilibrium conditions, the magnitude of the in-scattering rate is adjusted to preserve current continuity.) An analogous treatment for the wave equation can be implemented with an idea due to Büttiker [But86]. As shown in Fig. 5, we conceptually attach a floating contact to each node. Carriers are removed from the device and injected into the floating contact where they are thermalized and re-injected into the device.

In nanoMOS, out-scattering is treated phenomenologically by adding an imaginary potential to the Schrödinger equation [Dat00]

$$\Sigma_S(i,j) = i\eta, \tag{26}$$

which works if one sets η to zero below the band edge so that tunnel currents are not unphysically enhanced. When scattering is present, the retarded Green's function in eq. (20) becomes

$$\mathbf{G} = [E\mathbf{I} - \mathbf{H} - \Sigma_1 - \Sigma_2 - \Sigma_S]^{-1}. \tag{27}$$

Equation (26) does a reasonable job of describing scattering in MOSFETs if the scattering strength is calibrated to an appropriate mobility. The diagonal self-energy results because the scattering potential is assumed to be a δ-function at each node.

We have described the process of out-scattering electrons to a probe where they are thermalized. The next step is to re-inject them into the device. Following the prescription for filling states from the two real contacts, eq. (9), we have

$$n_m(x,E_\ell) = n_1(x) + n_2(x) + n_S(x)$$
$$= F(E_F - E_\ell)\frac{A_1^m(x,E_\ell)}{2\pi} + F(E_F - qV_D - E_\ell)\frac{A_2^m(x,E_\ell)}{2\pi} + F(E_P - E_\ell)\frac{A_s^m(x,E_\ell)}{2\pi}. \quad (28)$$

The spectral density, $A_S(x,E_\ell)$, gives the local density-of-states everywhere along the channel due to injection at from the probe located at $x = x_S$. We have assumed a single scatterer at x_S with a Fermi level, E_S, but in practice, probes are placed at each node and each one has a different Fermi level. The electron density at x_i due to injection from all reservoirs (the source and drain contacts as well as the floating probes used to introduce scattering) is

$$n_m(x_i, E_\ell) = \sum_n F(E_{Fn} - E_\ell)\frac{A_n(x_i, E_\ell)}{2\pi}, \quad (29)$$

where the sum runs over all the reservoirs, each with its own Fermi level. The source and drain Fermi levels are fixed by the applied source-drain bias; the Fermi level of each probe is determined by requiring that current is conserved. The current in subband, *m*, at reservoir, *k*, is given by an expression analogous to eq. (11b),

$$I_m^k(E_\ell) = \frac{2q}{h}\sum_l T_m^{k-l}(E_\ell)\left[F(E_{Fk} - E_\ell) - F(E_{Fl} - E_\ell)\right] \quad (30)$$

where the sum over *l* is over the remaining reservoirs. The transmission coefficient between any two reservoirs is computed in a manned analogous to eq. (12). The source and drain contacts carry current, but the Büttiker probes do not, so for Büttiker probe, *k*

$$\int I_m^k(E_\ell)dE_\ell = 0. \quad (31)$$

Equation (31) gives a set of constraints on the Fermi levels of each probe. This set of constraining equations is solved iteratively for the Fermi level of the probe [Ren01a]. The resulting position dependent Fermi level behaves in much the same way as the quasi-Fermi level in conventional semiconductor theory.

V. RESULTS

Several studies of nanoscale MOSFET device physics and design issues that make use of nanoMOS simulations have already been published [Ren00, Ren01a, Ren01b, Lun02]. Our purpose in this section is to present some simulation results that illustrate the capabilities of the program. The simulated device (recall Fig. 1) is an idealized structure for which we do not treat the flared-out S/D contacts that would be present in an actual device. The n-type source and drain regions are doped at 2 x 10^{20} cm^{-3}, the channel is intrinsic, and the S/D junctions are abrupt. No gate-to-S/D overlap is assumed. The oxide thickness is 1.0nm for both top and bottom gates, and the silicon film thickness is 3.0 nm. The gate workfunction was set to 4.22 eV in order to produce an off-current of ≈ 10μA/μm, consistent with the 2016 node of the International Technology Roadmap for Semiconductors [ITR01]. A power supply of V_{DD} = 0.4V was used.

Figure 6 is a plot of the first two subband minima vs. position in the on-state. Also shown is a plot of the conduction band minimum vs. position. Note that the conduction band varies in two dimensions (in the *x*-direction along the channel from the source to the drain as well in the *z*-direction across the thickness of the silicon body). The subband minima, however, are determined by solving the Schrödinger equation in the *z* direction, and they vary only in the *x*-direction. This uncoupled mode space approach is what allows us to treat transport in each subband with a 1D Green's function approach.

Figure 7, a plot of the electron density within the device under on-state conditions, shows the quantum confinement of carriers in the *z*-direction. The profile varies approximately as $\sin^2(z)$ in the S/D regions, which indicates that most electrons reside in the first subband (primed and unprimed). Classical Boltzmann and NEGF simulations are compared in Fig. 8, which shows the

integrated electron density (per cm^2) vs. position in the off-state (V_{GS} = 0, V_{DS} = 0.4V). As expected, quantum mechanical tunneling into the source-channel barrier increases the carrier density in the channel. A careful examination of the figure also reveals that the quantum mechanical carrier density just outside the channel is slightly reduced, an effect first observed by Willander [Wil98]. Figure 9 compares the Log(I_D) vs. V_{GS} characteristics from ballistic quantum and classical nanoMOS simulations. As expected, quantum mechanical tunneling increases the off-current, but note that it also decreases the on-current. As discussed by Ren [Ren00], this occurs because MOS electrostatics demands a certain charge at the top of the barrier, but for the quantum simulation, some of the charge arises from carrier tunneling under the barrier, which reduces their velocity. The classical and quantum mechanical ballistic common source characteristics are compared in Fig. 10a. Note that the term classical or quantum refers to the treatment of transport along the channel (the *x*-direction). Because we use a mode space approach for the *z*-direction, the effect of quantum confinement on the threshold voltage is included in both simulations. Figure 10a shows that even for a 10nm channel length, MOSFETs are expected to behave classically. Quantum mechanics increases the threshold voltage and decreases the on-current at a given threshold voltage, but no quantum oscillations are observed. In Fig 10b., we compare the common source characteristics with and without scattering. The Büttiker probe strength has been adjusted to obtain a low field mobility of 50 cm^2/V-s in the heavily doped source drain regions. In the channel, a doping dependent mobility model is used to adjust the probe self-energy. It is clear from Fig 10b that the on-current is significantly reduced due to scattering in the heavily doped S/D extensions.

The ballistic on-state source to drain transmission coefficient vs. energy is plotted in Fig. 11. Recall that there is a set of ellipsoids with their longitudinal effective mass oriented along the confinement direction, which gives rise to the so-called unprimed series of subbands. Two more sets of ellipsoids have their transverse effective masses oriented along the confinement direction and give rise to the primed series of subbands. The location of the first unprimed subband minimum is indicated with a dashed line in Fig. 11, and the location of the first primed subband minimum is indicated with a dash-dot line. The transmission coefficient begins to increase below the first subband because of tunneling then smoothly approached unity (for each of the two valleys) above the subband minimum. For higher energies, the transmission coefficient

increases again. At high energies, the transmission coefficient approaches three, because it is one for each of the three sets of valleys. (valley degeneracy is not accounted for in Fig. 11)

It is instructive to look at the floating boundary conditions more closely. To explain the floating boundary condition, we modify our idealized 3nm body DG MOSFET structure by appending heavily doped (5 x 10^{20} cm^{-3}) n^{++} regions to the left and right ends of the device. The profile of the first subband from the source to the drain, for the modified device structure under a high drain bias and for a high and low gate bias is (V_{GS} = 0.6 and 0V, V_{DS} = 0.6V) is shown in Fig. 12a. The presence of the heavily doped n^{++} regions, creates a large potential barrier for electrons injected from the source and drain contacts. This barrier gives rise to strong reflections, which maintain a near equilibrium distribution in the n^{++} regions even when a large bias is applied to the drain and gate. Therefore, a fixed potential boundary condition based on charge neutrality can be used when solving Poisson's equation for this modified device. The n^{++} regions maintain a near equilibrium distribution even under bias.

On examining the subband profile (Fig. 12a) of the modified device at high gate bias, it is clear that the subband is unchanged in the heavily doped n^{++} regions, but floats to a lower value in the n^{+} source region of the intrinsic device. This observation can be explained by examining Fig 3. At low gate biases, both the positive and the negative halves of the distribution in the n^{+} source are predominantly filled by the source Fermi level. When the gate bias is increased to higher and higher values, the number of source injected electrons reflected off the source-to-channel barrier is reduced because the source-to-channel barrier height decreases. Although nearly one half of the distribution is unoccupied at the source, 2D electrostatics requires that charge neutrality has to be maintained (integrated doping equals the 2D hole density) at every point within the source. To achieve charge neutrality, the electron density residing in the positive velocity states nearly doubles between the low and high gate bias conditions. To accomodate this increased charge the subband (potential) floats to a lower energy as seen in Fig. 12a (dotted line).

Carriers injected from the heavily doped n^{++} region are predominantly backscattered by the built in barrier, so the potential in this region is unchanged with gate bias. Now, if we remove the n^{++} regions but use the floating boundary condition to terminate the n^{+} regions, we observe

the potential behavior plotted in Fig. 12b (solid line). Figure 12b, demonstrates that the subband under bias for the device with floating boundaries (n^+-i-n^+) is identical to that of the device with fixed boundaries (n^{++}-n^+-i-n^+-n^{++}) within the region of interest (n^+-i-n^+). This indicates that the floating boundary condition does capture the effect of coupling a ballistic device to a scattering contact.

The energy-resolved local density of states (LDOS) under ballistic on-state conditions is plotted in Fig. 13a. The states below the barrier are due to tunneling, and the strong influence of quantum mechanical reflections is apparent. The LDOS for the higher (primed) subband is larger (darker in the grayscale plot) because of the four-fold degeneracy of the conduction band valleys. Figure 13b shows the same plot in the presence of scattering, which is seen to reduce the contrast of the interference pattern. Finally, Fig. 14 shows the energy resolved ballistic electron density which results in filling up the LDOS according to the Fermi levels in the source and drain contacts. Carrier tunneling below the source barrier is observed as well as oscillations in the source and drain, which are due to reflections from the barrier.

VI. DISCUSSION

Although the nanoMOS program has proven to be a useful tool for exploring device physics and design, it is necessarily based on a number of simplifying assumptions. The mode space approach is particularly efficient for fully-depleted (FD), ultra-thin-body (UTB), SOI MOSFETs, for which only a few, uncoupled modes need to be treated. In a bulk MOSFET under high drain bias, however, confinement is lost near the drain and numerous modes much be treated This makes a real-space discretization more suitable [Jov00, Svi01]. To treat the flared our contact in a FD UTB SOI MOSFET, mode coupling must also be included. The transition from the thick contact to the thin body could introduce resistance. For this problem one could use a coupled mode space approach [Dam02b] or a real-space discretization.

The treatment of scattering in nanoMOS is a simple, phenomenological one. A well-defined prescription for including scattering exists, but the computational burden rapidly gets out of hand if one resolves scattering between transverse modes [Ven02a]. When using a phenomenological

approach, one must be careful to mimic the correct physics. Our use of Büttiker probes that relax the energy of scattered carriers mimics the relaxation of longitudinal energy that occurs in MOSFETs [Lun02]. The strength of the NEGF approach lies in its ability to treat quantum confinement, reflection, and tunneling and our phenomenological approach provides nanoMOS with a way to include the first order effects of scattering. Monte Carlo simulations can include a much more detailed treatment of scattering, but it is necessary to include quantum effects phenomenologically [Tsu01, Shi00]. The two approaches complement each other; the NEGF approach is preferable when quantum transport is the first order issue and Monte Carlo simulation when scattering is the key issue.

The nanoMOS program was written in a scripting language (Matlab®) to permit rapid development and debugging. Since much of the computation occurs in compiled matrix routines, the performance penalty is slight. For larger problems, toolboxes to parallelize the scripts are being developed [Fal02]. When the time comes for a production CAD program, nanoMOS may have to be re-written in C or Fortran, but the scripting language approach greatly facilitates program development and provides adequate performance for use as a research tool.

VII. Conclusion

The program, nanoMOS 2.0, simulates quantum transport in fully-depleted, ultra-thin-body SOI MOSFETs with an efficiency that permits its use on a workstation. We have described the methods and approaches used in the program as well as the simplifying approximations that deliver its computational efficiency. The NEGF approach used in nanoMOS provides a solid base for simulating electronic devices at the nanoscale which can be extended to increase the fidelity of the physics (e.g. improving the bandstructure and the treatment of scattering) or to treat much different devices such as carbon nanotube transistors or molecular conductors. The nanoMOS program is a step towards a new generation of simulation tools that will allow device engineers to explore new classes of electronic devices. The program is available for use through the WWW or for access to its source code [nan02].

Acknowledgment—This work was supported by the Semiconductor Research Corporation under contract number NJ-99-724 and in part by an ARO Defense University Research Initiatives in Nanotechnology (DURINT) grant administered by D. Woolard.

REFERENCES

[Ase02] A. Asenov, A.R. Brown, and J.R. Watling, "The use of Quantum Potentials for Confinement in Semiconductor Devices," 5th Intern. Conf. on Modeling and Simulation of Microsystems, Tech. Proc., pp. 490-493, San Juan, Puerto Rico, April 21-25, 2002.

[Bac01] A. Bachtold, P. Hadley, and C. Dekker, "Logic circuits with carbon nanotube transistors," *Science* 294, 1317, 2001

[Bla82] J.S. Blakemore, "Approximations for Fermi-Dirac Integrals, Especially the Functions $F_{1/2}(\eta)$ to Describe Electron Density in a Semiconductor," *Solid-State Electron.*, vol. 25, p. 1067, 1982.

[But86] M. Büttiker, "Four-Terminal Phase Coherent Conductance," *Phys. Rev. Lett.*, vol. 57, p. 1761, 1986.

[Cah87] M. Cahay, M. McLennan, S. Datta, and M.S. Lundstrom, "Importance of Space-Charge Effects in Resonant Tunneling Devices," *Appl. Phys. Lett.*, vol. 50, p. 612, March 9, 1987.

[Dat97] S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge Univ. Press, Cambridge, U.K., 1997.

[Dam02] P. Damle, A. W. Ghosh and S. Datta, "First principles analysis of molecular conduction using quantum chemistry software," Chem. Phys., vol. 281, p. 171, 2002.

[Dam02b] P. Damle, A. W. Ghosh and S. Datta, "Nanoscale device modeling," to appear as a book chapter in *Molecular Nanoelectronics*, edited by Mark Reed, 2002.

[Dat00] S. Datta "Nanoscale Device Modeling: the Green's Function Method" *Superlattices and Microstructures,* vol. 28, p. 253, 2000.

[Der01] V. Derycke, R. Martel, J. Appenzeller, and Ph. Avouris, " Carbon Nanotube Inter- and Intramolecular Logic Gates," *Nano Lett.*, vol. 1, p. 453, 2001.

[Fal02] http://www.ece.purdue.edu/celab, Computational electronics group, Purdue University.

[ITR01] International Technology Roadmap for Semiconductors (ITRS), 2001 Edition, Semiconductor Industry Association, http://public.itrs.net.

[Jov00] D. Jovanovic and R. Venugopal, in Presented at the 7th International Workshop on Computational Electronics (University of Glasgow, UK, 2000)

[Jpl02] http://www-hpc.jpl.nasa.gov/PEP/gekco/nemo/nemo.html

[Kel65] L. V. Keldysh, "Diagram technique for nonequilibrium processes," *Sov. Phys. JETP*, vol. 20, p. 1018, 1965.

[Kno02] J. Knoch, B. Lengeler and J. Appenzeller, " Quantum Simulations of an Ultrashort Channel Single-Gated n-MOSFET on SOI," *IEEE Trans. Electron Dev.*, vol. 49, p. 1212, 2002.

[Lak92] R. Lake and S. Datta, “Nonequilibrium Green’s function method applied to double barrier resonant-tunneling diodes,” *Phys. Rev B*, vol. 45, 6670, 1992.

[Lan76] D. C. Langreth, “Linear and Non-Linear Electron Transport in Solids,” *NATO Advanced Study Inst. Series B*, vol. 17, p. 3, Plenum, New York, 1976.

[Lun02] M.S. Lundstrom and Z. Ren, Essential Physics of Nanoscale MOSFETs,” *IEEE Trans. Electron Dev*., vol. 49, p. 133, 2002.

[McE02] P.L. McEuen, M.S. Fuhrer, and H. Park, “Single-Walled Carbon Nanotube Electronics,” *IEEE Trans. Nanotechnology*, vol. 1, p. 78, 2002.

[nan02] Nanotechnology Simulation Hub, http://www.nanohub.purdue.edu

[Nek92] M. Nekovee, B. Geurts, H. M. J. Bootsand M. F. H. Schuurmans,”Failure of extended-moment-equation approaches to describe ballistic transport in submicrometer structures,” *Phys. Rev. B,* vol. 45, 6643, 1992.

[Pau02] M. Paulsson, F. Zahid and S. Datta, “Resistance of a Molecule,” In the LosAlamos Archives at www.arxiv.org/abs/cond-mat/0208183, 2002.

[Raf98] C.S. Rafferty, B. Biegel, Z. Yu, M.G. Acona, J. Bude, and R.W. Dutton, “Multi-dimensional Quantum Effects Simulation using a Density Gradient Model and Script Level Programming Technique,” *SISPAD ’98*, Eds. De Meyer and Biesemans, p. 137, 1998.

[Ren00] Z. Ren, R. Venugopal, S. Datta,, M.S. Lundstrom, D. Jovanovic, and J.G. Fossum, “The Ballistic Nanotransistor: A Simulation Study,” *Int. Electron Dev. Meeting, Tech. Digest*, p. 715, 2000

[Ren01a] Zhibin Ren, “Nanoscale MOSFETs: Physics, Simulation, and Design*,” Ph.D. Thesis*, Purdue University, West Lafayette, IN, Dec. 2001.

[Ren01b] Z. Ren, R. Venugopal, S. Datta, and M.S. Lundstrom, “Examination of Design and Manufacturing Issues in a 10 nm Double Gate MOSFET using Nonequilibrium Green’s Function Simulation,” *Int. Electron Dev. Meeting, Tech. Digest*, p. 107 Dec. 3-5, 2001.

[Rhe02] J.H. Rhew and M.S. Lundstrom, “Benchmarking Macroscopic Transport Models for Nanotransistor TCAD,” to appear in *J. of Computational Electronics*, 2002.

[Shi00] L. Shifren, A. Akis, and D. K. Ferry, “Correspondence Between Quantum and Classical Motion: Comparing Bohmian Mechanics with a Smoothed Effective Potential,” *Phy. Lett. A*, **vol.** 274, p. 75, 2000.

[Svi00] A. Svizhenko, M. Anantram, and T. Govindan, in Presented at the 7th International Workshop on Computational Electronics (University of Glasgow, UK, 2000).

[Svi01] A. Svizhenko, M. Anantram, T. Govindan, B. Biegel, and R. Venugopal, “Nano-transistor Modeling: Two dimensional Green’s Function Method” *J. Appl. Phys.,* vol. 91, p. 2343, 2002.

[Tia98] W. Tian, S. Datta, S. Hong, R. Riefenberger, J. I. Henderson and C. P. Kubiak, "Conductance spectra of molecular wires," *J. Chem. Phys.*, vol. 109, p. 2874, 1998.

[Tsu01] H. Tsuchiya, U. Ravaioli, " Particle Monte-Carlo simulation of quantum phenomena in semiconductor nanostructures," J. Appl. Phys., vol. 89, p. 4023, 2001.

[Ven89] F. Venturi, R. K. Smith, E. C. Sangiorgi, M. R. Pinto, and B. Ricco, "A General Purpose Device Simulator Coupling Poisson and Monte Carlo Transport with Applications to Deep Submicron MOSFET's," *IEEE Trans. on Computer-Aided Design*, **vol.** 8, p. 360, 1989.

[Ven02] R. Venugopal, Z. Ren, S. Datta, M.S. Lundstrom, and D. Jovanovic, "Simulating Quantum Transport in Nanoscale MOSFETs: Real vs. Mode Space Approaches," to appear in the *J. Appl. Phys*., 2002.

[Ven02a] R. Venugopal, S. Datta and M. Lundstrom, unpublished work, 2002.

[Wil98] Y. Fu, M. Karlsteen, M. Willander, N. Collaeert and K. De Meyer, "Quantum transport and I-V characteristics of quantum sized field-effect-transistors,", Superlattices and Microstructures, vol. 24(2) , p.111, 1998.

List of Figures

Fig. 1 The double gate MOSFET structure examined in this work. An oxide thickness of 1 nm, and a body thickness of 3 nm have been used to highlight quantum effects within this device geometry. The power supply is 0.4 V and the gate work function is adjusted to obtain an I_{off} (10 µA/µm) consistent with the ITRS specifications for a high performance transistor. The S/D doping is 2 x 10^{20} cm^{-3} and the transistor is assumed to be wide (Y-dimension is treated as infinite).

Fig. 2 A sketch of the generic subband energy vs. position along the channel (x). Also shown are the semi-infinite contacts (bounded by open rectangles) attached to the source and drain regions and the boundary conditions for injection of a unit amplitude from the source end.

Fig. 3 Illustration of why floating boundary conditions are assumed for the potential at the contacts . The empty circles on the source side, below the dotted line (source to channel barrier) represent source injected electrons reflected by the barrier. The drain injected electrons are represented by filled circles.

Fig. 4 Summary of the ballistic solution scheme

Fig. 5 Illustration of the concept of Büttiker probes. The probe self-energy is adjusted to tune scattering rates and the probe Fermi-levels to obtain zero net current at each scatterer

Fig. 6 The conduction and first two subband minima, as a function of position in the on-state (V_{GS}=V_{DS}=0.4V). Note that the conduction band is a function of both x and z, while the subband minima are functions of x alone. The lower subband minima is for the unprimed valleys with two fold degeneracy, while the higher minima is for the primed valleys with four fold degeneracy

Fig. 7 The total 3D electron density, $n(x,z)$, in the on-state. The thin silicon body is volume inverted, and the electron density goes to zero at the top and bottom oxide/silicon interface (0, 3nm). Quantum effects due to confinement are accurately captured by nanoMOS

Fig. 8 The areal electron density (cm^{-2}) along the channel is plotted in the off-state ($V_{GS}=0$, $V_{DS}=0.4$ V) from both the quantum (solid line) and the classical (dashed line) ballistic transport models at room temperature. Quantum mechanical tunneling through the

source-to channel barrier results in a higher channel charge density in case of the quantum model. Charge neutrality is achieved in the S/D regions in both cases.

Fig. 9 The I_{DS} vs. V_{GS} characteristics for the model device from both the quantum (solid line) and classical (dashed line) ballistic transport models. The ballistic off-current is higher from the quantum model due to source-to-channel tunneling.

Fig. 10 **(a)** The I_{DS} vs. V_{DS} characteristics for the model device from both the quantum (solid line) and classical (dashed line) ballistic transport models. The ballistic on-current is lower from the quantum model as compared to the classical model. **(b)** The I_{DS} vs. V_{DS} characteristics for the model device from both the quantum ballistic (dashed line) and quantum dissipative (solid line) transport models

Fig. 11 The ballistic source-to-drain transmission is plotted as a function of longitudinal energy (E_l) in the on-state. The top of the source-to-channel barrier for both the unprimed (dashed) and the primed (dashed with dots) subbands is also indicated. Note that the transmission is non-zero for longitudinal energies below the top of the subband barrier (tunneling) and increases smoothly to unity. The primed subband transmission goes to two instead of one due to contributions from two sets of valleys.

Fig. 12 **(a)** Subband profile, $E_1(x)$, for the device in Fig. 1 with an n^{++} source extension added under low (solid line) and high (dashed line) gate bias (VGS = 0 and 0.6V) at a drain bias of 0.6V. The potential (subband) floats to a lower energy in the n^{+} region, but remains unchanged in the n^{++} region**.** **(b)** Subband profile, $E_1(x)$, for the device in Fig. 1 with the n^{++} source extension and fixed boundary conditions (dashed line) and without the n^{++} extension and floating boundary conditions (solid line) at a high gate voltage.

Fig. 13 **(a)** The energy resolved Local Density of States (LDOS) is plotted along the channel in the on-state from the quantum ballistic transport model. Light areas imply a low density of states, while dark regions indicate a high density of states. Coherent oscillations in the LDOS are the result of reflections from the barrier. Non-zero LDOS in the forbidden region (below the subband energy), leads to tunneling through the source-to-channel barrier. The LDOS from the primed (blue line) band is higher than the unprimed (red line) due to degeneracy. **(b)** The energy resolved Local Density of States (LDOS) is plotted along the channel in the on-state from the quantum diffusive transport model. Coherent oscillations in the LDOS are washed out as a result of scattering. A small, potential drop in the source and drain regions is also discernible.

Fig. 14 The energy resolved electron density is plotted along the channel in the on-state from the quantum ballistic transport model. In the heavily doped S/D regions, multiple subbands are occupied. However much of the current in the channel is from the first subband. The ballistic source injected charge that propagates to the drain without any energy relaxation is clearly observed.

Appendix: Green's Function Formulation

Although we described nanoMOS from a wave function perspective in the text, nanoMOS is actually based on a Green's function formulation of the same problem. The non-equilibrium Green's function formalism offers advantages when more complicated geometries or basis functions are considered and when scattering needs to be included in a more rigorous manner. In this appendix we translate the wavefunction approach of the text into the NEGF formalism. Several references provide a fuller discussion of the method [Dat00].

To simulate a MOSFET, we need to compute the electron density and the current. In the wavefunction picture, eq. (9), we need $\phi^*(x)\phi(x)$ at each node, where ϕ is determined by the Greens function according to eq. (19). To get $\phi^*(x)\phi(x)$, consider the generalization

$$\phi^+\phi = \mathbf{G}\gamma_1\gamma_1^+\mathbf{G}, \tag{A1}$$

where

$$\phi = \begin{bmatrix} \phi_0 \\ \vdots \\ \\ \\ \phi_{\mathrm{N}-1} \end{bmatrix} \tag{A2}$$

is an N x 1 vector giving the value of the wavefunction at each node, and the superscript, +, denotes the Hermetian transpose. The matrix product in eq. (A2) is

$$\phi^+\phi = \begin{bmatrix} \phi_0^*\phi_0 & \phi_0^*\phi_1 & \cdots & \phi_0^*\phi_{N-1} \\ \phi_1^*\phi_0 & \phi_1^*\phi_1 & & \\ \vdots & & \ddots & \\ & & & \\ \phi_{N-1}^*\phi_0 & & & \phi_{N-1}^*\phi_{N-1} \end{bmatrix}. \tag{A3}$$

By defining an N x N matrix,

$$\Gamma_1 \equiv \gamma_1 \gamma_1^+ = \begin{bmatrix} 2t_0 \sin ka & 0 & \cdots & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \\ & & & \\ 0 & & & 0 \end{bmatrix} = i\left(\Sigma_1 - \Sigma_1^+\right), \tag{A4}$$

(where Σ_1 is the self-energy describing the connection to the source contact as given by eq. (17a)) we can use eq. (A1) to generalize the spectral function to

$$\mathbf{A}_1 = \mathbf{G}\Gamma_1 \mathbf{G}^+, \tag{A5}$$

which is an N x N matrix. Instead of the electron density as in eq. (9), we find the density matrix due to source injection as

$$\rho_S = \int_0^{E_{Top}} F\left(E_F - E_\ell\right) \frac{\mathbf{A}_1(x, E_\ell)}{2\pi}. \tag{A6}$$

Analogous expressions exist for injection from the drain, so the total density matrix is $\rho = \rho_1 + \rho_2$. In summary, from the Green's function, eq. (20), we find the density matrix whose diagonal elements give the electron density at each node.

To find the current, eq. (11b) still applies, but we need to express the transmission coefficient, $T^{1-2}\left(E_\ell\right)$, in terms of the Green's function. Assuming injection from the source as in eqs. (5), we find the incident and transmitted currents as

$$J_{inc} \equiv 1\upsilon_0 \tag{A7a}$$

and

$$J_{trans} \equiv |t|^2 \upsilon_{N-1} . \tag{A7b}$$

The velocities at the injection and exit points are

$$\upsilon_0 = \frac{1}{\hbar}\frac{dE_k}{dk_x}\bigg|_{x=0} = \frac{2at_0}{\hbar}\sin k_0 a = \left(\frac{a}{\hbar}\right)\Gamma_S(1,1) \tag{A8a}$$

$$\upsilon_{N-1} = \frac{1}{\hbar}\frac{dE_k}{dk_x}\bigg|_{x=L} = \frac{2a\,t_0}{\hbar}\sin k_{N-1}a = \left(\frac{a}{\hbar}\right)\Gamma_D(N-1,N-1). \tag{A8b}$$

Using these results, we find

$$T^{1-2}(E_\ell) = \frac{J_{trans}}{J_{inc}} = |t|^2\,\frac{\upsilon_{N-1}}{\upsilon_0} = |\phi_{N-1}|^2\,\frac{\Gamma_2(N-1,N-1)}{\Gamma_1(1,1)}. \tag{A9}$$

Since

$$\phi_{N-1} = \mathbf{G}(N-1,1)\Gamma_1(1,1), \tag{A10}$$

we have

$$T^{1-2}(E_\ell) = |\mathbf{G}(\mathrm{N}-1,1)|^2\,\Gamma_1(1,1)\Gamma_2(N-1,N-1), \tag{A11}$$

which can be expressed as

$$T^{1-2}(E_\ell) = \mathrm{trace}\left(\Gamma_2\mathbf{G}\Gamma_1\mathbf{G}^+\right) = \mathrm{trace}\left(\Gamma_1\mathbf{G}\Gamma_2\mathbf{G}^+\right). \tag{A12}$$

To summarize, after computing the Green's function, we find the transmission coefficient from eq. (A12), which is then used with eq. (11b) to find the current. Note that Eq. (A12) can be generalized to find the transmission between any two probes as

$$T^{k-l}(E_\ell) = \text{trace}\left(\Gamma_k \mathbf{G} \Gamma_l \mathbf{G}^+\right)., \tag{A13}$$

which is useful when Büttiker probes are used to include scattering.

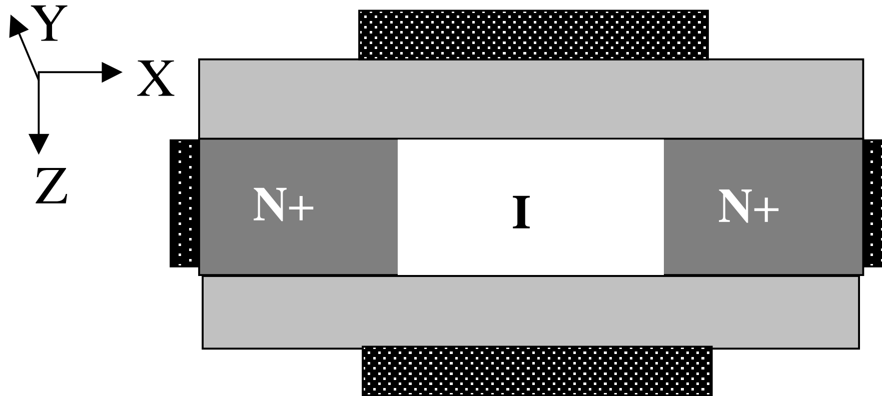


Fig. 1: The double gate MOSFET structure examined in this work. An oxide thickness of 1 nm, and a body thickness of 3 nm have been used to highlight quantum effects within this device geometry. The power supply is 0.4 V and the gate work function is adjusted to obtain an I_{off} (10 µA/µm) consistent with the ITRS specifications for a high performance transistor. The S/D doping is 2 x 10^{20} cm^{-3} and the transistor is assumed to be wide (Y-dimension is treated as infinite).

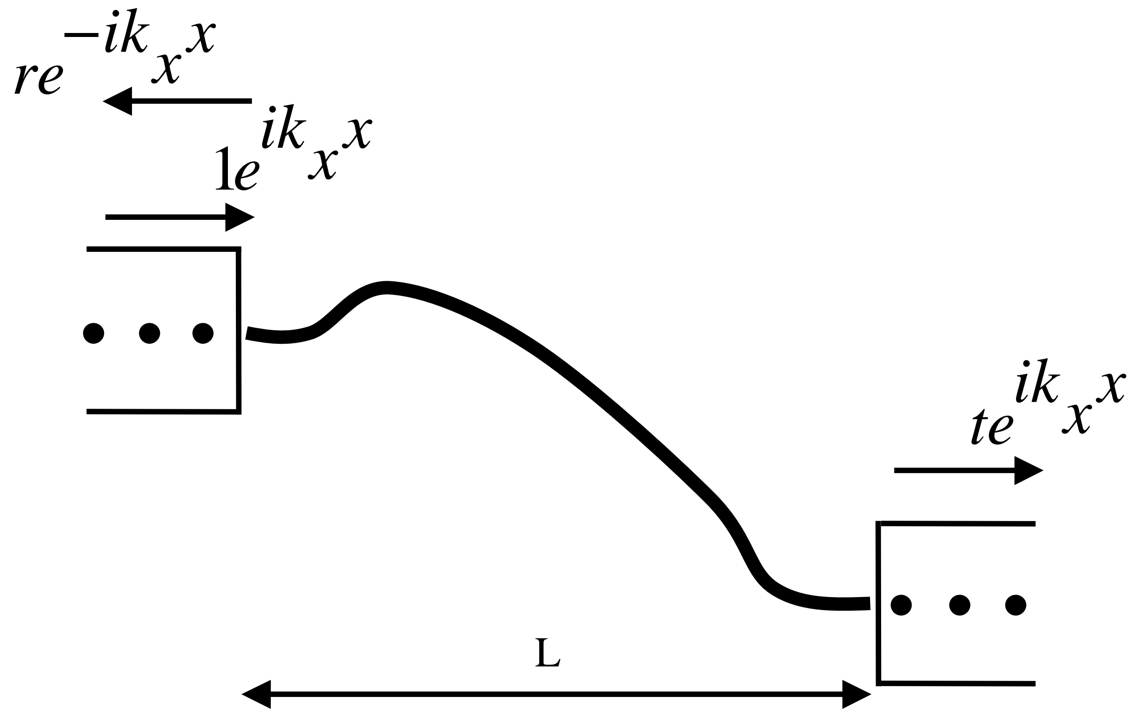


Fig. 2: A sketch of the generic subband energy vs. position along the channel (x). Also shown are the semi-infinite contacts (bounded by open rectangles) attached to the source and drain regions and the boundary conditions for injection of a unit amplitude from the source end.

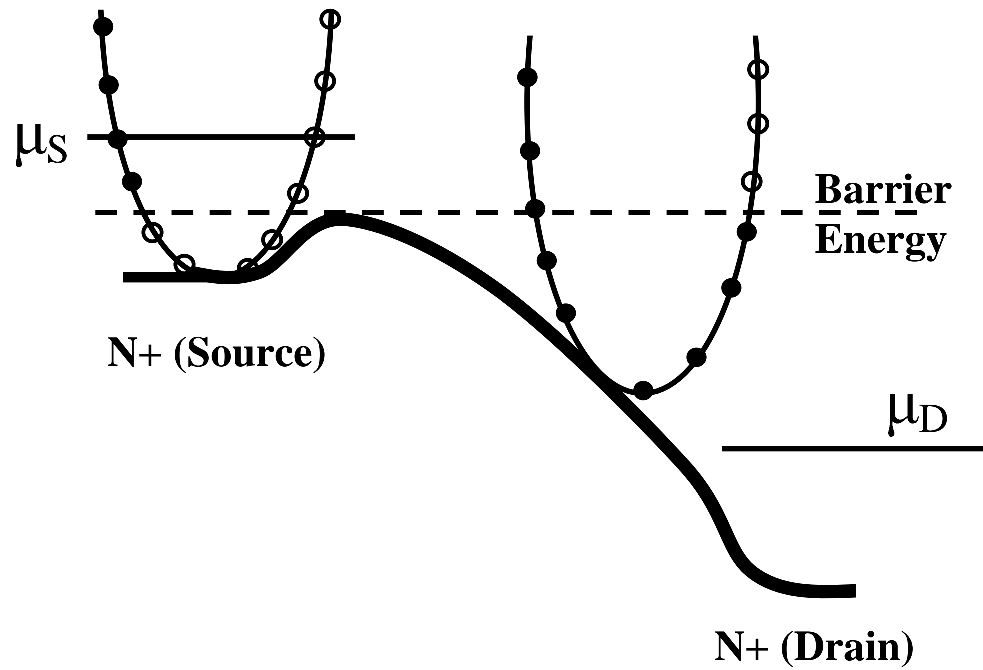


Fig 3: Illustration of why floating boundary conditions are assumed for the potential at the contacts . The empty circles on the source side, below the dotted line (source to channel barrier) represent source injected electrons reflected by the barrier. The drain injected electrons are represented by filled circles.

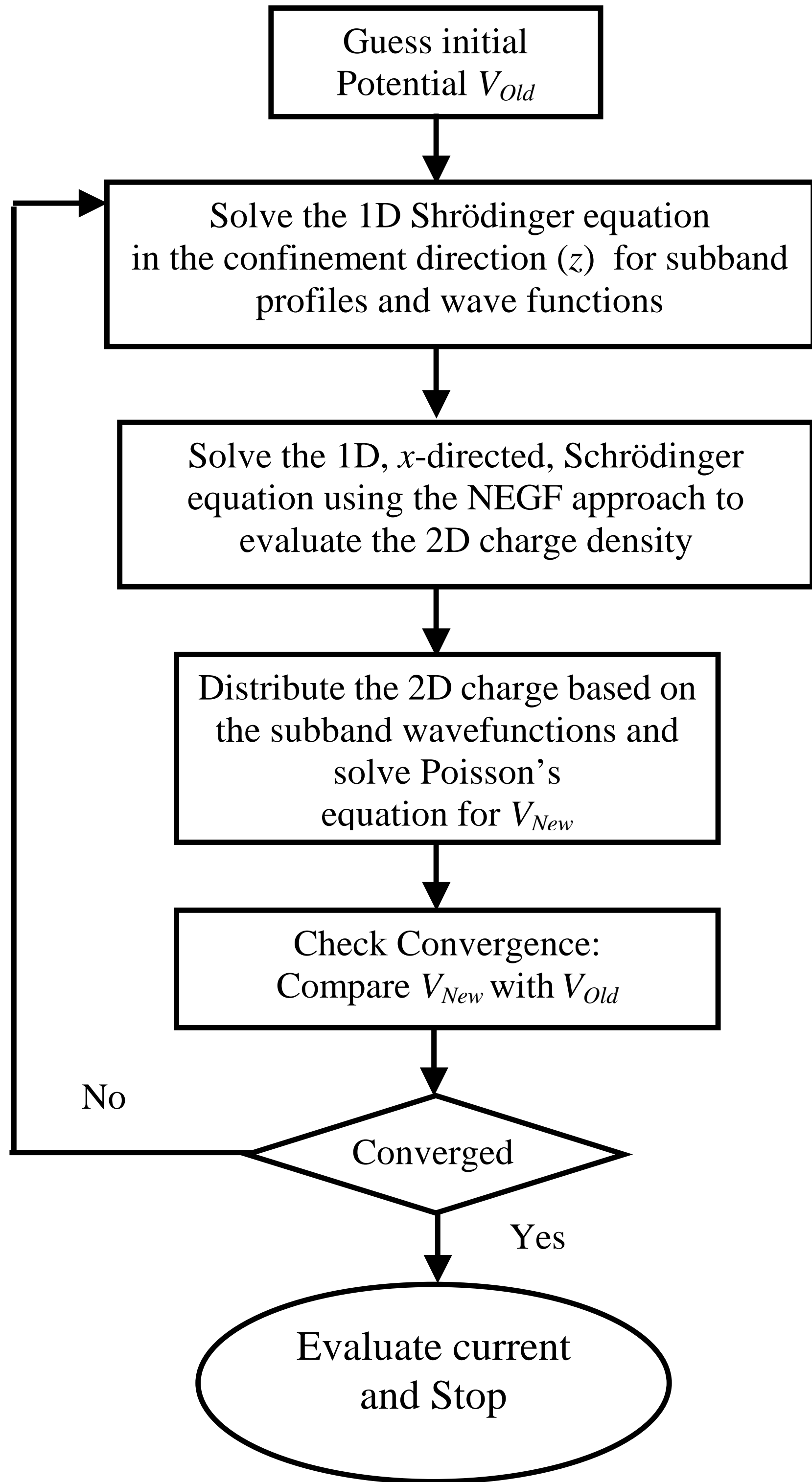


Fig 4: Summary of the ballistic solution scheme

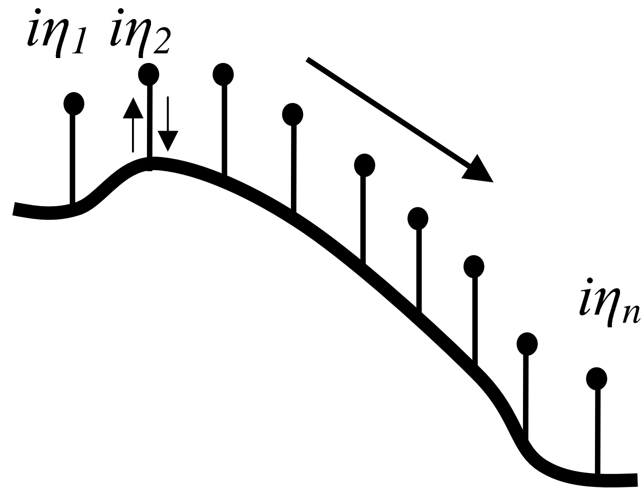


Fig 5: Illustration of the concept of Büttiker probes. The probe self-energy is adjusted to tune scattering rates and the probe Fermi-levels to obtain zero net current at each scatterer

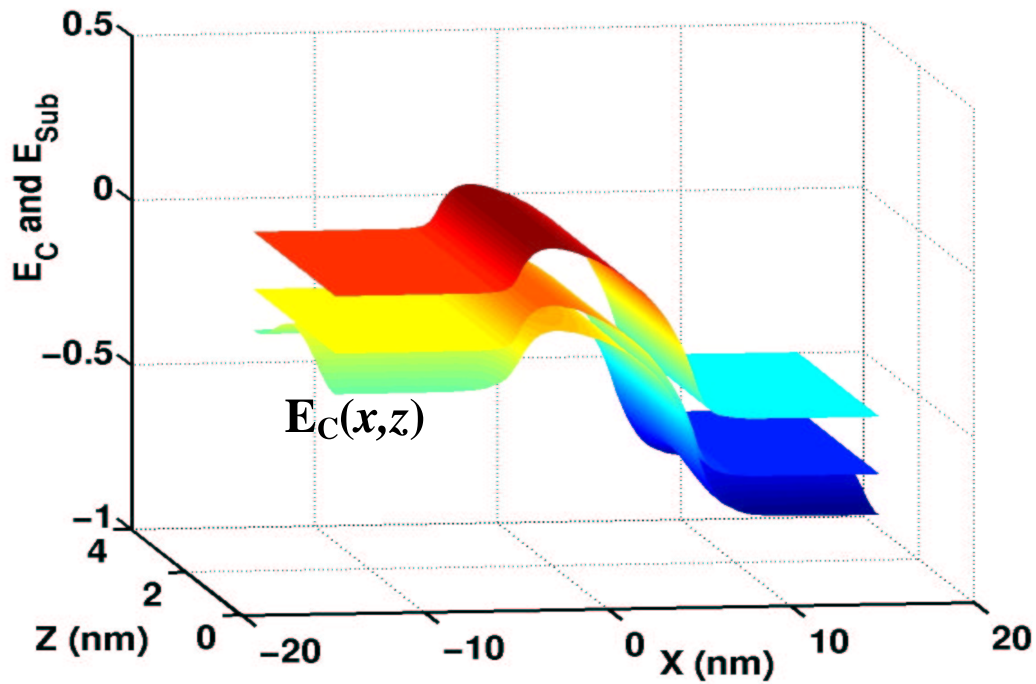


Fig. 6: The conduction and first two subband minima, as a function of position in the on-state (V_{GS}=V_{DS}=0.4V). Note that the conduction band is a function of both x and z, while the subband minima are functions of x alone. The lower subband minima is for the unprimed valleys with two fold degeneracy, while the higher minima is for the primed valleys with four fold degeneracy.

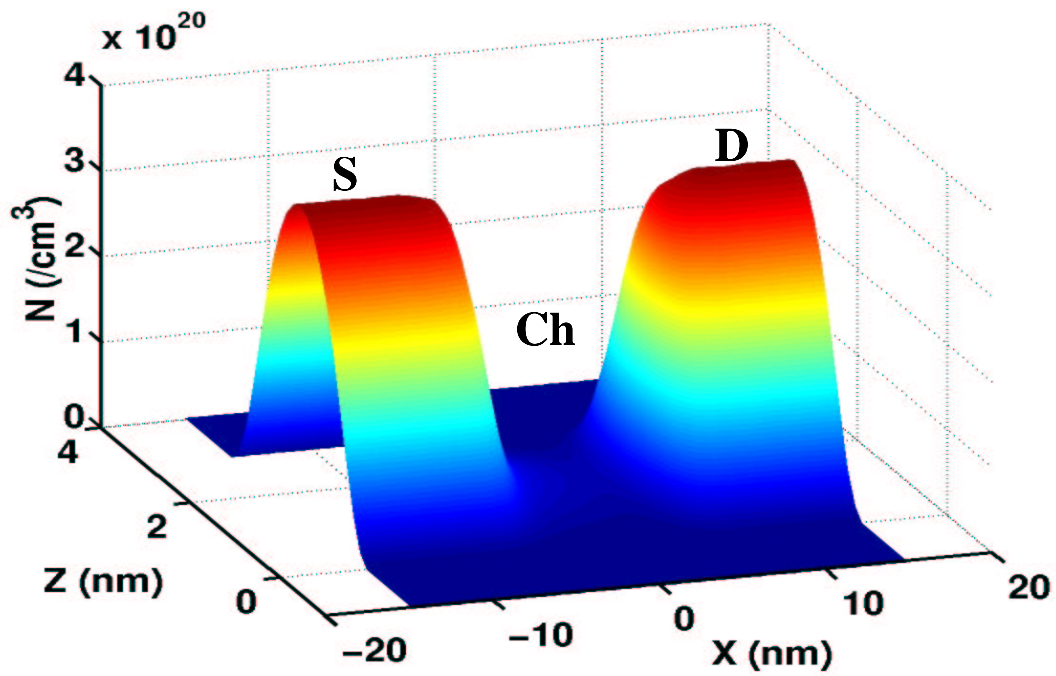


Fig. 7: The total 3D electron density, *n(x,z)*, in the on-state. The thin silicon body is volume inverted, and the electron density goes to zero at the top and bottom oxide/silicon interface (0, 3nm). Quantum effects due to confinement are accurately captured by nanoMOS.

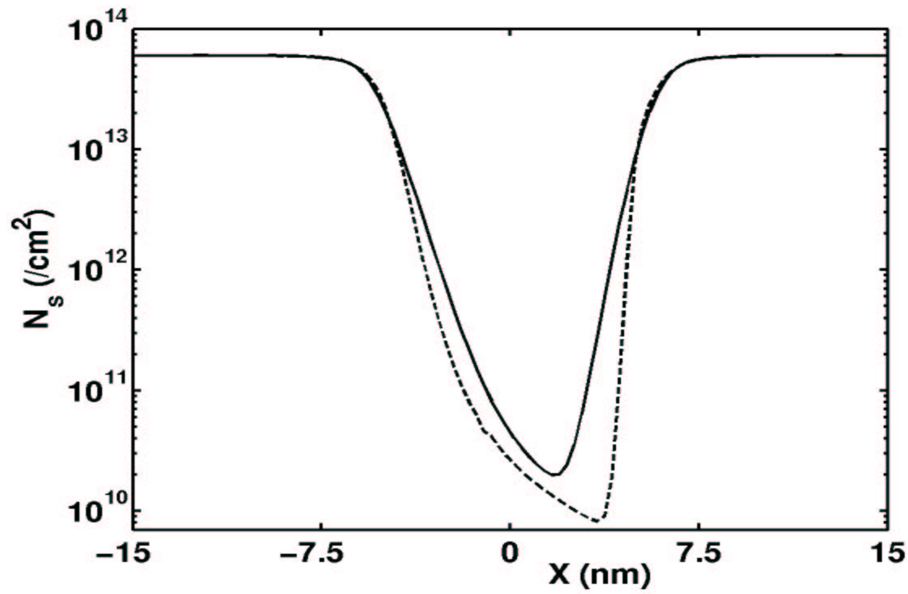


Fig. 8: The areal electron density (cm^{-2}) along the channel is plotted in the off-state ($V_{GS}=0$, $V_{DS}=0.4$ V) from both the quantum (solid line) and the classical (dashed line) ballistic transport models at room temperature. Quantum mechanical tunneling through the source-to channel barrier results in a higher channel charge density in case of the quantum model. Charge neutrality is achieved in the S/D regions in both cases.

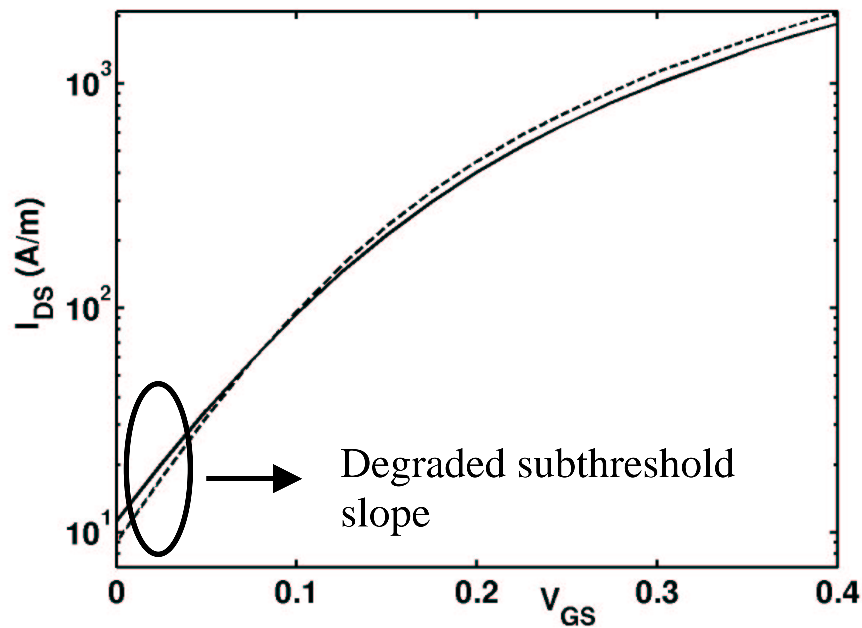
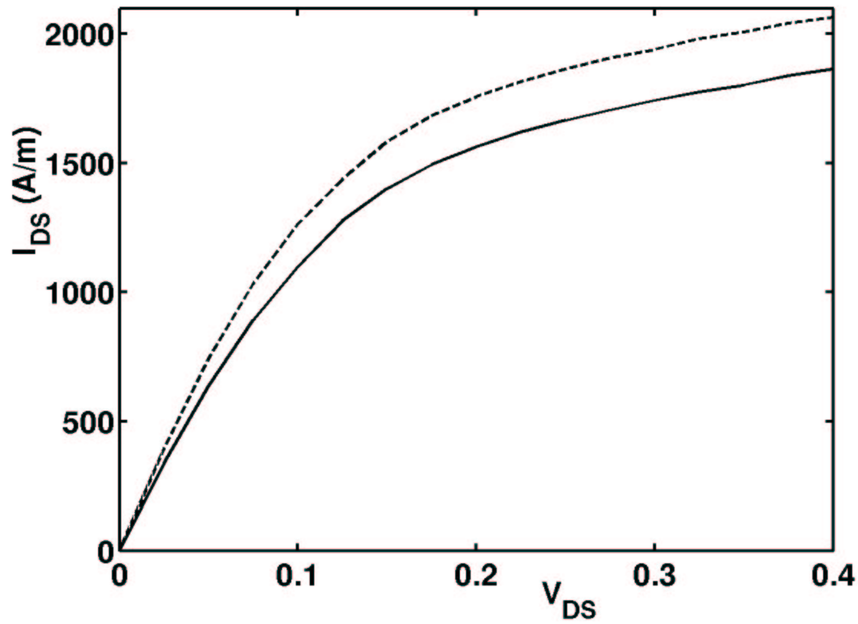
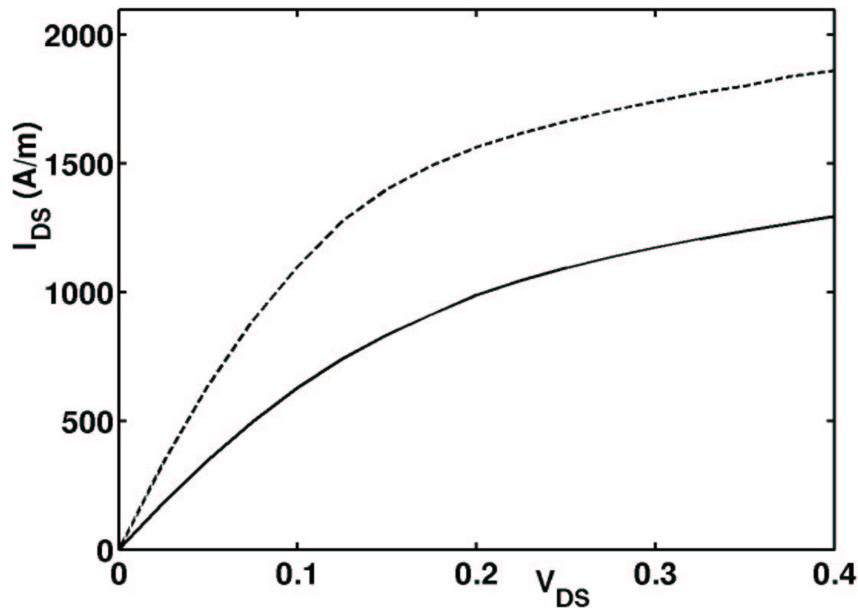


Fig. 9: The I_{DS} vs. V_{GS} characteristics for the model device from both the quantum (solid line) and classical (dashed line) ballistic transport models. The ballistic off-current is higher from the quantum model due to source-to-channel tunneling.



(a)



(b)

Fig. 10: **(a)** The I_{DS} vs. V_{DS} characteristics for the model device from both the quantum (solid line) and classical (dashed line) ballistic transport models. The ballistic on-current is lower from the quantum model as compared to the classical model. **(b)** The I_{DS} vs. V_{DS} characteristics for the model device from both the quantum ballistic (dashed line) and quantum dissipative (solid line) transport models

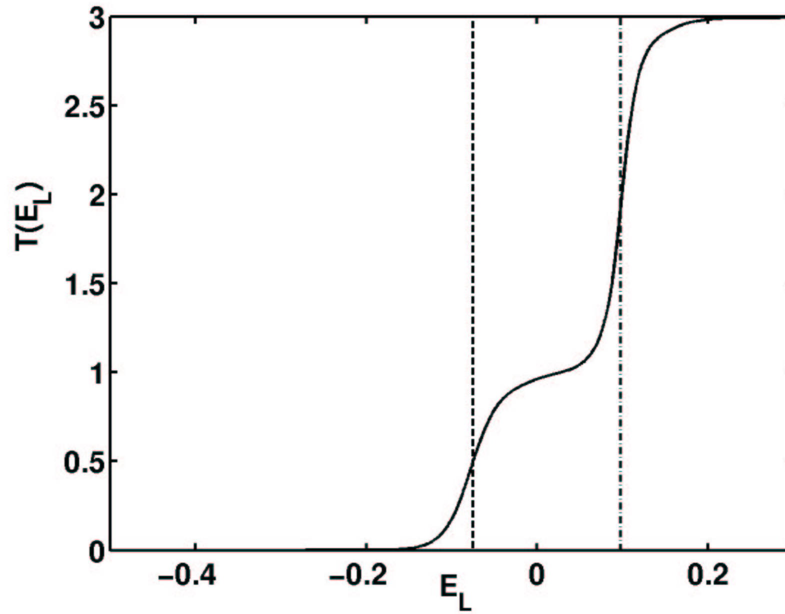
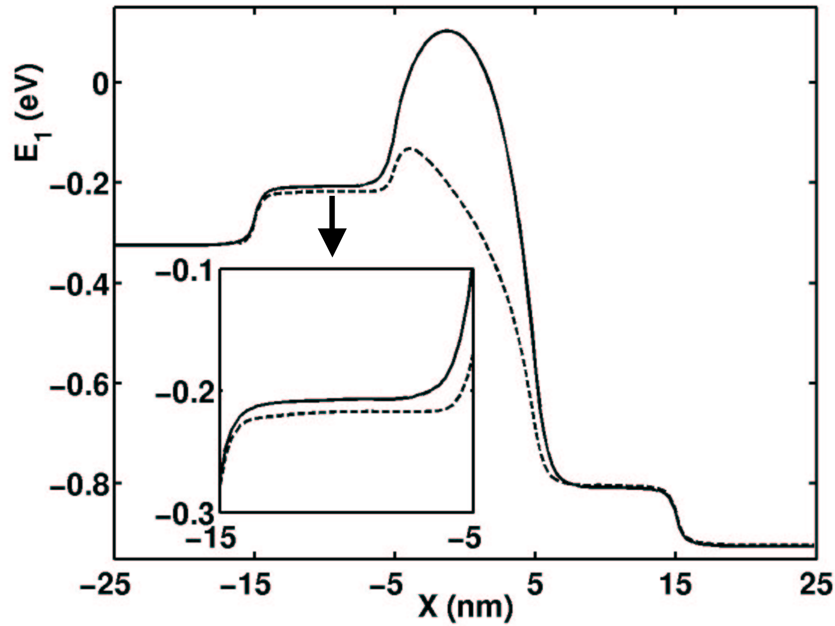
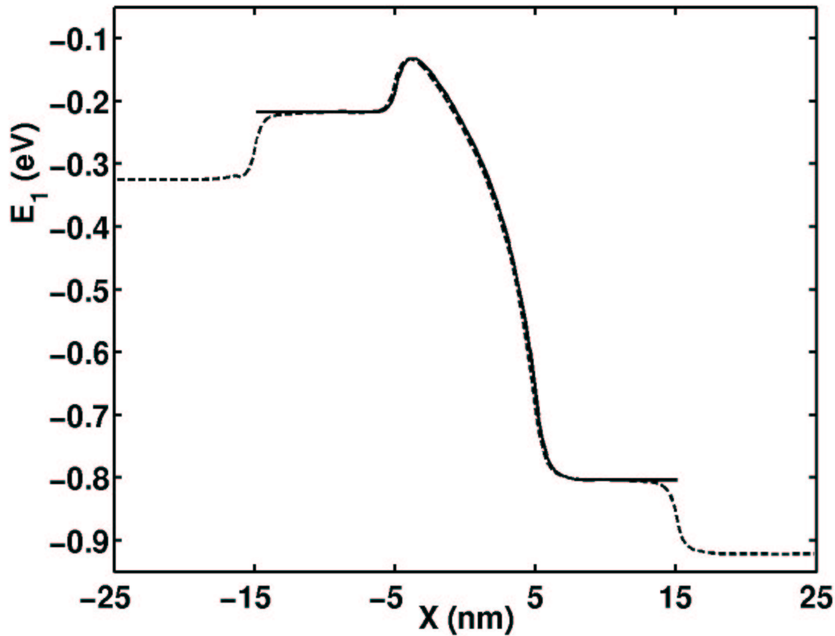


Fig. 11: The ballistic source-to-drain transmission is plotted as a function of longitudinal energy (E_l) in the on-state. The top of the source-to-channel barrier for both the unprimed (dashed) and the primed (dashed with dots) subbands is also indicated. Note that the transmission is non-zero at longitudinal energies below the top of the subband barrier (tunneling) and increases smoothly to unity. The primed subband transmission goes to two instead of one due to contributions from two sets of valleys.

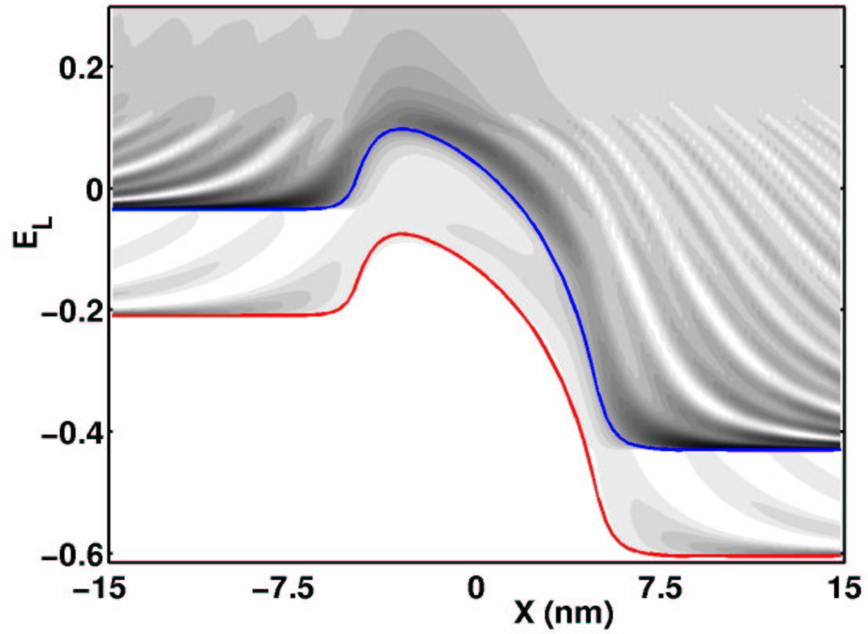


(a)

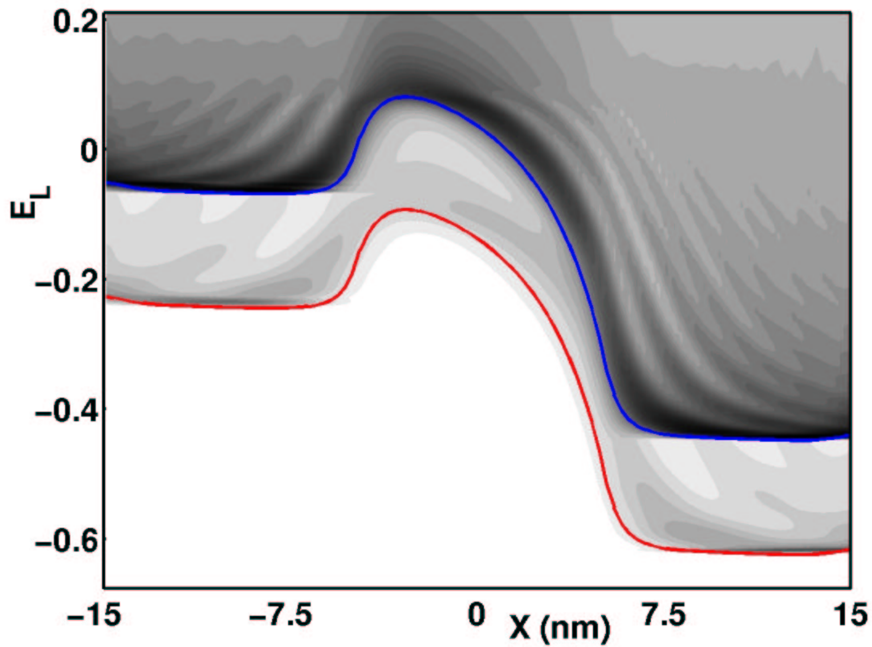


(b)

Fig 12: (a) Subband profile, $E_1(x)$, for the device in Fig. 1 with an *n*++ source extension added under low (solid line) and high (dashed line) gate bias (VGS = 0 and 0.6V) at a drain bias of 0.6V. The potential (subband) floats to a lower energy in the *n*+ region, but remains unchanged in the *n*++ region**. (b)** Subband profile, $E_1(x)$, for the device in Fig. 1 with the *n*++ source extension and fixed boundary conditions (dashed line) and without the *n*++ extension and floating boundary conditions (solid line) at a high gate voltage.



(a)



(b)

Fig. 13: (a) The energy resolved Local Density of States (LDOS) is plotted along the channel in the on-state from the quantum ballistic transport model. Light areas imply a low density of states, while dark regions indicate a high density of states. Coherent oscillations in the LDOS are the result of reflections from the barrier. Non-zero LDOS in the forbidden region (below the subband energy), leads to tunneling through the source-to-channel barrier. The LDOS from the primed (blue line) band is higher than the unprimed (red line) due to degeneracy. **(b)** The energy resolved Local Density of States (LDOS) is plotted along the channel in the on-state from the quantum diffusive transport model. Coherent oscillations in the LDOS are washed out as a result of scattering. A small, potential drop in the source and drain regions is also discernible.

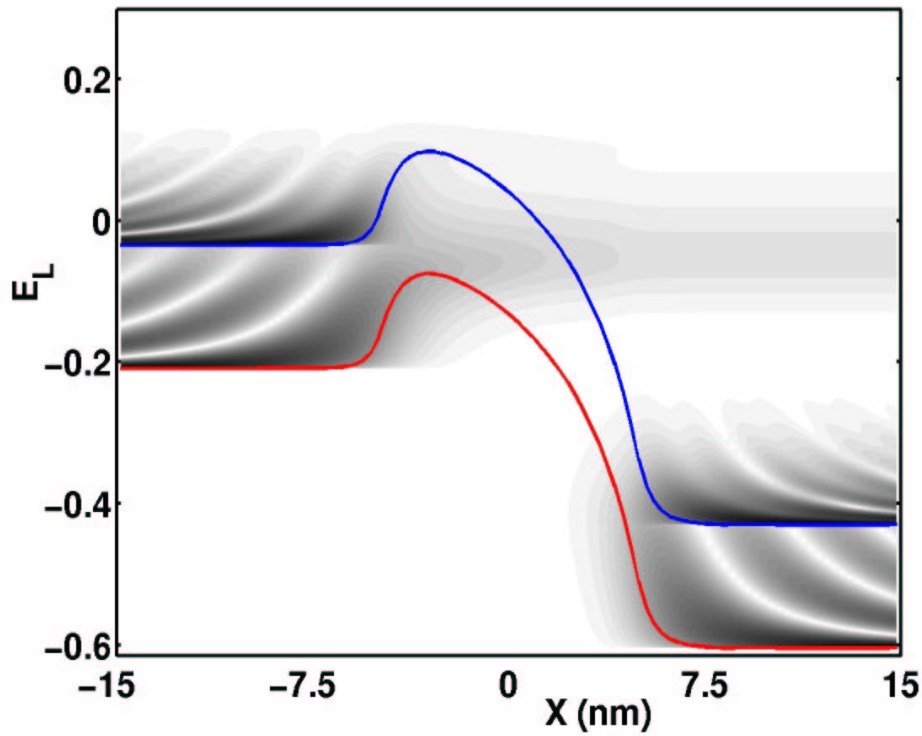


Fig. 14: The energy resolved electron density is plotted along the channel in the on-state from the quantum ballistic transport model. In the heavily doped S/D regions, multiple subbands are occupied. However much of the current in the channel is from the first subband. The ballistic source injected charge that propagates to the drain without any energy relaxation is clearly observed.