A Quantum Mechanical Analysis of Channel Access Geometry and Series Resistance in nanoscale transistors

R. Venugopal,* S. Goasguen, S. Datta, and M. S. Lundstrom

School of Electrical and Computer Engineering,

Purdue University, 1285 Electrical Engineering Building,

West Lafayette, Indiana 47907-1285

(Dated: October 16, 2003)

Abstract

In this paper, we apply a two-dimensional quantum mechanical simulation scheme to study the effect of channel access geometries on device performance. This simulation scheme solves the non-equilibrium Green's function equations self-consistently with Poisson's equation and treats the effect of scattering using a simple approximation inspired by Büttiker. It is based on an expansion of the device Hamiltonian in coupled mode-space. Simulation results are used to highlight quantum effects and discuss the importance of scattering when examining the transport properties of nanoscale transistors with differing channel access geometries. Additionally, an efficient domain decomposition scheme for evaluating the performance of nanoscale transistors is also presented. This paper highlights the importance of scattering in understanding the performance of transistors with different channel access geometries.

*Electronic address: venugopr@ecn.purdue.edu

I. INTRODUCTION

The push towards achieving small transistors with the requisite device characteristics has resulted in the exploration of novel device structures that have been recently reported in the literature [1] [2]. The channel length for some of these devices is of the order of 10 nm or lower. As critical transistor dimensions continue to shrink, the role of channel access geometry must be understood. Therefore, the primary objective of this paper is to understand and highlight the effects of differing access geometries on device performance. We achieve this objective by using a simulation scheme based on the non-equilibrium Green's function formalism (NEGF), which captures quantum effects and treats scattering within a single modeling framework [3] [4]. This technique is applied to examine device physics and design issues at the 10 nm scale.

Several device designs with single, double or tri-gate geometries are being considered as candidates for scaling silicon devices to the limit [5] [6] [7]. Despite the nature of the gate, all of these geometries share a common characteristic- they are composed of a thin silicon body sandwiched between insulators and coupled to large source/drain (S/D) reservoirs. One such device, a double-gate (DG), silicon-on-insulator (SOI) transistor, is illustrated in Fig. 1b. Within the thin silicon body, carriers are strongly quantum confined (in one and possibly two dimensions) and reside in a discrete set of subbands. This subband structure affects important device metrics such as threshold voltage and gate capacitance. As electrons propagate from the intrinsic device (the region within the dashed box in Fig. 1b) into the S/D regions (which are typically wide in order to reduce the resistance), the subband structure is relaxed, and there is significant mixing of electrons in different modes. Therefore, modeling such structures requires the capability to seamlessly couple these wide and narrow regions, while capturing all of the quantum effects within the narrow region accurately. The effects of scattering also need to be accounted for in both the wide and the narrow regions, because scattering which is present in real devices, permits electrons from several modes in the wide region to couple into a few modes within the narrow region. In our modeling framework scattering is simulated phenomenologically using the concept of Büttiker probes, often used in mesoscopic physics [8] [9]. Our treatment of scattering enables us to map quantum mechanical parameters to equivalent low-field mobilities in different regions of the device. Such a mapping is useful because low-field mobilities continue to be relevant even at the

short length scales considered in this paper [10] [11]. The overall simulation technique used in this study solves a coupled set of NEGF and Poisson equations [12] [13].

Electron transport from wide to narrow regions has been extensively examined in the physics literature through the study of quantum point contacts [14] [15]. These studies, (which were nonself-consistent) compared the low bias conductance behavior (at temperatures close to 0 K) of abrupt wide-to-narrow devices against devices where the channel access geometry was gradually flared out (also referred to as adiabatic). They concluded that for specific thicknesses of the narrow region, the low bias conductance is unaffected by the channel access geometry. When studying the low bias conductance properties around the Fermi surface (at low temperatures), it should be noted that if the Fermi level is well above the subband edge within the narrow region (achieved by changing the thickness), the large S/D reservoirs act as "reflectionless contacts" [9]. The term reflectionless implies that electrons at the Fermi surface can transmit from the narrow to the wide regions with negligible probability of reflection. In such cases, devices with differing channel access geometries could exhibit the same low bias conductance around the Fermi energy. However, electrons close to the subband edge (long wavelengths) may experience large reflections.

In nanoscale transistors the current spectrum is distributed over a wide range of energies (and not just around the Fermi energy) due to the high voltages that are applied. Also, the alignment of the Fermi energy relative to the subband edge is determined by self-consistent electrostatics. Therefore the large S/D reservoirs cannot be assumed as reflectionless contacts, thus implying that their geometry may affect device characteristics. Moreover, a 10-20 % difference in the series resistance due to differing channel access geometries is an important effect that needs to be emphasized when evaluating different device designs. Also, the effect of scattering, which can be safely neglected in QPC's operating at low temperatures cannot be ignored in transistors within which temperatures exceed 300 K. We will highlight the importance of scattering by comparing results from both ballistic (no scattering) as well as scattering simulations for nanoscale, silicon transistors with differing channel access geometries. Due to the aforementioned reasons, conclusions from the mesoscopic physics literature need to be carefully examined for nanoscale silicon transistors operating at room temperature.

Device design and optimization through the use of simulations requires the ability to quickly perform a large number of simulations within a reasonable amount of time. Even for

the idealized device structure with the dimensions illustrated in Fig. 1b, the simulation time could be very large. Computational burden is greatly increased because a large number of subbands need to be included when simulating the flared out S/D reservoirs quantum mechanically. Therefore, an approximate treatment of the S/D reservoirs which are usually much larger than those in Fig. 1b is clearly useful. We examine one such approximation wherein the device in Fig. 1b is decomposed into an intrinsic component (region within the dashed box), and a parasitic component (the S/D reservoirs). The intrinsic device is explicitly simulated, and the extrinsic characteristics which include the effect of the S/D reservoirs are estimated from the intrinsic current vs. voltage (I-V) characteristics through the addition of appropriate parasitic resistances. We present a recipe which can be applied to examine design issues in nanoscale transistors and highlight the appropriate boundary conditions that need to be used when simulating the intrinsic device.

The paper is divided into the following sections: Section II presents the solution scheme. Section III presents simulation results which explain the role of self-consistency, scattering and channel access geometry on device performance. Section IV discusses device design issues and Section V summarizes key findings.

II. THE MODELING SCHEME

The simulated device structures are shown in Fig. 1. A uniform rectangular grid with a grid spacing of a, along the x direction and b, along the z direction is used. The width of the device is assumed to be large, and all potentials (including the scattering potential) are assumed to be translationally invariant along the width (y dimension). The Fermi levels at the ends of the device are specified by the applied bias and a single band effective mass Hamiltonian that assumes an ellipsoidal, parabolic bulk bandstructure is used to model electron transport.

We begin by expanding the 3D effective mass Hamiltonian for the device in terms of $\delta(x-x')\delta(z-z')$ (real-space basis) and $e^{ik_j y}/\sqrt{W}$, where the plane wavefunction, $e^{ik_j y}/\sqrt{W}$, represents the device width. The quantum number, k_j, corresponds to the eigenenergy, $E_{k_j} = \hbar^2 k_j^2/2m_y^*$, where m_y^* is the electron effective mass in the y direction. The real-space delta functions, $\delta(x-x')$ and $\delta(z-z')$ with eigenvalues x' and z' respectively, combined with $e^{ik_j y}/\sqrt{W}$, form a complete and orthogonal expansion function set. Since all potentials

are assumed to be invariant along the device width, on expansion, the Hamiltonian for each k_j is,

$$h[x,z] = \left[\begin{array}{c|ccc|c} \alpha_0 & \beta & 0 & \cdots & \cdots \\ \hline \beta & \alpha_1 & \ddots & 0 & \cdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \cdots & 0 & \ddots & \alpha_{N_X} & \beta \\ \hline \cdots & \cdots & 0 & \beta & \alpha_{N_{X+1}} \end{array}\right] \tag{1}$$

This block tri-diagonal Hamiltonian can be separated into two parts; 1) A part representing the device (the matrix within the box) and 2) A part representing the semi-infinite leads. If we think of the device as being composed of vertical slices (refer Fig. 1a) adjoining each other, the α's represent the discretized, one-dimensional (1D) effective mass Hamiltonian along the z direction within each slice and the β's represent coupling between adjacent slices. The α's, with indices less than one, represent successive slices going into the source contact while those with indices greater than N_X represent successive slices into the drain. There are N_X slices within the active device region. The α's and β's are themselves tridiagonal matrices and are,

$$\alpha[x] = \begin{bmatrix} 2t_x + 2t_z - qV_1(x) & -t_z & 0 & \cdots & \cdots \\ -t_z & 2t_x + 2t_z - qV_2(x) & \ddots & \cdots & \cdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \cdots & 0 & \ddots & \ddots & -t_z \\ \cdots & \cdots & 0 & -t_z & 2t_x + 2t_z - qV_{N_Z}(x) \end{bmatrix} \tag{2}$$

$$\beta = \begin{bmatrix} -t_x & 0 & \cdots \\ 0 & -t_x & \cdots \\ \cdots & 0 & -t_x \end{bmatrix} \tag{3}$$

The V's (in Eq. 2), represent the potential along a vertical slice at site x and t_x and t_z, the coupling energies between adjacent grid points in x and z respectively. These site coupling energies are given by,

$$t_x = \frac{\hbar^2}{2m_x^* a^2} \quad \text{and} \quad t_z = \frac{\hbar^2}{2m_z^* b^2} \tag{4}$$

There is no restriction on the solution domain and it can be easily extended to include the insulator regions provided changes in the electron effective mass is correctly accounted for within the insulator and at the silicon/insulator interface when discretizing the effective mass Hamiltonian. The overall size of the discretized active device Hamiltonian in a real-space basis for each conduction band valley is $(N_X \times N_Z)^2$.

The choice of a real-space basis, although natural, leads to a computational problem. If we use a grid spacing of a=0.25 nm and b=0.1 nm to discretize the Hamiltonian for the device pictured in Fig. 1b, then N_X=200 and N_Z=62. Therefore the size of the active device Hamiltonian is $\sim 12000 \times 12000$. The computational burden of performing a large number of matrix operations on such a Hamiltonian is prohibitively expensive. Therefore we opt for a basis transformation where the new basis is designed to ensure that the size of the Hamiltonian is reduced considerably while capturing all of the physics relevant to our problem. The use of this basis has been described in [12] [16].

We note that even for the device shown in Fig. 1b, where the wide regions are 6.2 nm deep, electrons are strongly quantum confined in the z dimension and reside in a discrete set of subbands. At room temperature only a few low energy subbands are occupied even within the wide regions of the device shown in Fig. 1b. Therefore, the real-space Hamiltonian can be significantly reduced in size by transforming the basis from real-space to one comprising the lowest few modes (mode-space) without any loss of accuracy [12]. Mathematically, the new basis is obtained by solving for the eigenfunctions of α (refer Eq. 2), which represents a discretized one-dimensional, z directed effective mass equation within each vertical slice of the device (Fig. 1a). If we consider a simple example where N_Z=3, N_X=3 and assume that the lowest two modes are all that need to be included, the basis transformation matrix (mode-space basis in terms of the real-space basis) looks like,

$$
U = \begin{array}{c}
\begin{array}{cccccc} \multicolumn{3}{c}{\text{mode 1}} & \multicolumn{3}{c}{\text{mode 2}} \end{array} \\
\begin{array}{cccccc} 1 & 2 & 3 & 1 & 2 & 3 \end{array} \\
\left[\begin{array}{ccc|ccc}
\times & 0 & 0 & \times & 0 & 0 \\
\times & 0 & 0 & \times & 0 & 0 \\
\times & 0 & 0 & \times & 0 & 0 \\
0 & \times & 0 & 0 & \times & 0 \\
0 & \times & 0 & 0 & \times & 0 \\
0 & \times & 0 & 0 & \times & 0 \\
0 & 0 & \times & 0 & 0 & \times \\
0 & 0 & \times & 0 & 0 & \times \\
0 & 0 & \times & 0 & 0 & \times
\end{array}\right]
\end{array}
\tag{5}
$$

In Eq. 5, the crosses mark non-zero entries corresponding to the mode eigenfunctions which are expressed in a real-space representation. The size of the basis transformation matrix is $(N_X \times N_Z) \times$ ($N_X \times$ number of modes). Note that each column of U expresses the modes in real-space and that the columns are ordered into blocks of size N_X (in this case 3) as labeled in Eq. 5. Different columns of U are orthonormal to each other, thus indicating that U is a unitary transformation matrix. The device Hamiltonian in the new basis is

$$h_{\text{mode}} = U^{\dagger} h[x, z] U \tag{6}$$

The size of active device Hamiltonian in mode-space is $(N_X \times \text{number of modes})^2$. For the device structure in Fig. 1b, we use the lowest five modes to effect the basis transformation, thus resulting in a Hamiltonian matrix which is 1000×1000 in size (the original Hamiltonian in real-space had a size of 12000×12000 for each conduction band valley). When expressed in mode-space, the Hamiltonian has clear physical meaning. Diagonal blocks (size N_X^2) represent coupling within a mode, while off-diagonal blocks represent coupling between modes. For device structures such as those in Fig. 1a (in which the confining potential along z retains its shape as one moved from the source to the drain), it has been demonstrated that the off-diagonal blocks within the mode-space can be ignored and individual modes can be treated as decoupled [12]. However, when the confining potential changes abruptly as electrons flow from a wide to narrow region (Fig. 1b), the off-diagonal blocks of the Hamiltonian are significantly greater than zero. These non-zero elements in the mode-

space Hamiltonian indicate that different modes couple as a result of changes in the device geometry, even in the absence of real scattering (geometric scattering).

Once the mode-space Hamiltonian is computed, we can evaluate the Green's function in mode-space at each injection energy (denoted by E_L and referred to as the longitudinal energy).

$$G_{\text{mode}} = \left[E_L I - h_{\text{mode}} - \Sigma_{\text{S/D}}(E_L) - \Sigma_{\text{Scatt}}(E_L)\right]^{-1} \tag{7}$$

In Eq. 7, $\Sigma_{\text{S/D}}$, represents the self-energy matrix which accounts for the coupling between the active device and the source-drain (S/D) contacts and Σ_{Scatt}, is the self-energy matrix which models the effect of scattering within the device. The procedure to set up the self-energies is discussed in the subsequent paragraph. Once the Green's function has been evaluated, all the quantities of interest such as the local charge density and the terminal currents can be derived using the procedures listed in [4] [17] [12] [18]. These quantities which are computed in mode-space, can then be mapped onto the real-space basis by applying the inverse of the transformation used in Eq. 6.

Deep inside the S/D contacts the electrostatic potential is assumed to be invariant (along x). Therefore, diagonal blocks of the Hamiltonian (α's in Eq. 1) repeat themselves within the source (α's with indices less than 1) and the drain (α's with indices greater than N_X). Since the mode-space basis is composed of the eigenvectors of α, changing the basis from real to mode-space diagonalizes α within the S/D regions.

$$\alpha_{\text{mode}} = \begin{bmatrix} \epsilon_1 + 2t_x & 0 & \cdots & \cdots & \cdots \\ 0 & \epsilon_2 + 2t_x & 0 & \ddots & \ddots \\ \cdots & 0 & \epsilon_3 + 2t_x & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \epsilon_{N_Z} + 2t_x \end{bmatrix} \tag{8}$$

The ϵ's in Eq. 8 represent the cutoff energies of the various modes within the contacts. The coupling matrix β, in Eq. 6, remains unchanged as it is proportional to the identity matrix (this matrix couples adjacent grid point along the x dimension within the same mode). Therefore, we can visualize the contacts as N_Z separate 1D conductors in parallel. In this representation the self-energy matrix can be expressed by extending the results in

[4]. For the simple example that we considered to illustrate the basis transformation (Eq. 5), the self-energy matrix is,

$$\Sigma_{\mathrm{S/D}}(E_L) = \begin{array}{c} \\ 1 \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \text{mode 1} \\ 1 \quad 2 \quad 3 \end{array} \begin{array}{c} \text{mode 2} \\ 1 \quad 2 \quad 3 \end{array} \left[\begin{array}{ccc|ccc} -t_x e^{ik_1^1 a} & 0 & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & 0 & \cdots & \cdots & \cdots \\ \cdots & 0 & -t_x e^{ik_3^1 a} & 0 & \cdots & \cdots \\ \hline \cdots & \cdots & 0 & -t_x e^{ik_1^2 a} & 0 & \cdots \\ \cdots & \cdots & 0 & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & 0 & -t_x e^{ik_3^2 a} \end{array}\right] \tag{9}$$

where $i{=}\sqrt{-1}$. The superscript on the k's in Eq. 9 refer to the mode index (the wave vector is specified with respect to the cutoff energy of the mode) and the subscript to the site index along x. Since the first and the last slices of the active device are the only regions which couple to the S/D contacts, the self-energy matrix representing the contacts has non-zero entries only at sites 1 and 3 (or alternatively 1 and N_X for the general case). It should be noted that the contact self-energy matrix has the same size as the mode-space Hamiltonian.

Scattering within our modeling framework is treated phenomenologically using Büttiker probes [19] [8]. Within this phenomenological framework scattering processes can be viewed as another set of contacts which couple to the device just like the source and drain. This implies that the self-energy describing scattering has the same form as Eq. 9. However, the fundamental difference between the real S/D reservoirs and those represented by Büttiker probes is that the probes can only change the electron energy and momentum, but not the electron number within the device. Also, these Büttiker probes (contacts) do not have a well defined Fermi level like the S/D. One can view each Büttiker probe as extracting electrons from the device, perturbing the energy of those electrons and reinjecting an equal number back into the device with a different energy distribution. The coupling energy between the device and the probes (analogous to t_x in Eq. 9) can be adjusted to vary the scattering strength smoothly from zero (ballistic) to a high value (diffusive). This strength can also be specified to mimic an average low field mobility as discussed in [18]. The probe self-energy in terms of the coupling strength for the example used to illustrate Eqs. 5 and 9 is,

$$\Sigma_{\text{Scatt}}(E_L) = \begin{array}{c} \\ 1 \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \text{mode 1} \\ \begin{array}{ccc} 1 & 2 & 3 \end{array} \end{array} \begin{array}{c} \text{mode 2} \\ \begin{array}{ccc} 1 & 2 & 3 \end{array} \end{array} \left[\begin{array}{ccc|ccc} -\frac{|P_1^1|^2}{t_x}e^{ik_1^1 a} & 0 & \cdots & \cdots & \cdots & \cdots \\ \cdots & -\frac{|P_2^1|^2}{t_x}e^{ik_1^1 a} & 0 & \cdots & \cdots & \cdots \\ \cdots & 0 & -\frac{|P_3^1|^2}{t_x}e^{ik_1^1 a} & 0 & \cdots & \cdots \\ \hline \cdots & \cdots & 0 & -\frac{|P_1^2|^2}{t_x}e^{ik_1^1 a} & 0 & \cdots \\ \cdots & \cdots & \cdots & 0 & -\frac{|P_2^2|^2}{t_x}e^{ik_1^1 a} & 0 \\ \cdots & \cdots & \cdots & \cdots & 0 & -\frac{|P_3^2|^2}{t_x}e^{ik_1^1 a} \end{array}\right] \tag{10}$$

where the coupling energy $|P|^2/t_x$, is adjusted to mimic the same low field mobility for electrons in each mode. The subscripts and superscripts on the probe strength in Eq. 10 have the same interpretation as in Eq. 9. The Fermi level of the probes at the ends of the device (μ_1 and μ_{N_X} in Fig. 1b) is fixed by the applied bias, while that of the probes in the interior of the device (μ_2 upto μ_{N_X-1} in Fig. 1b) is adjusted to ensure current continuity (net source injected current equals the net drain collected current). Since the probe Fermi levels are adjusted by forcing the net current at each probe to equal zero ($\int I_{\text{probe}}(E_L)dE_L = 0$), the scattering model presented in this work causes a relaxation of the channel directed (or longitudinal) electron energies [4] [18] and also results in strong intervalley scattering (scattering between subbands from different valleys). Note that the Fermi potential of all the probes within a single vertical slice is shorted as shown in Fig. 1b. Such a treatment is justified because there is no net current flow along the z dimension. The probe Fermi levels (derived from current continuity) are a direct measure of how applied voltage drops from the source to the drain (along x). These Fermi levels are visualized and discussed in Sec. III.

The calculation cycle begins with a guess for the two-dimensional (2D) potential $V(x, z)$. Knowing the potential profile, the real-space Hamiltonian (Eq. 1) is transformed to mode-space using Eqs. 5 and 6. Büttiker probes are introduced (Eq. 10) and a Newton method is used to adjust the probe Fermi levels to ensure current continuity. These Fermi levels are then used to compute the density matrix in mode-space for each conduction band valley. These density matrices are then transformed (inverse of the transformation in Eq. 6) to real-space and in that representation their diagonal entries yield the three-dimensional (3D) charge density at each node of the (x, z) grid. The 3D charge (summed over all valleys and

spin) is used to solve Poisson's equation to obtain a new potential profile, and the entire calculation cycle is repeated until a self-consistent solution is achieved.

In order to accurately sample the injection from the S/D leads, a very fine longitudinal energy grid with a spacing of $\Delta E_L = 0.5$mV or less needs to be used. For typical values of power supply voltages listed in the International Technology Roadmap for Semiconductors (ITRS) [20], the energy grid is composed of $\sim$1500 points. Manipulating matrices of size 1000 $\times$ 1000, $\sim$4500 times (3 $\times$ 1500 as there are three conduction band valleys) is a computationally challenging task (because all the columns of the Green's function in Eq. 7 need to be computed). Thus, even when solved in mode-space, the computational time per bias point for the device structure in Fig. 1b is $\sim$7 hours on 45, 1.2GHz processors of a Linux Cluster. All calculations in this work were performed on a Linux cluster using Matlab-6.1. Parallelization of the energy grid was achieved using a Message Passing Toolbox (MPI) interface for Matlab, developed at the University of Granada in Spain [21].

III. RESULTS

The simulated device structures are illustrated in Fig. 1. Both devices are ultra thin body, fully depleted, symmetric, n-MOSFETs with S/D regions doped at $10^{20}\,\mathrm{cm}^{-3}$ and an intrinsic channel. Their gate lengths are 10 nm, and there is no gate-to-S/D overlap. The junctions are abrupt, and the oxide thickness for both top and bottom gates is 1 nm. A body thickness of 1.5 nm, and a power supply (V_{DD}) of 0.4 V has been used in this simulation study. The oxide regions are treated as infinite potential barriers for electrons. The effect of channel access geometry on device performance is examined by comparing the characteristics of a device with abruptly flared out S/D regions (Fig. 1b), against those of a device with uniformly narrow S/D regions (Fig 1a) while fixing the overall device length (50 nm). A S/D mobility of 55 cm^2/V-s and a channel mobility of 200 cm^2/V-s is used to capture the effects of scattering. Henceforth, we shall refer to the device with the flared out geometry as the wide-to-narrow (WN) device and the device without the flared out S/D as the uniformly-narrow (UN) device.

A. The WN Device

In order to highlight quantum effects, especially within the WN device (Fig 1b), we plot the self-consistently calculated subband energies (obtained by solving for the eigenvalues of α in Eq. 2) and charge densities along the slices labeled A-A, B-B and C-C in Fig. 2 (in the on-state, V_{GS}=V_{DS}=0.4 V). For the WN device, the thickness of the wide S/D is 6.2 nm and these regions are terminated using hard-wall (infinite potential) boundary conditions along the z dimension. Therefore, charge in these regions is quantized. However, unlike the narrow region which being extremely thin, exhibits single subband occupancy, charge within the wide S/D is distributed over several subbands. Away from the WN constriction (section A-A), the shape of the electrostatic potential (along z) and hence the modes (eigenfunctions of α in Eq. 2) do not change appreciably along the x dimension. Therefore, different modes within section A-A can be treated as decoupled because the off-diagonal terms in the mode-space Hamiltonian (Eq. 6) are close to zero [12]. Along slice A-A, the electron population from each mode can be viewed as being distributed according to the individual mode eigenfunctions. In slice A-A, all of the charge resides in the lowest five unprimed (valleys with a heavy effective mass along z, and two fold degenerate) and the lowest two primed subbands (valleys with a light effective mass along z, and four fold degenerate). Since the mode eigenfunctions for the low energy unprimed modes are evanescent at the device center (note that these energies are below the conduction band maximum in Fig. 2a), the spatial charge distribution for the unprimed valleys is peaked away from the centerline of the device. On the other hand, the charge distribution for the primed valleys is parabolic, with a peak at the center of the wide regions (all of the primed mode energies are above the conduction band maximum as shown in Fig. 2a). The net charge (summed over all the bands) as expected, exhibits a uniform distribution except at the ends, where it smoothly drops to zero conforming to the hard-wall boundary conditions.

The internal picture along slice B-B (one node to the left of the constriction) is much different from that along A-A (Fig. 1b). The shape of the modes (wrt. z) change rapidly along the x dimension as electrons propagate from a wide (6.2 nm) to a narrow (1.5 nm) quantum well [12]. Therefore, along slice B-B, individual mode energies and mode eigenfuctions do not provide an accurate description of the effective potential energy and spatial distribution of electrons. Non-zero values of the off-diagonal elements within the coupled mode-space

Hamiltonian (Eq. 6) results in intra mode scattering (scattering between modes belonging to the same conduction band valley) even in the ballistic limit. In fact the charge density distribution (along z) is peaked at the device center for both the primed and unprimed valleys (Fig. 2b, center). Electrons from the unprimed valleys respond to the channel directed electric field with a light effective mass ($0.19\,\mathrm{m_0}$) and hence have large wavelengths (inversely proportional to the effective mass), while those from the primed valleys respond with a heavy effective mass (on average) and have short wavelengths. Therefore, the horizontal distance (along x, away from the oxide regions) over which the electron density increases (from zero) to the values along section A-A is greater for the unprimed valleys (as illustrated in Fig 2b, center). Since the electron waves have propagating solutions only within the narrow constriction, individual valley and the net charge distributions along B-B exhibit sharply increasing trends between the top and bottom silicon-oxide interfaces ($z = 0$ and $z = 1.5\,\mathrm{nm}$). At the constriction (section C-C), there is only one propagating mode (from the unprimed valleys) within the energy range over which electrons are injected into the narrow region. All the other modes are evanescent and decay exponentially (note that the electon density from the primed valleys is much below that from the unprimed valley along section C-C). The electron distribution within section C-C is peaked at the device center and conforms to the hard-wall boundary conditions used to terminate the Hamiltonian at the oxide-silicon interface.

The internal pictures presented so far clearly indicate that the solution scheme captures all of the quantum effects within the WN geometry accurately. In order to discuss the role of scattering, we plot the subband energies along section B-B in Fig 2a (right). The subband energy of the lowest unprimed mode at the constriction (C-C) is also superposed on this plot (marked line). Note that although these energies represent a rough estimate of the potential energy for electrons (due to the mode coupling issues discussed earlier), it is clear that scattering (especially intervalley) is an important aspect of transport within the WN structure. Scattering enhances the ability of the evanescent modes to transfer charge into the propagating modes at the constriction. If the effect of scattering were not included, there would be no interaction between the primed and unprimed modes at sections B-B and C-C despite their being energetically very close. Although these internal plots correspond to the on-state ($V_{GS} = V_{DS} = V_{DD}$) of operation, the overall picture within the WN device is similar for all states of biasing. The internal picture throughout the UN device resembles that along section C-C (the UN device exhibits single mode occupancy everywhere).

Interesting features of the solution scheme are illustrated in Fig. 3 where the energy (channel directed energy) resolved local density of states (LDOS) within the WN device is visualized in the on-state (with scattering) for both the unprimed and primed valleys. White areas in the figure represent a large density of states, while dark regions indicate a low LDOS. The profile of the first unprimed subband is also superposed on the plots in order to provide a rough estimate of the minimum potential energy for electrons in different regions of the device (note that the subband profile is discontinuous at the WN interface, because of an abrupt change in the width of the quantum well from 6.2 nm to 1.5 nm). The well demarcated subband structure in the LDOS spectrum away from the WN interface is gradually lost due to mode coupling as one approaches the WN interface. Reflections at the WN interface ($x =$ -10 nm and $x = 20$ nm) result in a reduced LDOS at the constriction, which then sharply increases into the large S/D reservoirs (also evident in the 2D charge density plotted in Fig. 6b). Figure 3a clearly indicates that only the unprimed valleys contribute to the LDOS within the narrow region. Modes belonging to the primed valleys are completely reflected and decay exponentially within the narrow region. These modes indirectly inject current into the narrow region through inelastic scattering processes. Inelastic scattering also destroys coherent oscillations in the LDOS and the LDOS decays to zero within the forbidden regions below the first unprimed subband energy.

B. Effect of differing channel access geometry

Figure 4 presents the self-consistently simulated I_{DS} vs. V_{GS} and I_{DS} vs. V_{DS} characteristics for both the UN and WN device geometries, including the effects of scattering. For the same gate work function both geometries exhibit identical off-currents (~ 7 A/m), subthreshold slopes (subthreshold swing is ~ 100 mV/dec) and threshold voltages (Fig. 4a). Also, for the same off-current, the simulated on-current for the device with a WN geometry is only $\sim 10\%$ lower than that of the device with a UN geometry. Note that this difference in the on-current represents the worst case scenario. In reality, the geometry of the S/D regions would not be hyper-abrupt, but would flare out gradually in order to reduce the parasitic gate-to-S/D capacitances. For such geometries it is possible that the differences in the on-current between UN and WN devices could be even lower than 10%. The output characteristics for both device structures are weakly saturated because of the low power

supply voltage used in this study and due to series resistance effects . Simulated I-V characteristics for the WN device can be understood by viewing the large S/D regions as ideal reservoirs. These regions which maintain a near equilibrium distribution (due to multiple subband occupancy and scattering), inject and extract electrons from the narrow region while maintaining macroscopic charge neutrality. Therefore, current flow depends on the ease with which electrons can enter (exit) and propagate through the narrow region. We compare the transport properties of the UN and WN devices by focusing on the internal picture once again.

The conduction band profiles (averaged along z) in the off-state (V_{DS}=0.4 V, V_{GS}=0 V) for the WN and UN devices are compared in Fig. 5a (left). Note that although the conduction band profiles for the UN and WN devices match within the narrow region, there is a large built-in potential at the interface between the wide and narrow regions of the WN device (Fig. 5a, left). To explain the origin of this built-in potential we note that both the wide and the narrow S/D regions are degenerately doped ($10^{20}\,\mathrm{cm}^{-3}$). For the same degenerate bulk doping value, the equilibrium Fermi level within a neutral narrow semiconductor is much higher than the conduction band edge of a comparably doped, neutral wide semiconductor, because only one subband is occupied in the narrow semiconductor while several are occupied in the wide semiconductor. Therefore, when the wide and narrow portions are intimately connected, electrons spill from the narrow into the wide region (regions of high to low Fermi energy), and a built-in potential develops to equalize the Fermi levels.

The corresponding subband profiles (eigenvalues of α in Eq. 2) are visualized in Fig. 5a (right). It should be remembered that within the WN device, these subbands have clear physical meaning away from the WN constrictions. They have been visualized in order to qualitatively understand the off-state behavior of the I-V characteristics. Note once again that these subband energies are discontinuous at the interface between the wide and the narrow regions due to an abrupt change in the width of the quantum well (6.2 nm to 1.5 nm). Since the first subband is all that matters within the narrow region (x = -10 to 20 nm), high energy modes are not plotted within this region in Fig. 2a (right). Note that scattering (both geometric and real) between different modes at the WN interface ensures that electrons get injected into the single mode propagating within the narrow region (this mode self-consistently lines up between the third unprimed and the first primed subbands as shown in Figs. 2a and 5a). However, these electrons are reflected off the large source-

to-channel barrier (within the narrow region) in the off-state. Such a barrier is also present within the UN device (where the subband has a clear meaning everywhere) as seen from Fig. 2a (right). Since the height of this barrier (governed by the gate work function) is the same for both the UN and WN devices, it is clear that their off-currents should be identical. In Sec. II, we mentioned that the Fermi level of each Büttiker probe (which represents an isolated scattering center) is adjusted to ensure current continuity. Therefore, these Fermi potentials can be treated as a measure of how the applied source-to-drain voltage drops within the device. In the off-state, the channel is unconducting and we expect all of the applied voltage to drop in the channel region irrespective of the device geometry. This is clearly evident from Fig. 5b, where the scatterer Fermi levels are plotted along the channel for both, the UN and WN geometries.

The self-consistently calculated conduction and subband profiles in the on-state (with scattering) are plotted in Fig. 6a. In the on-state, the potential barrier for electrons is supressed (compared to Fig 5a) due to the large voltage applied to the gate electrodes. The on-current of the WN device depends on the ability of the large S/D reservoirs to inject (extract) electrons from the narrow region and the transport properties of the narrow region itself. As mentioned earlier inter subband scattering aids this injection and extraction process. For the UN device, the S/D reservoirs are natural extensions of the intrinsic device ($x =$ -10 nm to 20 nm). However, this is not the case within the WN device as evidenced by the 2D charge density (integrated along z) plots in Fig. 6b. In our discussion of the off-state we mentioned that an intimate connection between comparably doped, neutral, wide and narrow n^+ regions results in a tranfer of charge from the narrow to the wide region. Such a charge transfer occurs at the wide-to-narrow source and drain interfaces within the WN device. This charge transfer results in a depletion region which extends into S/D regions of the intrinsic device (Fig. 6b). As the drain voltage is increased from zero to $V_{DD} = 0.4\,V$, the $n^+(\text{wide}) - n^+(\text{narrow})$ junction at the source becomes reverse biased thereby extending the depletion region further into the intrinsic source. On the other hand, the positive drain voltage forward biases the $n^+(\text{narrow}) - n^+(\text{wide})$ junction on the drain side thus shrinking the depletion region as shown in Fig. 6b. Note that transport properties of the UN device are identical to those of the intrinsic WN device because they have same S/D extension and channel mobilities and also exhibit single mode occupancy. Therefore, observed differences ($\sim 10\%$ with scattering turned on) in the on-current between the UN and WN devices (for

the same off-current) are due to the depletion effects coupled with quantum mechanical reflections at the WN interface.

The effect of self-consistency is also evident in Figure 6b. Self-consistent electrostatics causes macroscopic charge neutrality (net 2D charge equals the net 2D doping) to be achieved within both, the large S/D reservoirs as well as the narrow S/D extensions (away from the WN interface) of the WN device. Due to the abruptness of the S/D-to-channel junctions and the high doping within these narrow S/D extensions, the effect of fringing fields from the gate within the WN device has little or no effect on the on-state charge density. In fact, the 2D charge profile within the UN device matches that of the WN device almost identically in the narrow region as shown in Fig. 6b.

In the on-state the channel is conducting, so a significant fraction of the applied source-to-drain voltage drops within the S/D reservoirs and extensions. As illustrated in Fig. 7a, where the Büttiker probe Fermi potentials are plotted in the on-state (V_{GS}=0.4 V, V_{DS}=50 mV), the applied voltage drops linearly within the S/D regions of the UN device. Within the WN device, however, almost all of the voltage drop in the heavily doped S/D regions occurs at the interface between the wide and the narrow regions. We can interpret the nature of this voltage drop by deriving a sheet resistivity ($\frac{\partial \mu}{\partial x}/\frac{I_{DS}}{W}$) at low V_{DS}, which is plotted in Fig 7b [22]. Both the UN and WN devices exhibit the following resistances: 1) quantum contact resistance 2) S/D resistance 3) tip resistance and 4) channel resistance. In addition, the WN device exhibits a large quantum-mechanical spreading resistance at the interface between the wide and the narrow regions. Note that within the S/D reservoirs of the WN device, the large number of propagating modes causes the sheet resistivity to be lower than than that of the UN device (single mode conduction). Within the intrinsic portion of the WN device, however, the nature of the voltage drop as well as the sheet resistance looks identical to that of the UN device (same mobilities and single mode occupancy).

Next, we examine the effect of the wide-to-narrow aspect ratio. If we progressively increase the thickness of the S/D regions (beyond 1.5 nm, which corresponds to the UN device) and examine the on-current, we obtain the trend shown in the inset of Fig. 7b. As the thickness of the S/D regions is increased, the bulk S/D resistance decreases (number of propagating modes increases), while the spreading resistance increases (quantum mechanical reflections), thus resulting in an optimum value of the on-current ($\sim$5 % higher than that of the UN device). A further increase in the thickness of the S/D regions causes the increasing

spreading resistance to dominate, thus resulting in a decrease in the on-current. However, beyond a certain thickness value (wide-to-narrow aspect ratios exceeding 4), the spreading resistance and hence the on-current saturates at a value which is $\sim 10\,\%$ less than that of the UN device. This behavior of the resistivity and the on-current in the presence of scattering clearly demonstrates that an understanding of the intrinsic device (Fig. 1b) is sufficient to explain the transport properties of devices with different S/D geometries.

Figure 8a plots the self-consistently calculated current spectrum at the source and drain contacts (slices 1 and N_X in Fig. 1) as a function of the channel directed energy for both the UN and the WN devices in the on-state (while reading the plot it should be noted that the source injected current is positive for electrons entering the device, while the drain collected current is negative for electrons leaving the device). Since Büttiker probes simulate the effect of inelastic scattering, we observe that the drain collected current is relaxed in energy when compared to the source injected current for both, the UN and the WN devices. Within the UN device which exhibits single mode occupancy, the source injected current smoothly increases for energies above the subband edge and finally tails to zero as the current at the source end is cutoff by the source Fermi function. Within the WN device, current at the source is injected into several subbands (their energies are indicated by the vertical lines in Fig. 8a). Therefore the current spectrum for the WN device exhibits a stepped behavior, with each step occurring at the energy at which a specific mode starts propagating. There is a monotonic increase in the current spectrum (till it is cut off by the source Fermi function) because both the primed and the unprimed modes within the wide source reservoir can inject current into the single unprimed mode within the narrow region due to scattering.

If we compare this picture to the current spectrum obtained by ballistically simulating the WN device non-selfconsistently (where the 2D potential profile is fixed from the scattering simulations), we observe large differences in the current spectrum (Fig. 8b) for the UN and WN devices. Since scattering is turned off, there is no energy relaxation of the source injected current. Therefore, the source and drain current spectra are symmetric. Also, the source current spectrum for the WN device does not increase monotonically. Coherent quantum mechanical reflections at the WN interface generate strong oscillations in the current spectrum and greatly degrade the current flowing through the WN device when compared to the UN device (for example, note the clear reduction in current at energies around the first primed mode when compared against Fig. 8a). A ballistic simulation (non-selfconsistent)

of the I-V characteristics for the UN and WN devices indicates that the computed current values could differ by as much as $\sim 250\,\%$ (as opposed to $\sim 10\,\%$, when scattering is turned on) due to coherent reflections caused by an abrupt change in the device geometry (Fig. 9). Such differences have been recently reported in the literature [23] [24]. These results clearly emphasize the importance of including the effects of scattering when performing a detailed design analysis of nanoscale transistors operating at room temperature.

IV. DISCUSSION

In this section we discuss an approach which can be applied to design and assess the performance of nanoscale silicon transistors efficiently. Device design and optimization requires the ability to perform a large number of simulations within a reasonable time period. Even for the small device dimensions considered in this paper (Fig. 1b), the simulation time per bias point on a Linux cluster was very large ($\sim$7 hrs). The computational burden was large because we explicitly included the large S/D reservoirs within our quantum simulation domain. In real devices these regions have an area which is much larger than $6.2\,\text{nm} \times 10\,\text{nm}$ in order to reduce the silicide resistance (not considered in our study), and are virtually impossible to handle within a quantum mechanical framework. Therefore, we examine the possibility of decomposing the simulation domain into an instrinsic (region within the dashed box in Fig. 1b and parasitic components (large S/D reservoirs). Knowing the characteristics of the intrinsic device and the parasitic resistances, the actual device performance (extrinsic characteristics) can be obtained by solving the following equations.

$$\begin{aligned} V_{GS} &= V_{GS}^{intrinsic} + I_{DS} R_S \\ V_{DS} &= V_{DS}^{intrinsic} + I_{DS} \left[R_S + R_D \right] \end{aligned} \tag{11}$$

where V_{DS} and V_{GS} are the extrinsic voltages. Note that in order to simulate the intrinsic device only a few modes need to be included when expanding the device Hamiltonian in mode-space (Eq. 6). Also note that within the intrinsic device each mode can be treated as decoupled [12]. Therefore, the computational burden for quantum mechanically simulating the intrinsic device is extremely small. The parasitic resistances of the extrinsic S/Ds can be extracted using a classical simulator and used in Eq. 11.

Simulation of the intrinsic device requires an appropriate set of boundary conditions.

Within the intrinsic device the transport equation (NEGF) is solved by imposing fixed boundary conditions on the Fermi potentials of the source and drain contacts (specified by V_{DS}). However, we impose floating boundary conditions when solving Poisson's equation. This boundary condition is realized by setting,

$$\hat{x} \cdot \vec{\nabla} V = 0 \tag{12}$$

at the source and drain ends of the simulation domain. Conventional transport models use fixed potential boundary conditions assuming equilibrium statistics and charge neutrality at the contacts. Under non-equilibrium conditions, equilibrium statistics no longer apply within the intrinsic device [25], and our use of the floating boundary condition helps us capture the effect of coupling the intrinsic device to the large scattering dominated contacts. Note that even if the potential is allowed to float, it cannot float to any arbitrary value. The potential floats relative to the source/drain Fermi levels in order to achieve charge neutrality (Fig. 6b). To explain the floating boundary condition, we plot the averaged conduction band profile for the device shown in Fig. 1b under equilibrium ($V_{GS} = 0.4V, V_{DS} = 0V$) and non-equilibrium ($V_{GS} = 0.4\,V, V_{DS} = 0.4\,V$) conditions (with scattering) in Fig. 10. The large S/D regions maintain a near equilibrium distribution even when a large bias is applied to the drain, because of scattering between the many modes which contribute to conduction in this region. Therefore, the electrostatic potential within this region is unaffected even when a large bias is applied to the drain as shown in Fig. 10. However, within the intrinsic device the conduction band floats to lower values as the drain bias is increased [26] (Fig. 10). Under equilibrium conditions, both the $+k_x$ and the $-k_x$ states in the intrinsic source are filled by a single Fermi level resulting in zero net current. As the drain bias is increased to higher and higher values, the drain injected half of the electron distribution is suppressed within the intrinsic source. To achieve charge neutrality in this region (Fig. 6b), the conduction band floats to lower energies, an effect that is captured through the imposition of floating boundary conditions [26].

In order to verify our domain decomposition algorithm we simulate a resistor (using NEGF) with ideal, non-uniform contacts in the linear response regime (inset of Figure 11a). The resistor has the dimensions of the wide S/D reservoirs for the device in Fig 1b. The left contact to the resistor is 6.2 nm deep, while the right contact is 1.5 nm deep. The dimension of the right contact is chosen to simulate the coupling of the source reservoir to the narrow

intrinsic device. All the regions of the resistor which are not covered by contacts, are terminated using hard-wall boundary conditions. Simulated charge density and Büttiker probe potentials for this resistor are plotted in Fig. 11a. As expected, reflections at the right edge causes the charge density to drop at the right end of the resistor. The 2D charge density gradually increases to the charge neutral value over a few Debye lengths away from the right contact. Since the charge density is non-uniform, the applied voltage drops non-linearly. The quasi-Fermi potential (of the Büttiker probes) drops rapidly in regions where the charge density is reduced in order to conserve current. The nature of this voltage drop is similar to the actual profile of the Büttiker probe potentials illustrated in Fig 7a (for the WN device). Based on our calculations in the linear response region, the estimated parasitic source (drain) resistance was $\sim 57\,\mathrm{ohm} - \mu\mathrm{m}$. This resistance value was used to estimate the I-V characteristics of the WN device shown in Fig 1b, using the intrinsic I-V and Eq. 11. A comparison (Fig. 11) between the I-V characteristics extracted using our domain decomposition algorithm and the actual I-V (from a quantum simulation of the entire device) characteristics presented in Fig 4b (for the WN device) indicates that our domain decomposition scheme provides a fairly accurate and efficient technique for analyzing and designing nanoscale transistors [27] [28]. The S/D reservoirs and the intrinsic device can be treated independently.

V. SUMMARY

We presented a computationally efficient method to quantum mechanically simulate electron transport within nanoscale transistors including the effect of scattering. This modeling scheme, which is based on an expansion of the effective mass Hamiltonian in coupled mode-space, was applied to nanoscale transistors with differing channel access geometries. Our simulation study indicated that devices with very different channel access geometries exhibit nearly identical current-voltage characteristics (no more than $\sim 10\,\%$ differences in the worst case) once the effect of scattering was included. However, when simulated ballistically, devices with different access geometries yielded largely differing current-voltage characteristics due to coherent reflections (as much as $\sim 250\,\%$ in the worst case).

We then proposed a domain decomposition algorithm which can be applied to efficiently assess and design nanoscale transistors at the end of the roadmap. This algorithm divided

the device into intrinsic and parasitic components, simulated each component independently, and extracted the overall device characteristics through a simple interpolation scheme. A comparison of the extracted data against data obtained from rigorous quantum mechanical simulations of the entire device clearly showed that this domain decomposition algorithm is an accurate method for analyzing the transport properties of transistors at the end of the roadmap.

VI. ACKNOWLEDEGEMENTS

The authors are indebted to Junghoon Rhew, Avik Ghosh and Magnus Paulsson for useful discussions. Ramesh Venugopal is supported by an Intel Ph.D Fellowship and the SRC.

REFERENCES

[1] B. Doris, M. Ieong, T. Kanarsky, R. Roy, O. Documaci, Z. Ren, F. Jamin, L. Shi, W. Natzle, H. Huang, et al., in *IEDM Tech. Digest* (2002), pp. 267–270.

[2] www.intel.co/labs (2001).

[3] S. Datta, J. Phys. Condens. Matter **2**, 8023 (1990).

[4] S. Datta, Superlattices and Microstructures **28**, 253 (2000).

[5] H. Majima, Y. Saito, and T. Hiramoto, in *IEDM Tech. Digest* (2001), pp. 733–736.

[6] M. Je, S. Han, I. Kim, and H. Shin, Solid-State Electronics **44**, 2207 (2000).

[7] Y. Cui, Z. Zhong, D. Wang, W. Wang, and C. Lieber, Nano Lett. **3**, 149 (2003).

[8] M. Buttiker, Phys. Rev. Lett. **57**, 1761 (1986).

[9] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press,UK, 1997).

[10] M. Lundstrom, IEEE Electron Dev. Lett. **3**, 293 (2001).

[11] R. Ohba and T. Mizuno, IEEE Trans. Electron Dev. **3**, 338 (2001).

[12] R. Venugopal, Z. Ren, D. Jovanovic, S. Datta, and M. Lundstrom, J. Appl. Phys. **92**, 3730 (2002).

[13] A. Svizhenko, M. Anantram, T. Govindan, B. Biegel, and R. Venugopal, J. Appl. Phys. **91**, 2343 (2002).

[14] A. Szafer and A. Stone, Phys. Rev. Lett. **62**, 300 (1989).

[15] S. Nonoyama, K. Ishibashi, Y. Aoyagi, and S. Namba, Appl. Phys. Lett. **60**, 234 (1992).

[16] P. Damle, A. Ghosh, and S. Datta, in *Molecular Electronics* (American Scientific Publishers, Stevenson Ranch, California, 2003).

[17] R. K. Lake and S. Datta, Phys. Rev. B **45**, 6670 (1992).

[18] R. Venugopal, M. Paulsson, S. Goasguen, S. Datta, and M. Lundstrom, J. Appl. Phys. **93**, 5613 (2003).

[19] C. Texier and M. Buttiker, Phys. Rev. B **62**, 7454 (2000).

[20] *www.itrs.net* (2001).

[21] J. Fernandez, in *Lecture notes in Computer Science* (Springer-Verlag, 2002).

[22] Y. Taur and T. Ning, *Fundamentals of VLSI Devices* (Cambridge University Press Cambridge, UK, 1998).

[23] S. Laux, A. Kumar, and M. Fischetti, in *IEDM Tech. Digest* (2002), pp. 715–718.

[24] S. Laux, A. Kumar, and M. Fischetti, IEEE Trans. Nanotechnology **1**, 255 (2002).

[25] J. Rhew, Z. Ren, and M. Lundstrom, Solid-State Electronics **46**, 1899 (2002).

[26] R. Venugopal, Z. Ren, and M. Lundstrom, *to appear in ieee transactions on nanoelectronics* (2003).

[27] Z. Ren, R. Venugopal, S. Datta, and M. Lundstrom, in *IEDM Tech. Digest* (2001), pp. 107–110.

[28] S. Hasan, J. Wang and M. Lundstrom, submitted to IEEE. Trans. Electron. Dev. (2003).

FIGURE CAPTIONS

Fig.1: (a) An ultra-thin body DG MOSFET structure with a S/D doping of $10^{20}\,\mathrm{cm}^{-3}$ and an intrinsic channel (channel thickness = 1.5 nm). The S/D regions are 20 nm in length, the gate length is 10 nm and the S/D-channel junctions are abrupt. A slice of the device within which a 1D, z directed effective mass equation is solved, is also indicated. (b) A thin body DG MOSFET with abruptly flared out S/D regions. The length of the flared regions is 10 nm, their thickness is 6.2 nm and the overall device length is 50 nm (same as that for device a). The region within the dashed box is the intrinsic device.

Fig 2: (a) The conduction and subband profiles along slices A-A (left) and B-B (right) for the WN device (Fig. 1b). Unprimed mode energies are illustrated using solid horizontal lines while primed mode energies are represented by dashed horizontal lines. The first unprimed mode within slice C-C is also superposed (right) and marked (triangles). (b) The charge distribution along slices A-A, B-B and C-C is presented. The net charge (dashed and dotted line) is obtained by summing charge contributions from the unprimed (solid line) and primed (dashed line) valleys.

Fig. 3: (a) The energy resolved LDOS spectrum within the WN device (fig. 1b) in the on-state, for the unprimed valleys. Note that all of the states within the narrow region are due to the unprimed valleys in the energy range of interest. (b) The energy resolved LDOS spectrum within the WN device from the primed valleys. States belonging to the primed valleys are reflected at the WN interface and decay exponentially within the narrow region. The first unprimed subband (dashed line) is also superposed on the plots. This subband is discontinuous at the WN interface due to an abrupt change in the width of the quantum well (from 6.2 nm to 1.5 nm).

Fig 4: (a) Simulated I_{DS} vs. V_{GS} characteristics for the UN (dashed line) and WN (solid line) devices at V_{DS}=0.4 V (including the effect of scattering). Both devices exhibit the same off-current (~ 7 A/m) and subthreshold slopes. (b) Simulated I_{DS} vs. V_{DS} characteristics for the UN (dashed line) and WN (solid line) devices at V_{GS}=0.4 V. For the same off-current, the simulated on-current for the WN device is $\sim 10\,\%$ lower than that of the UN device.

Fig 5: (a) The averaged conduction band profile (left) in the off-state (including the effect of scattering) for the UN (dashed line) and WN (solid line) devices. Note the large built-in potential at the WN interface. The corresponding subband profiles are visualized on the right. Many subbands (dots) are occupied in the S/D regions of the WN device.

These modes inject current into the single propagating mode within the narrow region due to scattering (modes line up as shown in Fig. 2b). The UN device exhibits single mode occupancy everywhere (dashed line). (b) The Büttiker probe potentials within the UN (triangles) and WN (solid line) devices indicates that all of the applied voltage drops within the unconducting channel in the off-state.

Fig 6: (a) The average conduction band profile (left) in the on-state (including the effect of scattering) for the UN (dashed line) and WN (solid line) devices. Note that the source-to-channel barrier is depressed due to the large gate voltage and that differences in the on-current between the UN and WN devices reflects the ability of the large S/D regions (in the WN device) to inject and extract electrons from the narrow region. (b) The 2D charge density within the UN (dashed line) and WN (dotted and solid lines) devices in the on-state. The S/D regions within both, the UN and WN devices exhibit charge neutrality (2D charge equals the net 2D doping).

Fig 7: (a) The profile of the Büttiker probe potentials within the UN (dashed line) and WN (solid line) devices in the linear response regime (V_{GS}=0.4 V and V_{DS}=0.05 V). Note that the applied voltage drops nonlinearly within the S/D regions of the WN device. Most of the voltage drops at the WN interface. (b) Derived sheet resistivity plots within both, the UN (dashed line) and WN (solid line) devices indicate the following resistances: 1) quantum contact resistance 2) S/D resistance 3) tip resistance and 4) channel resistance. In addition, the WN device also exhibits a large spreading resistance. This resistance which degrades the on-current of the WN device when compared to the UN device, saturates as the thickness of the S/D regions is increased beyond a certain value (inset).

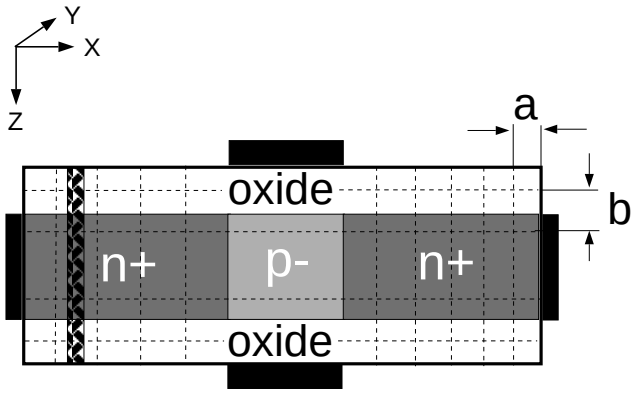
Fig 8: (a) The current spectrum within the UN (dashed line) and WN (solid line) devices in the on-state with scattering. Inelastic scattering causes the source injected current spectrum to relax in energy. Within the UN device current at the source is injected into a single mode. Within the WN device the source current is injected into several unprimed (solid vertical lines) and primed (dashed vertical lines) modes. Scattering, at the WN interface enables these modes to inject current into a single mode within the narrow region. (b) The ballistically (non-selfconsistent) calculated current spectrum within the UN (dashed line) and WN (solid line) devices. In the absence of scattering, there is no energy relaxation of source injected current spectrum and S/D current spectra are symmetric. Coherent reflections, due to abrupt changes in device geometry greatly degrade current flow within the WN

device when compared to the UN device.

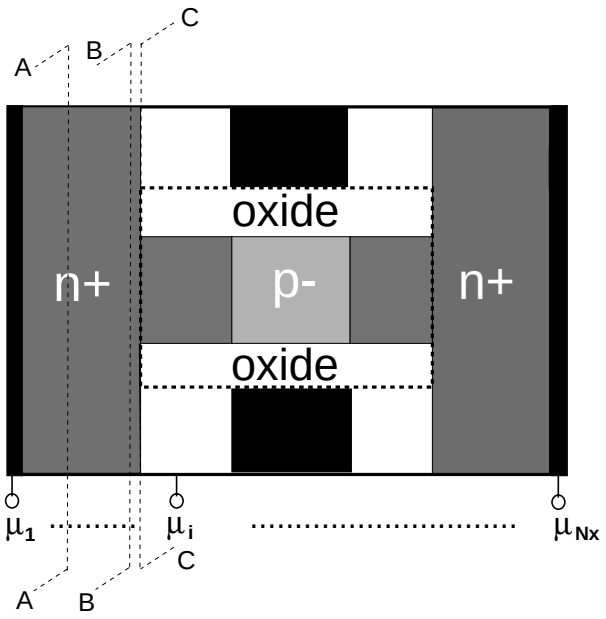
Fig 9: Ballistically simulated (non-selfconsistent, potential profiles obtained from self-consistent scattering simulations are used) I-V characteristics for the UN (dashed line) and WN (solid line) devices indicate that the on-currents for the two device structures could differ by as much as $\sim 250\,\%$. In contrast, once scattering is included, the I-V characteristics for the two devices differ by no more that $\sim 10\,\%$.

Fig 10: The averaged conduction band profile within the WN device (on-state) at equilibrium (solid line) and on the application of a large drain voltage (dashed line, V_{DS}=0.4 V). It is clear that the conduction band edge is unchanged within the large S/D reservoirs as they maintain a near equilibrium distribution even when a large drain bias is applied. The conduction band within the intrinsic source (Fig 1b) floats to lower values to enforce charge neutrality (fig 6b).

Fig 11: (a) Simulated 2D charge and Büttiker probe potential profiles within a resistor with non-uniform contacts (inset) in the linear response regime. Note that the nature of the applied voltage drop is similar to the drop with the S/D regions of the WN device pictured in Fig. 7a. The quasi-Fermi potential drops steeply in the regions where the 2D charge density decreases to enforce current continuity. (b) Quantum mechanically simulated I-V characteristics (solid line) are compared against the I-V characteristics obtained using the domain decomposition algorithm (dashed line with triangles). The I-V characteristics of the intrinsic device is also illustrated (dashed line).



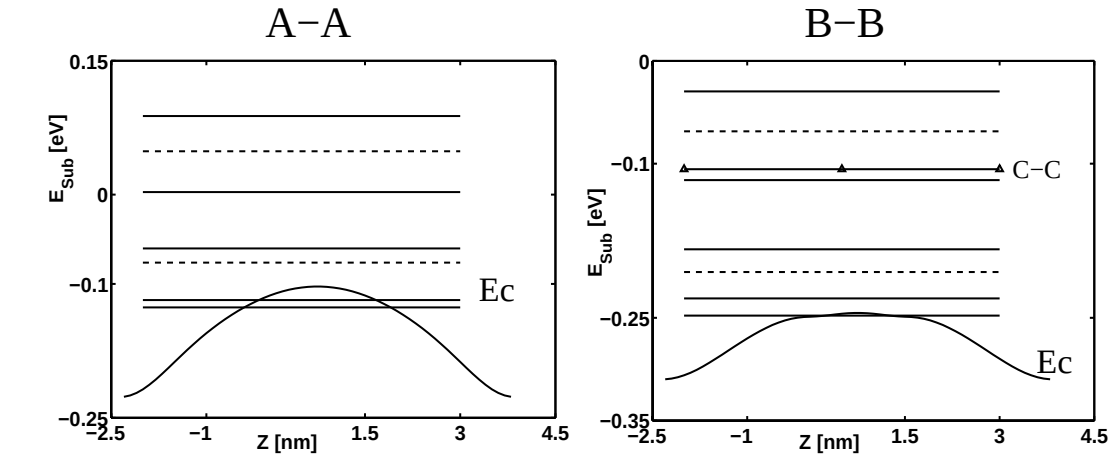
(a)



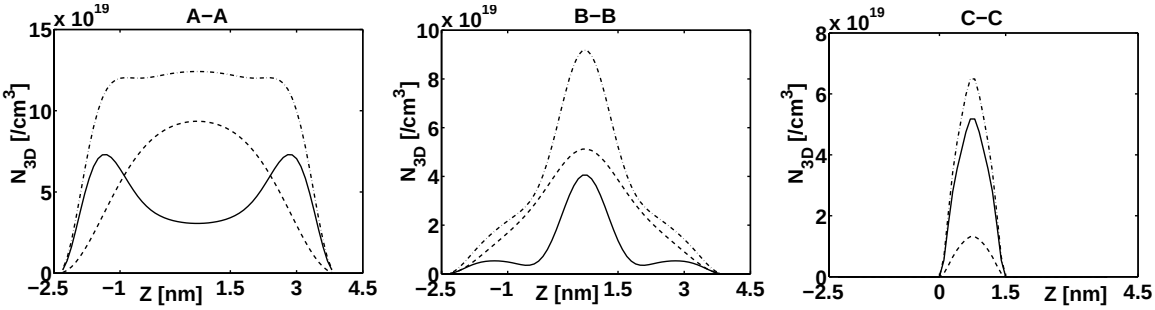
(b)

FIG. 1:

R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom



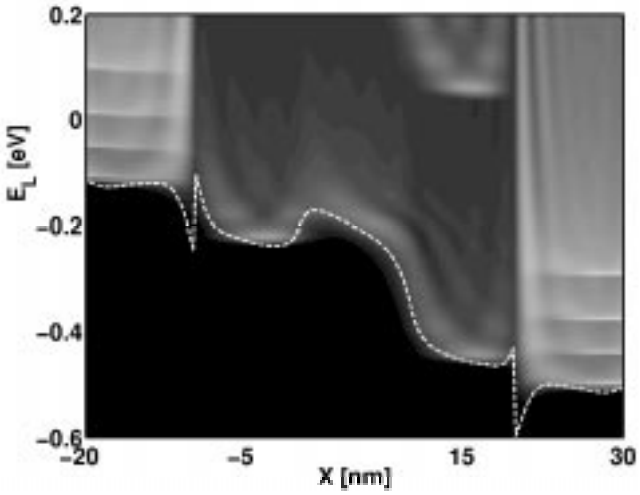
(a)



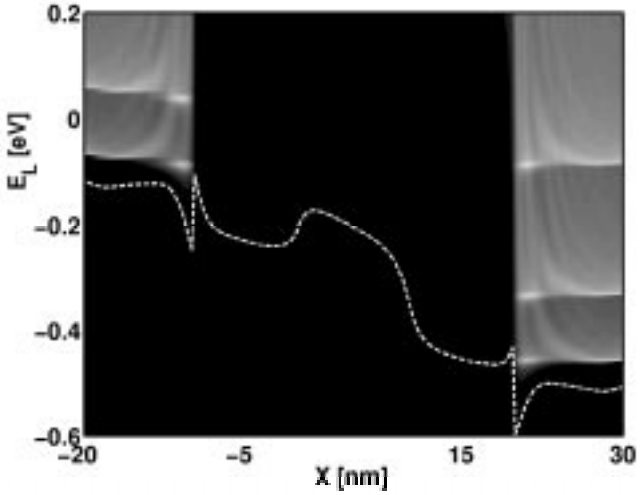
(b)

FIG. 2:

R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom



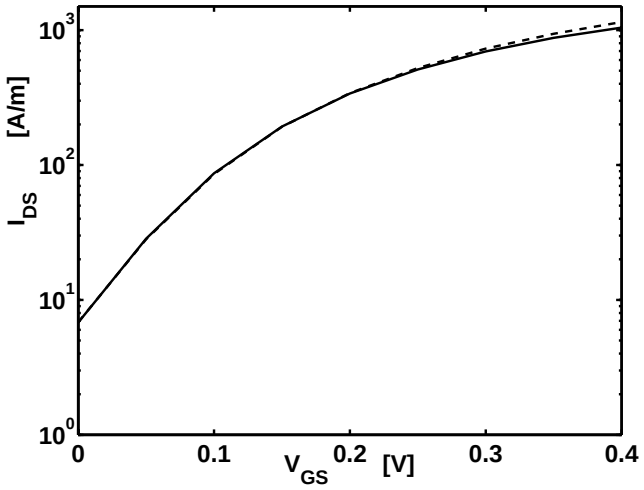
(a)



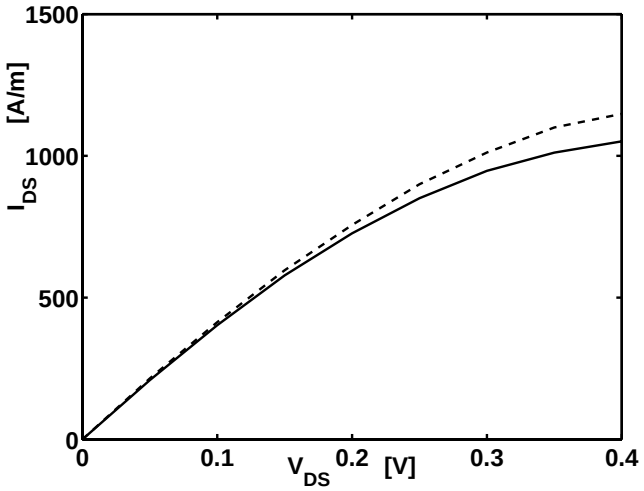
(b)

FIG. 3:

R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom



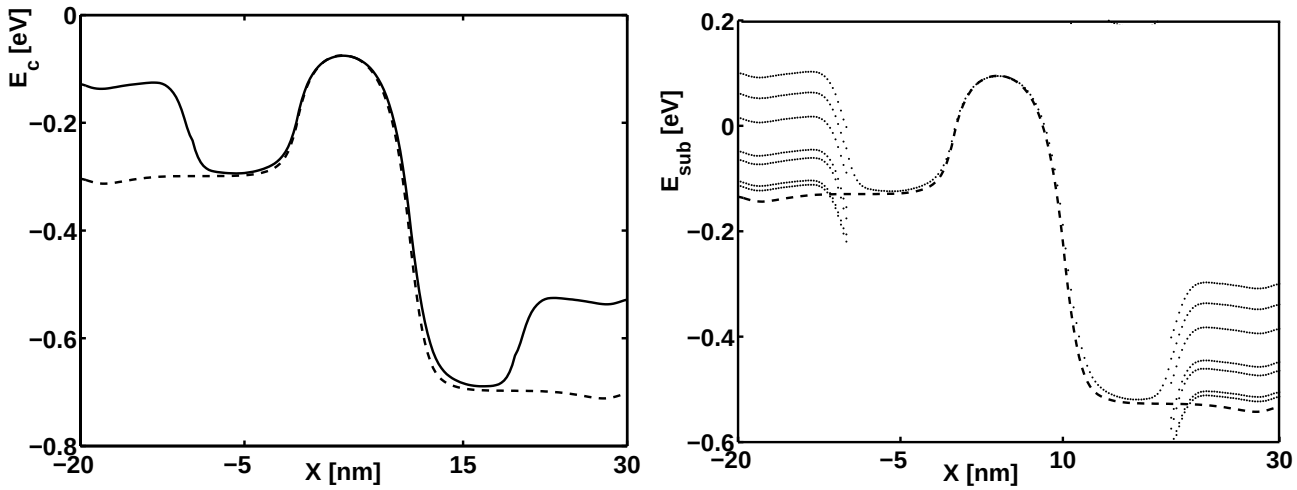
(a)



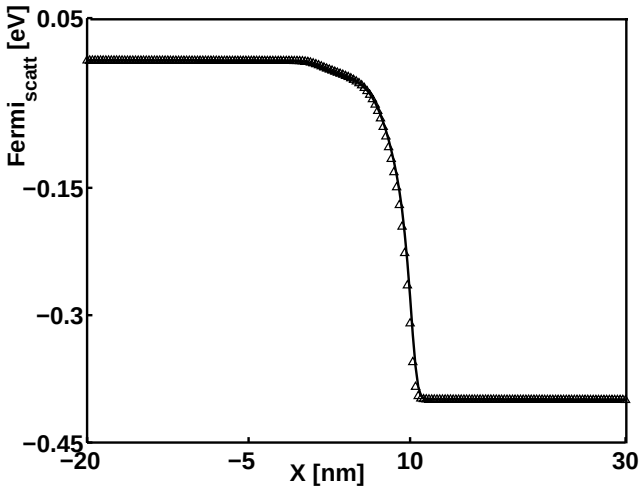
(b)

FIG. 4:

R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom



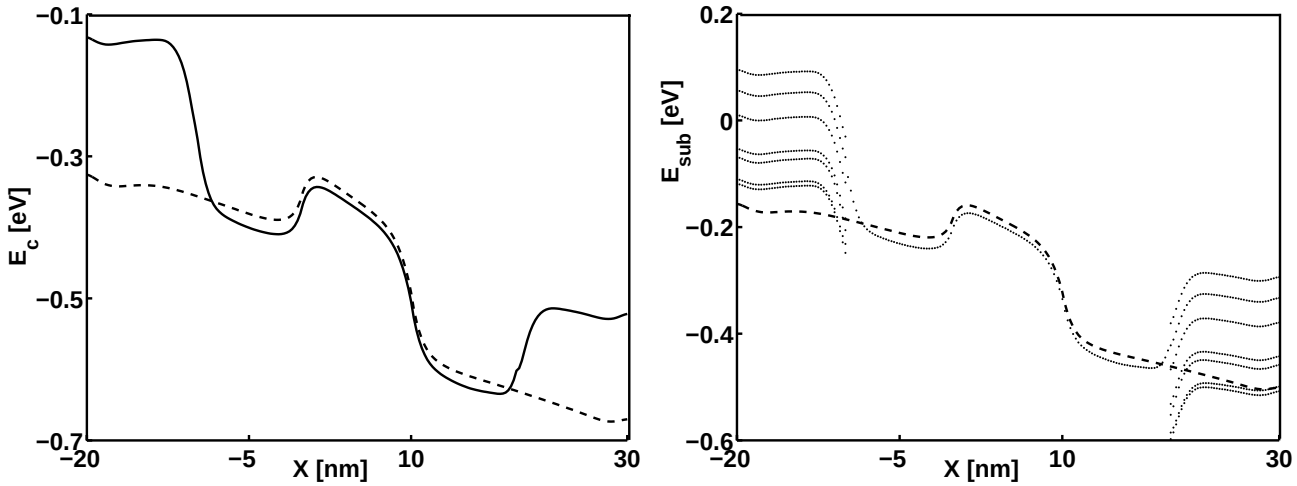
(a)



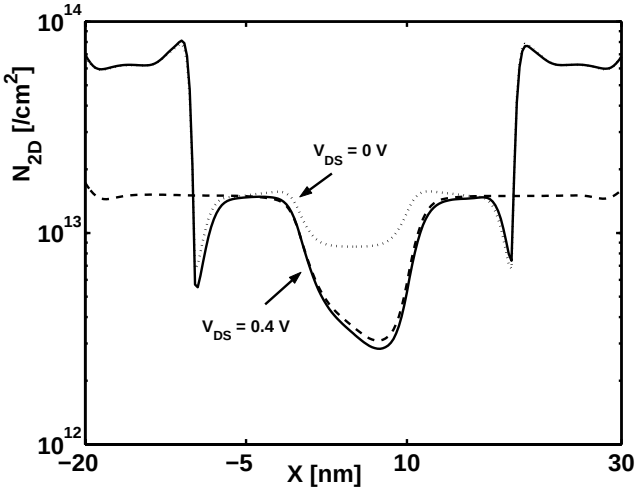
(b)

FIG. 5:

R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom



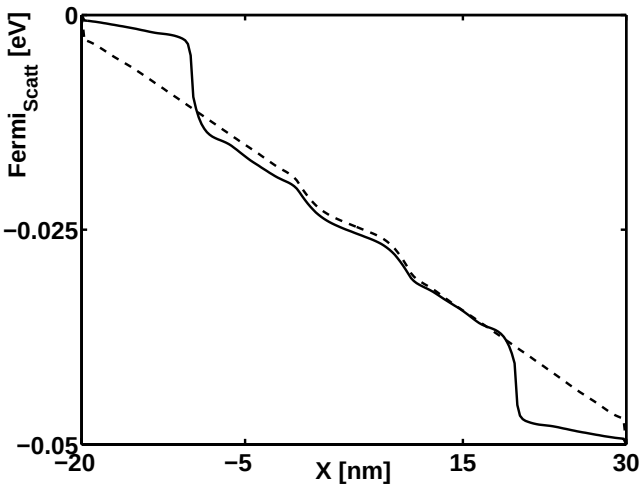
(a)



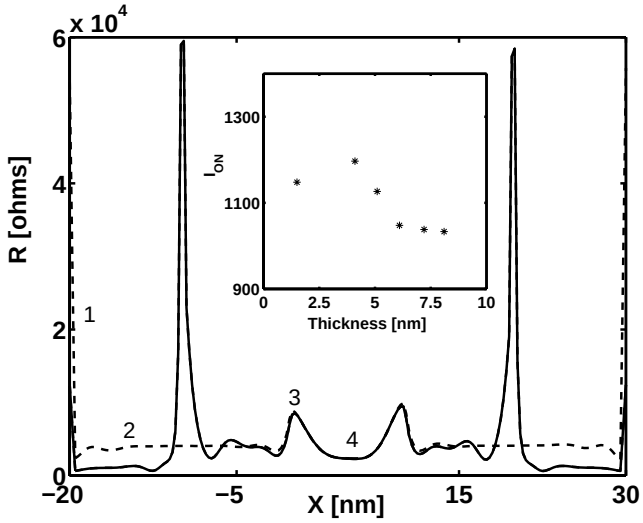
(b)

FIG. 6:

R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom



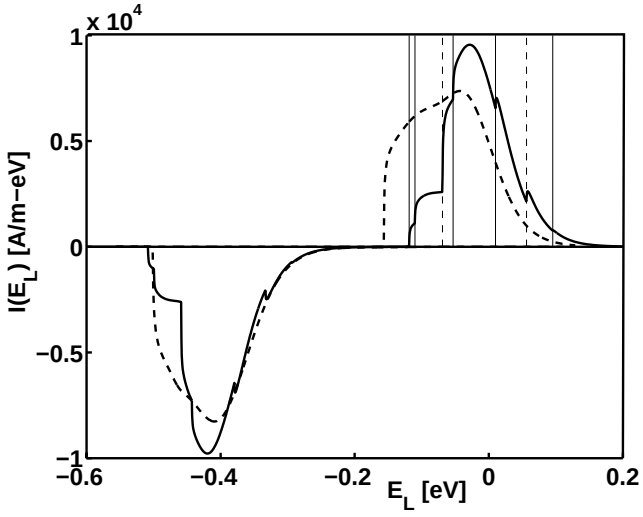
(a)



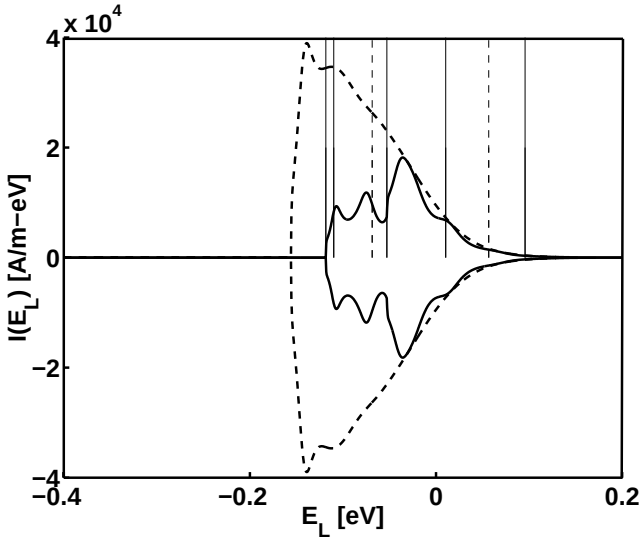
(b)

FIG. 7:

R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom



(a)



(b)

FIG. 8:

R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom

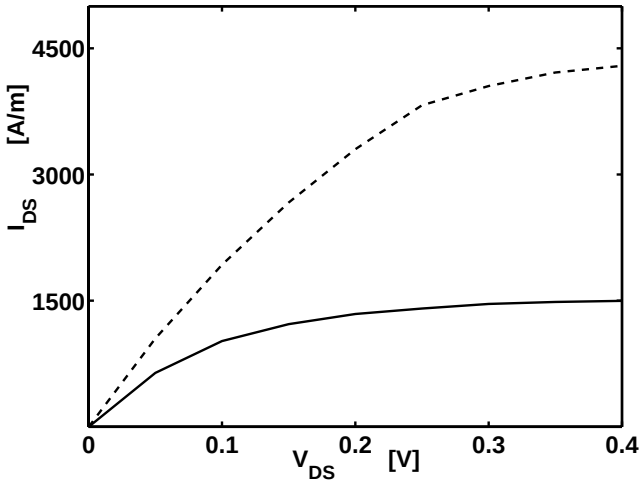


FIG. 9:

R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom

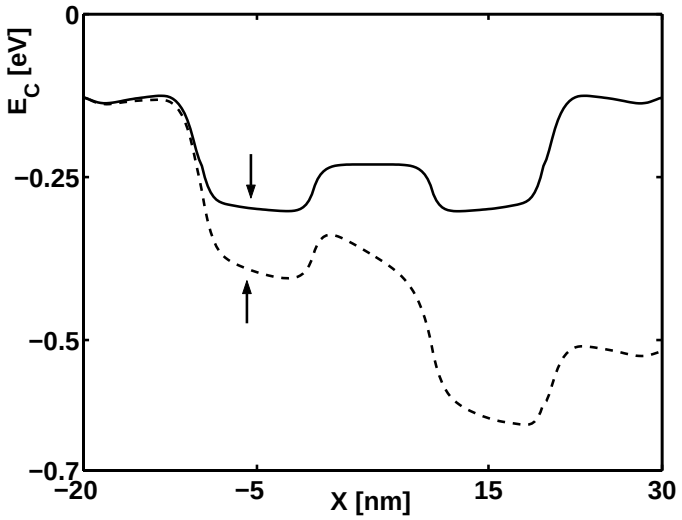
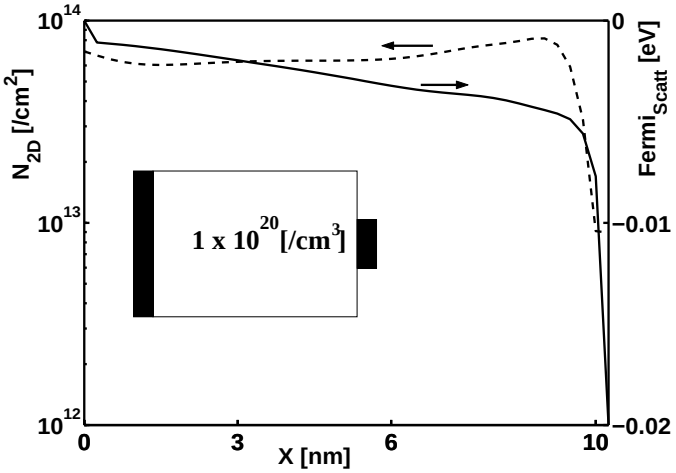
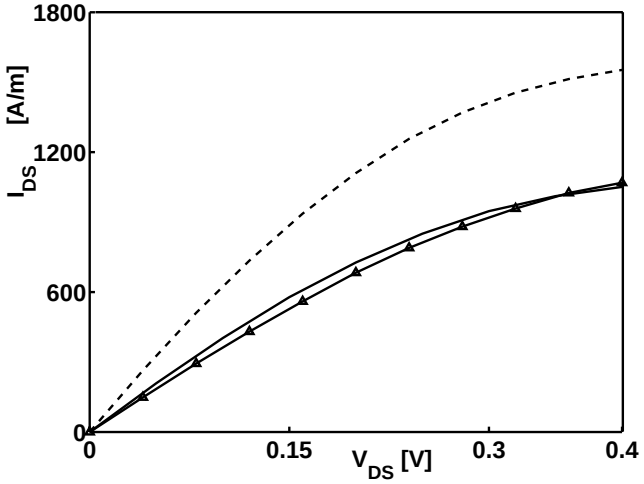


FIG. 10:

R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom



(a)



(b)

FIG. 11:

R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom