

Quantum transport in two- and three-dimensional nanoscale transistors: Coupled mode effects in the nonequilibrium Green's function formalism

Mathieu Luisier,^{a)} Andreas Schenk, and Wolfgang Fichtner
Integrated Systems Laboratory, Gloriastrasse 35, ETH Zurich, 8092 Zurich, Switzerland

(Received 27 March 2006; accepted 26 June 2006; published online 29 August 2006)

In this article, we study the coupled mode space approach to nonequilibrium Green's function (NEGF) simulation. When the lateral confinement of nanoscale devices changes abruptly and the correlation functions arising from coupled mode effects are improperly evaluated in the current and charge density calculations, it becomes difficult to solve nonequilibrium Green's function equations self-consistently with Poisson's equation because discrepancies appear in the charge distribution. To avoid this complication, two- and three-dimensional structures with a constant or a slightly varying confinement are often considered in the NEGF coupled mode space approach. We present a rigorous derivation of the method starting from the definition of the Green's function and its expansion in a coupled mode space, where current and charge density equations fully account for the coupling effects. Excellent agreement with real space calculation demonstrates the strength of the method and its applicability to the simulation of two- and three-dimensional nanoscale metal-oxide-semiconductor field-effect transistors with abruptly flared-out source/drain contacts in the ballistic limit. © 2006 American Institute of Physics. [DOI: [10.1063/1.2244522](https://doi.org/10.1063/1.2244522)]

I. INTRODUCTION

Recently reported device structures have metal-oxide-semiconductor (MOS) channel lengths in the order of 10 nm or even smaller.¹ To properly describe and model current flow in such devices, it becomes necessary to abandon classical concepts and to include quantum transport phenomena.² With appropriate simplifications to manage the computational burden, the nonequilibrium Green's function (NEGF) formalism provides a suitable framework for simulating quantum transport in nanoscale devices.^{3,4} Nevertheless, even with an effective mass level description of the device Hamiltonian, the required computational resources become large for two- and three-dimensional metal-oxide-semiconductor field-effect transistors (MOSFETs). This becomes particularly true if scattering effects are included via self-energies.⁵ One technique to reduce the computational burden is to work in mode space, i.e., to separate the Hamiltonian into its longitudinal and transverse directions. This approximation is justified in nanoscale MOSFETs as quantum confinement is strong and only the lowest energetic transverse modes are typically occupied. In the effective mass approximation, these modes will form an incomplete but sufficient basis to express all the device physical quantities such as current and charge densities.

The NEGF mode space approach with uncoupled modes has been used in the NANOMOS program⁶ that simulates double-gate MOSFETs. This method has also been examined in various MOSFET structures and geometries:⁷ they concluded that the absence of coupled mode effects does not affect the results as long as the transverse potential profile along the channel remains uniform, without any size varia-

tions. A real space calculation of the same devices confirmed these observations and the validity of the approach.

However, strong mode coupling is expected when the shape of the transverse modes varies along the channel. This is the case, for instance, in a device with a squeezed channel,⁷ with abruptly flared-out source/drain contacts,⁸ or if surface roughness is included.⁹ For the nanoscale transistors reported in the literature, source/drain contacts are often wider than the channel (in order to reduce the access resistance). Therefore, we concentrate our efforts on the simulation of abrupt transitions from a wide to a narrow region under ballistic conditions to illustrate the rigorous treatment of coupled mode effects. Venugopal *et al.*⁸ and Damle *et al.*¹⁰ explicitly described a useful real space to mode space transformation (as well as the reverse transformation) to solve the NEGF quantum transport problem and to obtain the physical observables. However, if some mode correlation terms are omitted or underestimated in the calculation of carrier and current densities,^{9,11} the results are incomplete: wide to narrow region interfaces will contain an abrupt spike in the charge distribution and will violate current conservation. Note that a wave function approach¹² requires the explicit derivative of the transverse modes along the channel direction, which is not defined for abrupt changes of the lateral confinement.

The purpose of this paper is to give insight into the coupled mode approach within the NEGF formalism. In Sec. II, we present a formal derivation of the method and show how it simplifies the quantum transport simulation of nanoscale transistors with nonuniform structures. The starting point is the definition of the nonequilibrium Green's function as introduced by Kadanoff and Baym¹³ and its expansion in terms of a complete set of orthogonal eigenfunctions (modes) that solves Schrödinger's equation in the transverse

^{a)}Electronic mail: mluisier@iis.ee.ethz.ch

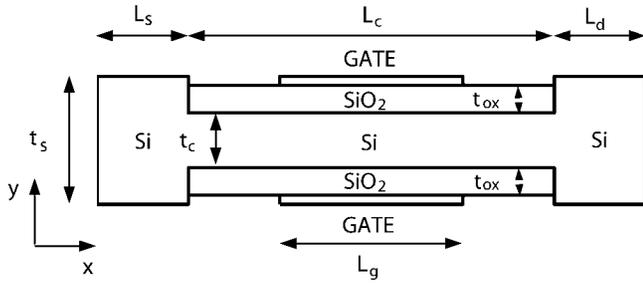


FIG. 1. Two-dimensional double-gate ultrathin body (UTB) Si MOSFET with flared-out source (left, width $t_s=7$ nm and length $L_s=5$ nm) and drain (right, width $t_s=7$ nm and length $L_d=5$ nm) contacts. The squeezed channel has a length $L_c=20$ nm, a width $t_c=3$ nm, and is controlled by a double-gate contact with $L_g=10$ nm. The two SiO₂ oxide layers surrounding the channel are 1.6 nm thick (t_{ox}). x is the transport direction, y is the confinement, and z (in-plane axis) is open.

direction(s). The strong quantum confinement in nanostructures requires only a few modes to accurately calculate current and charge densities, so that the complete basis can be reduced to its lowest occupied elements. Finally, the real space physical quantities are expressed in the resulting basis. We also highlight the differences with incomplete coupled mode space treatments. In Sec. III, we discuss two applications starting with a two-dimensional (2D) nanoscale MOSFET with flared-out source/drain contacts. The transition from wide contacts to a narrow channel leads to strong mode coupling effects. Without the correct inclusion of mode correlation terms in the carrier density calculation, self-consistency with Poisson's equation is difficult to achieve due to the charge discontinuity at the interface between contacts and channel. With the extended derivation presented here, however, convergence is readily achieved. Several relevant physical data of the device are presented, and excellent agreement with real space quantum transport simulations¹⁴ is demonstrated. The second structure is a three-dimensional (3D) extension of the 2D one: source and drain contacts are flared out in two directions leading to stronger coupling effects. It is also solved self-consistently in the ballistic regime. As both these structures are n doped, only the electron population is simulated. The hole density is much smaller and neglected in the transport calculation. In Sec. IV, the relation of this work to previous work is presented. It is also explained why the mode space approach essentially works for the effective mass approximation and not if more complicated band structure models are used. Finally, Sec. V summarizes the paper and its main contributions.

II. THEORY

The simulated MOSFET structures are represented in Figs. 1 and 2 for the 2D and 3D cases, respectively. For both devices, the channel direction corresponds to the x axis, quantum confinement to y axis, and z is open for the 2D nanoscale transistor and confined for the 3D one. A Hamiltonian H describes the device. It is coupled to two infinite reservoirs, the source and the drain, characterized by their Fermi level and the voltage applied to them. Because no leakage current through the gate contacts is assumed in these structures, effective transport occurs only along the x axis. In

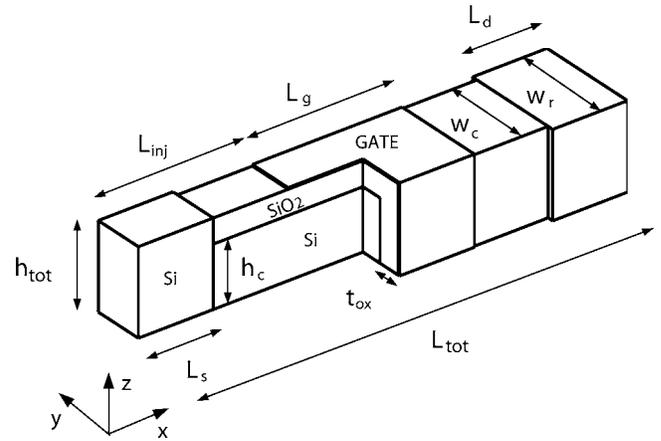


FIG. 2. Three-dimensional triple-gate Si nanowire MOSFET with flared-out source (left, width $w_s=7$ nm, height $h_{tot}=5.6$ nm, and length $L_s=5$ nm) and drain (right, width w_r , height h_{tot} , and length $L_d=5$ nm) contacts. The total length of the device L_{tot} is 30 nm, separated in two injection areas of length $L_{inj}=10$ nm embedding the triple-gate contact zone of length $L_g=10$ nm. The lower right corner of the device (on the source side) is removed in order to see the interior of the structure: the buried channel has a height $h_c=4$ nm and a thickness $t_c=3$ nm (half of the channel is visible). It is surrounded by three SiO₂ oxide layers with width $t_{ox}=1.6$ nm, so that $h_c+t_{ox}=h_{tot}$ and $t_c+2t_{ox}=w_c=6.2$ nm. x is the transport direction; y (width) and z (height) are confined.

this direction, the devices can be separated into slices, each of them being connected to their neighbors. The first and the last slices are coupled to the drain and to the source, respectively. In the following, we derive our theory for the two-dimensional case (a three-dimensional extension is obvious), where the third axis is assumed infinite and its carrier distribution modeled via the free electron wave function $e^{ik_z z}/\sqrt{L_z}$. Here L_z is a normalization constant in the z direction and $E_{k_z,ij}=\hbar^2 k_z^2/2m_{z,ij}^*$ is the portion of the total energy E coming from the open direction [$m_{z,ij}^*$ is the electron effective mass in this direction at grid point (x_i, y_j)]. The usual way to simulate such a nanostructure is to work in real space. This method is presented in Sec. II A. A more efficient algorithm, the coupled mode space approach, follows in Sec. II B.

A. Real space

In the real space approach, the x and y axes are discretized with a homogeneous finite difference grid (inhomogeneous grids are also possible) with distances Δ_x and Δ_y between their adjacent points. With N_x points in the x direction and N_y along the y axis, the effective mass Hamiltonian¹⁴ H becomes an $(N_x N_y) \times (N_x N_y)$ block tridiagonal matrix (for the rest of this article, matrices are denoted by bold indices and letters)

$$\mathbf{H} = \begin{pmatrix} \alpha_1 & \beta_{12} & 0 & \cdots & \cdots \\ \beta_{21} & \alpha_2 & \beta_{23} & 0 & \cdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \beta_{N_x-1, N_x-2} & \alpha_{N_x-1} & \beta_{N_x-1, N_x} \\ 0 & \cdots & 0 & \beta_{N_x, N_x-1} & \alpha_{N_x} \end{pmatrix}. \quad (1)$$

Each block matrix α_i is of size $N_y \times N_y$, containing all the connection information within a slice situated at x_i in the

transport direction, and $\beta_{ii+1} = \beta_{i+1i}^T$ represents the connection of a slice at x_i to the next slice at x_{i+1} . The tridiagonal matrix α_i and the diagonal matrix β_{i,i_2} are given by

$$\alpha_i = \begin{pmatrix} h_{ii11} & h_{ii12} & 0 & \cdots & \cdots \\ h_{ii21} & h_{ii22} & h_{ii23} & 0 & \cdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & h_{iiN_y-1N_y-2} & h_{iiN_y-1N_y-1} & h_{iiN_y-1N_y} \\ 0 & \cdots & 0 & h_{iiN_yN_y-1} & h_{iiN_yN_y} \end{pmatrix} \quad (2)$$

and

$$\beta_{i,i_2} = \begin{pmatrix} h_{i_1i_211} & 0 & \cdots & \cdots & \cdots \\ 0 & h_{i_1i_222} & 0 & \cdots & \cdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & h_{i_1i_2N_y-1N_y-1} & 0 \\ 0 & \cdots & \ddots & \ddots & h_{i_1i_2N_yN_y} \end{pmatrix}, \quad (3)$$

respectively. The definition of the different $h_{i_1i_2j_1j_2}$ [element (j_1, j_2) in the block (i_1, i_2)] is found in Appendix A. The source and drain contacts are incorporated into the Hamiltonian \mathbf{H} through self-energies Σ^{RD} and Σ^{RS} that modify the first diagonal block α_1 and the last block α_{N_x} . Their calculation can be achieved with an iterative algorithm,¹⁵ where a full matrix of size $(N_y \times N_y)$, for 2D structures, and $(N_y N_z) \times (N_y N_z)$ for 3D, must be inverted 30–40 times, considerably increasing the computational burden. Under ballistic conditions, the equation of motion for the retarded Green's function matrix \mathbf{G}^{R} is evaluated for steady state,¹⁶

$$(\mathbf{E} - \mathbf{H} - \Sigma^{\text{RB}})\mathbf{G}^{\text{R}} = \mathbf{I}, \quad (4)$$

and the lesser Green's function $G^<$ is obtained with¹⁶

$$\mathbf{G}^< = \mathbf{G}^{\text{R}}\Sigma^<\mathbf{B}\mathbf{G}^{\text{A}}. \quad (5)$$

All the matrices present in Eqs. (4) and (5) have a size $(N_x N_y) \times (N_x N_y)$ and describe only the electron population: \mathbf{E} is a diagonal matrix whose elements are $E_{ij} = E(x_i, y_j) = E - E_{k_z, ij}$ (E and $E_{k_z, ij}$ were introduced at the beginning of this section), \mathbf{I} is the identity matrix, and Σ^{RB} and $\Sigma^<\mathbf{B}$ are the boundary self-energies. Many different solution schemes can be found in the literature to solve Eqs. (4) and (5), such as direct methods¹⁶ or recursive algorithms.^{5,14} Despite the performance improvement resulting from these more advanced approaches, the computational costs are still very high due to the boundary condition calculations and the size of the block matrices $(N_y \times N_y)$ that need to be inverted N_x times for each energy point and for each conduction band minimum (three for Si) during the recursive algorithm steps. Once the system is solved, the 2D current density vector $[J_x(x_i, y_j) J_y(x_i, y_j)]^T$ and the charge distribution $n(x_i, y_j)$ are obtained with

$$n(x_i, y_j) = -\frac{i}{L_z \Delta_x \Delta_y k_z \sigma} \sum \int \frac{dE}{2\pi} G_{ijj}^<(k_z; E), \quad (6)$$

$$J_x(x_i, y_j) = -\frac{e}{\hbar L_z \Delta_y k_z \sigma} \sum \int \frac{dE}{2\pi} [h_{i+1ij} G_{ii+1jj}^<(k_z; E) - G_{i+1ij}^<(k_z; E) h_{ii+1jj}], \quad (7)$$

$$J_y(x_i, y_j) = -\frac{e}{\hbar L_z \Delta_x k_z \sigma} \sum \int \frac{dE}{2\pi} [h_{ii+j+1j} G_{ii+j+1j}^<(k_z; E) - G_{ii+j+1j}^<(k_z; E) h_{ii+j+1j}]. \quad (8)$$

The sum over σ represents the two spin polarizations; the integration over k_z can be done prior to NEGF calculation, directly in the contacts; and $G_{i_1i_2j_1j_2}^<(k_z; E)$ is the lesser Green's function at position (j_1, j_2) in the $(N_y \times N_y)$ block situated at (i_1, i_2) in $\mathbf{G}^<$.

B. The coupled mode space approach

The coupled mode space approach, while keeping all the relevant physics, considerably simplifies the high computational burden of a real space simulation. According to Appendix B, the discretized real space Green's function $G_{i_1i_2j_1j_2}(E, k_z)$ can be expanded in a coupled mode (eigenfunction) space as

$$G_{i_1i_2j_1j_2}(E, k_z) = \sum_{n,m} G_{i_1i_2nm}(E, k_z) \phi_n^{i_1}(y_{j_1}) \phi_m^{i_2*}(y_{j_2}). \quad (9)$$

The vectors $\phi_n^i = [\phi_n^i(y_1) \cdots \phi_n^i(y_{N_y})]^T$, constructed with the N_y y_j components of the n th mode, obey the following rules:

$$\alpha_i \phi_n^i = E_n \phi_n^i,$$

$$\int dy \phi_n^i(y) \phi_m^{*i}(y) = \delta_{n,m}, \quad (10)$$

$$\sum_n \phi_n^i(y_1) \phi_n^{*i}(y_2) = \delta(y_1 - y_2).$$

The eigenvectors ϕ_n^i of the matrix α_i form a complete orthogonal basis as shown in Eq. (10). As the total number of modes corresponds to the dimension of α_i (N_y), the size of the problem does not change if the nonequilibrium Green's function is expanded in an eigenmode basis. However, in nanostructures with strong confinement, only a few low energy modes are populated depending on the geometry of the device, the effective mass in the direction of confinement, and the doping concentration. Consequently, only a reduced number of modes N_m need to be considered, with the property $N_m \ll N_y$: increasing the number of modes to a value superior to N_m must not change the carrier and current densities any more. The total mode space Green's function matrix (lesser or retarded) \mathbf{G}_{ms} has the size $(N_x N_m) \times (N_x N_m)$ instead of $(N_x N_y) \times (N_x N_y)$ for its real space counterpart \mathbf{G}_{rs} . To find a transformation from \mathbf{G}_{rs} to \mathbf{G}_{ms} , $\mathbf{v}^i = [\phi_1^i \cdots \phi_{N_m}^i]$ matrices of size $(N_y \times N_m)$ are initially created. At position x_i , \mathbf{v}^i contains all the modes necessary to expand the real space Green's function localized there. Note that N_m can vary along the device transport axis: in the MOSFET from Fig. 1, the number of considered modes in the middle of the channel can be smaller than in the flared-out drain and source re-

gions. For the formalism derivation, however, N_m denotes the number of considered modes at each position. It is always much smaller than N_y . In a second step, a transformation matrix \mathbf{U} , with size $(N_x N_y) \times (N_x N_m)$, is generated: it contains the $N_x \mathbf{v}^i$'s defined above

$$\mathbf{U} = \frac{1}{\sqrt{\text{norm}}} \begin{pmatrix} \mathbf{v}^1 & 0 & \cdots & 0 \\ 0 & \mathbf{v}^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & \ddots & \mathbf{v}^{N_x} \end{pmatrix} = \frac{\tilde{\mathbf{U}}}{\sqrt{\text{norm}}}, \quad (11)$$

where $1/\sqrt{\text{norm}}$ is a normalization constant such that the matrix product $\mathbf{U}^T \cdot \mathbf{U}$ equals the identity matrix \mathbf{I} of size $(N_x N_m) \times (N_x N_m)$. Finally, $\tilde{\mathbf{U}}$ relates \mathbf{G}_{rs} to \mathbf{G}_{ms} by

$$\mathbf{G}_{rs} = \tilde{\mathbf{U}} \cdot \mathbf{G}_{ms} \cdot \tilde{\mathbf{U}}^T. \quad (12)$$

Equation (12) is the matrix generalization of (9) and is used to simplify the steady state equations of motion (4) and (5) for the retarded and lesser Green's functions, respectively. Multiplying them on the left with \mathbf{U}^T and on the right with \mathbf{U} (both left-hand-side and right-hand-side arguments), and replacing \mathbf{G}_{rs} by Eq. (12), we obtain the following system of equations:

$$(\mathbf{E} - \mathbf{H}_{ms} - \sum_{ms}^{\text{RB}}) \mathbf{G}_{ms}^{\text{R}} = \mathbf{I}, \quad (13)$$

$$\mathbf{G}_{ms}^{\text{<}} = \mathbf{G}_{ms}^{\text{R}} \sum_{ms}^{\text{<B}} \mathbf{G}_{ms}^{\text{A}}.$$

Each matrix has size $(N_x N_m) \times (N_x N_m)$. The mode space version of the Hamiltonian is given by the block tridiagonal matrix $\mathbf{H}_{ms} = \mathbf{U}^T \cdot \mathbf{H} \cdot \mathbf{U}$. The blocks α_i from Eq. (1) are replaced by $\alpha_i^{\text{diag}} = \mathbf{v}^{iT} \cdot \alpha^i \cdot \mathbf{v}^i$ [diagonal $N_m \times N_m$ matrices because of the properties from Eq. (10)] and $\beta_{ii\pm 1}$ by $\gamma_{ii\pm 1} = \mathbf{v}^{iT} \cdot \beta^{ii\pm 1} \cdot \mathbf{v}^{i\pm 1}$ with the same size as α_i^{diag} . Since the modes ϕ_n^i do not necessarily have the same shape all along the device transport axis, in general $\mathbf{v}^{iT} \cdot \mathbf{v}^{i\pm 1} \neq \mathbf{I}$ (an $N_m \times N_m$ identity matrix) and the $\gamma_{ii\pm 1}$ blocks may be full: this represents the coupled mode effect, whose absence signifies that an equation of motion is solved for each mode independently from the others. The boundary self-energies \sum_{ms}^{RB} and $\sum_{ms}^{\text{<B}}$ are directly computed in mode space⁷ and not with $\sum_{ms}^{\text{ms}} = \mathbf{U}^T \cdot \sum_{ms}^{\text{B}} \cdot \mathbf{U}$ because their real space calculation is computationally inefficient as mentioned earlier.

Solving the NEGF in the coupled mode space from Eq. (13) presents a substantial improvement over the real space calculation from Eqs. (4) and (5): the size of the linear system decreases from $(N_x N_y) \times (N_x N_y)$ to $(N_x N_m) \times (N_x N_m)$. There is a gain of (N_y/N_m) in the size of the blocks building the Hamiltonian \mathbf{H}_{ms} , and therefore in the size of the matrices that need to be inverted in the recursive algorithm.¹⁴ After the solution of the standard Eq. (13) is obtained, carrier and current densities can be computed: Venugopal *et al.* pointed out that the mode space to real space transformation in Eq. (12) must be used for that purpose,⁸ the diagonal elements of $\mathbf{G}_{rs}^{\text{<}}$ being the carrier density at each grid point. This corresponds to the insertion of Eq. (9) into Eq. (6) and leads to

$$n(x_i, y_j) = - \frac{i}{L_z \Delta_x \Delta_y} \sum_{nm} \sum_{k_z, \sigma} \int \frac{dE}{2\pi} G_{iimm}^{\text{<}}(k_z; E) \phi_n^i(y_j) \phi_m^{i*}(y_j), \quad (14)$$

$$J_x(x_i, y_j) = - \frac{2e}{\hbar L_z \Delta_y} \sum_{nm} \sum_{k_z, \sigma} \int \frac{dE}{2\pi} \text{Re}[h_{i+1ij} G_{ii+1nm}^{\text{<}}(k_z; E) \phi_n^i(y_j) \phi_m^{i+1*}(y_j)], \quad (15)$$

$$J_y(x_i, y_j) = - \frac{2e}{\hbar L_z \Delta_x} \sum_{nm} \sum_{k_z, \sigma} \int \frac{dE}{2\pi} \text{Re}[h_{ij+1j} G_{iimm}^{\text{<}}(k_z; E) \phi_n^i(y_j) \phi_m^j(y_{j+1})]. \quad (16)$$

All the correlation elements G_{iimm} (introduced in Appendix B) from the diagonal blocks of \mathbf{G}_{ms} are involved in the calculation of the carrier density $n(x_i, y_j)$ and the y current component $J_y(x_i, y_j)$. For the x current component $J_x(x_i, y_j)$ all the correlation elements G_{ii+1nm} from the first nondiagonal blocks of \mathbf{G}_{ms} must be taken into account.

An incomplete calculation consists in keeping only the diagonal mode space expansion coefficients $G_{iimm}^{\text{<}}$ of the non-equilibrium Green's function (although $G_{iimm}^{\text{<}} \neq 0$ for $n \neq m$) and then applying a transformation similar to Eq. (12). This modifies Eq. (14) to^{9,11}

$$n(x_i, y_j) = - \frac{i}{L_z \Delta_x \Delta_y} \sum_{k_z, \sigma, n} \int \frac{dE}{2\pi} G_{iimm}^{\text{<}}(k_z; E) |\phi_n^i(y_j)|^2. \quad (17)$$

Working in coupled mode space does not only mean solving the system of Eqs. (13), but also adapting the calculation of the physical quantities to the corresponding space. The correlations G_{iimm} resulting from coupled mode effects play a non-negligible role in the calculation of all the observables: in Sec. III, a two- and a three-dimensional idealized MOSFET example will show that the inclusion of all the mode correlation effects ensures a correct electron density and current conservation. Note that a spatially independent equation is also possible for the effective current density in the transport direction [$J_x(x, y)$ integrated over the direction of confinement y]

$$\tilde{J}_x = \frac{e}{\hbar L_z} \sum_{k_z, \sigma} \int \frac{dE}{2\pi} T(E, k_z) [f^s(E) - f^d(E)], \quad (18)$$

where $f^s(E)$ and $f^d(E)$ are the electron distributions in the source and drain, respectively. Equation (18) is correct in coupled mode space because the transmission $T(E, k_z)$ (Ref. 5) contains all the correlation effects. The term \tilde{J}_x , however, is not a spatially resolved quantity and does not allow to control current conservation.

III. APPLICATIONS

We simulate the two silicon (Si) idealized devices from Figs. 1 and 2. Both have n -doped flared-out source and drain contacts with a donor concentration $N_d=10^{20} \text{ cm}^{-3}$. The 20 nm long channel remains undoped and is surrounded by SiO_2 layers of thickness $t_{\text{ox}}=1.6 \text{ nm}$. The band gap offset between semiconductor and oxide for electrons (holes are neglected throughout the simulations) is 3.04 eV (40% of the Si– SiO_2 band gap difference). The relative dielectric constant of Si is set to $\epsilon_{\text{Si}}=11.9$ and that of the oxide to $\epsilon_{\text{ox}}=3.9$. Because the Si bulk band structure has six different conduction band minima, three valley configurations, each of them being degenerated two times, are considered, assuming a transverse effective mass $m_t=0.19m_0$ and a longitudinal $m_l=0.98m_0$. The SiO_2 effective mass is isotropic with an approximated value $m_{\text{ox}}=0.5m_0$. We use $\chi_{\text{Si}}=4.05 \text{ eV}$ for the electron affinity in Si and a work function $\phi_m=4.25 \text{ eV}$ for the metal gate contacts. The Hartree potential is taken into account by the Poisson equation, which is solved self-consistently with the coupled mode space NEGF [Eq. (13)]. The x , y , and z axes (if necessary) are discretized with a constant grid spacing of $\Delta=0.2 \text{ nm}$. Electron flow through the gate contacts is not considered. All simulations are performed at room temperature, i.e., $T=300 \text{ K}$ determines the electron distribution in the contacts. As scattering is not included, neither via self-energies⁵ nor directly in the Hamiltonian H ,⁹ only the ballistic limit of these devices is studied.

A. Two-dimensional device: Si UTB

Figure 1 shows a Si double-gate ultrathin body (UTB) MOSFET with flared-out contacts. This structure has source and drain lengths $L_s=5 \text{ nm}$ and $L_d=5 \text{ nm}$, respectively, a channel length $L_c=20 \text{ nm}$, and two gate contacts situated in the middle of the structure with $L_g=10 \text{ nm}$. The device width measures $t_s=7 \text{ nm}$ in the extended contact regions, while the channel is squeezed to $t_c=3 \text{ nm}$ (it starts at $y=2 \text{ nm}$ and ends at $y=5 \text{ nm}$). Two oxide layers ($t_{\text{ox}}=1.6 \text{ nm}$) surround the channel and minimize current leakage through the double gate, so that this effect is not considered in the simulations. Effective transport occurs along the x axis ($0 \leq x \leq 30 \text{ nm}$) and the y axis ($0 \leq y \leq 7 \text{ nm}$) represents the confinement direction. It is obvious that the abrupt transitions from the wide source and drain contacts to the narrow channel cause strong mode interactions at the interface between the different regions.

The first step to simulate the two-dimensional device from Fig. 1 consists in solving Schrödinger's equation in the y direction for the N_x discretization points x_i and for the three different conduction band valleys. This corresponds to the solution of Eq. (10), where the lowest occupied modes are kept to expand the Green's function. Typically, if the transistor in Fig. 1 has an effective mass in the y direction equal to m_t , six modes are required in the contacts and two are sufficient in the channel. Otherwise, if the direction of confinement is aligned with m_l , 12 modes are necessary in the contacts but only four in the channel. The resulting incomplete

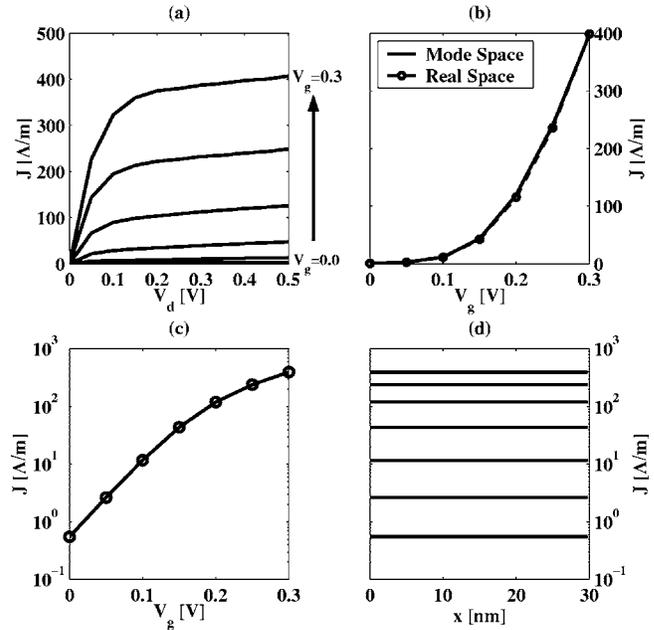


FIG. 3. Current characteristics for the two-dimensional MOSFET structure from Fig. 1. (a) Total drain-source current density J_D (unit: A/m) in transport direction as function of drain voltage V_{ds} for different gate bias V_{gs} from 0.0 to 0.3 V in steps of 0.05 V (seven curves). (b) J_D as function of V_{gs} for $V_{ds}=0.4 \text{ V}$: comparison of coupled mode space (CMS, solid line) and real space solutions (RS, circled dashed line). (c) Same as (b) but in log scale for the seven different gate bias points (circles). (d) Same as (b) and (c) but the seven different J_D are plotted as function of position x .

basis has the useful property to be energy independent, requiring only one solution of Schrödinger's equation at each self-consistent Poisson iteration.

Subsequently, the transformation matrix \mathbf{U} from Eq. (11) is constructed with the selected modes, the Hamiltonian \mathbf{H} is transformed into its mode space representation and the system of Eqs. (13) is solved for each total energy E .⁵ Finally, the desired physical quantities, electron and current densities, are obtained with Eqs. (14)–(16). To verify the simulation results, the real space system from Eqs. (6)–(8) is solved¹⁴ with the same discretizations and material parameters.

In Fig. 3, the current-voltage characteristics of the ultrathin body MOSFET are presented. Subplot (a) shows the current density in the transport direction integrated over the y direction (confinement), as function of the drain voltage V_{ds} and for different applied gate potentials V_{gs} (from 0.0 to 0.3 V with a constant step of 0.05 V). The current density J , for a given V_{gs} , increases with V_{ds} , but then saturates. Subplots (b) and (c) in Fig. 3 show the usual MOSFET transfer characteristic for $V_{ds}=0.4 \text{ V}$ and V_{gs} ramped from 0.0 to 0.3 V in a normal and in a logarithmic scale, respectively. An excellent agreement between coupled mode space (CMS, solid line) and real space (RS, dashed line) is achieved in subplot (b). Subplot (d) presents the seven current densities from (b) and (c) as function of their position along the UTB transport direction, proving that current is conserved despite the squeezed channel. Current continuity was never confirmed in the past treatments of ballistic CMS simulations^{9,11} because the spatially independent Eq. (18) was used.

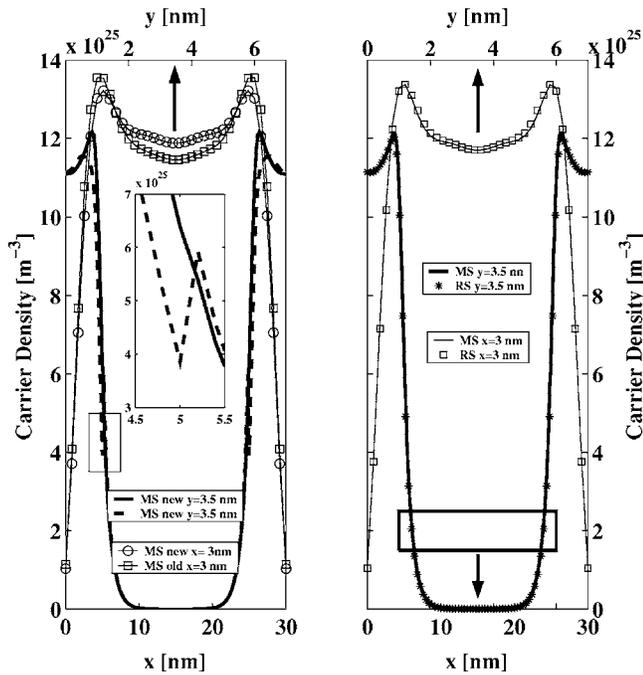


FIG. 4. Electron density profile in the UTB from Fig. 1 at $V_{ds}=V_{gs}=0$ V. (Left) Comparison of two different CMS solutions along the line $y=3.5$ nm [device axis of symmetry, $n(x,y=3.5)$, solid and dashed lines] and along the line $x=3$ nm (lines with symbols) obtained without mode correlations (lines labeled MS old) and with them (lines labeled MS new). The inset is an enlargement of the interface between the flared-out source and the channel. (Right) Comparison for $y=3.5$ nm [$n(x,y=3.5)$, bottom x axis] of CMS (solid line) and RS (stars) calculations and for $x=3$ nm [$n(x=3,y)$, in the source, top y axis] of CMS (thin line) and RS (squares) simulations.

According to Eq. (17), in the absence of correlation terms, the electron density exhibits a nonphysical behavior at the interface between wide and narrow regions, as illustrated in Fig. 4: a cut along the transport axis is extracted for $V_{ds}=V_{gs}=0$. The solid line results from the correct Eq. (14) and the dashed line from the incomplete Eq. (17). The inset clearly shows the appearance of a spurious spike in the electron distribution at the interface between the flared-out source and the squeezed channel. The same discrepancy occurs on the other side of the channel. It is evident that mode correlation effects smooth the charge density, enabling a self-consistent solution of Green's function and Poisson's equations. Consequently, it is necessary to compute the carrier density as given by Eq. (14) if strong mode coupling effects are present. On the right side of Fig. 4, a comparison between CMS and RS for two different electron density cuts, one along the line $y=3.5$ nm (solid line and stars) and the other along $x=3$ nm (dashed line and squares), confirms the validity of the coupled mode space approach when no bias is applied to the nanoscale transistor.

When the device is strongly biased, we expect more coupling effects because the electrons injected in the source interact with many modes before reaching the drain. Therefore, Figs. 5 and 6 represent electron and current densities (J_x) for $V_{gs}=0.25$ V and $V_{ds}=0.4$ V along different cut lines. Both CMS (lines) and RS (symbols) solutions are depicted for the lines $y=3.5$ nm (symmetry axis), $x=3$ nm (in the source), and $x=15$ nm (channel center). Excellent agreement (less than 2% relative difference between CMS and RS) is

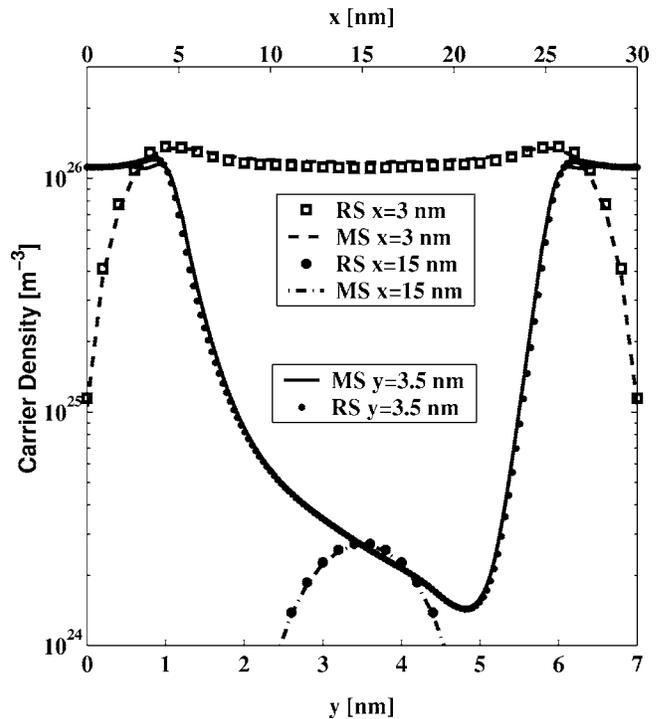


FIG. 5. Electron density profile in the MOSFET structure from Fig. 1 at $V_{ds}=0.4$ V and $V_{gs}=0.25$ V. Comparison of coupled mode space (solid line) and real space (dots) charge profiles in the cut $y=3.5$ nm [$n(x,y=3.5)$, top x axis], of CMS (dashed line) and RS (squares) calculations for $x=3$ nm [$n(x=3,y)$, bottom y axis], and of CMS (dashed-dotted line) and RS (full circles) solutions for $x=15$ nm [$n(x=15,y)$, center of the channel, bottom y axis].

achieved between the two curves. We see that the source-channel ($x=5$ nm) and channel-drain ($x=25$ nm) interfaces are smooth for both carrier and current densities, confirming that coupled mode effects are treated correctly. For high bias too, the accuracy of the CMS approach is verified.

Electrons occupy the full width of the contacts (dashed line and squares in Fig. 5), but are confined in the channel (dashed-dotted line and full circles). Current is concentrated in the channel (dashed line and crosses in Fig. 6) and does not widen much in the contacts (dashed-dotted line and squares). Furthermore, the charge penetration into the oxide layers is low due to the high band gap offset between Si and SiO₂. The different confinement behaviors of the carrier and the current densities are shown in the contour plots of Fig. 7: the electron population is strong everywhere in the flared-out source and drain contacts, but it narrows at the beginning and at the end of the channel, and it is very low in the middle. The current density J_x is somewhat wider in the contacts, but the effect is much smaller than for the carrier density because the electrons present in the protruding contact regions cannot flow directly forward due to the potential barrier blocking their movement. To better understand this phenomenon, a vector field plot of the current density is presented in Fig. 8 for the same V_{ds} - V_{gs} configuration. The source electrons (left) situated in the protruding upper and lower parts of the contact regions (i.e., with coordinates $0 \leq y \leq 2$ nm, $5 \leq y \leq 7$ nm, and $0 \leq x \leq 5$ nm) must flow along a longer path than the others that are essentially confined to regions that prolong the channel.

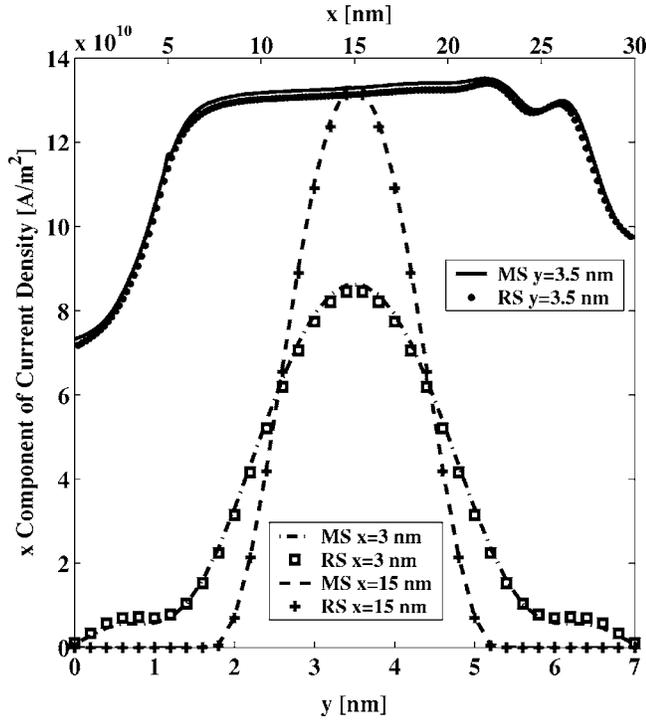


FIG. 6. Effective current density profile along the transport axis in the MOSFET structure from Fig. 1 at $V_{ds}=0.4$ V and $V_{gs}=0.25$ V. Comparison of CMS (solid line) and RS (dots) simulations for the line $y=3.5$ nm [$J_x(x,y=3.5)$, top x axis], of CMS (dashed line) and RS (crosses) current profiles for $x=3$ nm [$J_x(x=3,y)$, bottom y axis], and of CMS (dashed-dotted line) and RS (squares) calculations for $x=15$ nm [$J_x(x=15,y)$, bottom y axis].

Finally, an energy dependent quantity, the transmission coefficient from source to drain, is calculated in coupled mode space and in real space for $V_{gs}=0.25$ V and $V_{ds}=0.4$ V. The results are depicted in Fig. 9 for two different

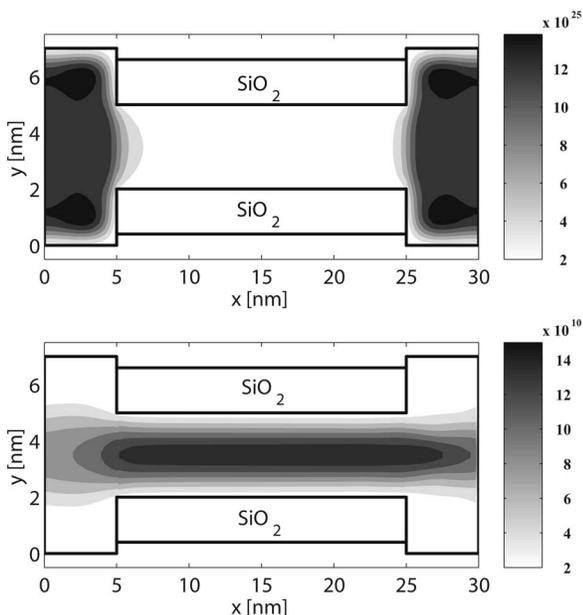


FIG. 7. Contour plot of the electron density [top, $n(x,y)$, unit: m^{-3}] and of the current density in the transport direction [bottom, $J_x(x,y)$, unit: A/m^2] for the UTB from Fig. 1 at $V_{ds}=0.4$ V and $V_{gs}=0.25$ V in coupled mode space.

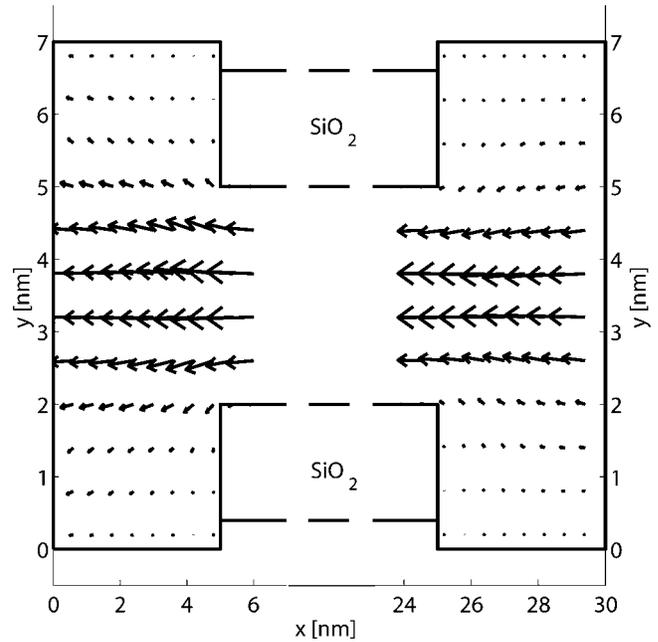


FIG. 8. Two-dimensional vector field plot of the current density in the MOSFET structure from Fig. 1 at $V_{ds}=0.4$ V and $V_{gs}=0.25$ V, calculated in coupled mode space.

effective masses in the y direction: (1) m_y is aligned with the longitudinal mass m_l (solid line and circles) and (2) m_y is aligned with the transverse one m_t (dashed line and squares). The CMS and RS curves are almost identical at low energy and slightly diverge at high energy because an incomplete basis is used to expand the Green's function in mode space. For higher energies, the incompleteness of the basis becomes more important since only the lowest energetic eigenmodes of the α_i 's were kept to form the basis.

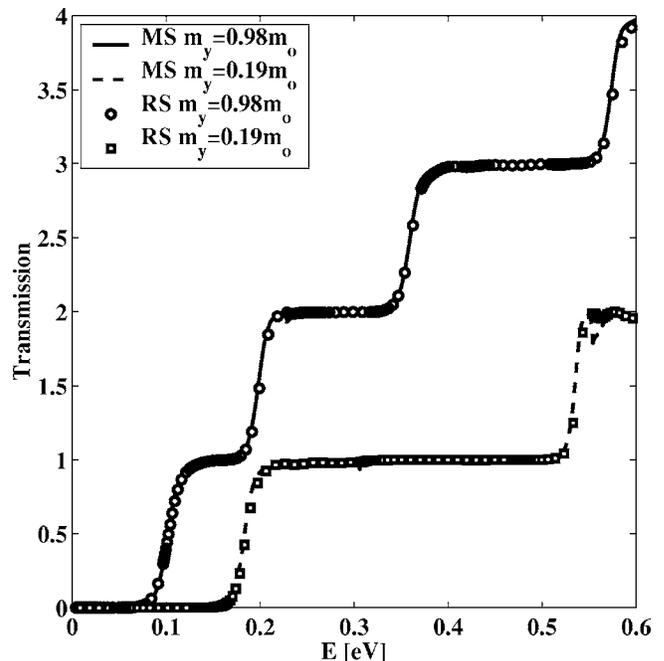


FIG. 9. Transmission coefficient from source to drain for the device from Fig. 1. CMS solutions (solid and dashed lines) are compared to RS calculations (circles and squares) for two different effective masses in the y direction; the longitudinal $m_y=0.98m_0$ and the transverse $m_y=0.19m_0$.

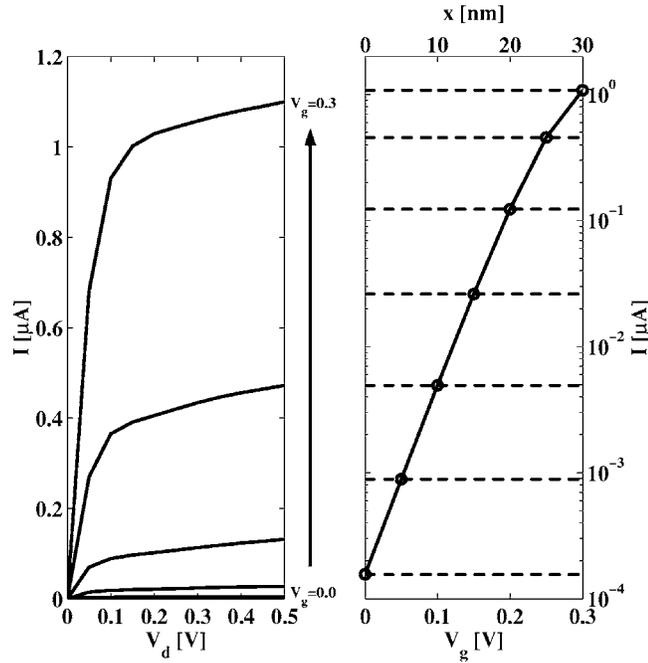


FIG. 10. Drain-source current I_D (μA) of the three-dimensional structure from Fig. 2 in CMS. (Left) I_D vs drain bias V_{ds} for seven different gate voltages V_{gs} going from 0.0 to 0.3 V in steps of 0.05 V. (Right) I_D as function of V_{gs} for $V_{ds}=0.4$ V in a log scale (bottom V_{gs} axis) and as function of position along the channel (top x axis).

B. Three-dimensional device: Si nanowire

The second structure we simulate in coupled mode space is the three-dimensional Si nanowire from Fig. 2. It has a total length $L_{\text{tot}}=30$ nm, composed of two flared-out contacts, the source (length $L_s=5$ nm, height $h_{\text{tot}}=5.6$ nm, and width $w_r=7$ nm) and the drain (length $L_d=5$ nm and same height and width as source), and a buried channel (thickness $t_c=3$ nm, height $h_c=4$ nm, length $L_c=20$ nm, and extending from $x=5$ nm to $x=25$ nm, from $y=2$ nm to $y=5$ nm, and from $z=0$ nm to $z=4$ nm) surrounded by three 1.6 nm thick SiO_2 layers ($t_{\text{ox}}=1.6$ nm). The buried oxide below the device is modeled by hard wall (Dirichlet) boundary conditions. The triple-gate contact is situated in the middle of the channel, covering the left, right, and upper sides of the nanowire together with the SiO_2 layers. The lower right corner of the device on the source side is artificially removed to offer a view on the epitaxial structure: the transition from the wide, highly doped source and drain to the squeezed channel is perturbed in the y (total width: $0 \leq y \leq 7$ nm and channel: $2 \leq y \leq 5$ nm) and z (total height: $0 \leq z \leq 5.6$ nm and channel: $0 \leq z \leq 4$ nm) directions. Therefore, strong mode coupling effects are expected at $x=5$ nm (source-channel interface) and $x=25$ nm (channel-drain interface). Effective transport occurs only along the x axis ($0 \leq x \leq 30$ nm) due to the absence of leakage current through the gate contacts.

The simulation scheme for the 3D MOSFET from Fig. 2 follows the procedure outlined in Sec. III A. However, there are two elements that increase the computational burden: the eigenfunctions generating the incomplete basis are two-dimensionally confined and second, Poisson's equation must be solved in 3D. Therefore, if the previously open direction z is now discretized with N_z points, the transformation matrix

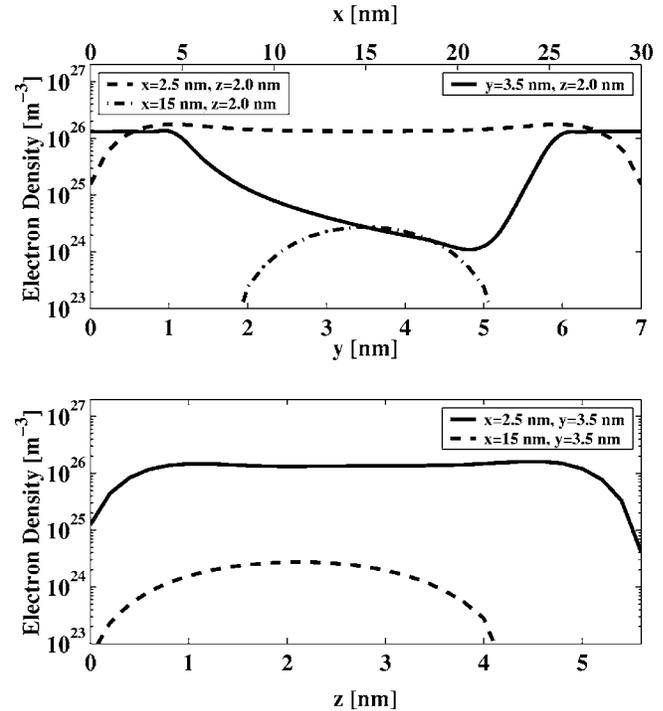


FIG. 11. Electron density profile for the nanowire from Fig. 2 at $V_{ds}=0.4$ V and $V_{gs}=0.25$ V calculated in CMS. (Up) The solution in the cuts $y=3.5$ nm (symmetry axis), $z=2.0$ nm (middle of channel height) [$n(x, y=3.5, z=2.0)$, solid line, top x axis], $x=2.5$ nm (middle of the drain), $z=2.0$ nm [$n(x=2.5, y, z=2.0)$, dashed line, bottom y axis], and $x=15$ nm (channel center), $z=2.0$ nm [$n(x=15, y, z=2.0)$, dashed-dotted line, bottom y axis] are shown. (Down) The lines $x=2.5$ nm, $y=3.5$ nm [$n(x=2.5, y=3.5, z)$, solid line] and $x=15$ nm, $y=3.5$ nm [$n(x=15, y=3.5, z)$, dashed line] are presented.

\mathbf{U} defined in Eq. (11) is of size $(N_x N_y N_z) \times (N_x N_m)$, but the mode space Hamiltonian \mathbf{H}_{ms} remains of size $(N_x N_m) \times (N_x N_m)$. It is worth noting that the number of modes N_m that need to be considered in 3D is usually larger than in 2D because the energy spacing between them is lower. Carrier and current densities are obtained with Eqs. (14)–(16), where the modes $\phi_n^i(y_j)$ become $\phi_n^i(y_j, z_k)$ to account for the third dimension z . In addition, J_z , the current in the z direction, is calculated similarly to J_x and J_y . No real space solution of the 3D device has been attempted due to the high computational resources required.

Figure 10 shows the simulation results for the current-voltage characteristics in the nanowire MOSFET: on the left, the total current I_D (integral of the current density J_x over the nanowire cross section, unit in microamperes) from drain to source as function of drain voltage V_{ds} is depicted for seven different gate voltages V_{gs} going from 0.0 to 0.3 V (steps of 0.05 V). On the right, I_D is calculated for one given drain voltage $V_{ds}=0.4$ V and different V_{gs} 's (solid line, bottom V_{gs} axis). At the same time, these seven different currents are represented as function of the x axis (top axis) to prove that current conservation is ensured (dashed lines). This could not be achieved if the correlation terms were not included in the current density calculation.

Figures 11 and 12 present carrier and effective current density (J_x) cuts along x , y , and z axis for $V_{gs}=0.25$ V and $V_{ds}=0.4$ V. From the line cut $y=3.5$ nm, $z=2.0$ nm (symme-

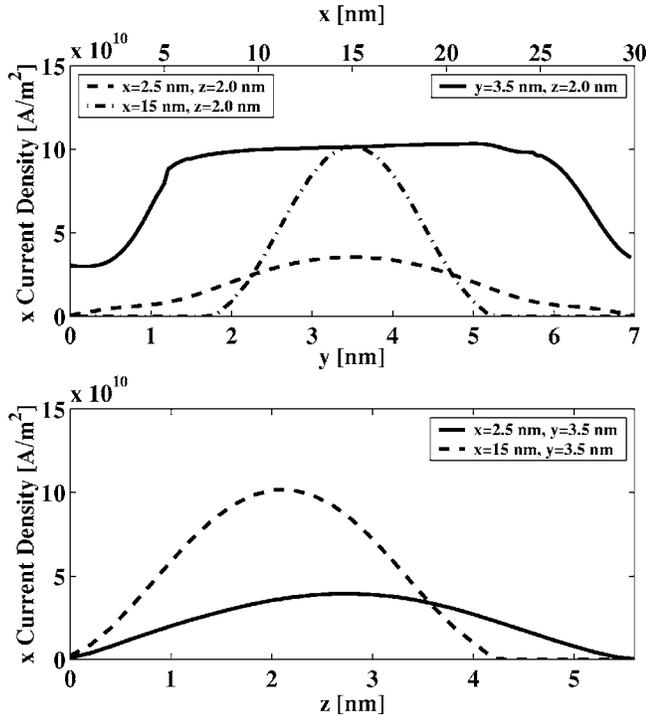


FIG. 12. Effective current density profile for the MOSFET from Fig. 2 at $V_{ds}=0.4$ V and $V_{gs}=0.25$ V in CMS. (Up) The results for $y=3.5$ nm, $z=2.0$ nm [$J_x(x,y=3.5,z=2.0)$, solid line, top x axis], $x=2.5$ nm, $z=2.0$ nm [$J_x(x=2.5,y,z=2.0)$, dashed line, bottom y axis], and $x=15$ nm, $z=2.0$ nm [$J_x(x=15,y,z=2.0)$, dashed-dotted line, bottom y axis] are plotted. (Down) The lines $x=2.5$ nm, $y=3.5$ nm [$J_x(x=2.5,y=3.5,z)$, solid line] and $x=15$ nm, $y=3.5$ nm [$J_x(x=15,y=3.5,z)$, dashed line] are shown.

try axis and middle of channel height, solid line, upper subplots, referred to x axis), we see that there is no discrepancy at the interface between the wide contacts and the channel so that the transitions are smooth: no spurious spike appears where the mode coupling effect is strong. The cut at $x=2.5$ nm, $z=2.0$ nm (in the source, dashed line, upper subplots, y axis) shows that electrons occupy the full width of the contacts (here the source), but that the current contribution from the flared-out regions is lower than the one from the middle, as in Fig. 6. This is not surprising, because we find exactly the UTB structure in Fig. 1 if we cut the nanowire in Fig. 2 at $z=2.0$ nm. Therefore, the electron and current densities along the line $x=15$ nm, $z=2.0$ nm (channel center, dashed-dotted line, upper subplots, y axis) show the same characteristics as the two-dimensional device: they are strongly confined in the channel with almost no penetration into the oxide layers. Finally, a cut along the z axis is given in the lower subplot of Figs. 11 and 12. Because the 4 nm high channel is covered by an oxide layer with $t_{ox}=1.6$ nm contrary to the source and drain contacts (height of 5.6 nm with pure n -doped Si), z is also a direction of confinement. While electrons and current fill the total device height in the contacts (solid line; $x=2.5$ nm, $y=3.5$ nm), they are squeezed in the channel (dashed line; $x=15$ nm, $y=3.5$ nm). This double confinement (y and z) can be efficiently treated in coupled mode space, but only with mode correlation terms.

The above observations are confirmed by Figs. 13 and 14. Figure 13 shows the contour plot of the electron density

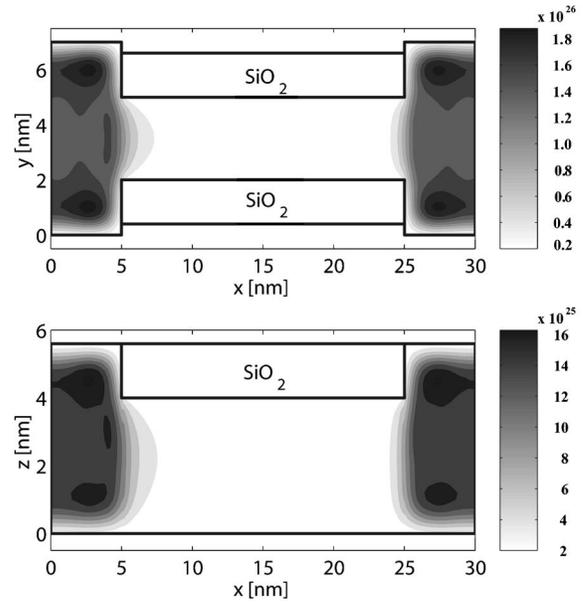


FIG. 13. Contour plot of the electron density (unit: m^{-3}) for the device from Fig. 2 at $V_{ds}=0.4$ V and $V_{gs}=0.25$ V, in the planes $z=2.8$ nm [middle of total height h_{tot} , $n(x,y,z=2.8)$, top] and $y=3.5$ nm [$n(x,y=3.5,z)$, bottom].

in the planes $z=2.0$ nm (upper subplot) and $y=3.5$ nm (lower plot). Mobile charge is confined due to the different oxide layers surrounding the nanowire channel, but electrons extend everywhere in the source and drain regions. Figure 14 illustrates the current path in the source region. On the left, the current density is plotted in the plane $z=2.0$ nm, on the right for $y=3.5$ nm. In both subplots, we observe that the current density extends everywhere: coming from the squeezed channel, the particle flow widens in all open directions when it reaches the source region.

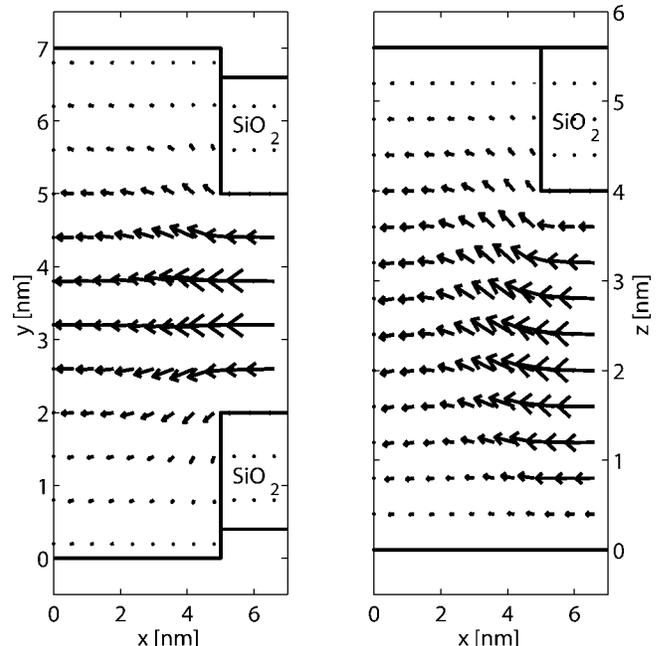


FIG. 14. 2D vector field of the current density in the nanowire from Fig. 2 at $V_{ds}=0.4$ V and $V_{gs}=0.25$ V in CMS. (Left) Cut in the plane $z=2.8$ nm. (Right) Cut in the plane $y=3.5$ nm.

IV. DISCUSSION

In this section, we discuss the advantages and disadvantages of the coupled mode space approach, the improvement of this work in comparison to previous studies, and the possibility of including scattering and more sophisticated band structure models in the formalism. The choice of an incomplete basis to expand the nonequilibrium Green's function leads to a substantial reduction of the computational burden: a factor (N_y/N_m) , where N_m is the number of considered modes and N_y the number of discretization points in the direction of confinement, is gained for two-dimensional structures. This speed-up is even larger for three-dimensional transistors. However, the mode space method has two important drawbacks. It works only for structures with a strong confinement, where only few modes are occupied, so that $N_m \ll N_y$. Otherwise, if many modes are involved in the calculations of the carrier and current densities, there is no reason to work in mode space because the time required to solve the Schrödinger equation in the confined direction(s) is not compensated by the reduction of the matrix size we have to deal with. In this case, a real space solution becomes advantageous. In addition, the real space solution takes the gate leakage current into account,¹⁴ which is difficult in mode space, constituting its second restriction: to obtain the eigenfunctions in the confinement direction, Dirichlet boundary conditions are applied, but in the gate region, the inclusion of a leakage current implies open boundary conditions. A combination of both approaches is not evident. We also want to stress that coupled mode space calculations do not only deliver current in the transport axis, but that they also describe the particle flow in the other directions. This is not an effective current, however, because it vanishes when integrated over the entire device structure.

This paper emphasizes that working in coupled mode space, as explained in Sec. II, does not only imply solving an equation of motion for the Green's function with mode coupling effects, but also making the necessary changes in the calculations of current and carrier densities. The omission of this important step causes inexact results, such as a nonphysical spike in the carrier distribution at the interface between a wide and a narrow regions, as shown in the inset of Fig. 4. In the case of simple double-gate transistors with uniform potential, simulation results¹⁸ show that the absence of correlation terms does not modify the charge density and the total current [current density integrated over the confinement direction(s)], but modifies its spatial distribution. The correlation terms $G_{inm}^<$ in Eq. (14) become also strong in the presence of elastic or inelastic scattering that supports the

transition of particles from one mode to the other: it reinforces the necessity of treating NEGF mode coupling effects rigorously. It is thus highly recommended to always compute electron and current densities as proposed in Eqs. (14)–(16).

A mode space approach (coupled or not) is not adapted to include accurate band structure models in the device simulations: in the effective mass approximation, electron and current densities can be expressed in an energy-independent and incomplete basis. This is not the case for $k \cdot p$ or tight-binding methods that require more basis components, different for each injection energy: in effect, due to the nonparabolicity of the semi-infinite reservoir band structures, the injected, reflected, or transmitted states have different wave functions for each wave vector¹⁹ (or injection energy).

V. SUMMARY AND CONCLUSION

In this paper, we presented a revisited derivation of the coupled mode space approach in the nonequilibrium Green's function formalism. It was pointed out that the omission of some important mode correlation terms in the carrier and current density calculations leads to nonphysical discrepancies at positions where strong mode coupling effects occur, for example, at the interface between a wide and a narrow region. We proved in two applications that our complete NEGF treatment removes these inconsistencies: a two-dimensional ultrathin body and a three-dimensional nanowire, both with flared-out contacts and a squeezed channel, where the coupling effects are crucial, were studied. A self-consistent solution of Poisson's and NEGF equations was made possible without losing any physical accuracy. A real space solution confirmed the exactitude of the coupled mode space method for two-dimensional nanostructures.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Fund (SNF), Project No. 200021-109393 (NEQUATTRO). One of the authors (M.L.) would like to thank Professor M. Lundstrom, the Network for Computational Nanotechnology for hosting his extended visit to Purdue, where this work was performed, and Dr. R. Venugopal for useful discussions.

APPENDIX A: DEFINITIONS

The two-dimensional device from Fig. 1 is discretized in the x and y directions in terms of a product of δ functions centered at (x_i, y_j) , $\delta(x-x_i)\delta(y-y_j)$, where x_i and y_j are the selected grid points. The effective mass Hamiltonian H in this basis has discrete elements $h_{i_1 i_2 j_1 j_2}$ (i_1 and i_2 refer to x_{i_1} and x_{i_2} , while j_1 and j_2 refer to y_{j_1} and y_{j_2}) defined as

$$h_{i_1 i_2 j_1 j_2} = \begin{cases} h_{i_1 i_1 j_1 j_1} = t_{j_1 j_1}^{i_1 i_1+1} + t_{j_1 j_1}^{i_1 i_1-1} + t_{j_1 j_1+1}^{i_1 i_1} + t_{j_1 j_1-1}^{i_1 i_1} + V_{i_1 j_1}, & i_2 = i_1, \quad j_2 = j_1 \\ h_{i_1 i_1 \pm 1 j_1 j_1} = -t_{j_1 j_1}^{i_1 i_1 \pm 1}, & i_2 = i_1 \pm 1, \quad j_2 = j_1 \\ h_{i_1 i_1 j_1 j_1 \pm 1} = -t_{j_1 j_1 \pm 1}^{i_1 i_1}, & i_2 = i_1, \quad j_2 = j_1 \pm 1 \\ 0 & \text{otherwise,} \end{cases} \quad (\text{A1})$$

where the coupling elements $t_{j_1 j_2}^{i_1 i_2}$ have the following form:

$$t_{j_1 j_2}^{i_1 i_2} = \begin{cases} t_{j_1 j_1}^{i_1 i_1 \pm 1} = \frac{2\hbar^2}{(m_{x,i_1 j_1}^* + m_{x,i_1 \pm 1 j_1}^*) \Delta x_{i_1}^\pm (\Delta x_{i_1}^+ + \Delta x_{i_1}^-)}, & i_2 = i_1 \pm 1, \quad j_2 = j_1 \\ t_{j_1 j_1 \pm 1}^{i_1 i_1} = \frac{2\hbar^2}{(m_{y,i_1 j_1}^* + m_{y,i_1 j_1 \pm 1}^*) \Delta y_{j_1}^\pm (\Delta y_{j_1}^+ + \Delta y_{j_1}^-)}, & i_2 = i_1, \quad j_2 = j_1 \pm 1 \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A2})$$

$V_{ij}=V(x_i,y_j)$ is the quantum confinement potential, $m_{x,ij}^*$ the effective mass along the x axis at position (x_i,y_j) , $m_{y,ij}$ the effective mass in the confined direction y at (x_i,y_j) , $\Delta x_{i_1}^\pm = |x_{i_1 \pm 1} - x_{i_1}|$, and $\Delta y_{j_1}^\pm = |y_{j_1 \pm 1} - y_{j_1}|$. The three-dimensional generalization of these definitions is straightforward and will not be shown here.

APPENDIX B: EIGENFUNCTION EXPANSION

The nonequilibrium Green's function $G(r_1 t_1; r_2 t_2)$, as introduced by Kadanoff and Baym,¹³ has the following form:

$$G(r_1 t_1; r_2 t_2) = -\frac{i}{\hbar} \langle T \{ \hat{\psi}(r_1 t_1) \hat{\psi}^\dagger(r_2 t_2) \} \rangle, \quad (\text{B1})$$

where $\hat{\psi}^\dagger(r,t)$ and $\hat{\psi}(r,t)$ are time-dependent creation and annihilation operators at position r , respectively, and T is a time-ordering operator on an imaginary contour. $\hat{\psi}^\dagger(r,t)$ and $\hat{\psi}(r,t)$ can be represented in any basis as long as it is orthogonal and complete. A possible expansion for a 2D structure is

$$\begin{aligned} \hat{\psi}(r,t) &= \sum_{n,k_z} e^{ik_z z} \hat{c}_{n,k_z}(t) v_n(x,y), \\ \hat{\psi}^\dagger(r,t) &= \sum_{n,k_z} e^{-ik_z z} \hat{c}_{n,k_z}^\dagger(t) v_n^*(x,y). \end{aligned} \quad (\text{B2})$$

The operators $\hat{c}_{n,k_z}^\dagger(t)$ and $\hat{c}_{n,k_z}(t)$ create and annihilate a particle in a state with quantum numbers n and k_z at time t . For the device represented in Fig. 1, the basis function $v_n(x,y)$ could belong to its complete set of eigenfunctions. However, because open boundary conditions dominate the transport x axis, while y is the confinement direction, it is justified to separate $v_n(x,y)$ into a product $u(x) \cdot \phi_n(x,y)$, where $u(x)$ is a function localized at x and $\phi_n(x,y)$ is the n th eigenfunction of the α matrices defined in Eq. (2). $\phi_n(x,y)$ still depends on x because the confinement is not the same all along the transport direction. The function $u(x)$ is combined with $\hat{c}_{n,k_z}^\dagger(t)$ and $\hat{c}_{n,k_z}(t)$ to form two operators $\hat{c}_{n,k_z}^\dagger(x,t)$ and $\hat{c}_{n,k_z}(x,t)$ that create and annihilate an electron at position x and time t in a state characterized by its quantum numbers n and k_z . Inserted into Eq. (B1), this gives

$$\begin{aligned} G(r_1 t_1; r_2 t_2) &= -\frac{i}{\hbar} \sum_{k_z, n, m} \langle T \{ \hat{c}_{n,k_z}(x_1 t_1) \hat{c}_{m,k_z}^\dagger(x_2 t_2) \} \rangle \\ &\quad \times e^{ik_z(z_1 - z_2)} \phi_n(x_1, y_1) \phi_m^*(x_2, y_2) \end{aligned} \quad (\text{B3})$$

$$\begin{aligned} &= \sum_{k_z, n, m} G_{nm}(x_1 t_1; x_2 t_2; k_z) \\ &\quad \times e^{ik_z(z_1 - z_2)} \phi_n(x_1, y_1) \phi_m^*(x_2, y_2). \end{aligned} \quad (\text{B4})$$

The term $G_{nm}(x_1 t_1; x_2 t_2; k_z)$ is the time-dependent mode space (or eigenfunction) expansion coefficient of the nonequilibrium Green's function. As working with two times t_1 and t_2 lying on an imaginary contour is too difficult, we apply the Langreth theorem¹⁷ on $G_{nm}(x_1 t_1; x_2 t_2; k_z)$ to obtain two Green's functions with real-time arguments, the retarded $G_{nm}^R(x_1 t_1; x_2 t_2; k_z)$ and the lesser $G_{nm}^<(x_1 t_1; x_2 t_2; k_z)$. Furthermore, we want to calculate the steady state solution of the MOSFET devices, where only the difference between t_1 and t_2 is relevant so that a Fourier transform into energy E can be operated. To complete the simplification, x is discretized into x_i , or just index i , and y becomes y_j or just j , leading to $G_{i_1 i_2 j_1 j_2}^{R/<}(E, k_z)$ defined as

$$G_{i_1 i_2 j_1 j_2}^{R/<}(E, k_z) = \int d(t_1 - t_2) G_{nm}^{R/<}(x_{i_1} t_1; x_{i_2} t_2; k_z) e^{iE(t_1 - t_2)/\hbar}. \quad (\text{B5})$$

With the knowledge of the mode space expansion coefficients of the nonequilibrium Green's function $G_{i_1 i_2 j_1 j_2}^{R/<}(E, k_z)$, the original real space Green's function $G_{i_1 i_2 j_1 j_2}^{R/<}(E, k_z)$ can be reconstructed together with current and carrier densities.

- ¹B. Doris *et al.*, Tech. Dig. - Int. Electron Devices Meet. **2002**, 267.
- ²Z. Ren, R. Venugopal, S. Datta, M. S. Lundstrom, D. Jovanovic, and J. G. Fossum, Tech. Dig. - Int. Electron Devices Meet. **2000**, 715.
- ³S. Datta, J. Phys.: Condens. Matter **2**, 8023 (1990).
- ⁴S. Datta, Superlattices Microstruct. **28**, 253 (2000).
- ⁵R. Lake, G. Klimeck, R. Chris Bowen, and D. Jovanovic, J. Appl. Phys. **81**, 7845 (1997).
- ⁶Z. Ren, R. Venugopal, S. Goasguen, S. Datta, and M. S. Lundstrom, IEEE Trans. Nanotechnol. **50**, 1914 (2003).
- ⁷R. Venugopal, Z. Ren, S. Datta, M. S. Lundstrom, and D. Jovanovic, J. Appl. Phys. **92**, 3730 (2002).
- ⁸R. Venugopal, S. Goasguen, S. Datta, and M. S. Lundstrom, J. Appl. Phys. **95**, 292 (2004).
- ⁹J. Wang, E. Polizzi, A. W. Ghosh, S. Datta, and M. S. Lundstrom, Appl. Phys. Lett. **87**, 043101 (2005).
- ¹⁰P. S. Damle, A. W. Ghosh, and S. Datta, *Molecular Nanoelectronics* (American Scientific Publishers, Stevenson Ranch, California, 2003), p. 115.
- ¹¹J. Wang, E. Polizzi, and M. S. Lundstrom, J. Appl. Phys. **96**, 2192 (2004).
- ¹²E. Polizzi and N. Ben Abdallah, J. Comput. Phys. **202**, 150 (2005).
- ¹³L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Benjamin, New York, 1962).
- ¹⁴A. Svizhenko, M. P. Anantram, T. R. Govindan, R. Biegel, and R. Venugopal, J. Appl. Phys. **91**, 2343 (2002).

¹⁵M. P. Lopez Sancho, J. M. Lopez Sancho, and J. Rubio, *J. Phys. F: Met. Phys.* **15**, 851 (1985).

¹⁶S. Datta, *Electronic Transport in Mesoscopic Systems* (Press Syndicate University of Cambridge, Cambridge, 1995).

¹⁷H. Haug and A.-P. Jauho, *Quantum Kinetics in Transport and Optics of*

Semiconductors (Springer, Berlin, 1989).

¹⁸To be posted on <http://www.nanohub.org>

¹⁹M. Luisier, A. Schenk, W. Fichtner, and G. Klimeck, *Phys. Rev. B* (submitted).