Atomistic Simulation of Nanowire Transistors

Mathieu Luisier^{1,*} and Andreas Schenk^{1,2}

 ¹Integrated Systems Laboratory, Gloriastrasse 35, ETH Zurich, CH-8092 Zurich, Switzerland
 ²Synopsys Switzerland LLC., Affolternstrasse 52, CH-8050 Zurich, Switzerland
 E-mail*:mluisier@iis.ee.ethz.ch, Phone: +41 44 632 5437, Fax: +41 44 632 1194

(Dated: August 23, 2007)

Abstract

At the nanometer scale, the computer simulation of electronic transport cannot be conceived without including quantum-mechanical effects as well as the atomic granularity of the simulation domain. In this review we present a three-dimensional quantum transport simulator based on the $sp^3d^5s^*$ semi-empirical tight-binding (TB) method that fulfills these two requirements. The integration of the multi-band TB model into a transport code is only possible, if open boundary conditions (OBCs) are introduced. The available procedures to calculate OBCs in three-dimensional structures are computationally too intensive, since they take the form of a generalized eigenvalue problem or require iterative solvers. Therefore, an improved method based on the scattering-boundary approach is reviewed in this work. It significantly reduces the computational burden associated with the OBCs calculation. Furthermore, it can be formulated either in the Non-Equilibrium Green's Function or in the Wave Function formalism, it works for any channel orientation, material composition, or cross section shape, and it is self-consistently coupled to the three-dimensional computation of the electrostatic potential in the device. These features allow the analysis and the comparison of nanowire field-effect transistors (FETs) with transport along the [100], [110], [111], and [112] crystal axis. Hence, their ON-current, subthreshold swing, and interface roughness sensitivity are investigated.

Keywords: Quantum Transport, Tight-Binding, Non-Equilibrium Green's Function, Numerical Simulation

I. INTRODUCTION

The constant request for performance improvement of electronic devices is mainly satisfied by reducing the size of their active components, the transistors. Thirty years of aggressive scaling have pushed their dimensions close to the atomic range. Recently reported structures of metal-oxide-semiconductor field-effect transistors (MOSFETs) already have channel lengths in the order of 10 nm or even smaller¹. Such short channels are hard to fabricate, induce source-to-drain tunneling, and require very thin oxide layers causing significant gate leakage currents. The search for advanced device architectures that could overcome these difficulties has just started. It is part of the new research field called nanoelectronics.

Semiconductor nanowires (NWs) may play an important role in the future of nanoelectronics, since they can act both as active devices (transistors) and as wire connectors. Recently, several groups have grown Si², GaAs³, Ge⁴, or SiGe⁵ NWs with different channel orientations and cross section shapes. Field-effect transistors (FETs) with a triangular⁶, a rectangular⁷, or a cylindrical⁸ wire as channel have been reported in the literature. Nanostructures with even more exotic cross sections, such as T-shape wires, have found practical applications, for example, in the optoelectronics area⁹.

The fabrication of novel devices is a long and expensive process. A well-established way of accelerating the production and of reducing the costs consists in replacing experimental development techniques by technology computer aided design (TCAD). In this review, we present the physics and the numerical implementation of a simulator designed for post-CMOS nanowire FETs. Modeling the current flow in these devices requires to abandon classical concepts and to include quantum-mechanical phenomena¹⁰. In effect, these nanodevices exhibit significant quantization effects that a classical theory cannot capture. This implies the application of a full-band transport model to simulate transistors with cross sections not exceeding 5 nm \times 5 nm¹¹. For that purpose we choose the $sp^3d^5s^*$ semi-empirical tight-binding (TB) method^{12,13} because of its capability of reproducing the full band structure of semiconductors with a diamond (Si)¹⁴, zincblende (GaAs)¹⁵, or wurtzite (GaN)¹⁶ structure. Furthermore, the atomic granularity of the simulation domain is automatically taken into account in the TB method.

In Section II we describe the incorporation of the TB bandstructure model into an atomistic quantum transport solver based either on the Non-Equilibrium Green's Function (NEGF)^{17,18} or on the Wave Function (WF) formalism¹⁹. The calculation of the open boundary conditions (OBCs) and their coupling to NEGF or WF are briefly reviewed²⁰. It is shown how to compute carrier and current densities in both formalisms. Our simulator is not only capable to treat the usual [100] transport direction and square cross sections²¹⁻²⁴, but any crystal orientation such as [110], [111], or [112], and any wire shape (e. g. triangular, circular, T, hexagonal, ...).

Simulation results are presented in Section III. First, the band structure and transmission of infinite ideal wires are used to demonstrate the validity of the physical models and their correct implementation. Then, simulations of Si nanowire transistors with different channel orientation, gate lengths, and atomic surface construction are presented. The full-band current characteristics are compared to results obtained with the effective mass approximation, the performances of n-doped and p-doped FETs are analyzed, and source-to-drain tunneling is studied. Although our simulator is essentially designed for ballistic transport²⁵, the inclusion of elastic scattering like alloy disorder²⁶ or interface roughness²⁷ is possible.

Section IV deals with the numerical implementation of the transport model. The parallelization of the code is discussed as well as the libraries used to compute the eigenvalue problems and to solve the linear systems described in Section II. Sparse and full, real and complex matrices are present at the same time, but require distinct treatments. Benchmarks of the open boundary conditions and transport parameter calculations are given. For different samples, corresponding to different matrix sizes, the linear and eigenvalue solvers are tested on a single CPU or on many shared/distributed memory processors.

Finally, Section V summarizes and concludes this review. We highlight the main achievements of this work and suggest some possible improvements in the field of atomistic simulations of nanowire transistors.

II. THEORY

In this Section, the algorithm to obtain open boundary conditions in nanowire structures is summarized. It is then used to compute carrier and current densities in the NEGF and in the WF formalisms. The method is based on a scattering-boundary approach. It was first derived in Ref.²⁰. Figure 1 shows the schematic view of a nanowire with length L_w (scattering-boundary region) without its semi-infinite reservoirs. Effective transport occurs along the x-axis while y and z are directions of confinement. Each atom is characterized by a set of orbitals. In the $sp^3d^5s^*$ tight-binding model, ten different orbitals are kept. Each of them is twofold degenerate, if spin orbit coupling is considered.

Independent of the chosen formalism, the Schrödinger equation is explicitly or implicitly solved with proper boundary conditions. An incident electron, coming from the left or the right reservoir, with energy E measured relative to the top of the valence band, can be scattered into reflected states that propagate back to their origin or into transmitted states that propagate to the other contact(s). The device equation can be written as

$$H_0|\psi_E\rangle = E|\psi_E\rangle.\tag{1}$$

The Hamiltonian H_0 contains the lattice and the electrostatic potentials. The scattering wave function $|\psi_E\rangle$ can be expanded in terms of orthogonalized Löwdin atomic orbitals $\phi_{\sigma}(\mathbf{r})$ of type σ (s, p, d, or excited s^{*})

$$\psi(\mathbf{r}; E) = \sum_{\sigma, i, j, k} C^{\sigma}_{ijk}(E) \phi_{\sigma}(\mathbf{r} - \mathbf{R}_{ijk})$$
$$|\psi_E\rangle = \sum_{\sigma, i, j, k} C^{\sigma}_{ijk}(E) |ijk, \sigma\rangle,$$
(2)

where $C_{ijk}^{\sigma}(E)$ is the expansion coefficient for an orbital σ of an atom situated at $\mathbf{R} = (x_i, y_j, z_k)$ in the nanowire. To solve Eq. (1), one works in a slab basis²⁸. A slab represents the minimal number of atomic layers required to generate an infinite nanowire, if it is translated in the transport direction. For example, a slab is composed of four atomic layers if x is aligned with [100] and six layers for [111]. A slab has a width Δ . A nanowire

with length L_w is, therefore, composed of L_w/Δ slabs that represent its central scattering region. In this basis, the scalar $C_{ijk}^{\sigma}(E)$ becomes a vector $C_i(\mathbf{R}_s, E)$, where *i* denotes the *i*th wire slab and \mathbf{R}_s the position and the orbital type of an atom residing inside this slab. Considering only coupling to the nearest neighbors, disregarding three-center integrals¹², and left-multiplying Eq. (1) with $\langle ijk, \sigma |$ at each position and for each orbital, the matrix equation

$$(E - H_{ii})C_i(\mathbf{R}_s, E) - H_{ii+1}C_{i+1}(\mathbf{R}_s, E) - H_{ii-1}C_{i-1}(\mathbf{R}_s, E) = 0$$
(3)

has to be solved for each slab index *i*. Its element $E - H_{ii}$ describes the on-site energy and the bond connections within a slab *i*, H_{ii+1} is the coupling to the next slab, and H_{ii-1} the coupling to the previous one. If t_b is the tight-binding order (10 without spin, 20 with) and each slab contains N atoms, the size of these square matrices is t_bN . Equation (3) is valid in the device as well as in the semi-infinite left and right contacts, where a scattering boundary ansatz is applied to find the coefficients $C_i(\mathbf{R}_s, E)$. For brevity, only the left contact is treated, the derivation of the other(s) is obvious

$$C_i(\mathbf{R}_s, E) = \frac{1}{\sqrt{N_x}} \sum_n \left(a_n e^{ik_n(E)x_i} \varphi_{i,n}^+(\mathbf{R}_s, E) + b_n e^{-ik_n(E)x_i} \varphi_{i,n}^-(\mathbf{R}_s, E) \right).$$
(4)

Here, N_x is a normalization constant, a_n is the injection coefficient for the n^{th} state $\varphi_{i,n}^+(\mathbf{R}_s, E)$ transmitted through the device (i.e. propagating from the left to the right), and b_n is the coefficient for the n^{th} state $\varphi_{i,n}^-(\mathbf{R}_s, E)$ reflected back into the contact, respectively. After having inserted Eq. (4) into Eq. (3) and separated the transmitted and the reflected parts of the coefficient $C_i(\mathbf{R}_s, E)$, the fact is exploited that both resulting contributions must add up to zero for each quantization level n. In effect, the contacts are assumed infinite and

$$\left(E - H_{ii} - H_{ii+1}e^{\pm ik_n(E)\Delta} - H_{ii-1}e^{\mp ik_n(E)\Delta}\right)\varphi_{i,n}^{\pm}(\mathbf{R}_s, E) = 0.$$
(5)

Only one of these two equations needs to be solved, because the solution of the other equation is automatically taken into account²⁹.

A well established procedure of calculating the $k_n(E)$'s and the $\varphi_n(\mathbf{R}_s, E)$'s consists in writing Eq. (5) as a complex non-hermitian (with spin-orbit coupling) or as a real nonsymmetric (without) generalized eigenvalue problem (GEVP)^{23,24} of size $2t_bN$

$$\begin{pmatrix} E - H_{ii} & -H_{ii+1} \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \varphi_i \\ \varphi_{i+1} \end{pmatrix} = e^{-ik\Delta} \begin{pmatrix} H_{ii-1} & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \varphi_i \\ \varphi_{i+1} \end{pmatrix},$$
(6)

where the variables E and \mathbf{R}_s are omitted for brevity. Despite the fact that the numerical solution of Eq. (6) avoids matrix inversion³⁰, it is still computationally intensive to obtain the desired variables for a large nanowire. However, going back to Eq. (5), a better insight into the physical structure of the matrices H_{ii} , H_{ii+1} , and H_{ii-1} leads to a simplified procedure to evaluate the OBCs. The detailed derivation can be found in Ref.²⁰. Here we restrict ourselves to the final step of the calculation. A matrix M is formed from H_{ii} , H_{ii+1} , and H_{ii-1} by applying a shift and invert spectral transformation. The size of M is equal to that of the H's. Then the $k_n(E)$'s and the $\varphi_n(\mathbf{R}_s, E)$'s are the solutions of an ordinary eigenvalue problem (EVP)

$$M\varphi(\mathbf{R}_s, E) = \frac{1}{e^{-ik(E)\Delta} - 1}\varphi(\mathbf{R}_s, E).$$
(7)

The matrix M is always well-defined and has an advantageous structure

$$M = \begin{pmatrix} 0 & 0 & M_{02} & M_{03} \\ 0 & 0 & M_{12} & M_{13} \\ 0 & 0 & M_{22} & M_{23} \\ 0 & 0 & M_{32} & M_{33} \end{pmatrix}.$$
(8)

It is thus not necessary to consider the whole matrix M in the eigenvalue problem defined in Eq. (7), but only its lower right corner. When the transport direction is aligned with the [100] crystal axis, the complex non-hermitian (spin-coupling) or real non-symmetric (no spin) eigenvalue problem to be solved is of size $t_b N/2$, which is a significant improvement compared to Eq. (6) whose matrices have a size $2t_b N$. Apart from a gain of a factor 4 in the size of the blocks, it is not required to solve a GEVP, but only an ordinary EVP. The method presented here is more efficient than what was published previously^{23,24} (see also Table III for quantitative results). Furthermore, the approach works for all the crystal orientations, even if one atomic layer is connected not only to the next layer but to several consecutive layers. The matrix M is always constructed in a way to minimize the size of the EVP.

It remains to determine the electron (hole) density $n(\mathbf{r})$ $(p(\mathbf{r}))$ and the current density $\mathbf{J}(\mathbf{r})$ by coupling the OBCs to a NEGF or WF solver. Due to the strong localization of the Löwdin orbital functions $\phi_{\sigma}(\mathbf{r} - \mathbf{R}_{i})$ used in the tight-binding model, $n(\mathbf{r})$, $p(\mathbf{r})$, and $\mathbf{J}(\mathbf{r})$ are represented by δ -functions centered around the atom positions \mathbf{R}_{i}

$$n(\mathbf{r},t) = \sum_{i} n_{i} \,\delta(\mathbf{r} - \mathbf{R}_{i}),\tag{9}$$

$$p(\mathbf{r},t) = \sum_{i} p_i \,\delta(\mathbf{r} - \mathbf{R}_i),\tag{10}$$

$$\mathbf{J}(\mathbf{r},t) = \sum_{i,j} I_{ij} \,\delta(\mathbf{r} - \mathbf{r}_i)(\mathbf{R}_j - \mathbf{R}_i).$$
(11)

Bold letters denote vectors, n_i (p_i) is the number of electrons (holes) at \mathbf{R}_i , I_{ij} is the current flowing from one atom situated at \mathbf{R}_i to its nearest neighbor at \mathbf{R}_j (in principle four different j per i) along the connecting bond. The ansatz for the current and the charge densities is proved in Sections II A and II B.

A Wave Function Solver

In the Wave Function formalism, the transport problem takes the form of a linear system that has to be solved for each injection energy E^{20}

$$(\mathbf{E} - \mathbf{H}) \cdot \mathbf{C} = \mathbf{I}_{\mathbf{0}}.\tag{12}$$

The Hamiltonian **H** is a square matrix of size $N t_b \times N t_b$, where N is the total number of atoms constituting the nanowire and t_b the tight-binding order. The boundary conditions are incorporated in its first and in its last diagonal blocks. The injection matrix $\mathbf{I_0}$ has the size $(N t_b) \times (p n_p)$. The parameter p denotes the number of injection ports in the device (two in our case, the left and right contacts). From each port, n_p states caused by the two-dimensional wire confinement are injected. The unknown matrix **C** has the same size as $\mathbf{I_0}$ and contains the coefficients introduced in Eq. (3). Its individual elements depend on the energy E (or wave vector k(E)) on the port, on the injected state (index p and n_p , respectively), on the orbital type σ , and on the atom position \mathbf{R}_i . They are labeled $C_{i,p,n_p}^{\sigma}(k(E))$, where k(E) is sometimes omitted and σ sometimes cast into i for brevity.

The electron density in the nanowire $n(\mathbf{r})$ ($p(\mathbf{r})$ is obtained analogously) is then given by

$$n(\mathbf{r}) = \langle \mathbf{r} | \psi \rangle \langle \psi | \mathbf{r} \rangle$$

$$= \frac{1}{N_x} \sum_{n_p, p, \sigma_1, \sigma_2} \sum_{i,j} \sum_k C_{i,p,n_p}^{\sigma_1 *}(k) C_{j,p,n_p}^{\sigma_2}(k) f(E_{p,n}(k) - \mu_p) \phi_{\sigma_1}^*(\mathbf{r} - \mathbf{R}_i) \phi_{\sigma_2}(\mathbf{r} - \mathbf{R}_j)$$

$$= \frac{\Delta}{2\pi} \sum_{n_p, p, \sigma} \sum_i \int_{0}^{\pi/\Delta} dk |C_{i,p,n_p}^{\sigma}(k)|^2 f(E_{p,n_p}(k) - \mu_p) \delta(\mathbf{r} - \mathbf{R}_i)$$

$$= \sum_i \underbrace{\frac{\Delta}{2\pi} \sum_{n_p, p, \sigma} \int dE |C_{i,p,n_p}^{\sigma}(k_{p,n_p})|^2 \left| \frac{dE}{dk_{p,n_p}} \right|^{-1} f(E - \mu_p) \delta(\mathbf{r} - \mathbf{R}_i)}_{n_i}$$
(13)

In the contacts, Fermi distributions with a chemical potential μ_p are assumed. The injection energy (from port p) of a state n with wave vector k is $E_{p,n_p}(k)$.

The position dependent current density $\mathbf{J}(\mathbf{r})$ is calculated with³¹

$$\mathbf{J}(\mathbf{r}) = i \frac{e\hbar}{2m_0} \lim_{\mathbf{r}' \to \mathbf{r}} \left(\nabla_{\mathbf{r}'} - \nabla_{\mathbf{r}} \right) \langle \mathbf{r} | \psi \rangle \langle \psi | \mathbf{r}' \rangle
= i \frac{e}{2\hbar} \lim_{\mathbf{r}' \to \mathbf{r}} \left(\left[H_0, \mathbf{r} \right] - \left[H_0, \mathbf{r}' \right] \right) \langle \mathbf{r} | \psi \rangle \langle \psi | \mathbf{r}' \rangle,$$
(14)

where the Hamiltonian H_0 is the same as in Eq. (1). After some straight forward steps, where the orthogonality properties of the tight-binding basis as well as the same simplifications as in Eq. (13) are recalled, the following expression is obtained for $\mathbf{J}(\mathbf{r})$

$$\mathbf{J}(\mathbf{r}) = \sum_{i_{1},i_{2}} i \frac{e}{2\hbar} \sum_{p,n_{p}} \frac{\Delta}{2\pi} \int dE \left(H_{i_{1}i_{2}} C_{i_{2},p,n_{p}} C^{*}_{i_{1},p,n_{p}} - C_{i_{1},p,n_{p}} C^{*}_{i_{2},p,n_{p}} H_{i_{2}i_{1}} \right) \times \left| \frac{dE}{dk_{p,n_{p}}} \right|^{-1} f(E - \mu_{p}) (\mathbf{R}_{i_{2}} - \mathbf{R}_{i_{1}}) \delta(\mathbf{r} - \mathbf{R}_{i_{1}}).$$
(15)

The matrix elements H_{ij} represent the tight-binding coupling between an orbital σ_i on an atom situated at \mathbf{R}_i and an orbital σ_j on an atom situated at \mathbf{R}_j . The vector $\mathbf{R}_j - \mathbf{R}_i$

characterizes the bond connecting two atoms. In the ballistic regime, the current can be alternatively computed with the two-terminal Landauer formula for the non-interacting case ³². It requires the knowledge of the position-independent transmission coefficient $T(E)^{33}$. Hence, it does not allow to check current continuity.

B NEGF Solver

Recently, the Non-Equilibrium Green's Function (NEGF) formalism has become very popular for the simulation of quantum transport in nanodevices. In Ref.²⁰, it is shown that the scattering OBCs obtained above allow to calculate the retarded boundary self-energies Σ^{RB} of the NEGF formalism. The use of iterative solvers³⁴ for the OBCs calculation (20 to 40 inversions of a very dense or full matrix of size $t_b N$) is less efficient than the EVP approach.

The stationary retarded $G_{ij}^{R}(E)$ and lesser $G_{ij}^{\leq}(E)$ Green's functions are calculated with a recursive algorithm^{35,36} involving the Hamiltonian matrix **H** of Eq. (12). The indices *i* and *j* include the atom position **R** and the orbital type σ . With the same approximation as in the Wave Function case, one obtains

$$n(\mathbf{r}) = -i \int \frac{\mathrm{d}E}{2\pi} G^{<}(\mathbf{rr}; E)$$

= $-i \sum_{j} \int \frac{\mathrm{d}E}{2\pi} G^{<}_{jj}(E) \delta(\mathbf{r} - \mathbf{R_j})$ (16)

for the electron density and

$$\mathbf{J}(\mathbf{r}) = \frac{e\hbar}{2m_0} \lim_{\mathbf{r}' \to \mathbf{r}} \int \frac{\mathrm{d}E}{2\pi} \left(\nabla_{\mathbf{r}'} - \nabla_{\mathbf{r}} \right) G^<(\mathbf{rr}'; E)$$
$$= \frac{e}{2\hbar} \int \frac{\mathrm{d}E}{2\pi} \left(H_{i_1 i_2} G^<_{i_2 i_1}(E) - G^<_{i_1 i_2}(E) H_{i_2 i_1} \right) (\mathbf{R}_{\mathbf{i_2}} - \mathbf{R}_{\mathbf{i_1}}) \delta(\mathbf{r} - \mathbf{R}_{\mathbf{i_1}})$$
(17)

for the current density³¹. Equation (17) has exactly the same structure as Eq. (15) and yields a direct relation between the WF coefficients $C_{i,n,p}^{\sigma}(k)$ and the lesser Green's functions $G_{ij}^{\leq}(E)$. However, the NEGF current density is often calculated with the Landauer-Büttiker formalism^{32,35} and not with Eq. (17).

III. RESULTS

In this Section, simulations of Si and GaAs nanowires with different cross sections, crystal orientations, and lengths are presented. The contacts are assumed ideal (semi-infinite prolongation of the first and of the last slab of the nanowire). The central scattering region of the device has a length L_w . According to Fig. 1, x is the transport direction, y and z are directions of confinement and delimit the device cross section. The $sp^3d^5s^*$ tight-binding parameters were optimized by T. B. Boykin et al. to reproduce Si¹⁴ and GaAs¹⁵ bulk bandstructures. The orbital on-site energies as well as the connections to the nearest neighbors are assumed to remain the same in nanowires. Hard wall boundary conditions are applied to the surface atoms by "passivating" their dangling-bonds³⁷.

A Ideal Nanowires

First we consider ideal nanowires whose active part of length L_w cannot be distinguished from the left and the right semi-infinite reservoirs. The electrostatic potential along these structures is constant, if no external bias is applied, so that they can be considered as infinitely long. Hence, all the nanowire slabs are identical and it is possible to compute the resulting wire bandstructure. Since y and z are directions of confinement, there is only one free direction, x, reducing the crystal Brillouin Zone to a one-dimensional quantity depending on k.

Figures 2 (a) and (d) depict the cross section of two different infinite nanowires. The visible atoms represent a wire slab projected into one single plane. If a slab has a width Δ , an infinite ideal wire is generated by repeating the slab structure along the *x*-direction for each position $x_n = n \times \Delta$ (*n* is an integer).

In the subplots (b) and (e) of Fig. 2 the conduction subbands (calculated without spinorbit coupling) corresponding to the infinite nanowires shown in (a) and (e) are presented. Half of the one-dimensional Brillouin zone is drawn due to symmetry with respect to k=0. The wave vectors are normalized with their maximum value $k_{max} = \pi/\Delta$.

The plots (c) and (f) show the electron transmission of the nanowires whose cross section and conduction subbands are given in (a), (d) and (b), (e), respectively. The OBCs are calculated with the eigenvalue method proposed in this paper. They are then coupled to a Wave Function solver as described in Section II A. Spin-orbit coupling is neglected, because it does not affect much the conduction band properties as it will be proved later in this Section. In infinite structures the transmission coefficients T(E) can be directly extracted from their band structure. In effect, the semi-infinite left and right reservoirs as well as the central channel they are connected to form a uniform nanowire whose electrical properties do not vary in transport direction. In this case, if there are *n* available incident modes at energy E^* (i.e. modes with a positive velocity in the left reservoir and a negative velocity in the right reservoir), one has $T(E^*) = n$. This is a good way to check, whether the results obtained with the procedure outlined in Section II are correct.

The two examples depicted in Fig. 2 are studied now. The first structure (a) is a square Si nanowire (2.1 nm × 2.1 nm) where the transport axis is aligned with [100], y with [010] and z with [001]. A slab contains four atomic layers (width Δ =0.543 nm = lattice constant of bulk Si) or 128 atoms, the active part of the wire has a length L_w =32.6 nm (60 slabs), and therefore 7680 atoms. The square matrix M involved in the calculation of the OBCs has the size N_M =1280, but the arising eigenvalue problem has a reduced size of N_{EVP} =640. This implies that 50% of N_M is sufficient to calculate the k's and the φ 's.

From the band structure plot in Fig. 2 (b), one can infer the transmission given in plot (c). For example, there are four subbands starting at around E=1.515 eV. Exactly at the same energy four channels turn on in the transmission. Due to the two-dimensional confinement, these four conduction subbands have a minimum at k=0. Four of the six split conduction band valleys, labeled "unprimed valleys", are projected back to the Γ point in Si [100]-oriented nanowires. At k=0.35 and E=1.657 eV a subband minimum is observed. At this energy point the transmission increases from 4 to 6 (channels), because a state with a positive velocity appears at $k = 0.35 + \delta$, but also on the other side of the Brillouin zone, at

 $k = -0.35 + \delta$. These two bands come from the two remaining "primed valleys". When one branch of these two bands merges with the other bands at k=0.22 or k=-0.22 and E=1.67eV the transmission encounters a double step down (from 6 to 4) marking the turn-off of two channels. Following this procedure, the complete transmission can be explained, proving that the results obtained in the quantum transport calculation are correct.

The subplots (d), (e), and (f) in Fig. 2 show the results for a GaAs circular nanowire (diameter of 2.5 nm) with [111] as transport direction, $[1\bar{1}0]$ for y, and $[11\bar{2}]$ for z. The device is composed of 33 slabs (width $\Delta=0.979$ nm, made up of six atomic layers and 218 atoms whose half are visible and half hidden behind). The size of the matrix M in Eq. (8) is $N_M=2180$, but an eigenvalue problem with $N_{EVP}=730$ has to be solved for the boundary conditions (33.5% of N_M). The behavior of the transmission can also be explained by the wire band structure. GaAs is a direct band gap material so that the first subband is not degenerate and it has its minimum at the Γ point (E=2 eV). At the same energy the first channel turns on in the transmission plot. The second band is more interesting since it has a minimum at $k=\pm 1$ and E=2.11 eV, exactly where the second channel turns on. It comes from the upper bulk valley situated at the L-point in [111]-direction and E=1.709 eV in the three-dimensional Brillouin zone. Due to the confinement, this valley becomes aligned with the wire transport axis x and causes subbands with minima at $k=\pm 1$.

More examples of ideal nanowires with other shapes (triangular, hexagonal, ...) and other crystal orientations ([110] or [113]) are given in Ref.²⁰.

B Ballistic Transport

In this Section the ideal Si nanowires are supplemented with a triple-gate contact to form field-effect transistors (FETs). We first assume that electrons or holes can only be injected or collected from the left (source) and the right (drain) contacts, not from the gate. The room temperature IV-characteristics of n-FETs and p-FETs with channel orientation along the [100], [110], [111], and [112] crystal axis are simulated. The full-band (FB) charge and current densities are calculated as described in Section II. They are solved self-consistently with the three-dimensional (3D) electrostatic potential (Poisson equation) of the device. Elementary 'top-of-the-barrier' approaches³⁸ based on the band structure of ideal nanowires reduce the electrostatic potential to a one-dimensional quantity and consider semi-classical transport. But they also give a good insight into the device IV-characteristics. However, their application is restricted to the ballistic transport of perfect atomic structures (no spatial variation of the nanowire cross section), whose energy dispersion relation (E - k) can be calculated. Our simulator furthermore includes short channel effects such as drain-induced barrier lowering, source-to-drain tunneling, lateral potential variation, and allows to treat Si-SiO₂ interface roughness at the nanowire surface. Energy relaxation of the atom positions is out of the scope of this work. To emphasize the importance of full-band simulations, the IV-characteristics of a [100] and a [110] channel-oriented wire are compared to effective mass (EM) calculations obtained by a three-dimensional coupled mode-space NEGF solver^{39,40}.

To solve the three-dimensional Poisson equation in the nanowire FETs without loosing the δ -character of the carrier density, the finite element method (FEM) is chosen^{41,42}. A Delaunay grid is constructed by imposing the constraint that no atom is contained in the volume of any tetrahedron mesh element. The atom positions are given as input parameter to the Qhull mesh generator⁴³, which uses them as fixed nodes. The δ -functions thus disappear when the electron density $n(\mathbf{r})$ is integrated with the FEM test functions.

The SiO₂ oxide layers do not participate to the transport calculation (only poor tightbinding representation available), but they are included to obtain the 3D electrostatic potential. Consequently, the oxide grid points carry no charge and hard wall boundary conditions are applied to the Si surface atoms. This means that the Si dangling bonds are passivated by increasing the on-site energies of the *s*- and *p*-orbitals³⁷.

The upper part of Fig. 3 shows the schematic view of a triple-gate nanowire transistor of length $L_s + L_g + L_d \approx 32.5$ nm, composed of n-doped (p-doped) source and drain contacts with $N_D = 10^{20}$ cm⁻³ ($N_A = 10^{20}$ cm⁻³), and a three-part gate (top, left, and right). The transistor is deposed on a buried oxide and is surrounded by three oxide layers of thickness $t_{ox}=1$ nm. The source, the drain, and the gate measure $L_s=9.7$ nm, $L_d=9.7$ nm, and $L_g=13$ nm, respectively. As in Fig. 1, x is the transport direction, y and z are directions of confinement. The cross section of the nanotransistor (neglecting the buried oxide) has the size $L_y=4.1$ nm times $L_z=3.1$ nm. In the lower part of Fig. 3, the cross sections corresponding to transport along the (a) [100], (b) [110], (c) [111], and (d) [112] crystal axis are given. Points represent Si atoms while lines model atomic bonds. The SiO₂ layers are not described in the tight-binding formalism. The nanowire channel (only Si atoms) is a 2.1 nm x 2.1 nm square. A potential V_g is applied to the gate (work function $\phi_m=4.15$ eV for n-FETs and $\phi_m=5.0$ eV for p-FETS), $V_s=0$ V to the source and V_d to the drain.

The [100] channel is composed of 60 slabs of width $\Delta=0.543$ nm (transistor length= $60 \times \Delta=32.6$ nm), each slab has 4 atomic layers and is made out of 128 atoms. For the [110] channel, 84 slabs of width $\Delta=0.384$ nm are counted, each of them with 2 atomic layers and 88 atoms. The [111] channel contains 34 slabs ($\Delta=0.941$ nm) with 6 atomic layers (210 atoms) per slab. Finally 49 slabs ($\Delta=0.665$ nm) are present in the [112] wire, each of them has 6 atomic layers and 154 atoms.

For each channel type there are more than 7000 atoms constituting the nanowire. This means that the transport problem is described by matrices of size N>140000 if spin-orbit coupling is included or of size N>70000 if it is omitted. Consequently, at each injection energy E we have to deal with large complex matrices. However, the spin-orbit coupling energy in Si is relatively small such that this effect can be neglected. To justify this approximation the transmission through the conduction and the valence bands of infinite Si nanowires with the same size and channel orientation as in Fig. 3 are plotted in Fig. 4. The black solid lines represent the transmission with spin-orbit coupling, the dashed gray lines the transmission without spin-orbit coupling multiplied by a factor 2. In the energy range significant for charge and current density calculations (\approx 150 meV from the band edge) the curves with and without spin interaction are almost identical, except for the valence band of the [100] nanowire. Nevertheless this small discrepancy will not affect the conclusions drawn later in this Section.

In Fig. 5, the FB (black lines) and the EM (gray lines) $I_d - V_{gs}$ transfer characteristics at $V_{ds}=0.4$ V are compared for n-FETs with a [100] (lines with circles) and a [110] (lines without symbols) oriented channel. In both cases the current in the subthreshold regime is underestimated by the effective mass approximation. This leads to an increase of the transistor threshold voltage V_{th} . However, for the [100] nanowire, the EM ON-current (defined as drain current I_d at $V_{gs}=0.6$ V and $V_{ds}=0.4$ V) is larger than its FB counterpart, as can be seen in the inset of Fig. 5. This is the same conclusion as in Ref.¹¹.

Figure 6 shows the 3D full-band $I_d - V_{gs}$ ($V_{ds} = \pm 0.4$ V) transfer characteristics of n-FETs (right) and p-FETs (left) with [100] (black dashed line), [110] (black line with circles), [111] (black solid line), and [112] (gray solid line) as channel orientation. The comparison is done for the same gate metal work function $\phi_m = 4.15$ eV for the n-doped structures and $\phi_m = 5.0$ eV for the p-doped ones. This means that ϕ_m is not adjusted to obtain the same OFF-current ($I_{OFF} = I_d$ at $V_{gs} = 0$ V and $V_{ds} = \pm 0.4$ V) as done in Ref.³⁸.

As also predicted by the 'top-of-the-barrier' approach³⁸, the channel orientation [110] offers the highest ON-current in the case of n-FETs. For I_{ON} defined at voltages V_{gs} =0.6 V and V_{ds} =0.4 V we find a 1.8 times larger value than for [100]-orientation, a 2.6 times higher current compared to [112], and a 6.5 times higher current compared to [111]. For p-FETs, the trends are also the same in both approaches. The highest ON-current (taken at V_{gs} =-0.6 V and V_{ds} =-0.4 V) is obtained for the [110] channel (15.7 μ A). It is 1.5 times higher than for [111], 1.7 times higher than for [112], and 6.8 times higher than for [100]-orientation. The values are summarized in Table I. The prominence of the [110] channel for both n-FETs and p-FETs is a consequence of the good compromise between the electron velocity (proportional to the inverse of the effective mass) and the number of available states (proportional to the effective mass) that is reached by the lowest subbands of the [110] wire. The lowest [111] conduction subbands, for example, have too heavy electrons that cannot flow efficiently from source to drain.

C Interface Roughness

Nanowires with a perfect crystal structure and a well-defined Si-SiO₂ interface will probably never be grown. The 3D self-consistent inclusion of the electrostatic potential in the transport calculations makes it possible to simulate wires with a cross section (or slab contour) that varies from source to drain^{25,27}. The example of a non-ideal wire slab (channel orientation x=[100]), taken under the gate at the slab position 35 over 60, is given in Fig. 7. The atom positions in the middle of the slab are all occupied while some atoms are missing or are added to the Si-SiO₂ interface. Therefore, slab 35 includes 123 atoms instead of 128 as in the ideal case in Fig. 3. This means that L_y and L_z slightly vary from their original value, but not L_s , L_g , and L_d .

The interface roughness is generated by randomly distributing the atoms at the Si-SiO₂ interface according to an exponential autocovariance function $\Gamma_s(x)^{44,45}$

$$\Gamma_s(x) = \Delta_m^2 \cdot e^{-|x|/L_m} , \qquad (18)$$

where Δ_m is the root mean square of the rough Si surface, x is the distance between two Si atoms at the Si-SiO₂ interface, and L_m is the correlation length of the roughness. As a first assumption, the root mean square of the rough surface $\Delta_m=0.14$ nm and its correlation length $L_m=0.7$ nm are the same for all the surface types ([100], [110], [111], and [112]).

Due to the rough Si-SiO₂ interface all the nanowire slabs are different and it is not possible to first construct a two-dimensional mesh with triangle elements on each slab and then to interpolate between the slab meshes to obtain tetrahedra. A full 3D FEM mesh, as projected in Fig. 7, must be generated to calculate the electrostatic potential $V(\mathbf{r})$. The mesh elements are finer around the atoms than in the oxide layers, where $V(\mathbf{r})$ varies linearly due to the absence of charge (the second derivative of $V(\mathbf{r})$ vanishes).

In Fig. 8 (taken from Ref.²⁷), the full-band $I_d - V_{gs}$ (transfer) characteristics at $V_{ds}=0.4$ V of a [100] (solid line with circles and dotted line) and a [110] (solid and dashed lines) structure with interface roughness are compared to the results of ideal structures. For the

simulated realizations of the random Si-SiO₂ interface, the [100] drain current I_d at $V_{gs}=0.4$ V is reduced to about 30% of its ideal value, the [110] drain current at $V_{gs}=0.4$ to 65% of the ideal I_d . The drain current in the subthreshold region is seriously deteriorated by the variation of the nanowire cross section. This is due to a mobility decrease caused by interface roughness scattering. At the same time, the oxide thickness t_{ox} is modulated, leading to a shift of the threshold voltage V_{th} , but the swing S remains close to S=60 mV/decade (the current is only controlled by the barrier height).

However, another random distribution of the interface atoms leads to a different transfer characteristics. Since the subthreshold swing S is not much affected by the interface roughness, several I_d samples can be calculated at a fixed V_{gs} and the variations of the threshold voltage V_{th} are extrapolated from the results. For each channel orientation, 105 different Si-SiO₂ interface realizations are simulated, the mean drain current \tilde{I}_d is calculated, and the variation of V_{th} is obtained for each sample I_d as

$$\Delta V_{th} = S \cdot \log_{10} \left(\tilde{I}_d / I_d \right). \tag{19}$$

The histogram distributions of ΔV_{th} are shown in Fig. 9 (taken from Ref.²⁷) for (a) the [110], (b) the [100], (c) the [111], and (d) the [112] n-FET nanowires. The standard deviation σ of ΔV_{th} is also given for each channel configuration.

To design a circuit, all the involved transistors should have the same characteristics although they have slightly different structures. Hence the standard deviation of ΔV_{th} should be as small as possible. The lowest value of σ is obtained for a channel oriented along the [110] crystal axis (σ =9.8 [mV]). [100] follows with σ =13.8 [mV], then [112] (σ =16.8 [mV]), and [111] (σ =24.8 [mV]).

All the I_d samples calculated to generate the histograms in Fig. 9 must fulfill current continuity. This very important property in the simulation of ballistic transistors should not be altered by interface roughness. To demonstrate its validity, the current I_d flowing from one nanowire slab to the other is reported in Fig. 10 for $V_{ds}=0.4$ V and for different V_{gs} . The current $I_d(x_n)$ is calculated in a non-ideal [100] nanowire with Eq. (15) by summing all the contributions from one slab to the next. Even when a nanowire transistor is composed of slabs with different shapes and with a different number of atoms, the current is conserved in our approach.

Finally, Fig. 11 shows how interface roughness affects the electron distribution in a [100] transistor. Because of the δ -character of the electron density $n(\mathbf{r})$, only the total number of electrons in each wire slab can be plotted. The number of electrons per slab with interface roughness depends on three different effects: the electrostatic potential, the confinement in the y and z directions, and the conduction band edge lowering or increase caused by the modifications of the slab shapes (gray line in Fig. 11). For example, in slab 27, the conduction band edge is lowered by the roughness so that more electrons are present in the non-ideal structure (black solid line) than in the ideal one (black dashed line). In slab 33, the opposite effect occurs. As a consequence of the conduction band increase, a reduction of the electron population is observed.

D Source-to-Drain Tunneling

Source-to-drain tunneling is investigated for the n-doped Si triple-gate nanowire transistors depicted in Fig. 3. For that purpose we keep the same dimensions L_s , L_d , L_y , and L_z as there, but we let the gate length L_g vary from 4 nm to 13 nm. Interface roughness is no more taken into account. The full-band $I_d - V_{gs}$ characteristics are computed for the [100], [110], [111], and [112] channel orientations and for $L_g=4$, 7, 10, and 13 nm. The subthreshold swing S is then extracted from these results to determine the scaling limit of nanowire transistors⁴⁶. If the physical gate L_g is long, the current will flow above the gate-induced potential barrier only. However, as the gate length of the transistors decreases, tunneling from source to drain becomes important for electrons. It increases the subthreshold swing S and the OFF-current, and therefore degrades the transistor performances. Ideally, S should be equal to 60 mV/decade at room temperature.

In order to illustrate the behavior of source-to-drain tunneling, the energy- and position-

resolved drain current is depicted in Figs. 12 for the [110] nanowire with different channel lengths L_g . Scattering has no significant effect on the subthreshold slope. Hence, it is not included in the simulation, which conserves the spectral current along the device. The shorter the gate becomes, the higher the probability that an electron travels from source to drain through the barrier. This can be seen in the transfer characteristics $I_d - V_{gs}$ in Fig. 13 and in the corresponding subthreshold swing curves in Fig. 14. Values of S are also reported in Table II, where the * refers to results obtained with the effective mass approximation.

First, the importance of a full-band treatment of the source-to-drain tunneling is apparent. As can be seen in Fig. 12, for short gate lengths the current essentially comes from the tunneling of electrons through the potential barrier. In this case states situated well below the conduction band edge start to play a dominant role. Since these states are a mixture of electron and hole states, their properties can only be accurately reproduced by a FB model. This is particularly true for the $sp^3d^5s^*$ tight-binding model which yields the correct electron and hole effective mass at the same time¹⁴. The results in Table II confirm that the effective mass approximation yields large errors in the simulated source-to-drain tunneling current. At $L_g=4$ nm, the EM subthreshold swing S is 12.6% higher than the FB one for the [100] wire and 18.6% smaller for the [110] wire.

Surprisingly, the subthreshold swing S is almost the same for the [100] and the [110] nanowires in the effective mass approximation. This can be explained by the nature of the subbands that are responsible for the tunneling current. For both the [100] and [110] configurations, there are two out of the six split valleys with a transport effective mass m_x equal to the transverse mass $m_t=0.19m_o$, a confinement mass $m_y=m_t$ and $m_z=m_l=0.92m_o$ ⁴⁰. These two valleys generate the lowest conduction subbands (because $m_z=m_l$) which have a small tunneling mass $(m_x=m_t)$.

The study of the FB results shows that source-to-drain tunneling drastically increases the subthreshold swing S when L_g shrinks to 4 nm. For all the transport directions considered in this work, S becomes larger than 80 mV/decade. A gate length L_g as short as 7 nm does not affect too much the transistor performances for a channel along the [100], [111], and

[112] crystal axis. S remains close to its ideal value of 60 mV/decade.

The highest S at $L_g=4nm$ and $L_g=7$ nm are obtained for a [110] oriented channel. The electron effective mass in transport direction, m^* , is responsible for such a behavior. While a low m^* , as for the [110] orientation, is advantageous for a high ON-current (see Fig. 6), it limits the gate length scaling by increasing the tunneling probability, the OFF-current, and the subthreshold swing. On the other side, a large effective mass m^* in transport direction restricts source-to-drain tunneling, but does not allow to achieve high ON-current.

IV. IMPLEMENTATION ISSUES

To calculate the transmission through ideal nanowires or to simulate the output characteristics of a nanotransistor, electrons must be injected into the device from all the contacts and at each possible energy. Typically, 500 energy points are necessary for each Poisson self-consistent iteration. Therefore, the sparse linear system in Eq. (12) has to be solved 500 times, if the Wave Function formalism is chosen. Since there is no coupling for different energies, the solutions for all energies can be obtained in parallel. The message passing interface $(MPI)^{47}$ is called to distribute them on different CPUs. A total of N available CPUs corresponds to a speed up factor of N. This is the first level of parallelism implemented in our transport simulator. However, it is not always possible to solve Eq. (12) on a single machine, if, for example, the size of the matrix $(\mathbf{E} - \mathbf{H})$ is too large or if the memory per processor is too low. In this case a pool of M CPUs is dedicated to one energy point. With N CPUs, N/M energy points can be treated at the same time. This is the second level of parallelism. Assuming that T_0 is the time necessary to solve a linear system on a single CPU, M CPUs working on the same system seldom reduce the time T_0 to T_0/M . Hence, the speed-up factor of the second level of parallelism is always smaller than N (except for particular occasions).

At a given energy point the first task consists in calculating the OBCs in order to construct $(\mathbf{E} - \mathbf{H})$. In Section II we stated that our approach is more efficient than the usual methods

 23,24,30,33,34,48,49 classified into three categories. (i) There are approaches which do not work for nanowire described by the $sp^3d^5s^*$ tight-binding model^{33,48}. (ii) There are the methods based on the solution of a generalized eigenvalue problem (GEVP)^{23,24,30,49} as in Eq. (6). (iii) An iterative algorithm³⁴ is used to calculate the OBCs. The transport code, written in C++, calls Fortran LAPACK⁵⁰ routines to solve the normal and generalized eigenvalue problems and to invert or manipulate full matrices.

In Table III, we report the OBCs computational time for the two ideal nanowires depicted in Fig. 2 (without spin-orbit coupling). The generalized eigenvalue problem, the iterative procedure (labeled "Sancho-Rubio"), and our EVP method are tested on a 64-bit Sun Fire X4600 with 8x2.8 GHz Dual Core Opteron processors. The second and the third column of the table characterize the contact structure (number of layers and of atoms N_A). Hence, the GEVP is of size $2 \times N_A \times t_b$, where t_b is the tight-binding order, and full matrices of size $N_A \times t_b$ have to be inverted in the iterative scheme. The fourth column gives the size of the normal eigenvalue problem involved in our method. This represents a significant reduction, shown in the fifth column of the table, in comparison to the size of the original matrix M $N_A \times t_b$. In the two last columns of Table III the CPU time (on a single processor) of the different methods is presented. The benchmarks refer to the calculation of one contact only, the source or the drain. While the computational burden associated to the approaches based on eigenvalue problem is energy-independent, the iterative algorithm can be more or less efficient depending on the injection energy. In any case the method we reviewed in Section II is about two orders of magnitudes faster than the others. Further comparisons for smaller structures can be found in Ref.²⁰.

Once the OBCs calculation are obtained, the sparse matrix $(\mathbf{E} - \mathbf{H})$ in Eq. (12) is assembled and factorized. This task is either processed directly by sparse linear solvers like Umfpack 5.0.1⁵¹, PARDISO⁵², SuperLU_{dist} 2.0⁵³, MUMPS 4.6.3⁵⁴ (all based on a *LU* factorization), or indirectly by using a recursive Green's Function (RGF) algorithm^{35,36}. The packages SuperLU_{dist} 2.0 and MUMPS 4.6.3 are designed for distributed memory computers using MPI⁴⁷ to communicate, PARDISO works on shared memory machines, while Umfpack 5.0.1 and RGF are sequential.

To compare the different methods, Eq. (12) is solved for the [100] Si nanowire in Fig. 2. Different wire sizes are studied. The length (60 Slabs or 32.6 nm) remains constant, but the original cross section $(2.1 \times 2.1 \text{ nm}^2)$, equivalent to $4a_0 \times 4a_0)$ is enlarged to reach $5a_0 \times 5a_0$ and $6a_0 \times 6a_0$. For the latter case the parallel performance and scalability of the linear solvers are tested. The results are summarized in Table IV. All simulations were run on a 64-bit Sun Fire X4600 with 8x2.8 GHz Dual Core Opteron processors. Columns 2 to 6 reproduce the time (in seconds) to factorize and solve Eq. (12) with Umfpack 5.0.1, PARDISO, SuperLU_{dist} 2.0, MUMPS 4.6.3, and RGF, respectively. The seventh column is the time necessary to obtain the OBCs for one contact with the normal EVP approach. The size of the sparse matrix ($\mathbf{E} - \mathbf{H}$) is found in the last column (spin-orbit coupling not included).

For the smallest nanowire, Umfpack 5.0.1 and MUMPS 4.6.3 are the fastest solvers, but PARDISO is efficient, too. The RGF algorithm which requires the inversion of full block matrices is about 4 times slower than the fastest package. As the nanowire size increases the performance of Umfpack 5.0.1 decreases. For the $6 \times 6 \times 60$ wire, MUMPS 4.6.3 is almost three time faster than Umfpack 5.0.1, while they have the same speed for the $4 \times 4 \times 60$ structure. RGF once again cannot compete with MUMPS 4.6.3 or PARDISO which are 7.27 and 5.84 times more efficient, respectively. The analysis of the parallel scalability shows that $SuperLU_{dist}$ 2.0 is better than the others. This must be carefully considered since the factorization and solve time on 2 CPUs is more than two times faster than on a single CPU. This could originate from the fact that this package was designed for parallel use only. Despite its good scalability SuperLU_{dist} 2.0 remains about two times slower than PARDISO and MUMPS 4.6.3, which have similar scalability properties. Hence, the best packages tested in our transport simulator are PARDISO and MUMPS 4.6.3. The latter has the advantage to run on both distributed and shared memory machines. On the other side, the least attractive approach is the recursive Green's function algorithm. For ballistic transport, it is less efficient than a LU factorization with adequate sparse solvers. This trend could be modified by the inclusion of scattering into the devices.

V. CONCLUSION

In this review, a full-band approach to simulate nanowire transistors at the atomic scale is presented. The quantum transport solver is based on the semi-empirical $sp^3d^5s^3$ tightbinding model. It includes an atomistic description of the device structure, an improved procedure to calculate open boundary conditions, and different linear solvers working either with the Wave Function or with the Non-Equilibrium Green's Function formalism. Charge and current densities are self-consistently coupled to the calculation of the Poisson equation. The simulation of aggressively scaled triple-gate nanowire field-effect transistors with different channel orientations is achieved. We compare the ON-current, the interface roughness sensitivity, the subthreshold swing, and the scalability of transistors whose transport direction is aligned with the [100], [110], [111], and [112] crystal axis. It is revealed that some properties, like a low transport effective mass, can be good for certain aspects (ONcurrent, interface roughness), but bad for others (source-to-drain tunneling). Despite the promising results obtained with the current simulator, some physical issues should be further investigated for future improvements.

In the nanowire structures depicted in Fig. 3, the doping concentration is equally distributed in the source and drain regions. However, it would be more meaningful to treat each doping atom individually and to carefully describe them in the transport calculation. Accurate tight-binding parameters are required for that purpose. This holds likewise for the SiO_2 oxide layers that appear only in the calculation of the electrostatic potential. In this way, no electron penetration into the oxide is allowed and the surface Si atoms are artificially passivated. Furthermore, gate leakage currents, which are a performance limiting factor for very thin oxide layers, cannot be taken into account. As a consequence, the OFF-current of the transistors is not correctly modeled.

When the nanowire cross sections are small, as in our examples, the surface atoms relax

from their "equilibrium" position in order to minimize the total energy of the system. This is particularly true when single atoms are added or removed at the surface. These effects are not considered in this work, but they are certainly important, especially in the case of an atomistic description of SiO_2 .

Scattering mechanisms may play a considerable role in a nanowire transistor, but they are not treated in this paper, with the exception of interface roughness. This effect can be directly included in the Hamilton matrix, which is not the case for inelastic interactions such as electron-optical phonon scattering. This complicated effect must be probably included via self-energies as described in Ref.³⁵, for example. Furthermore, the use of confined phonons instead of bulk phonons becomes necessary at the atomic scale.

Our full-band quantum transport simulator has also the capability of treating mechanical strain⁵⁵ and other materials than Si or GaAs (Ge, InAs, SiGe, ...). However, it is limited by the dimension of the nanowire cross section and its inherent computational burden. Each injection energy requires the solution of two eigenvalue problems for the open boundary conditions and of one linear system for the transport through the nanowire. These tasks can be distributed on parallel CPUs, each of them working on a different energy point, but for large cross sections, due to memory restriction, a group of CPUs has to be dedicated to one single energy. If a computer cluster with several hundreds of CPUs is available, it is possible to simulate any transistor structure. On the other hand, as the cross section of the structure grows, the effective mass approximation becomes more and more accurate and requires much less computational power. The user will have to determine which method is the more indicated to his application.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Fond (SNF), project 200021-109393 (NEQUATTRO).

REFERENCES

- ¹ B. Doris et al., IEDM Tech. Dig. 2002, 267 (2002).
- ² Y. Cui, L. J. Lauhon, M. S. Gudiksen, J. Wang, and C. M. Lieber, Appl. Phys. Lett. 78, 2214 (2001).
- ³ A. I. Persson, M. W. Larsson, S. Steinström, B. J. Ohlsson, L. Samuelson, and L. R. Wallenberg, Nature Materials 3, 677 (2004).
- ⁴ A. B. Greytak, L. J. Lauhon, M. S. Gudiksen, and C. M. Lieber, Appl. Phys. Lett. 84, 4176 (2004).
- ⁵ J. Xiang, W. Lu, Y. Hu, Y. Wu, H. Yan, and C. M. Lieber, Nature 441, 489 (2006).
- ⁶ T. Saito, T. Saraya, T. Inukai, H. Majima, T. Nagumo, and T. Hiramoto, IEICE Trans. Electron. E85-C, 1073 (2002).
- ⁷ H. Majima, Y. Saito, and T. Hiramoto, IEDM, Tech. Dig. 2001, 733 (2001).
- ⁸ Y. Cui, Z. Zhong, D. Wang, J. Wang, and C. M. Lieber, Nano Lett. 3, 149 (2003).
- ⁹ Y. Hayamizu, M. Yoshita, S. Watanabe, H. Akiyama, L. N. Pfeiffer, and K. W. West, Appl. Phys. Lett. 81, 4937 (2002).
- ¹⁰ Z. Ren, R. Venugopal, S. Datta, M. S. Lundstrom, D. Jovanovic, and J. G. Fossum, IEDM Tech. Dig. 2000, 715 (2000).
- ¹¹ J. Wang, A. Rahman, A. Ghosh, G. Klimeck, and M. S. Lundstrom, IEEE Trans. on Elec. Dev. 52, 1589 (2005).
- ¹² J. C. Slater and G. F. Koster, Phys. Rev. 94, 1498 (1954).
- ¹³ J. -M. Jancu, R. Scholz, F. Baltram, and F. Bassani, Phys. Rev. B 57, 6493 (1998).
- ¹⁴ T. B. Boykin, G. Klimeck, and F. Oyafuso, Phys. Rev. B 69, 115201 (2004).
- ¹⁵ T. B. Boykin, N. Kharche, G. Klimeck, and M. Korkusinski, J. Phys.: Condens. Matter 19, 036203 (2007).
- ¹⁶ J. M. Jancu, F. Bassani, F. D. Sala. and R. Scholz, Appl. Phys. Lett. 81, 4838 (2002).
- ¹⁷ S. Datta, J. Phys.: Condens. Matter 2, 8023 (1990).
- ¹⁸ S. Datta, Superlattices Microstruct. 28, 253 (2000).

- ¹⁹ W. R. Frensley, Rev. Mod. Phys. 62, 745 (1990).
- ²⁰ M. Luisier, G. Klimeck, A. Schenk, and W. Fichtner, Phys. Rev. B 74, 205323 (2006).
- ²¹ Y. -J. Ko, M. Shin, S. Lee, K. W. Park, J. of Appl. Phys. 89, 374 (2001).
- ²² Y. Zheng, C. Rivas, R. Lake, K. Alam, T. B. Boykin, and G. Klimeck, IEEE Trans. on Elec. Dev. 52, 1097 (2005).
- ²³ M. Städele, B. R. Tuttle, and K. Hess, J. of Appl. Phys. 89, 348 (2001).
- ²⁴ C. Rivas and R. Lake, Phys. Stat. Sol. B 239, 94 (2003).
- ²⁵ M. Luisier, A. Schenk, and W. Fichtner, IEDM, Tech. Dig. 2006, 811 (2006).
- ²⁶ T. B. Boykin, M. Luisier, A. Schenk, N. Kharche, and G. Klimeck, IEEE Trans. on Nanotech. 6, 43 (2007).
- ²⁷ M. Luisier, A. Schenk, and W. Fichtner, Appl. Phy. Lett. 90, 102103 (2007).
- ²⁸ Y. -C. Chang and J. N. Schulman, Phys. Rev. B 25, 3975 (1982).
- ²⁹ T. B. Boykin, Phys. Rev. B 54, 8107 (1996).
- ³⁰ T. B. Boykin, Phys. Rev. B 54, 7670 (1996).
- ³¹ P. Danielewicz, Annals of Phys. 152, 239 (1984).
- ³² Y. Meir and N. S. Wingreen, Phys. Rev. Lett. 16, 2512 (1992).
- ³³ D. Z. -Y. Ting, E. T. Yu, and T. C. McGill, Phys. Rev. B 45, 3583 (1992).
- ³⁴ M. P. Lopez Sancho, J. M. Lopez Sancho, and J. Rubio, J. Phys. F: Met. Phys. 15, 851 (1985).
- ³⁵ R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, J. of Appl. Phys. 81, 7845 (1997).
- ³⁶ A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel, and R. Venugopal, J. of Appl. Phys. 91, 2343 (2002).
- ³⁷ S. Lee, F. Oyafuso, P. von Allmen, and G. Klimeck, Phys. Rev. B 69, 045316 (2004).
- ³⁸ J. Wang, A. Rahman, G. Klimeck, and M. Lundstrom, IEDM, Tech. Dig. 2005, 530 (2005).
- ³⁹ M. Luisier, A. Schenk, and W. Fichtner, J. Appl. Phys. 100, 043713 (2006).
- ⁴⁰ M. Bescond, N. Cavassilas, K. Kalna, K. Nehari, L. Raymond, J. L. Autran, M. Lannoo, and A. Asenov, IEDM Tech. Digest. 2005, 526 (2005).

- ⁴¹ P. M. Gresho and R. L. Sani, Incompressible Flow and the Finite Element Method: Isothermal Laminar Flow, John Wiley and Sons, New York (2000).
- ⁴² J. N. Reddy and D. K. Gartling, The Finite Element Method in Heat Transfer and Fluid Dynamics, CRC Press, Boca Raton (1994).
- ⁴³ C. B. Barber, D. P. Dobkin, and H. T. Huhdanpaa, ACM Trans. on Mathematical Software 22, 469 (1996).
- ⁴⁴ S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy, and O. L. Krivanek, Phys. Rev. B 32, 8171-8186, (1985).
- ⁴⁵ J. Wang, E. Polizzi, A. W. Ghosh, S. Datta, and M. S. Lundstrom, Appl. Phys. Lett. 87, 043101 (2005).
- ⁴⁶ M. Luisier, A. Schenk, and W. Fichtner, to appear in the proceeding of the SISPAD 07 in Vienna, Austria (2007).
- ⁴⁷ W. Gropp, E. Lusk, N. Doss, and A. Skjellum, Parallel Computing 22, 789 (1996).
- ⁴⁸ R. C. Bowen, W. R. Frensley, G. Klimeck, and R. Lake, Phys. Rev. B 52, 2754 (1995).
- ⁴⁹ C. Strahberger and P. Vogel, Physica B 272, 160 (1999).
- ⁵⁰ E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, LAPACK User's Guide, Third Edition, SIAM, Philadelphia (1999).
- ⁵¹ T. A. Davis, ACM Trans. on Math. Software 30, 165 (2004).
- ⁵² O. Schenk and K. Gärtner, J. of Future Generation Computer Systems 20, 475 (2004).
- ⁵³ X. S. Li and J. W. Demmel, ACM Trans. on Math. Software 29, 110 (2003).
- ⁵⁴ P. R. Amestoy, I. S. Duff, and J.-Y. L'Excellent, Comput. Methods in Appl. Mech. Eng. 184, 501 (2000).
- ⁵⁵ T. B. Boykin, G. Klimeck, R. C. Bowen, and F. Oyafuso, Phys. Rev. B 66, 125207 (2002).

LIST OF FIGURES

1	Schematic view of an atomistic nanowire. x is the transport axis, y and z	
	are directions of confinement, L_w represents the wire length. Each atom is	
	considered as well as the connections to its four nearest neighbors, except at	
	the surface.	31
2	Cross section (a and d), conduction subbands (b and e), and electron trans-	
	mission coefficient of two infinite nanowires. The first row refers to a 2.1 nm	
	\times 2.1 nm Si square wire (transport in [100]), the second row to a circular	
	GaAs wire with a radius of 1.25 nm (transport in [111]). The cross section	
	corresponds to the projection of a wire slab with atoms in different planes	
	into a single plane.	32
3	Schematic view of a triple-gate nanowire transistor. The 2.1×2.1 nm ² Si	
	channel is surrounded by three oxide layers of thickness $t_{ox}=1$ nm and is	
	deposed on a buried oxide. The source and the drain have a length $L_s = L_d = 9.7$	
	nm and the gate length L_g may vary between 4 nm and 13 nm. The cross	
	section of the nanotransistor (neglecting the buried oxide) has a size $L_y=4.1$	
	nm times $L_z=3.1$ nm. Along the transport direction x, the wire is composed	
	of 60 slabs for $x=[100]$ (a), 88 slabs for $x=[110]$ (b), 34 slabs for $x=[111]$	
	(c), and 49 slabs for $x=[112]$ (d). A slab corresponding to the different cross	
	sections is shown in the lower part of the figure. Each type of slab contains a	
	different number of atomic layers	33
4	Transmission through 2.1 nm \times 2.1 nm infinite nanowires with x aligned with	
	(a) $[100]$, (b) $[110]$, (c) $[111]$, and (d) $[112]$. Calculations are shown for the	
	conduction and the valence subbands with spin-orbit coupling (black solid	
	lines) and without (gray dashed lines, curves multiplied by a factor 2). \ldots	34

5	Full-band (FB, black lines) vs effective mass (EM, gray lines) transfer char-	
	acteristics of triple-gate nanowire transistors with a gate length $L_g{=}13~\mathrm{nm}$	
	and $V_{ds}=0.4$ V. The lines with circles correspond to the transport direction	
	$x=[100]$, the lines without symbols to $x=[110]$. The inset shows the $I_d - V_{gs}$	
	behavior of the transistor around $V_{gs}=0.6.$	35
6	Full-band transfer characteristics of triple-gate nanowire transistors with a	
	gate length $L_g=13$ nm. The left part of the plot compares the $I_d - V_{gs}$	
	$(V_{ds}=-0.4 \text{ V})$ of p-FETs with transport directions along $x=[100]$ (dashed line),	
	x=[110] (solid line with circles), $x=[111]$ (black line), and $x=[112]$ (gray line).	
	In the right part of the figure the $I_d - V_{gs}$ ($V_{ds}=0.4$ V) of n-FETs with the	
	same channel orientations are compared	36
7	Projection of the slab number 35 of a $[100]$ nanowire transistor with interface	
	roughness. Bullets are Si atoms, lines represent the projection of the three-	
	dimensional Delaunay finite element (FEM) mesh used to compute the self-	
	consistent electrostatic potential. \bigcirc 2006 IEEE	37
8	Full-band transfer characteristics $I_d - V_{gs}$ at $V_{ds} = 0.4$ V for perfect [100] and	
	$[110]$ nanowires (solid lines) and for $[100]$ and $[110]$ nanowires with ${\rm Si-SiO}_2$	
	interface roughness (dashed and dotted lines). \bigodot 2007 American Institute of	
	Physics	38
9	Histogram distribution of the 105 samples representing the variation of the	
	threshold voltage ΔV_{th} caused by interface roughness. The results for (a) the	
	$[110],(\mathrm{b})$ the $[100],(\mathrm{c})$ the $[111],\mathrm{and}(\mathrm{d})$ the $[112]$ Si nanowire transistors are	
	plotted. σ is the standard deviation of ΔV_{th} . © 2007 American Institute of	
	Physics	39
10	Full-band drain current $I_d(x)$ plotted along the x axis of the [100] nanowire	
	(L_g =13 nm, 60 slabs) at V_{ds} =0.4 V and V_{gs} =0.0 V, 0.2 V, 0.4 V, and 0.6 V.	
	Si-SiO ₂ interface roughness is present. \textcircled{C} 2006 IEEE	40

11	Number of electrons per nanowire slab at $V_{ds}=0.4$ V and $V_{gs}=0.5$ V for the	
	[100] n-FET nanowire with perfect $Si-SiO_2$ interface (black dashed line) and	
	with interface roughness (IR, black solid line). The gray line describes the	
	variation of the confinement potential due to IR	41
12	Source-to-drain spectral current in the nanowire transistor of Fig. 3 when \boldsymbol{x}	
	is aligned with [110]. The simulations were done for $L_g = 4$ nm, 7 nm, 10 nm,	
	and 13 nm, $V_{gs}=0$ V, and $V_{ds}=0.4$ V	42
13	Transfer characteristics $I_d - V_{gs}$ at $V_{ds} = 0.4$ V for the device of Fig. 3 with $x =$	
	(a) [100], (b) [110], (c) [111], (d) [112] and $L_g = 4$ nm, 7 nm, 10 nm, and 13 nm.	43
14	Subthreshold swing as function of the gate length L_g for the nanowire tran-	
	sistor of Fig. 3 with the transport direction x aligned with [100], [110], [111],	
	and [112]	44



FIG. 1:



FIG. 2:









FIG. 4:



FIG. 5:



FIG. 6:



FIG. 7:



FIG. 8:



FIG. 9:



FIG. 10:



FIG. 11:



FIG. 12:



FIG. 13:



FIG. 14:

LIST OF TABLES

Ι	ON-Current $I_{ON}=I_d$ at $V_{gs}=0.6$ V and $V_{ds}=0.4$ V (n-FETs) or $I_{ON}=I_d$ at	
	V_{gs} =-0.6 V and V_{ds} =-0.4 V (p-FETs) for the Si nanowire transistors depicted	
	in Fig. 2	46
II	Subthreshold swing $S [\text{mV/decade}]$ as function of the channel orientation and	
	of the gate length L_g . The results labeled * are obtained with the effective	
	mass approximation, the others with the full-band formalism. \ldots	47
III	Characteristics and OBCs computation time (in seconds) on a 64-bit Sun Fire	
	X4600 with $8x2.8$ GHz Dual Core Opteron processors for the two nanowires	
	presented in Fig. 2. The method described in this article to obtain OBCs	
	is compared to an iterative calculation, labeled "Sancho-Rubio" and to a	
	generalized eigenvalue problem (GEVP) approach. Spin-orbit coupling is not	
	considered	48
IV	Time (in seconds) required to calculate the device wave function (columns 2,	
	3, 4, and 5) or to obtained the Green's functions (column 6) for one energy	
	point once the open boundary conditions have been computed (column 7).	
	The parallel scalability of the solvers is available if possible. The size ${\cal N}$ of	
	the corresponding sparse matrix $(\mathbf{E} - \mathbf{H})$ in Eq. (12) is given in column 8.	
	It does not take spin-orbit coupling into account. In the first column, N_x ,	
	N_y , and N_z indicate the size of the nanowire along x, y , and z as function of	
	the lattice constant a_0 , respectively. All simulations were run on a 64-bit Sun	
	Fire X4600 with 8x2.8 GHz Dual Core Opteron processors	49

Channel	[100]	[110]	[111]	[112]
$\mathrm{n}\text{-}\mathrm{I}_{ON}[\mu\mathrm{A}]$	4.3	7.7	1.2	2.9
$\mathrm{p}\text{-}\mathrm{I}_{ON}[\mu\mathrm{A}]$	2.3	15.7	10.3	9.2

TABLE I:

	[100]	[100]*	[110]	[110]*	[111]	[112]
$L_g=4 \text{ nm}$	88.5	99.7	122.3	99.5	83.2	87.5
$L_g = 7 \text{ nm}$	65.7	68.9	78.8	68.8	64.7	64.9
$L_g=10 \text{ nm}$	61.5	62.8	62.9	62.2	61.0	61.4
$L_g=13 \text{ nm}$	60.6	60.5	62.1	60.4	60.4	60.6

TABLE II:

GEVP [s]	360	920
Sancho-Rubio [s]	250	009
our method [s]	3.1	2
% of normal size	50	33.5
EV problem size	640	082
Atoms per slab	128	218
Layers per slab	4	9
Transport	[100]	[111]

TABLE III:

172800	36.1	Х	$46/3.45 \ {\rm x}$	98/5.04 x	$50/3.96 \ { m x}$	Х	$60{\times}6{\times}6/8$
172800	36.1	Х	$68/2.33 \ {\rm x}$	115/4.3 x	$71/2.79 \ { m x}$	Х	$60{ imes}66{ imes}6/4$
172800	36.1	Х	101/1.57 x	202/2.44 x	$124/1.6 \ {\rm x}$	Х	$60{ imes}66{ imes}6/2$
172800	36.1	1156	159	494	198	461	$60{ imes}66{ imes}6/1$
120000	11.2	382	62	190	84	117	$60{ imes}5{ imes}5/1$
76800	3.1	108	25.7	44.8	30.3	27.4	$60{ imes}4{ imes}4/1$
Ν	Bound. Cond. [s]	RGF [s]	MUMPS 4.6.3 [s]	SuperLU $_{dist}$ 2.0 [s]	PARDISO [s]	Umfpack 5.0.1 [s]	$N_x \times N_y \times N_z/{\rm CPUs}$

TABLE IV:



Mathieu Luisier was born in Le Châble, Switzerland. He received the degree of Dipl.-Ing. (with honors) in electrical engineering in 2003 and the doctoral degree in 2007 from the ETH Zurich, Switzerland. From July 2005 to December 2005 he was a visitor at the Network for Computational Nanotechnology, Purdue University, USA. His research interest is focused on quantum transport in nanoscale devices.



Andreas Schenk received the Dipl. Phys. degree and the Ph.D. from Humboldt University in Berlin (HUB) in 1981 and 1987, respectively. From 1987 till 1991 he was working on various aspects of the physics and simulation of optoelectronic devices. In 1991 he joined the Integrated Systems Laboratory (ISL) of ETH Zurich as a senior research/teaching assistant, where he qualified to give lectures at university in 1997 for "Physics and Modeling of Microelectronic Devices". In 2004 he became honorary professor at the ISL. His main activities are in the physics-based modeling of micro- and nanoelectronic devices: Quantum effects in ultra-short transistors, physics and simulation of single-electron devices, many-body effects, contacts, heterojunctions, degradation effects, as well as transport processes like resonant and sequential tunneling, generation-recombination, and noise. He is heading the device physics group at ISL. He authored and co-authored two books and 125 papers.