Atomistic simulation of nanowires in the $sp^3d^5s^*$ tight-binding formalism: From boundary conditions to strain calculations

Mathieu Luisier, Andreas Schenk, and Wolfgang Fichtner

Integrated Systems Laboratory, Gloriastrasse 35, ETH Zurich, 8092 Zurich, Switzerland

Gerhard Klimeck

Network for Computational Nanotechnology, Purdue University, West Lafayette, Indiana 47907, USA
and Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

As the active dimensions of metal-oxide field-effect transistors are approaching the atomic scale, the electronic properties of these "nanowire" devices must be treated on a quantum mechanical level. In this paper, the transmission coefficients and the density of states of biased and unbiased Si and GaAs nanowires are simulated using the $sp^3d^5s^*$ empirical tight-binding method. Each atom, as well as the connections to its nearest neighbors, is represented explicitly. The material parameters are optimized to reproduce bulk band-structure characteristics in various crystal directions and various strain conditions. A scattering boundary method to calculate the open boundary conditions in nanowire transistors is developed to reduce the computational burden. Existing methods such as iterative or generalized eigenvalue problem approaches are significantly more expensive than the transport simulation through the device. The algorithm can be coupled to nonequilibrium Green's function and wave function transport calculations. The speed improvement is even larger if the wire transport direction is different from [100]. Finally, it is demonstrated that strain effects can be easily included in the present nanowire simulations.

PACS number(s): 71.15.−m, 71.23.−k

I. INTRODUCTION

Semiconductor nanowires (NWs) may play an important role in the future of nanoelectronics. They can act both as active devices and as wire connectors. Lately, several groups have grown Si,[1] GaAs,[2] or Ge (Ref. [3]) NWs for different crystal orientations and cross sections. Field effect transistors, whose channels can be viewed as wires with triangular,[4] rectangular,[5] or cylindrical[6] cross section have been reported in the literature. Also, nanostructures with more exotic cross sections, such as T-shape wires, find practical applications, for example, in the optoelectronic field.[7]

If the nanowire cross section has a size comparable to the de Broglie wavelength, electronic transport exhibits significant quantization effects that are strongly dependent on the wire configuration (material, cross section, direction). At this atomistic scale, band-structure effects are crucial and their influence needs to be well understood in order to design new devices. This work provides insight into the electronic properties of NWs with different cross sections, growth directions, material compositions, and operation conditions. The considered nanowire cross sections are smaller than 5 $\times$ 5 nm², where full band-structure calculations play an important role.[8] The goal of this study is to present an improved method for quantum transport simulation in nanodevices in order to predict their performance limits.

The first step consists in finding an accurate band-structure model. The effective mass approximation holds in the vicinity of conduction band minima but does not always ensure a correct calculation of the quantization levels in nanostructures. The nearest neighbor $sp^3d^5s^*$ empirical tight-binding method, however, satisfies the accuracy condition because its parameters are optimized to reproduce the com-

plete bulk band structure.[9,10] Furthermore, its atomistic description of the simulation domain is advantageous at the nanometer scale. In this context, phenomena such as interface roughness, alloy disorder, heterostructures, surfaces, or impurity scattering can be treated rigorously.

The transition from infinite (bulk) to two-dimensionally confined structures (NWs) is computationally straightforward. The nanowires are constructed by translating their primitive unit cell across the device volume. The atomic on-site energies as well as the connections to the nearest neighbors are modeled with the bulk material parameters, and surface atoms are "passivated" by increasing the dangling-bond energy.[11]

In the second phase, the band-structure model is incorporated into a quantum transport solver. Nonequilibrium Green's function[12,13] (NEGF) or wave function[14] formalisms are well suited for that purpose, but both approaches suffer from the computational burden caused by the open boundary conditions (OBCs) calculation. For most of the applications, a nanowire can be separated into a transport (the x axis, for example) and two transverse directions (y and z). In these cases, the wire unit cell in the transport direction is a slab composed of different atomic layers (planes orthogonal to the transport direction). The number of atomic layers that make up the repeatable wire slab depends on the crystal orientation.

Each unit cell has N atoms. It is connected to the previous and to the next slab. The $sp^3d^5s^*$ tight-binding method includes ten orbitals without spin-orbit coupling and 20 with coupling, and the size of the block matrices involved in the calculation of the OBCs is either $10N$ or $20N$. Iterative algorithms[15] require the inversion of very dense or even full matrices of this size until convergence is achieved, typically

after 20–50 steps. In scattering boundary methods (SBMs),[16] the reservoir state eigenfunctions are calculated by solving a generalized eigenvalue problem (GEVP) of size $20N$ (without spin) or $40N$ (with spin). Obviously, if the wire cross section increases or if a special crystal orientation is chosen so that a wire slab contains many atoms, the efficiency of the iterative method and the SBM with GEVP degrades and a high amount of the computational time is dedicated to the boundary condition calculation.

In this paper, we investigate a different approach to treat the OBC problem. Starting from a scattering boundary ansatz, we take advantage of the atomic fitting in the nanowires and adapt the physical description of a slab. This method works not only for the usual [100] transport direction and square cross section,[12–14,16] but for any crystal orientation such as [110], [111], or [113], and any wire shape (e.g., triangular, circular, T, hexagonal, etc.). Furthermore, the computational burden increases proportionally with the cross section dimensions, as in the iterative and the GEVP approaches (staying nevertheless lower than in both cases), but not as function of the crystal orientation. A higher number of atomic layers (and therefore more atoms) in a slab with the same size does not necessarily lead to more effort to compute the OBCs as happens with other methods. The results can then be coupled to a NEGF (if incoherent scattering are included) or to a wave function solver (much more efficient in the ballistic case). In this study, it is shown that our method works for any input electrostatic potentials, for which no self-consistent adjustment has been performed (future work). Consequently, the results of our quantum transport simulator under nonequilibrium condition must be considered carefully. Since the charge neutrality is not ensured, the calculated transmission coefficients and density of states do not describe the true properties of the simulated systems, but the reaction to a fictitious electrostatic potential.

The paper is organized as follows. In Sec. II, our method is described to calculate the open boundary conditions for quantum transport in nanowires, the improvement compared to previous treatments is highlighted, and the coupling to different solution schemes is derived. An additional simplification based on the device symmetry is described in the Appendix. Section III presents results for Si and GaAs wires with different cross sections and crystal orientations and it examines the influence of applied bias and biaxial strain. The transmission coefficients as well as the density of states are calculated for these different nanowires. In Sec. IV, we discuss the numerical implementation, the advantages and disadvantages of a wave function and of a NEGF device calculation, and the different alternatives for the open boundary conditions. Section V concludes this work.

II. THEORY

In this section, we describe a computational procedure to obtain open boundary conditions in a two-dimensionally confined quantum transport problem and the coupling to a NEGF and a wave function ballistic solver. The procedure is based on a scattering boundary approach. Figure 1 shows the schematic view of a nanowire with wire length L_w without
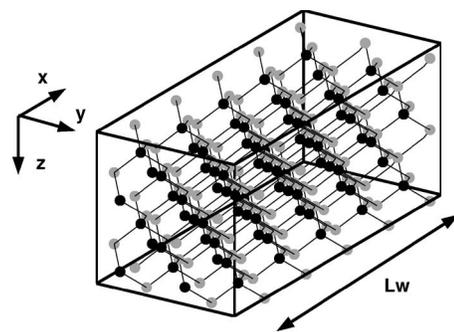
FIG. 1. Schematic view of an atomistic nanowire. x is the transport axis, y and z are confinement directions, L_w represents the wire length. Each atom is considered as well as the connections to its four nearest neighbors, except at the surface.

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 method, ten different orbitals are kept. Each of them is two times degenerate if spin-orbit coupling is considered.

Independent of the underlying 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

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

The Hamiltonian H 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^*)

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

$$|\psi_E\rangle = \sum_{\sigma,i,j,k} C_{ijk}^\sigma(E)|ijk,\sigma\rangle, \tag{2}$$

where $C_{ijk}^\sigma(E)$ is the expansion coefficient for the orbital σ of an atom situated at $\mathbf{R}=(x_i,y_j,z_k)$ in the nanowire. To solve Eq. (1), we work in a slab basis.[17] 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 width Δ. A nanowire with length L_w is therefore composed of L_w/Δ cells that represent its central scattering region. In this basis, the scalar $C_{ijk}^\sigma(E)$ becomes a vector $C_i(R_s,E)$, where i denotes the ith wire slab and R_s the position and the orbital type of an atom localized inside of it. Considering only connections to the nearest neighbors, disregarding three-center integrals,[18] and left-multiplying Eq. (1) with $\langle ijk,\sigma|$ at each position and for each orbital, we obtain the matrix equation

$$(E - H_{ii})C_i(R_s,E) - H_{ii+1}C_{i+1}(R_s,E) - H_{ii-1}C_{i-1}(R_s,E) = 0 \tag{3}$$

or

$$D_{ii}C_i(R_s,E) + T_{ii+1}C_{i+1}(R_s,E) + T_{ii-1}C_{i-1}(R_s,E) = 0, \tag{4}$$

whose elements $E - H_{ii} \equiv D_{ii}$ describe the on-site energy and the bond connections within a slab i, $H_{ii+1} \equiv -T_{ii+1}$ is the coupling to the next slab, and $H_{ii-1} \equiv -T_{ii-1}$ is 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 $t_b N$. Equation (4) is valid in the device as well as in the semi-infinite left and right contacts, where we apply a scattering boundary ansatz for the coefficients $C_i(R_s,E)$ (i denotes a slab in the left lead). For brevity, only the left contact is treated, the derivation for the other(s) is obvious

$$C_i(R_s,E) = \frac{1}{\sqrt{N_x}} \sum_n \left[a_n e^{ik_n(E)x_i} \varphi_{i,n}^+(R_s,E) \right.$$
$$\left. + b_n e^{-ik_n(E)x_i} \varphi_{i,n}^-(R_s,E) \right]. \tag{5}$$

Here N_x is a normalization constant, a_n is the injection coefficient for a state $\varphi_{i,n}^+(R_s,E)$ transmitted through the device (i.e., flowing from left to right), and b_n is the coefficient for a state $\varphi_{i,n}^-(R_s,E)$ reflected back in the contact, respectively. $\varphi_{i,n}^\pm(R_s,E)$ (associated with $e^{ik_n(E)x_i}$) is the nth reservoir state and can be propagating or exponentially decaying. Inserting Eq. (5) into Eq. (4) and separating the transmitted and the reflected parts of the coefficient $C_i(R_s,E)$, we exploit the fact that both resulting contributions must vanish for each quantization level n since the contacts are assumed infinite,

$$(D_{ii} + T_{ii+1}e^{\pm ik_n(E)\Delta} + T_{ii-1}e^{\mp ik_n(E)\Delta})\varphi_{i,n}^\pm(R_s,E) = 0, \tag{6}$$

with the slab width Δ. Only one of these two equations needs to be solved because the solution of the other equation is automatically taken into account.[19] It is worth noting that Eq. (6) is exactly the equation solved for the band-structure calculation of an infinite wire, but with exchanged input and output variables. In the open boundary condition problem, one searches for all the wave vectors k corresponding to one injection energy E. In the band-structure case, one calculates the energy eigenvalues at one k point. To recognize if the calculated state is transmitted or reflected, two different approaches are required. For a propagating state [i.e., $k_n(E)$ has no imaginary part], the energy $E(k_n)$ is derived with respect to k_n because the particle velocity $v_n \propto dE(k_n)/dk_n$ is related to this quantity. In the left reservoir, a positive v_n means transmission, a negative reflection. For an exponentially decaying state, the imaginary part of k indicates the state nature. A positive imaginary part, or $|\exp[ik_n(E)\Delta]| < 1$, denotes a decaying transmission, while $|\exp[-ik_n(E)\Delta]| < 1$ corresponds to a decaying reflection. These operations are done after the $k_n(E)$ and $\varphi_{i,n}(R_s,E)$ are calculated at one given injection energy E. A well established procedure consists in transforming Eq. (6) to the following complex non-Hermitian (with spin-orbit coupling) or real nonsymmetric (without) generalized eigenvalue problem[14,16] of size $2t_b N$

$$\begin{pmatrix} D_{ii} & T_{ii+1} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \varphi_i \\ \varphi_{i+1} \end{pmatrix} = e^{ik\Delta} \begin{pmatrix} -T_{ii-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varphi_i \\ \varphi_{i+1} \end{pmatrix}, \tag{7}$$

where the variables E and R_s are omitted for brevity. Despite the fact that the numerical solution of Eq. (7) avoids matrix inversion,[20] it is still computationally intensive to obtain the desired k and φ values for a nanowire with a large cross section or a crystal orientation different from [100]. However, going back to Eq. (6), better insight into the physical structure of the matrices D_{ii}, T_{ii+1}, and T_{ii-1} leads to a simplified procedure to evaluate the OBCs. For the derivation, we use a transport axis in the [100] direction (the method is not limited to this specific orientation). In this case, a slab i in the left (right) reservoir is composed of four atomic layers, two entangled pairs of cations and anions. A cation (anion) layer is only connected to the previous and to the next anion (cation) layers. Therefore we can write Eq. (6) in an atomic layer basis instead of a slab basis:

$$\widetilde{D}_{00}\widetilde{\varphi}_0 + \widetilde{T}_{0-1}\widetilde{\varphi}_{-1} + \widetilde{T}_{01}\widetilde{\varphi}_1 = 0,$$

$$\widetilde{D}_{11}\widetilde{\varphi}_1 + \widetilde{T}_{10}\widetilde{\varphi}_0 + \widetilde{T}_{12}\widetilde{\varphi}_2 = 0,$$

$$\widetilde{D}_{22}\widetilde{\varphi}_2 + \widetilde{T}_{21}\widetilde{\varphi}_1 + \widetilde{T}_{23}\widetilde{\varphi}_3 = 0,$$

$$\widetilde{D}_{33}\widetilde{\varphi}_3 + \widetilde{T}_{32}\widetilde{\varphi}_2 + \widetilde{T}_{34}\widetilde{\varphi}_4 = 0. \tag{8}$$

In the layer basis, $\widetilde{D}_{ii}$ describes the one-site energy as well as the connections to the neighbor atoms inside layer i, $\widetilde{T}_{ii\pm1}$ the coupling to the next and to the previous layers, $\widetilde{\varphi}_i$ is the eigenfunction of layer i, and the total vector $\varphi = [\widetilde{\varphi}_0; \widetilde{\varphi}_1; \widetilde{\varphi}_2; \widetilde{\varphi}_3]$ is the same as in Eq. (6). With the scattering boundary ansatz from Eq. (5) and the assumption that all reservoir slabs are identical, we obtain that $\widetilde{\varphi}_{-1} = \widetilde{\varphi}_3 \times \exp(-ik\Delta)$, $\widetilde{\varphi}_4 = \widetilde{\varphi}_0 \times \exp(ik\Delta)$, and $\widetilde{T}_{34} = \widetilde{T}_{-10}$ so that Eq. (8) becomes

$$\begin{pmatrix} \widetilde{D}_{00} & \widetilde{T}_{01} & 0 & \widetilde{T}_{0-1}e^{-ik\Delta} \\ \widetilde{T}_{10} & \widetilde{D}_{11} & \widetilde{T}_{12} & 0 \\ 0 & \widetilde{T}_{21} & \widetilde{D}_{22} & \widetilde{T}_{23} \\ \widetilde{T}_{-10}e^{ik\Delta} & 0 & \widetilde{T}_{32} & \widetilde{D}_{33} \end{pmatrix} \begin{pmatrix} \widetilde{\varphi}_0 \\ \widetilde{\varphi}_1 \\ \widetilde{\varphi}_2 \\ \widetilde{\varphi}_3 \end{pmatrix} = 0. \tag{9}$$

We define two new variables $\mathcal{H}$ and $\mathcal{P}$ as

$$\mathcal{H} = \begin{pmatrix} \widetilde{D}_{00} & \widetilde{T}_{01} & 0 & 0 \\ \widetilde{T}_{10} & \widetilde{D}_{11} & \widetilde{T}_{12} & 0 \\ 0 & \widetilde{T}_{21} & \widetilde{D}_{22} & \widetilde{T}_{23} \\ \widetilde{T}_{-10} & 0 & 0 & 0 \end{pmatrix}, \tag{10}$$

$$\mathcal{P} = - \begin{pmatrix} 0 & 0 & 0 & \widetilde{T}_{0-1} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \widetilde{T}_{32} & \widetilde{D}_{33} \end{pmatrix}, \tag{11}$$

respectively, and use them to rewrite Eq. (9)

$$\mathcal{H}\varphi = e^{-ik\Delta}\mathcal{P}\varphi. \tag{12}$$

Equation (12) looks almost like an eigenvalue problem, but for that purpose the matrix $\mathcal{H}$ or $\mathcal{P}$ should be inverted so that on one side of the equation there is a matrix M and on the other an eigenvalue $\exp(-ik\Delta)$. This is not possible because $\mathcal{P}$ obviously cannot be inverted due to its structure. In addition, $\mathcal{H}$ is also singular as the connecting block $\widetilde{T}_{23}$ contains columns filled exclusively with zeros (nearest neighbor approximation). A slight modification of Eq. (12) can remove this deficiency and generate a normal eigenvalue problem (EVP)

$$\mathcal{H}\varphi = e^{-ik\Delta}\mathcal{P}\varphi + \mathcal{P}\varphi - \mathcal{P}\varphi,$$

$$(\mathcal{H} - \mathcal{P})\varphi = (e^{-ik\Delta} - 1)\mathcal{P}\varphi,$$

$$(\mathcal{H} - \mathcal{P})^{-1}\mathcal{P}\varphi = \frac{1}{e^{-ik\Delta} - 1}\varphi,$$

$$M\varphi = \lambda\varphi. \tag{13}$$

The matrix $(\mathcal{H} - \mathcal{P})$ can always be inverted, except for $E = E(k=0)$, which is one of its eigenvalues. To calculate the band structure at $k=0$ for an infinite wire whose slabs are described by Eq. (4), we have to find the eigenvalues of a matrix identical to $(\mathcal{H} - \mathcal{P})$. However, by adding a small imaginary part to the energy E or by artificially avoiding $E = E(k=0)$, the matrix M is always defined and has an advantageous structure:

$$M = \begin{pmatrix} 0 & 0 & M_{02} & M_{03} \\ 0 & 0 & M_{12} & M_{13} \\ 0 & 0 & M_{22} & M_{23} \\ 0 & 0 & M_{32} & M_{33} \end{pmatrix}. \tag{14}$$

It is thus not necessary to consider the whole matrix M in the eigenvalue problem defined in Eq. (13), but only its lower right corner and

$$\begin{pmatrix} M_{22} & M_{23} \\ M_{32} & M_{33} \end{pmatrix} \begin{pmatrix} \widetilde{\varphi}_2 \\ \widetilde{\varphi}_3 \end{pmatrix} = \frac{1}{e^{-ik\Delta} - 1} \begin{pmatrix} \widetilde{\varphi}_2 \\ \widetilde{\varphi}_3 \end{pmatrix}. \tag{15}$$

When the transport direction is aligned with the [100] crystal axis, the complex non-Hermitian (spin coupling) or real nonsymmetric (no-spin) eigenvalue problem to be solved is of size $t_bN/2$, which is a significant improvement compared to Eq. (7) whose matrices have size $2t_bN$. Apart from a gain of a factor 4 in the size of the blocks, we do not need to work with a GEVP, but with a normal EVP. A weakness resides in the inversion of the matrix $(\mathcal{H} - \mathcal{P})$ with size t_bN. This causes an additional computational effort, but its high sparsity de-

gree enables a fast inversion so that the method presented here is more efficient than what was published previously[14,16] (see also Table I for quantitative results).

Once the solution of Eq. (15) is obtained, the eigenfunctions of the first ($\widetilde{\varphi}_0$) and of the second ($\widetilde{\varphi}_1$) atomic layers are computed with

$$\begin{pmatrix} \widetilde{\varphi}_0 \\ \widetilde{\varphi}_1 \end{pmatrix} = -(e^{-ik\Delta} - 1)\begin{pmatrix} M_{02} & M_{03} \\ M_{12} & M_{12} \end{pmatrix} \begin{pmatrix} \widetilde{\varphi}_2 \\ \widetilde{\varphi}_3 \end{pmatrix}. \tag{16}$$

Note that the approach presented for a transport axis in the [100] direction works for all crystal orientations, as mentioned above, even if one atomic layer is connected not only to the next layer but to several consecutive layers. In this case, the matrices $\mathcal{H}$ and $\mathcal{P}$ are generated in such a way that they minimize the bandwidth of the matrix M: the size of the eigenvalue problem becomes as small as possible. An illustration of this principle will be given in Sec. III. Another possibility to reduce the size of M is to consider the symmetry properties of the simulated nanowires. An explanation is given in the Appendix.

A. Wave function solver

In the wave function formalism, one needs to couple the OBCs from Eqs. (15) and (16) to the Schrödinger equation expressed in a slab basis. It is assumed that the considered nanowire has N_s slabs starting from 1 to N_s, and that the injected electrons come from the left contact (slab 0),

$$\begin{pmatrix} L_{00} & L_{01} \\ L_{10} & D_{11} & T_{12} \\ & T_{21} & D_{22} & T_{23} \\ & & \ddots & \ddots & \ddots \\ & & & T_{N_sN_s-1} & D_{N_sN_s} & R_{N_sN_s+1} \\ & & & & R_{N_s+1N_s} & R_{N_s+1N_s+1} \end{pmatrix} \begin{pmatrix} C_0 \\ C_1 \\ C_2 \\ \vdots \\ C_{N_s} \\ C_{N_s+1} \end{pmatrix}$$

$$= \begin{pmatrix} I_0 \\ I_1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}. \tag{17}$$

To calculate the elements L_{00}, L_{01}, L_{10}, (coupling to the left reservoir) $R_{N_s+1N_s+1}, R_{N_sN_s+1}, R_{N_s+1N_s}$, (coupling to the right reservoir), I_0, and I_1 (injection mechanism), the lead eigenfunctions $\varphi_{i,n}(R_s, E)$ and wave vectors $k_n(E)$ are classified as functions of their properties. The N_p^+ states propagating from the left to the right (transmitted) are cast into the $(t_bN) \times N_p^+$ matrix φ_p^+, the N^+ transmitted states into φ^+ [size $(t_bN) \times N^+$], and finally the N^- reflected states into φ^- [size $(t_bN) \times N^-$]. The same notation is applied to $k_n(E)$. Furthermore, $\exp(i\mathbf{k}^{\pm}\Delta)$ is a diagonal matrix containing the contributions from all the transmitted $(+)$ or reflected wave vectors $(-)$ with $\mathbf{k}^+ = -\mathbf{k}^-$. We obtain

TABLE I. Table summarizing the material composition, crystal orientation, size, and atomic fitting of the nanowires presented in Fig. 2. The three last rows compare the OBC computational effort for the calculation method presented in this paper, for an iterative algorithm (Ref. 15) labeled "Sancho-Rubio," and for a generalized eigenvalue problem (GEVP) approach (Refs. 14 and 16). In the last two rows, the second set of numbers given after the CPU times refers to the speed degradation of the GEVP and Sancho-Rubio methods compared to our approach.

	Cross section					
	(a) Square	(b) Triangle	(c) Circle	(d) T	(e) Hexagon	(f) Square
Material	Si	Si	Si	GaAs	GaAs	GaAs
x axis	[100]	[111]	[100]	[100]	[110]	[113]
y axis	[010]	[$\bar{1}$10]	[011]	[010]	[$\bar{1}$10]	[$\bar{1}$10]
z axis	[001]	[11$\bar{2}$]	[0$\bar{1}$1]	[001]	[001]	[33$\bar{2}$]
Wire length (nm)	22.74	22.53	22.74	22.53	22.30	22.47
Number of atoms	1722	1440	2562	3520	4389	1680
Number of slabs	42	24	42	40	57	12
Number of layers	168	144	168	160	114	264
Cell width (nm)	0.543	0.9405	0.543	0.5653	0.3997	1.8749
Layers per slab	4	6	4	4	2	22
Atoms per slab	41	60	61	88	77	140
EV problem size	210	160	290	440	770	240
% of normal size	51.2	26.7	47.5	50	100	17.1
Present method (s)	0.21	0.24	0.61	1.71	5.61	1.16
Sancho-Rubio (s)	12.28/61×	26.57/110×	43.03/70×	125.03/73×	260.1/45×	126.61/109×
GEVP (s)	6.39/32×	17.05/70×	24.01/39×	74.28/43×	67.62/12×	212.09/182×

$$L_{00} = \varphi^{+\dagger}D_{00}\varphi^- + \varphi^{+\dagger}T_{10}\varphi^- e^{-i\mathbf{k}^-\Delta},$$

$$L_{01} = \varphi^{+\dagger}T_{01}, \quad L_{10} = T_{10}\varphi^-,$$

$$I_0 = -(\varphi^{+\dagger}D_{00}\varphi_{\mathbf{p}}^+ + \varphi^{+\dagger}T_{10}\varphi_{\mathbf{p}}^+ e^{-i\mathbf{k}_{\mathbf{p}}^+\Delta})A_{inj}(E),$$

$$I_1 = -T_{10}\varphi_{\mathbf{p}}^+ A_{inj}(E). \tag{18}$$

The definition of the matrices $R_{N_s+1N_s+1}$, $R_{N_sN_s+1}$, and $R_{N_s+1N_s}$ follows Eq. (18). If the states are injected from the right contact, the right-hand side of Eq. (17) will have two non-zero vectors I_{N_s} and I_{N_s+1}. The diagonal matrix A_{inj} is defined so that the elements $A_{inj}^\dagger \times A_{inj}$ represent the probability that a state injected from the left with energy E is occupied. Equation (17) must be solved for each injection energy and multiple right-hand sides containing all propagating states from the different device contacts. Consequently, the unknown C_i coefficients depend on the energy E [or wave vector $k(E)$], on the port and on the state they come from (indices p and n, respectively) for all i, plus on the orbital type σ, and on the atom position R_s in the slab i if $1 \leq i \leq N_s$. They are labeled $C_{i,p,n}^\sigma(R_s, k(E))$ for $1 \leq i \leq N_s$ and $C_{i,p,n}(k(E))$ else. The carrier density in the nanowire $n(\mathbf{r})$ is given by

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

$$n(x,r) = \frac{1}{N_x}\sum_{n,p,\sigma}\sum_{i,R_s}\sum_k |C_{i,p,n}^\sigma(R_s,k)|^2 f(E_{p,n}(k) - \mu_p)$$

$$\times \delta(x - x_i)\delta(r - R_s)$$

$$= \frac{\Delta}{2\pi}\sum_{n,p,\sigma}\sum_{i,R_s}\int_0^{\pi/\Delta} dk |C_{i,p,n}^\sigma(R_s,k)|^2 f(E_{p,n}(k) - \mu_p)$$

$$\times \delta(x - x_i)\delta(r - R_s). \tag{19}$$

In the contacts, we assume Fermi distributions with chemical potential μ_p, and the injection energy (from port p) $E_{p,n}(k)$ of a state n with wave vector k. The product of two orbital functions ϕ_σ introduced in Eq. (2) should be present in Eq. (19). We replace it by $\delta(x - x_i)\delta(r - R_s)$ because the resulting term is very localized around one atom situated at (x_i, R_s). Finally, we convert the sum over the wave vector k to an integral. In this formalism, the position-dependent density of states $Z_p(x, r, E)$ injected from port p is defined as

$$Z_p(x,r,E) = \frac{\Delta}{2\pi}\sum_{n,\sigma}\sum_{i,R_s}\int_0^{\pi/\Delta} dk |C_{i,p,n}^\sigma(R_s,k)|^2 \delta(E - E_{p,n}(k))$$

$$\times \delta(x - x_i)\delta(r - R_s). \tag{20}$$

The current density is calculated by using the two-terminal Landauer formula for the noninteracting case.[21] It requires the knowledge of the transmission coefficient $T(E)$ weighted

with the contact distribution functions. If an electron is injected from the left port $p=1$, $T(E)$ at the right contact is calculated with C_{N_s+1} [the situation described in Eq. (17)],

$$T(E) = \sum_{n,m} |C_{N_s+1,p=1,n}(k_m)|^2 \left| \frac{dE}{dk_m} \right| \left| \frac{dE}{dk_n} \right|^{-1}. \quad (21)$$

n and m are indices that run over all the propagating states in the left (slab 0) and the right (slab N_s+1) reservoirs, respectively, not over the exponentially decaying states.

B. NEGF solver

Recently, the nonequilibrium Green's function formalism has become very popular for the simulation of quantum transport in nanodevices. In the following, we show that the scattering OBCs obtained above can be used to calculate the retarded boundary self-energies Σ^{RB} of the NEGF formalism. The use of iterative solvers[15] for the OBC calculation (20–50 inversions of a very dense or full matrix with size $t_b N$) is less efficient than the approach proposed in this paper (one inversion of a sparse $t_b N$ matrix and solution of an eigenvalue problem with reduced size). In our method, Σ^{RB} is obtained after some simple steps, and not from the direct output of an iterative solver. The starting point is the calculation of the contact retarded Green's function[23] $g_{ij}^R(R_{s1}R_{s2},E)$ (i and j are slab indices, R_{s1} and R_{s2} are the atom positions inside the slabs, E refers to the energy, R_{s1}, R_{s2}, and E are neglected in the remainder of this section). We derive the equation for the left reservoir, where slab 0 is the last slab before the device and $T_{01}=0$, which is the condition for the calculation of g_{ij}^R. The following system of equations must be solved:[24]

$$D_{00}g_{00}^R + T_{0-1}g_{-10}^R = I,$$

$$D_{-1-1}g_{-10}^R + T_{-10}g_{00}^R + T_{-1-2}g_{-20}^R = 0. \quad (22)$$

With the invariance of the slabs in the reservoirs, $D_{-1-1} = D_{00}$ and $T_{-1-2}=T_{0-1}$. It is easy to prove that the following ansatz for g_{ij}^R (assuming the same notation conventions as in the wave function case):

$$g_{ij}^R = \varphi_0^- e^{i\mathbf{k}^- x_i} \tilde{g}^R e^{-i\mathbf{k}^- x_j} \varphi_0^{-\dagger} \quad (23)$$

satisfies the second part of Eq. (22) and that it can be inserted into the first part to calculate $\tilde{g}^R$ and then Σ_{11}^{RB}. After some straightforward algebra, we find

$$\tilde{g}^R = (\varphi_0^{-\dagger} D_{00} \varphi_0^- + \varphi_0^{-\dagger} T_{0-1} \varphi_0^- e^{-i\mathbf{k}^- x_j})^{-1},$$

$$\Sigma_{11}^{RB} = T_{10} g_{00}^R T_{01} = T_{10} \varphi_0^- \tilde{g}^R \varphi_0^{-\dagger} T_{01}. \quad (24)$$

The boundary self-energy $\Sigma_{N_sN_s}^{RB}$ of the right contact is obtained with a similar procedure. We calculate the carrier and the current density with a recursive algorithm[25] involving the Hamiltonian H that corresponds to the left-hand side matrix of Eq. (17) where the boundary condition terms L_{ij} and R_{ij} are removed and replaced by the boundary self-energies Σ_{11}^{RB} and $\Sigma_{N_sN_s}^{RB}$.
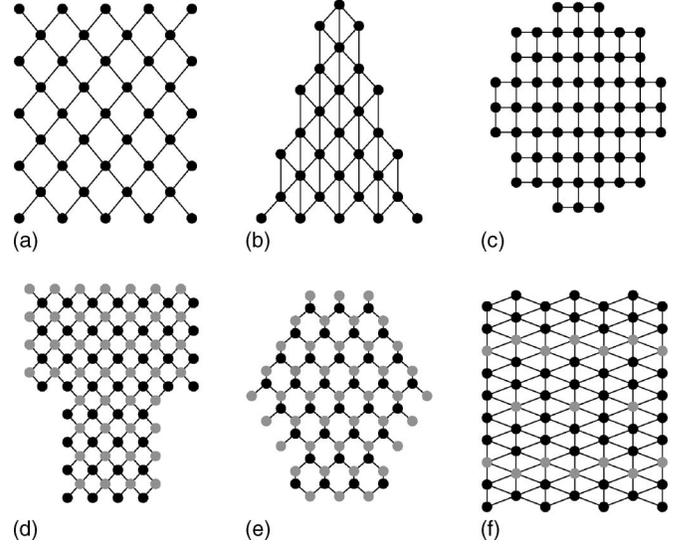
FIG. 2. Cross section of the different nanowires simulated in this work (data are summarized in Table I): (a) Si rectangular wire (transport in [100]), (b) Si triangular wire (transport in [111]), (c) Si circular wire (transport in [100]), (d) GaAs T-shape wire (transport in [100]), (e) GaAs hexagonal wire (transport in [110]), and (f) GaAs rectangular wire (transport in [113]). For the GaAs wires, the dark atoms are Ga and the light As. The cross section corresponds to the projection of a wire slab on a single plane (the atoms are situated in different planes).

III. RESULTS

In this section, we present simulations of Si and GaAs nanowires with different cross sections and crystal orientations (see Fig. 2), but with almost the same length $L_w = 22.5 \pm 0.5$ nm (given in Table I). L_w is the central scattering region of the device. According to Fig. 1, x is the transport direction; y and z are confinement directions and delimit the device cross section. The $sp^3d^5s^*$ tight-binding parameters were optimized by Boykin *et al.* to reproduce Si (Ref. 10) and GaAs (Ref. 27) bulk band structures. Figure 2 depicts different wire cross sections projected onto one atomic plane. They correspond to wire slabs, as introduced in Sec. II, that are composed of the minimal number of atomic layers (all the atoms contained in a plane orthogonal to the x axis), so that we generate an infinite wire by translating it.

In Fig. 3, the electron band structure of infinite wires (calculated without spin-orbit coupling) with the same cross sections as in Fig. 2 is 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 $k_{max}=\pi/\Delta$, where Δ is the length of a wire slab in the transport direction.

Figure 4 shows the electron transmission of the same nanowires as in Figs. 2 and 3 for three different operating conditions. The open boundary conditions are calculated with the eigenvalue method proposed in this paper. They are then coupled to a wave function solver as described in Sec. II A. Spin-orbit coupling is neglected because we consider electron transmission and its effects are small for the conduction band as we will show later in this section.
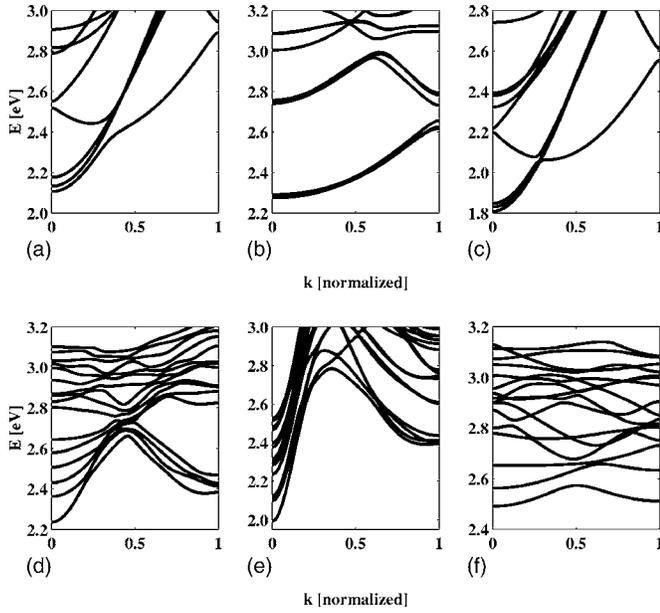
FIG. 3. Conduction subbands of infinite (in the transport direction) nanowires without an applied bias. The cross sections and the crystal orientations are the same as in Fig. 2: (a) Si rectangular wire, (b) Si triangular wire, (c) Si circular wire, (d) GaAs T-shape wire, (e) GaAs hexagonal wire, and (f) GaAs rectangular wire.

The unbiased device transmission coefficients $T(E)$ in Fig. 4 (dark solid line) corresponds to the band structure of infinite structures. In effect, the semi-infinite left and right reservoirs as well as the central channel are connected to form a uniform nanowire whose electrical properties do not vary in the 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 in the right reservoir), we have $T(E)=n$. This is a good way to check if the results obtained with the procedure outlined in Sec. II are correct.

When a linear bias of 0.1 V (dotted dark lines) and 0.2 V (light solid lines) is applied to the nanowires, the band bending starts 5 nm after the left contact and ends 5 nm before the right contact. $T(E)$ can no longer be related to the band structure of infinite wires, where the potential does not vary from the left to the right semi-infinite leads. However, in order for transmission to occur, an incident mode must have the same symmetry properties as the reservoir state that collects it on the other side of the device. For example, a mode whose probability density has one single maximum in the middle of the cross section (caused by the confinement) is injected from the left reservoir. Only states with different probability densities are available in the right reservoir. No transmission is possible at this energy because in the ballistic regime, a state cannot change its symmetry during a passage through the channel. For energy E, if n modes are present in the left reservoir and m in the right one, then $T(E) \leqslant \min(n,m)$. This principle is illustrated in Fig. 4.

We study now the six examples depicted in Fig. 2. The first structure (a) is a Si rectangular nanowire (1.2 $\times 1.2$ nm^2) where the transport axis is aligned with [100], y with [010], and z with [001]. A slab contains four atomic
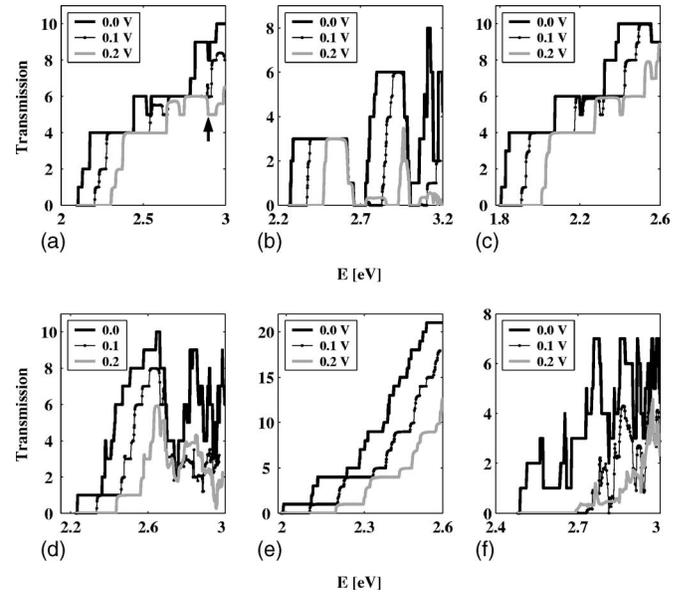
FIG. 4. Electron transmission through the nanowires described in Fig. 2 calculated without spin-orbit coupling. Three different situations are simulated: (dark line) without bias, (dotted line) a bias of 0.1 V is applied to the device, and (light line) a bias of 0.2 V is applied. The linear bias extends from 5 nm after the left contact to 5 nm before the right one; flat plateaus are left at both ends of the devices. (a) Si rectangular wire, (b) Si triangular wire, (c) Si circular wire, (d) GaAs T-shape wire, (e) GaAs hexagonal wire, and (f) GaAs rectangular wire.

layers (width $\Delta=0.543$ nm), 41 atoms, the wire 42 slabs, and therefore 1722 atoms. The square matrix M involved in the calculation of the OBCs has a size $N_M=410$ but the eigenvalue problem in Eq. (15) has a reduced size of $N_{EVP}=210$. This implies that 51.2% of N_M is sufficient to calculate the k's and the φ's. The remaining states (48.8%) have an infinite imaginary part and do not contribute to quantum transport.

From the band-structure calculation in Fig. 3(a), we can determine the unbiased transmission of Fig. 4. For example, the first band starts at $E=2.105$ eV, and so does the transmission. Due to the confinement effects, this conduction subband has a minimum at $k=0$ and [100]-oriented Si nanowires become direct band gap structures with four of the six split valleys projected to $k=0$. At $k=0$, the third band turning on at $E=2.176$ eV is doubly degenerate and induces an increase of step size 2 in the transmission $T(E)$. The fourth visible conduction band has a minimum at $k=0.25$ at $E=2.444$ eV. At this point, in the transmission plot, we observe a double step because a state with positive velocity appears at $k=0.25+\delta$, but also on the other side of the Brillouin zone, at $k=-0.25+\delta$ (with $\delta \rightarrow 0$). When one branch of this fourth band stops at $k=0$ at $E=2.521$ eV, the transmission encounters a step down (from 6 to 5) marking the turn-off of one channel. Following this procedure, the complete transmission can be explained in the unbiased case, proving that the results obtained in the quantum transport calculation are correct.

When bias is applied to the device, we note shifts of 0.1 and 0.2 eV in the transmission curves and a smoothing of the abrupt quantization steps at low energies. At high energy, in
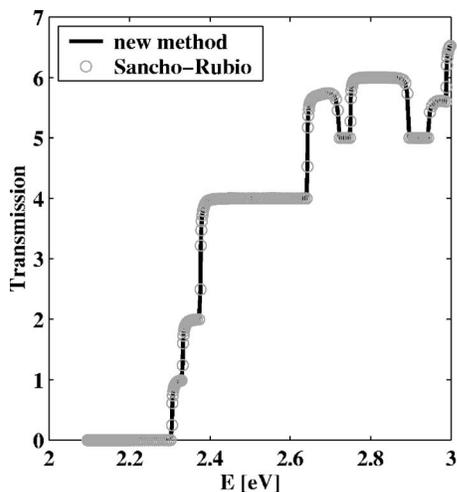
FIG. 5. Electron transmission through the Si rectangular [100]-oriented wire in Fig. 2 for an applied bias of 0.2 V. The results obtained with the OBC method presented in this paper (solid line) are compared to the results from the iterative algorithm proposed by Sancho and Rubio (Ref. 15) (bright circles).

FIG. 6. Electron density of states (DOS) for the same wire and the same conditions as in Fig. 5. The results obtained with the present method (lines) are compared to the results from the Sancho-Rubio iterative algorithm (Ref. 15) (symbols): the contribution to the total DOS coming from the left contact (dark line and circles) and coming from the right contact (light line and stars) are shown in the first device slab ($0 \leqslant x \leqslant 0.543$ nm).

addition to these effects, the transmission changes its physical behavior: at $E = 2.89$ eV [see arrow in Fig. 4(a)], we would expect that $T(E)$ with a bias of 0.2 V (light line) remains flat but it actually goes down. This arises from the fact that all the bands with $2 \leqslant n \leqslant 9$ are present in the left reservoir (1 has just disappeared and 10 has not been reached yet) and all the bands with $1 \leqslant m \leqslant 6$ are found in the right contact (first band situated at $E = 2.105 + 0.2$ eV; band 7 has not started at $E = 2.788 + 0.2$ eV). Because of the symmetry properties of the probability density, only the following (n,m) couples are possible: (2,2), (3,3), (4,4), (5,5), and (6,6). Thus the transmission at $E = 2.89$ eV with a bias of 0.2 V must be smaller than or equal to 5, as illustrated by the light line in Fig. 4(a) and cannot remain on plateau number 6.

To verify that our results are correct when bias is applied to the nanowires, we implement a second simulation model where the OBCs are calculated with an iterative method,[15] cast into self-energies, and quantum transport is solved in the nonequilibrium Green's function formalism.[23,25] The comparison of the wave function (present eigenvalue method for the boundaries, solid lines) and of the NEGF (iterative OBCs, labeled "Sancho-Rubio," symbols) solutions is shown in Figs. 5 and 6 for the Si rectangular wire from before (bias: 0.2 V). The electron transmission (Fig. 5) and the density of states in the first wire slab (Fig. 6) match perfectly, indicating that the OBC method we use in this paper works under all conditions. In Fig. 6, there are two different densities of states, one coming from the left contact (black) and weighted by the left electron distribution function when carrier density is calculated, the other (light) coming from the right contact and weighted by the right electron distribution function.

The second Si nanowire in Fig. 2(b) has a triangular cross section (base 1.6 nm, height 1.2 nm), [111] as transport direction, $[\bar{1}10]$ for y, and $[11\bar{2}]$ for z. The six atomic layers composing a wire slab ($\Delta = 0.9405$ nm) contain 60 atoms so that the matrix M has a size $N_M = 600$, but the OBC eigen-

value problem has the order $N_{EVP} = 160$ (26.7% of N_M). The wire is composed of 24 slabs or 144 atomic layers. Its band structure is represented in Fig. 3(b) for the infinite and unbiased case. No electron subbands appear in the energy range comprised between 2.656 eV and 2.731 eV. Consequently, the transmission must vanish for these energies, as we see in Fig. 4(b), even when bias is applied on the device. The left reservoir keeps a constant potential so that no transmission is possible between 2.656 and 2.731 eV. Furthermore, the potential of the right contact increases by 0.1 or 0.2 eV, leading to a second gap between 2.756 and 2.831 eV (bias 0.1 V) or 2.856 and 2.931 eV (bias 0.2 V) where the transmission also disappears, in agreement with the simulation results.

Figures 2(c), 3(c), and 4(c) present results for a Si circular nanowire (diameter of 1.7 nm) with [100] as transport direction, [011] for y, and $[0\bar{1}1]$ for z. The device has 42 slabs (width $\Delta = 0.543$ nm, made up of four atomic layers and 61 atoms). The size of the matrix M is $N_M = 610$, but an eigenvalue problem with $N_{EVP} = 290$ must be solved for the boundary conditions (47.5% of N_M). Its band structure and electron transmission are similar to those of the Si rectangular wire, but with a lower band gap due to the larger dimensions (2.26 nm² instead of 1.44 nm²).

A GaAs T-shape nanowire [height (max) 2.2 nm, height (min) 1.1 nm, width (max) 1.9 nm, width (min) 1.1 nm] is presented in Figs. 2(d), 3(d), and 4(d): [100] is the transport direction, [010] y, and [001] z. A slab is $\Delta = 0.5653$ nm wide and contains four atomic layers or 88 atoms (cations are dark, anions light). Therefore, M is an 880×880 square matrix in the absence of spin-orbit coupling, and the eigenvalue problem in Eq. (15) is of size $N_{EVP} = 440$ (50% of M). Although GaAs is a direct band gap material, the nanowire band structure in Fig. 3(d) has strong resemblance to the bulk case states situated around the X point in the three-dimensional Brillouin zone. Many subbands have a local

minimum at $k \simeq k_{max}$, where $E(k)$ is close to the value at $k = 0$. Thus they play an important role in the calculation of the transmission $T(E)$. For example, the minima of the first band are situated at $E(k=0)=2.236$ eV (global, corresponding to the first turn-on in the electron transmission) and at $E(k=0.92)=2.378$ eV [local X state causing the third, doubly-degenerate step in $T(E)$].

Figures 2(e), 3(e), and 4(e) are devoted to a GaAs hexagonal wire (height 2.4 nm, maximum width 2.4 nm, minimum width 1 nm) whose transport direction coincides with [110], y with [$\bar{1}$10], and z with [001]. There are two atomic layers (77 atoms) per slab of width $\Delta = 0.3997$ nm and 4389 atoms in the nanowire. This is a special case because the size of the matrix M involved in the OBC calculation is the same as the size of the reduced eigenvalue problem in Eq. (15): $N_M = N_{EVP} = 770$. This is due to the presence of only two atomic layers per slab. Both are connected to the neighboring slabs and need to be included in the matrix $\mathcal{P}$ defined in Eq. (11). All columns of $\mathcal{P}$ will have at least one element different from zero. The band structure and the electron transmission do not exhibit relevant features. A larger cross section compared to the other nanowires presented in this article tightens the electron subbands and the transmission grows faster. With bias, we observe a shift of the curves and a smoothing of the steps, but no additional effects.

The last GaAs nanowire of Figs. 2(f), 3(f), and 4(f) has a rectangular cross section (1.2×1.2 nm^2) and a special crystal orientation: x is aligned with [113], y with [$\bar{1}$10], and z with [33$\bar{2}$]. Therefore a wire slab ($\Delta = 1.8749$ nm) contains 22 atomic layers and 140 atoms, 3.4 times more than the [100] Si rectangular wire in Fig. 2(a) with the same dimensions. The matrix M, with size $N_M = 1400$, can be reduced according to Eq. (15) to $N_{EVP} = 240$, 17.1% of N_M. Our OBC method is particularly advantageous when a slab contains many atomic layers.

For nanowires, strain induced by lattice mismatch or growth conditions is significant and should not be omitted. Until now, we simulated perfect structures, but we aim to show that our approach still works if atoms are shifted from their original positions. For that purpose, we deform the hexagonal GaAs nanowire introduced in Fig. 2(d). In Fig. 7, the original unstrained cross section (dashed atom connections) is compared to its biaxially deformed counterpart (solid atom connections). The applied tension increases the cross section while the width of a wire slab decreases proportionally to the resulting compressive uniaxial strain in the transport direction.[28]

The strain parameters used in the simulation are found in the literature.[29] Si examples would be more meaningful, but according to our knowledge, there have been no published strain tight-binding parameters for this material that take both Harrison's scaling rule and the orthogonal character of the Löwdin's orbitals into account.[29] The calculated transmissions are shown in Fig. 8. We compare the unstrained (dark solid lines), tension (light solid lines), and compression (dark thin lines) cases for a flat potential (upper plot) and for a bias of 0.1 V (lower plot). The biaxial strain amounts to $\epsilon_{yy} = \epsilon_{zz} = \pm 0.015$, the uniaxial, calculated using van de Walle's[28] value for D^{110}, to $\epsilon_{xx} = \mp 0.0087$. A compressive
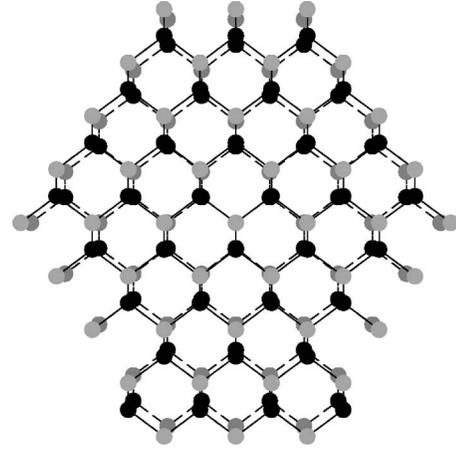
FIG. 7. Superposition of an unstrained (dashed atom connections) and of a strained (solid atom connections) GaAs hexagonal wire cross section. The unstrained case corresponds to wire (e) in Fig. 2. An homogeneous biaxial tension is applied in the strain case ($\epsilon_{yy} = \epsilon_{zz} > 0$).

strain pushes the conduction band edge up, leading to a turn-on of the first channel at a higher energy ($E = 2.0186$ eV) than without strain ($E = 1.989$ eV). In the same way, a tensile strain pushes the conduction band edge down, leading to a first channel turn-on at $E = 1.9475$ eV. From the turn-on of the second channel the change of the effective mass due to strain can compensate the lowering (tension) or increase (compression) of the conduction band edge.

For all the transmission curves in Fig. 4, spin-orbit coupling has been neglected, requiring ten atomic orbitals, one s, three p, five d, and one excited s^*. Spin degeneracy could be added by multiplying each curve by two for the spin-up and the spin-down contributions. However, it is legitimate to wonder if this simplification is justified or not. First, spin-orbit coupling is more important for GaAs ($\Delta_{SO} = 0.34$ eV) than for Si ($\Delta_{SO} = 0.044$ eV). Also, its influence is stronger if
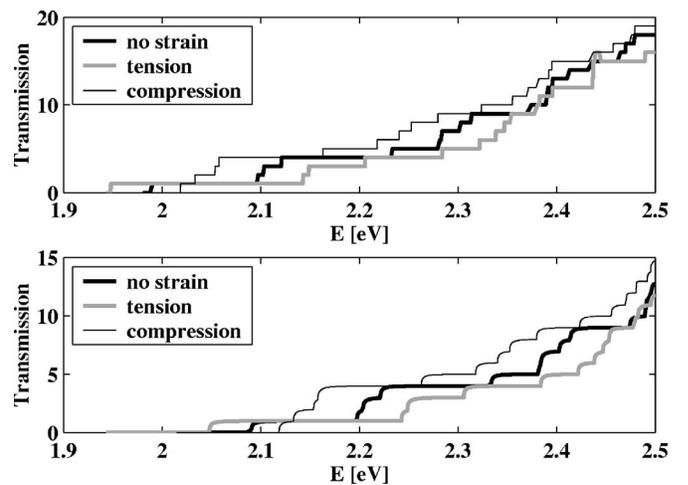
FIG. 8. Electron transmission for the GaAs hexagonal wire in Fig. 2 without strain (dark solid line), with a biaxial tension $\epsilon_{yy} = \epsilon_{zz} = 0.015$, $\epsilon_{xx} = -0.0087$ (light solid line), and for a biaxial compression $\epsilon_{yy} = \epsilon_{zz} = -0.015$, $\epsilon_{xx} = 0.0087$ (dark thin line). (Up) no bias. (Down) bias of 0.1 V.
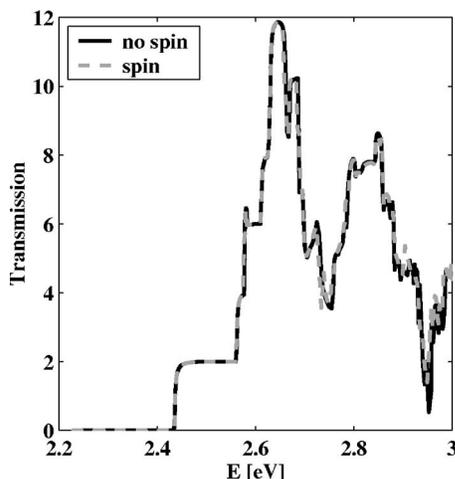
FIG. 9. Electron transmission through the GaAs T-shape wire in Fig. 2 for an applied bias of 0.2 V, calculated with (light dashed line) and without (dark solid line) spin orbit coupling.

a nonuniform potential is applied on the nanowire. We compare in Fig. 9 the electron transmission of the GaAs T-shape wire from Fig. 2(d) in the case of a 0.2 V bias with (gray dashed line) and without (dark solid line) spin-orbit coupling. The results match at low energies and slightly diverge at higher energies indicating that the neglect of spin-orbit coupling in Figs. 3–6 and 8 is a good approximation.

IV. DISCUSSION

In Sec. II we stated that our approach to calculate open boundary conditions is more efficient than other methods found in the literature.[14–16,20,22,24,30] To support this affirmation, we classify other available methods into three categories. First, there are the solutions that do not work for nanowires because they require the inversion of singular matrices to generate a generalized[22] or a normal[24] eigenvalue problem whose size is in any case larger than Eq. (15). Second, the transformation of Eq. (6) to a complex non-Hermitian (if spin-obit coupling is included) or real nonsymmetric (without spin-orbit coupling) generalized eigenvalue problem as in Eq. (7) is very popular[14,16,20,30] since it can be coupled to wave function[14] or nonequilibrium Green's function[16] transport solvers. Nevertheless, if a nanowire slab contains N atoms and t_b is the tight-binding order, the size of the GEVP amounts to $N_{GEVP}=2t_bN$, a considerable number for large cross sections and transport directions different from [100]. LAPACK functions[31] can be used to solve Eq. (7). Finally, the third category includes the OBC calculation with an iterative algorithm.[15] In this case, very dense or even full matrices of size $N_M=t_bN$ must be inverted 20–50 times to reach convergence. This approach works only to obtain the boundary self-energies in the NEGF formalism.

In Table I, we report the computational time for the generalized eigenvalue problem, the iterative scheme (labeled "Sancho-Rubio"), and our present method as well as the speed-up factor achieved. All methods have been implemented in MATLAB,[32] because it automatically calls LAPACK

routines to solve the eigenvalue problems and to invert matrices. Although the CPU times would be faster if the codes were written in C++, a relative comparison of the different methods makes sense in MATLAB, too. The benchmark examples are run on the same hardware platform. The CPU times in Table I three last lines) refer to the OBC calculation of one single contact at one given injection energy, without spin-orbit coupling (real EVP), and where the symmetry simplification presented in the Appendix is not applied. Note that the computational burden in the iterative algorithm case depends on the injection energy and could be therefore more important than what is given in Table I. The methods based on eigenvalue problems require the same effort for all the energies and among them our approach is always the most efficient with an OBC computational time at least one order of magnitude below the others.

For transport directions such as [111] or [113], the improvement is larger because a wire slab contains many atomic layers that can be disregarded in the OBC calculation (see the line labeled "% of normal size" that indicates what part of matrix M is effectively considered in the reduced EVP). Even for the less advantageous case where x is aligned with [110] (only two atomic layers per slab), Eq. (15) is the most appropriate solution. Rivas *et al.* mentioned[16] that the GEVP approach is more efficient than iterative methods[15] since it does not involve repetitive calculations. This is confirmed by our results (a factor of 1.5–4), except when x coincides with [113] where the GEVP is slower.

After we calculate the OBCs with Eq. (15), we have the possibility to couple them either to the linear system of equations (17) or to transform them to self-energies with Eq. (24) in order to simulate transport. In the wave function approach, we proceed to a LU factorization of Eq. (17) because it can be done in parallel[33] on several CPUs. In the NEGF formalism we utilize a recursive algorithm[23,25] that is slower than the LU factorization of Eq. (17), even on one single CPU, since it requires the inversion of N_S (number of wire slabs) matrices of size $N_M=t_bN$. However, Green's function facilitates the inclusion of inelastic scattering. To improve both the wave function and NEGF approaches we can reduce the bandwidth of the tight-binding Hamiltonian matrix. All the elements D_{ii}, T_{ii+1}, and T_{ii-1} are expressed in a slab basis, but there is no restriction to change it and to use an atomic layer basis, as we did for the boundary conditions. The block matrices that must be inverted in the NEGF recursive algorithms become then smaller and the LU factorization works better.

In this paper, we assume that the nanowire contacts are perfect with all the slabs in the semi-infinite left and right reservoirs identical to the first and the last wire slab, respectively. In reality, it is not possible to fabricate such nanowires. They can be grown as nanopillars on a bulk substrate[13,16] or embedded between two quantum well reservoirs.[6] Equation (15) can be modified to treat two-dimensional[34] (quantum well) or three-dimensional (bulk) reservoirs. Furthermore our OBC method enables the simulation of nanowires with larger cross sections than previously. To study the influence of alloy disorder, we have simulated an AlGaAs wire with a rectangular cross section of 6×6 nm^2, more than 1000 atoms per slab, and a total of

~46 000 atoms could be investigated with a full tight-binding band structure.[35]

V. SUMMARY AND CONCLUSION

This paper presents simulation approaches for nanowires with two-dimensional confinement. At this atomic scale, band-structure effects play an important role and must be treated carefully. We have developed a quantum transport simulator based on the $sp^3d^5s^*$ semiempirical tight-binding method in order to improve the understanding of the electrical behavior of such structures. Since the calculation of the open boundary conditions causes a significant part of the computational burden, a method involving the solution of a complex non-Hermitian or real nonsymmetric eigenvalue problem is proposed. It works for Si, GaAs, or any other material nanowires with different cross sections and crystal orientations, when a bias is applied or not, and when the structure is deformed by strain. A factor of 10–100 gain in speed compared to the other available methods could be gained in the evaluation of the OBCs, enabling the treatment of larger and more complicated structures. The coupling of this approach to a wave function and to a NEGF quantum transport solver has also been presented.

ACKNOWLEDGMENTS

APPENDIX: SYMMETRY PROPERTIES

Figure 10 shows the cross section of a rectangular GaAs wire whose transport direction is aligned with the [100] crystal axis. In fact, the cross section corresponds to the projection of a wire slab containing 41 atoms distributed over four atomic layers, two cation (dark atoms) and two of anion (light atoms) layers. By writing down Eq. (6), it is obvious
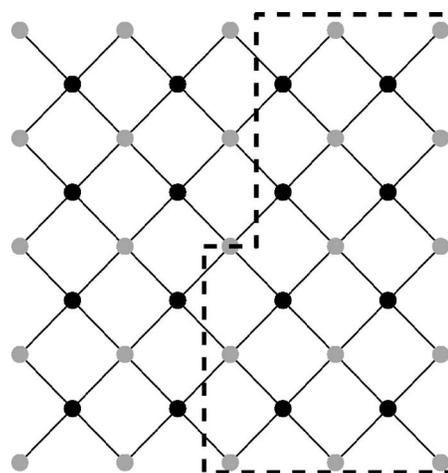
FIG. 10. Cross section of a rectangular GaAs wire with rotational symmetry properties around its center.

that the cross section of Fig. 10 has symmetry properties. For example, the atoms in the domain limited by the dashed line can be rotated by 180° around the wire center. We could think that the cross section has also a vertical, a horizontal, or a diagonal symmetry axis but this is not true since the reflected atoms would not belong to the same atomic plane as the atoms they replace. Therefore, we have to solve Eq. (15) twice, but with a size divided by 2, once for symmetric and once for antisymmetric rotation. The computational burden is significantly reduced, even more if the two cases are treated on parallel CPUs. This simplification works for the semi-infinite contacts if the quantum potential applied on them has the same symmetry as the atom positions, which is the case for double-gate or all-around-gate nanowire transistors. In the device itself, this property should not be used if we want to include alloy disorder[26] or interface roughness effects, which are completely random. If the transport direction is aligned with a different crystal orientation, the symmetry properties are different and must be studied case by case.

[1] Y. Cui, L. J. Lauhon, M. S. Gudiksen, J. Wang, and C. M. Lieber, Appl. Phys. Lett. **78**, 2214 (2001).

[2] A. I. Persson, M. W. Larsson, S. Steinström, B. J. Ohlsson, L. Samuelson, and L. R. Wallenberg, Nat. Mater. **3**, 677 (2004).

[3] A. B. Greytak, L. J. Lauhon, M. S. Gudiksen, and C. M. Lieber, Appl. Phys. Lett. **84**, 4176 (2004).

[4] T. Saito, T. Saraya, T. Inukai, H. Majima, T. Nagumo, and T. Hiramoto, IEICE Trans. Electron. **E85-C**, 1073 (2002).

[5] H. Majima, Y. Saito, and T. Hiramoto, Tech. Dig. - Int. Electron Devices Meet. **2001**, 733 (2001).

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

[7] Y. Hayamizu, M. Yoshita, S. Watanabe, H. Akiyama, L. N. Pfeiffer, and K. W. West, Appl. Phys. Lett. **81**, 4937 (2002).

[8] J. Wang, A. Rahman, A. Ghosh, G. Klimeck, and M. S. Lundstrom, IEEE Trans. Electron Devices **52**, 1589 (2005).

[9] J.-M. Jancu, R. Scholz, F. Beltram, and F. Bassani, Phys. Rev. B **57**, 6493 (1998).

[10] T. B. Boykin, G. Klimeck, and F. Oyafuso, Phys. Rev. B **69**, 115201 (2004).

[11] S. Lee, F. Oyafuso, P. von Allmen, and G. Klimeck, Phys. Rev. B **69**, 045316 (2004).

[12] Y.-J. Ko, M. Shin, S. Lee, and K. W. Park, J. Appl. Phys. **89**, 374 (2001).

[13] Y. Zheng, C. Rivas, R. Lake, K. Alam, T. B. Boykin, and G. Klimeck, IEEE Trans. Electron Devices **52**, 1097 (2005).

[14] M. Städele, B. R. Tuttle, and K. Hess, J. Appl. Phys. **89**, 348 (2001).

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

[16] C. Rivas and R. Lake, Phys. Status Solidi B **239**, 94 (2003).

[17] Y.-C. Chang and J. N. Schulman, Phys. Rev. B **25**, 3975 (1982).

[18] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).

[19] T. B. Boykin, Phys. Rev. B **54**, 8107 (1996).

[20] T. B. Boykin, Phys. Rev. B **54**, 7670 (1996).

[21] Y. Meir and N. S. Wingreen, Phys. Rev. Lett. **68**, 2512 (1992).

[22] D. Z.-Y. Ting, E. T. Yu, and T. C. McGill, Phys. Rev. B **45**, 3583 (1992).

[23] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, J. Appl. Phys. **81**, 7845 (1997).

[24] R. C. Bowen, W. R. Frensley, G. Klimeck, and R. K. Lake, Phys. Rev. B **52**, 2754 (1995).

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

[26] T. B. Boykin, M. Luisier, A. Schenk, N. Kharche, and G. Klimeck, IEEE Trans. Nanotechnol. (to be published).

[27] G. Klimeck, F. Oyafuso, T. B. Boykin, R. C. Bowen, and P. von Allmen, Comput. Model. Eng. Sci. **3**, 601 (2002).

[28] C. G. Van de Walle, Phys. Rev. B **39**, 1871 (1989).

[29] T. B. Boykin, G. Klimeck, R. C. Bowen, and F. Oyafuso, Phys. Rev. B **66**, 125207 (2002).

[30] C. Strahberger and P. Vogl, Physica B **272**, 160 (1999).

[31] 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*, 3rd ed. (SIAM, Philadelphia, 1999).

[32] Math Works, Inc., MATLAB, version 7.0.1 (Natick, MA, 2004).

[33] X. S. Li and J. W. Demmel, ACM Trans. Math. Softw. **29**, 110 (2003).

[34] M. Luisier, G. Klimeck, A. Schenk, and W. Fichtner, in Proceedings of IWCE11, 2006 (unpublished).

[35] M. Luisier, A. Schenk, N. Kharche, G. Klimeck, and T. B. Boykin (unpublished).