J Comput Electron (2008) 7: 309–314
DOI 10.1007/s10825-008-0183-9

Full-band quantum transport in nanowire transistors

Mathieu Luisier

Published online: 26 January 2008

Abstract Semiconductor nanowires may be the core components of next generation processors and memories. In effect, several groups already demonstrated the feasibility of Si or Ge nanowire field-effect transistors (FETs). However, the fabrication of novel devices is always a difficult and expensive process. The recourse to technology computer aided design can facilitate the development of new structures and help reducing the inherent costs. In this article a full-band quantum transport (QT) solver dedicated to nanowire transistors is presented. The semi-empirical $sp^3d^5s^*$ tight-binding (TB) method is chosen as bandstructure model for its accuracy to reproduce the bulk properties, for its straight forward extension to nanostructures, and for its atomic description of the simulation domain. The calculation of multi-band open boundary conditions (OBCs) and their integration into a three-dimensional Schrödinger-Poisson or Non-equilibrium Green's Function solver are fundamental in the development of a ballistic QT simulator. Different approaches are investigated and compared in this work. They all allow transport with any channel orientation, material composition, and cross section shape. The computational burden restricts most of them to the simulation of small nanowire structures. However, some advanced numerical techniques open promising perspectives towards realistic devices.

Keywords Nanowires · Tight-binding method · NEGF · Open boundary conditions

M. Luisier (✉)
Integrated Systems Laboratory, Gloriastrasse 35 ETH Zurich, 8092 Zurich, Switzerland
e-mail: mluisier@iis.ee.ethz.ch

1 Introduction

During the last thirty years the size of the transistors has been aggressively scaled to reach the atomic range. Some groups have already reported metal-oxide-semiconductor field-effect transistors (MOSFETs) with a gate length below 10 nm. The lateral dimensions of these transistors has also been reduced to form a kind of "wire" structure. The resulting nanowire (NW) devices may play an important role in the future of nanoelectronics [1]. This article presents an atomistic quantum transport solver designed for these post-CMOS FETs [2]. The quantization effects characterizing nanowires cannot be captured by classical models and require the development of a full-band (FB) simulator. This particularly concerns transistors whose cross section does not exceed 5 nm × 5 nm [3]. The $sp^3d^5s^*$ semi-empirical tight-binding (TB) method [4, 5] is an appropriate bandstructure model for the desired application. It has the advantage to fully account for the atomic granularity of the simulation domain.

The paper is organized as follows: Sect. 2 describes three procedures to compute open boundary conditions (OBCs) in a multi-band model and their integration into an atomistic quantum transport solver based either on the Non-Equilibrium Green's Function [6] (NEGF) or on the Wave Function [7] (WF) formalism. The results are used to compute carrier and current densities in nanowires with any crystal orientation such as [100], [110], or [111], and any wire shape (e. g. square, triangular, circular, ...).

Section 3 deals with the numerical implementation of the transport models. The methods to compute the OBCs and to solve the NEGF or WF problems are benchmarked on a single CPU and on many shared/distributed memory processors for selected nanowire structures.

Simulation results are presented in Sect. 4. First, the limit of the effective mass approximation (EMA) is investigated. Square Si nanowire transistors with transport along the [100] crystal axis are simulated as function of their cross section size. The corresponding FB and EMA current characteristics are then compared. It is followed by a study of circular Si nanowires with different crystal orientations. The optimization of the computational resources enables the treatment of structures with a cross section surface up to 16 nm^2.

Finally, this article is summarized and concluded in Sect. 5. Some indications regarding further improvement of the quantum transport models are also given.

2 Theory

This Section starts with a brief description of three FB algorithms to obtain open boundary conditions in nanowire structures, (1) an iterative solver [8], (2) a generalized eigenvalue problem approach [9], and (3) a shift-and-invert procedure resulting in an eigenvalue problem [10]. Figure 1 shows the schematic view of a nanowire transistor 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 (twenty with spin-orbit coupling).

To solve the Schrödinger equation in the WF or NEGF formalism, electrons with an energy E enter and leave the n-doped device in Fig. 1 at the source and drain contacts. The valence band is assumed completely filled. The resulting open system can be written as

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

if the Schrödinger wave functions are expanded in terms of orthogonalized Löwdin atomic orbitals and coefficient vectors C_i [10]. The index i denotes the i^{th} wire unit cell (or slab) of width Δ. The matrices $E - H_{ii}$ represent the on-site energy and the bond connections within a slab, $H_{ii\pm1}$ the coupling to the nearest-neighbors unit cells. 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 (1) has to be solved for each slab index i. The matrices H_{10} and H_{NN+1} couple the device to its semi-infinite contacts represented by the self-energies Σ_{11} and Σ_{NN} that are added to H_{11} and H_{NN}, respectively.

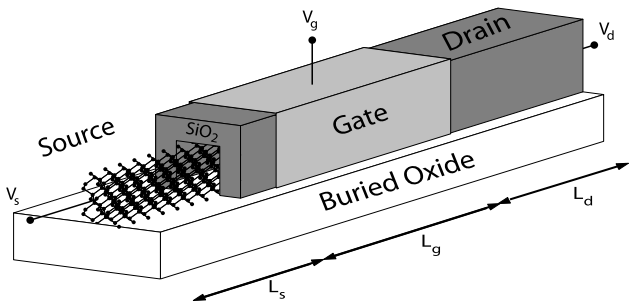

Fig. 1 Schematic view of a n-doped triple-gate nanowire transistor deposited on a buried oxide and surrounded by three oxide layers. The source is denuded in order to give an overview of the structure. The atomic granularity of the nanowire is fully accounted, but there is no electron penetration into the oxide

To calculate Σ_{11} iteratively, the following matrix equation has to be solved

$$\Sigma_{11} = H_{10}(E - H_{11} - \Sigma_{11})^{-1}H_{01}. \tag{2}$$

A similar expression is found for Σ_{NN}. Equation (2) hardly ever converges. Hence, the procedure is improved so that the modified iteration n corresponds to the iteration 2^n in (2) [8]. Nevertheless, 20 to 40 inversions of a dense or full matrix of size t_bN are required for each energy point and for each contact.

Another common procedure of calculating the Σ's consists in separating the boundary coefficients C_0 (or C_{N+1}) into a transmitted and a reflected part

$$C_0 = \sum_n \left(a_n e^{ik_n^+ x_0}\varphi_n^+ + b_n e^{ik_n^- x_0}\varphi_n^-\right). \tag{3}$$

and then in constructing a generalized eigenvalue problem (GEVP) of size $2t_bN$ based on them [9]. In (3) a_n is the injection coefficient for the n^{th} state φ_n^+ injected into the device and b_n the coefficient for the n^{th} state φ_n^- reflected back to the contact. All these quantities are energy dependent. The Σ's are derived from the $\varphi_n^\pm$ and the wave vectors $k_n^\pm$ [10].

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. [10]. Only the final step of the calculation is given here. 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's and the φ_n's are the solutions of an ordinary eigenvalue problem (EVP)

$$M\varphi_n = \frac{1}{e^{ik_n\Delta} - 1}\varphi_n. \tag{4}$$

The matrix M is always well-defined and has an advantageous structure. For example, if the transport direction x is aligned with [100], M looks like

$$M = \begin{pmatrix} 0 & 0 & M_{02} & M_{03} \\ 0 & 0 & M_{12} & M_{13} \\ 0 & 0 & M_{22} & M_{23} \\ 0 & 0 & M_{32} & M_{33} \end{pmatrix}. \tag{5}$$

Table 1 OBCs computational time (in seconds) for one single contact and energy point in square Si nanowires with x aligned with [100] and L_y and L_z given in the first column. The shift-and-invert method, the iterative procedure, and the generalized eigenvalue problem (GEVP) approach outlined in Sect. 2 are compared. Spin-orbit coupling is not considered

$L_y \times L_z$ nm^2	Atoms per slab	Shift-and-invert [s]	Iterative [s]	GEVP [s]
2.5×2.5	181	7.2	197	506
2.9×2.9	242	18.5	462	1490
3.3×3.3	313	39	1070	3930

It is thus not necessary to consider the whole matrix M in the eigenvalue problem defined in (4), but only its lower right corner, corresponding to a size of $t_b N/2$. The matrix M can be constructed for any crystal orientation and always has a size smaller or equal to $t_b N$. The knowledge of the contact's variables $\varphi_n^{\pm}$ and $k_n^{\pm}$ yields the Σ's [10]. The influence of the source and drain shape is discussed in Ref. [11].

Once the open boundary conditions are known, (1) can be cast into a linear system

$$(\mathbf{E} - \mathbf{H} - \mathbf{\Sigma}) \cdot \mathbf{C} = \mathbf{Inj}, \tag{6}$$

where the matrix **Inj** represents the injection mechanism, or into a NEGF problem

$$(\mathbf{E} - \mathbf{H} - \mathbf{\Sigma}) \cdot \mathbf{G^R} = \mathbf{I}, \tag{7}$$

where $\mathbf{I}$ is the identity matrix [10]. If there are N_A atoms in the nanowire, the Hamiltonian $\mathbf{H}$ is a square matrix of size $N_A\, t_b \times N_A\, t_b$ as the boundary self-energy $\mathbf{\Sigma}$. Its first and last diagonal blocks contain Σ_{11} and Σ_{NN}, respectively. Packages capable of solving the linear system of equations (6) are presented in the next Sect. 3. Equation (7) is solved with a recursive Green's function (RGF) algorithm [12].

It remains to determine the electron density $n(\mathbf{r})$ (the hole density is obtained in a similar way) and the current density $\mathbf{J}(\mathbf{r})$. Due to the strong localization of the Löwdin orbital functions, $n(\mathbf{r})$ and $\mathbf{J}(\mathbf{r})$ are represented by δ-functions centered around the atom positions $\mathbf{R_i}$ [2]. The calculation of the carrier density and of the electrostatic potential in the device are self-consistently coupled. A finite element grid is used for that purpose. Floating boundary conditions are applied to the source and drain ensuring charge neutrality in these regions.

3 Numerical Aspect

To calculate the density-of-states and the transmission in a nanowire, electrons are injected into the device from the source and the drain contacts. This procedure is repeated for each injection energy. Typically, the energy vector is composed of 500 points with a finer discretization around the contact eigenstates. This means that (6) or (7) is solved 500 times. In the absence of dissipative scattering in the device, all the energy points are independent and can be treated in parallel. The distribution of the tasks is realized by the message passing interface (MPI). A total of N available CPUs yields a speed up factor of N. However, it is not always possible to solve (6) or (7) on a single node, if, for example, the size of the matrix $(\mathbf{E} - \mathbf{H} - \mathbf{\Sigma})$ 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.

For a given energy point the first task consists in calculating the OBCs in order to construct $(\mathbf{E} - \mathbf{H} - \mathbf{\Sigma})$. In Table 1, the OBCs computational time for [100] square Si nanowires is reported. Spin-orbit coupling is neglected. The generalized eigenvalue problem, the iterative procedure, and the shift-and-invert method are tested on a 64-bit Sun Fire X4600 with 4×2.8 GHz Dual Core Opteron processors. The benchmark times refer to the calculation of one contact only, the source or the drain. The shift-and-invert method is about 25 to 100 times faster than the others and is preferred for the simulation of nanowire transistors.

After the calculation of the OBCs, the sparse matrix $(\mathbf{E} - \mathbf{H} - \mathbf{\Sigma})$ is assembled and factorized. This task is either processed directly by sparse linear solvers like Umfpack 5.0.1 [13], PARDISO [14], SuperLU$_{dist}$ 2.0 [15], MUMPS 4.6.3 [16], a recently developed basis compression algorithm (BACA) [17], or indirectly by using a recursive Green's Function (RGF) algorithm [12]. The packages SuperLU$_{dist}$ 2.0 and MUMPS 4.6.3 as well as the basis compression algorithm are designed for distributed memory computers, PARDISO works on shared memory machines, while Umfpack 5.0.1 and the RGF are sequential.

To compare the different methods, a [100] Si nanowire with a 3.3×3.3 nm^2 cross section and a total length $L = 35$ nm is solved. The sequential and parallel performances of the linear solvers are tested on the same computer architecture as before. The results are summarized in Table 2. Columns 3 to 8 reproduce the time (in seconds) to factorize and solve (6) or (7). The ninth column is the time necessary to obtain the OBCs for one contact with the shift-and-invert approach. The size of the sparse matrix $(\mathbf{E} - \mathbf{H} - \mathbf{\Sigma})$ is shown in the last column (spin-orbit coupling not included).

Table 2 Time (in seconds) required to calculate the wave function (columns 3, 4, 5, 6, and 7) or to obtain the Green's functions (column 8) for one energy point of a square nanowire once the open boundary conditions have been computed (column 9). The parallel scalability of the solvers is also shown. The size N of the Hamiltonian matrix is given in column 10 (no spin-orbit coupling). In the first column, L_x, L_y, and L_z indicate the size of the nanowire

$L_x \times L_y \times L_z$ [nm^3]	CPU	Umfpack [s]	PARDISO [s]	SuperLU$_{dist}$ [s]	MUMPS [s]	BACA [s]	RGF [s]	Bound. Cond. [s]	N
$35 \times 3.3 \times 3.3$	1	406	271/1x	560/1x	240/1x	105/1x	1418	39	197190
$35 \times 3.3 \times 3.3$	2	–	141/1.9x	258/2.2x	129/1.86x	54/1.94x	–	39	197190
$35 \times 3.3 \times 3.3$	4	–	84/3.2x	130/4.3x	76/3.15x	31/3.38x	–	39	197190
$35 \times 3.3 \times 3.3$	8	–	63/4.3x	112/5x	56/4.3x	21/5x	–	39	197190

On a single processor, the basis compression algorithm is at least two times faster than all the other methods. On the opposite side, the solution of the NEGF problem is inefficient. The NEGF formalism is not indicated in the case of ballistic transport, but becomes significant in the presence of scattering. The analysis of the parallel scalability shows that SuperLU$_{dist}$ 2.0 is better than the others on 2 and 4 processors. This must be carefully considered since the factorization and solve time on 2 (4) CPUs is more than two (four) times faster than on a single CPU. This could originate from the fact that this package was designed for parallel use only. On 8 processors, the scalability of the basis compression algorithm is as good as that of SuperLU$_{dist}$, but it factorizes and solves (6) five times faster. Hence, the best package tested in this study is the basis compression algorithm. Consequently, it is used in combination with the shift-and-invert OBCs calculation to produce the results of the next section.

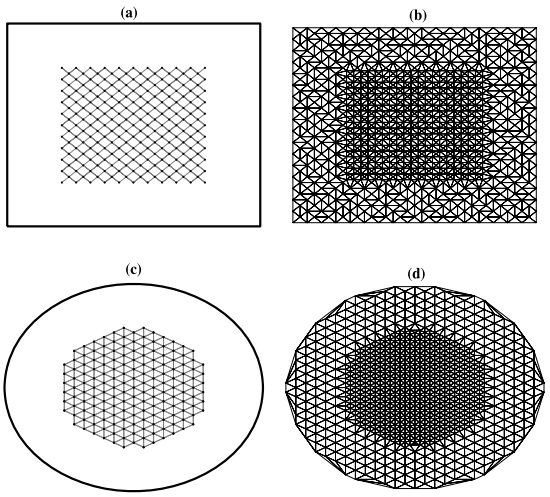

Fig. 2 Cross section (*left*) and corresponding finite element mesh projection (*right*) of a square and of a circular nanowire transistor surrounded by 1 nm oxide layers

4 Results

The room temperature IV-characteristics of square and circular Si field-effect nanowire transistors are simulated in this Section. The full-band charge and current densities are calculated as described in Sect. 2. They are solved self-consistently with the three-dimensional (3D) electrostatic potential (Poisson equation) of the device. To keep the δ-character of the carrier density, the finite element method (FEM) is chosen. A Delaunay grid is constructed by imposing the constraint that no atom is contained in the volume of any tetrahedron mesh element. A projection of the FEM grid is given in Fig. 2 for a [100] square nanowire and a [111] circular structure. As seen on the left part of the figure the SiO_2 oxide layers do not participate to the transport calculation, but are included in the 3D electrostatic potential. This is due to a poor tight-binding parametrization of SiO_2 (only crystalline structure) and to the difficulty of constructing Si–SiO_2 interfaces for other than [100]-grown nanowires, for corners, and for circular structures. Consequently, the oxide grid points carry no charge, a coarse mesh is used in these regions, and hard wall boundary conditions are applied to the Si surface atoms [18]. A perfect crystal structure is assumed in the entire nanowire. As a first approximation the bulk dielectric constants are used for Si ($\epsilon = 11.9$) and SiO_2 ($\epsilon = 3.9$).

The selected transistors (circular or square) are composed of 10 nm long source and drain (length L_s and L_d) extensions that are n-doped ($N_D = 10^{20}$ cm^{-3}). The doping charge is homogeneously distributed on the source and drain atoms. The undoped channel, the source, and the drain are surrounded by 1 nm thick oxide layers. A potential V_g is applied to the gate (work function $\phi_m = 4.25$ eV), $V_s = 0$ V to the source and V_d to the drain. Note that p-doped transistors are treated in a similar way, assuming that the conduction band is completely empty [2].

4.1 Square structure

In Fig. 3 the FB (black lines) and the EMA (gray lines) $I_d - V_{gs}$ transfer characteristics of square nanowire with

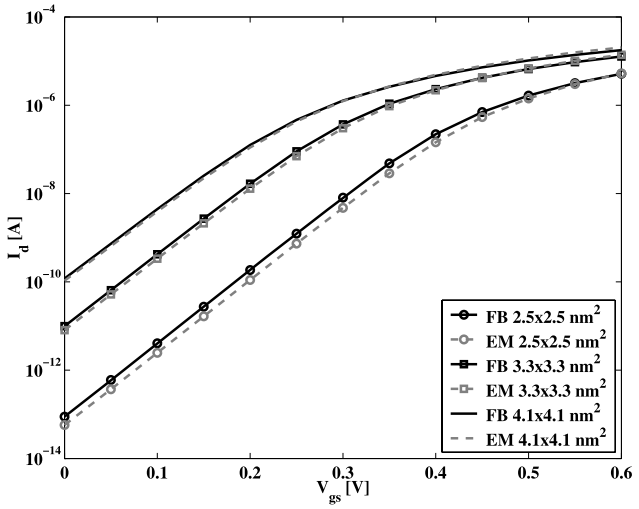

Fig. 3 Current transfer characteristics $I_d - V_{gs}$ at $V_{ds} = 0.6$ V for [100] square Si nanowires (length $L = 35$ nm, gate length $L_g = 15$ nm). FB (*black lines*) and EMA (*gray lines*) results are presented for three nanowire cross sections: 2.5×2.5 nm^2 (*lines* with *circles*), 3.3×3.3 nm^2 (*lines* with *squares*), and 4.1×4.1 nm^2 (*lines* without *symbols*). The simulations are done at room temperature for a gate work function $\phi_m = 4.25$ eV and a Si affinity $\chi_{Si} = 4.05$ eV

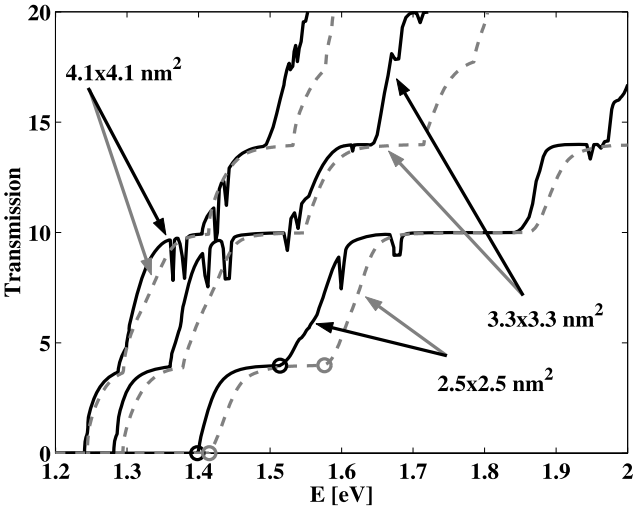

Fig. 4 FB (*black solid lines*) and EMA (*gray dashed lines*) transmission coefficient at $V_{gs} = V_{ds} = 0.6$ V corresponding to the three nanowires structures simulated in Fig. 3

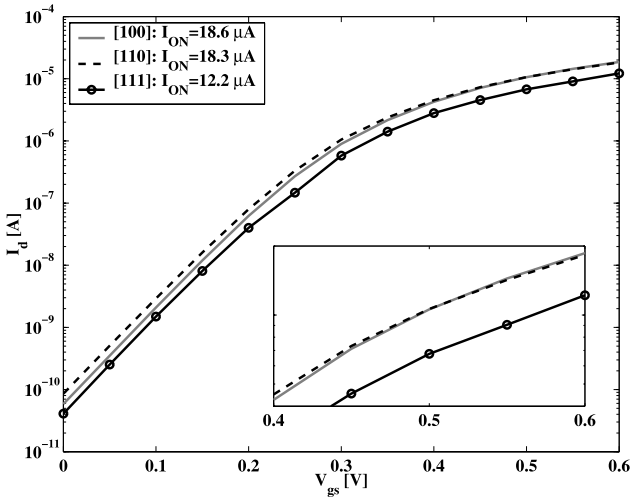

Fig. 5 Full-band current transfer characteristics $I_d - V_{gs}$ at $V_{ds} = 0.6$ V for circular Si nanowires (diameter $d = 4$ nm, length $L = 30$ nm, gate length $L_g = 10$ nm) with transport along the [100], [110], and [111] crystal axis. The simulations are done at room temperature for a gate work function $\phi_m = 4.25$ eV and a Si affinity $\chi_{Si} = 4.05$ eV. The ON-current I_{ON} is defined as I_d at $V_{gs} = V_{ds} = 0.6$ V. The region around I_{ON} is enlarged in the inset

different cross sections are compared. They are calculated at a fixed drain-source voltage $V_{ds} = 0.6$ V. All the transistors have a channel aligned with [100] and are controlled by a triple-gate of length $L_g = 15$ nm, as in Fig. 1. The EMA results are obtained with a longitudinal $m_l^* = 0.92 m_o$ and a transverse $m_t^* = 0.19 m_o$ effective mass extracted from the bulk bandstructure. The lines with circles denote a 2.5×2.5 nm^2 cross section (11403 atoms in the wire), the lines with squares a 3.3×3.3 nm^2 cross section (19719 atoms), and the lines without symbols a 4.1×4.1 nm^2 cross section (28350 atoms). These surfaces do not include the 1 nm oxide layer surrounding the nanowire.

Independently from the size of the transistor 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}. In the saturation regime the EMA ON-currents (defined as drain current I_d at $V_{gs} = V_{ds} = 0.6$ V) tend to be larger than their FB counterpart. However, the overall difference between the EMA and FB results decreases as the cross section of the nanowire increases so that both currents are about the same for the 4.1×4.1 nm^2 case. This is confirmed by the FB and EMA transmission coefficients plotted in Fig. 4 at $V_{gs} = V_{ds} = 0.6$ V. There are two important channel turn-on in each transmission curve, the first one involving four bands, the second one six bands. They are marked as circles for the 2.5×2.5 nm^2 cross section. The shift between the FB and EMA turn-on almost disappears for the 4.1×4.1 nm^2 structure, making the two currents approximatively the same.

4.2 Circular structure

Figure 5 shows the FB $I_d - V_{gs}$ ($V_{ds} = 0.6$ V) transfer characteristics of circular FETs with [100] (gray line), [110] (black dashed line), and [111] (black line with circles) as channel orientation. The three transistors have a gate-all-around architecture of length $L_g = 10$ nm and a diameter $d = 4$ nm without the oxide layers and $d = 6$ nm with. The [100] nanowire is composed of 18755 atoms, the [110] of 19344, and the [111] of 19264. The channel orientation that

offers the highest ON-current (I_d at $V_{gs} = V_{ds} = 0.6$ V) is [100] (18.6 μA), followed by [110] (18.3 μA), and finally [111] (12.2 μA). At the same time, the [100] direction profits from a lower OFF-current (I_d at $V_{gs} = 0$ V and $V_{ds} = 0.6$ V) than [110]. This behavior results from a crossing of the [110] and [100] current characteristics that can be observed in the inset of Fig. 5. The prominence of the [100] channel is a consequence of the compromise between the electron velocity (proportional to the inverse of the effective mass) and the number of available states (proportional to the effective mass) reached by its lowest conduction subbands. Source-to-drain tunneling does not affect the OFF-current since the subthreshold swing S of the three structures is very close to its ideal value of 60 mV/decade.

5 Conclusion

In this paper full-band quantum transport approaches to simulate nanowire transistors are presented. They are based on the semi-empirical $sp^3d^5s^3$ tight-binding model and on an atomistic description of the device structure. Different procedures to calculate open boundary conditions and to obtain carrier and current densities in the Wave Function or in the Non-Equilibrium Green's Function formalism are derived. The resulting simulator allows the self-consistent simulation of square or circular nanowire FETs with transport along the [100], [110], or [111] crystal axis.

As possible improvement to the current quantum transport simulator, an accurate description and random distribution of the doping atoms, a better treatment of the oxide layers, a relaxation of the surface atoms from their "equilibrium" position, and the inclusion of scattering mechanisms such as electron-phonon interactions should be mentioned.

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

References

1. Xiang, J., Lu, W., Hu, Y., Wu, Y., Yan, H., Lieber, C.M.: Nature **441**, 489 (2006)
2. Luisier, M., Schenk, A.: J. Comput. Theor. Nanosci. (2008, to appear)
3. Wang, J., Rahman, A., Ghosh, A., Klimeck, G., Lundstrom, M.S.: IEEE Trans. Electron. Devices **52**, 1589 (2005)
4. Slater, J.C., Koster, G.F.: Phys. Rev. **94**, 1498 (1954)
5. Boykin, T.B., Klimeck, G., Oyafuso, F.: Phys. Rev. B **69**, 115201 (2004)
6. Datta, S.: J. Phys.: Condens. Matter **2**, 8023 (1990)
7. Frensley, W.R.: Rev. Mod. Phys. **62**, 745 (1990)
8. Lopez Sancho, M.P., Lopez Sancho, J.M., Rubio, J.: J. Phys. F: Met. Phys. **15**, 851 (1985)
9. Rivas, C., Lake, R.: Phys. Stat. Sol. B **239**, 94 (2003)
10. Luisier, M., Klimeck, G., Schenk, A., Fichtner, W.: Phys. Rev. B **74**, 205323 (2006)
11. Luisier, M., Schenk, A., Fichtner, W., Klimeck, G.: J. Commun. Electron. **6**, 199 (2007)
12. Lake, R., Klimeck, G., Bowen, R.C., Jovanovic, D.: J. Appl. Phys. **81**, 7845 (1997)
13. Davis, T.A.: ACM Trans. Math. Softw. **30**, 165 (2004)
14. Schenk, O., Gärtner, K.: J. Future Gen. Comput. Syst. **20**, 475 (2004)
15. Li, X.S., Demmel, J.W.: ACM Trans. Math. Softw. **29**, 110 (2003)
16. Amestoy, P.R., Duff, I.S., L'Excellent, J.-Y.: Comput. Methods Appl. Mech. Eng. **184**, 501 (2000)
17. Boykin, T.B., Luisier, M., Klimeck, G.: unpublished
18. Lee, S., Oyafuso, F., von Allmen, P., Klimeck, G.: Phys. Rev. B **69**, 045316 (2004)