Computer-aided simulation of nanoelectronic devices: the importance of the choice of the physical model

PAOLO MARCONCINI
Università di Pisa
Dipartimento di Ingegneria dell’Informazione
Via Caruso 16, I-56122 Pisa
ITALY
paolo.marconcini@iet.unipi.it

Abstract: The importance of the choice of the physical model in computer-aided simulation of nanoelectronic devices is discussed. In particular, the effect of different alternatives on code performance and on the accuracy of the simulation results is discussed for some cases of the author’s research activity, stressing how the choice of physically equivalent descriptions can often determine a large variation in the efficiency of the simulation program.

Key–Words: Computer-aided simulation, nanoelectronic devices, physical model

1 Introduction

Nanoelectronics is the branch of electronics dealing with nanometer scale devices. It involves both traditional silicon CMOS electronics scaled to a size of a few nanometers, and more innovative devices (which could solve some of the problems resulting from the scaling of CMOS technology, e.g. excessive power dissipation and insufficient electrostatical control of the channel), such as devices obtained by confining a two-dimensional electron gas (2DEG) in semiconductor heterostructures [1–5], or graphene-based devices [6–11]. At this level of miniaturization, the physical implementation involves quite complex and expensive fabrication techniques and thus the prototyping activity has to be limited as much as possible. This increases the importance of the development of computer-aided design and simulation tools for these devices, both in order to fully understand their behavior at the physical level, and to optimize their electronic performance by means of an accurate design of their structure and processing steps, before their actual fabrication.

Since down at this scale quantum mechanical effect are fundamental to correctly describe the physical and electrical behavior, and thus a simple drift-diffusion model is generally not acceptable, the study of nanoelectronic devices requires quite a large computational effort. In order to reduce the simulation time, an optimization is needed, both from an algorithmic and from a modeling point of view.

From a computational point of view, a parallelization of the simulation codes (for example, using MPI libraries) is generally necessary and a proper and optimized implementation of the simulation codes, which can benefit from parallelization, has to be performed. We need to achieve a trade-off between the necessity, on the one hand, to preserve computed quantities useful in different parts of the code, and, on the other hand, to reduce memory occupation, avoiding paging problems, and to limit the number of read/write operations, in order to reduce the execution times. Moreover, since simulations require an intensive use of advanced numerical solution techniques, the exploitation of well-tested and optimized Fortran and C routines (such as Lapack [12] and Numerical Recipes [13] codes) is largely adopted and a proper choice of the numerical routines for the considered physical problem has to be made.

In this paper, instead, we will focus on the importance, for the optimization of these codes, of the choice of the physical model. In particular, referring to the present research activity of the author, it will be stressed how, while there is often a trade-off between the simplicity of the adopted representation (and therefore the performance of the code) and the accuracy of the obtained results, in many cases equivalent physical treatments can lead to a very different efficiency of the simulation code for particular devices. A proper choice of the physical model is therefore a fundamental aspect to perform optimized computer-aided simulations.

2 Semiclassical/full-quantum models

A first choice to be made is between a full-quantum model, which includes a complete quantum mechan-
Figure 1: Value (averaged over 50 sets of interbarrier distances) of the quantity $F$ (defined in the text) for a series of unevenly spaced tunnel barriers, as a function of the number of barriers. The three curves correspond to three different values of the barrier transparency $\Gamma$.

3 Many-body/single-electron models

Another important choice that one has to make while writing a simulation code for nanoelectronic devices is that between a many-body and a single-particle model. In the first one, a complete quantum description of the system, made up of all charge carriers in the device, is considered. An exact many-body calculation is however very difficult (or impossible) to perform (with the exception of systems with very few particles), unless appropriate approximations are made (mean-field methods, like Hartree, Hartree-Fock, Configuration-Interaction and DFT [16]). This kind of description is essential for structures where the interactions between particles play a fundamental role, as in devices (such as single-electron transistors) based on the Coulomb-blockade phenomenon [17]. In most of the cases, however, a single-electron description of transport is sufficient to correctly describe the system and allows a much faster simulation of nanoelectronic devices.

4 Spatial detail of the model

Another important element that can have a strong effect on the computational burden of the simulation code is the degree of spatial detail of the physical model. Inside a full-quantum description, a possibility is to solve the Schrödinger equation for an electron, considering all the atomistic details of the devices, including the potential of each single atom. Another one is to use a continuum, envelope function approach [18, 19], which neglects the atomic details of the structure and finds a slowly-varying function which multiplied by a function periodic with the lattice period gives the actual electron wave function. With this latter choice the equation to be solved is an effective mass Schrödinger equation in the case of common semiconductors, while is the Dirac equation in the case of graphene. If a device containing a very large number of atoms has to be simulated, a fully-atomistic simulation is generally impossible and the envelope function approach is preferable. On the contrary, if atomistic details are expected to play a fundamental role, especially in the case of small devices, an atomistic description is needed.

For example, we have analyzed the effect of a low concentration of boron doping on the transmission probability of a small ribbon of graphene [9, 20]. In this case, due to the reduced size of the structure, the atomistic details in correspondence of the boron atoms are fundamental to study the transport behavior of the device. In particular, the presence of quasi-bound states localized around boron impurities introduces a strong hole backscattering. However, while...
an atomistic description captures this effect, a continuum model based on the solution of the Dirac equation misses it and reproduces only the minor (in the case of narrow ribbons) electrostatic effect of negatively charged dopant atoms on electrons.

In order to simulate more complex structures, capturing the most important atomistic details without too large a computational burden, a possibility is to use a hierarchical simulation approach, in which from an atomistic study of a simplified structure a small number of parameters are extracted, which are then used in a less detailed description of the device, which can be performed faster. We have recently adopted this methodology to study silicon nanowire transistors.

5 Self-consistent solution and semi-analytical approximations

In the solution of the transport equation (Schrödinger or Dirac equation) we assume to know the value of the external potential. However, the value of the potential derives from the solution of the electrostatic (Poisson) equation, where also the charge density in the device (deriving from the solution of the transport equation) appears. Therefore a correct simulation of a nanoelectronic device requires the self-consistent solution of the transport and electrostatic equations, through an iterative procedure that consists in solving repeatedly the two equations till a solution that approximately satisfies both of them is found. While necessary to obtain an accurate result, this procedure is very resource-demanding from the computational point of view. In many cases an approximate solution can be sufficient to express an opinion on the correct functionality of a device, especially in the cases where a preliminary study on the dependence of the device characteristics on a large parameter space is needed. In these situations is can be useful to adopt, instead of a self-consistent solution of the transport and electrostatic equations, approximate procedures to compute the screened potential to be used in the transport equation.

For example, studying devices implemented with GaAs/AlGaAs heterostructures, confining the 2DEG located at the interface between GaAs and AlGaAs through the electrostatic action of polarized gates defined on the surface of the structure, we have observed and analyzed the tunneling enhancement effect generated on a tunnel barrier by the presence of a cavity surrounding it [4, 21], for several positions of the barrier inside the cavity. Since the number of configurations to be simulated was very large, we have in that case decided to use a technique based on the calculation of the linear response of the 2DEG to the potential of the gates [22]. The results, for the case of a tunnel barrier deriving from a gate with edge roughness, are reported in Fig. 2 and were well suited for this feasibility analysis. In that case, also an analysis of the screened effect of charged impurities located at a certain distance from the 2DEG was performed using a semi-analytic technique [23], without solving the complete Schrödinger/Poisson problem, which would be much more computationally demanding.

6 Solution domain and equivalent basis sets

Up to this point, we have considered a series of choices for which a trade-off between the speed of the computer simulations and the accuracy or detail of the obtained results exists.

However, while writing a simulation code for nanoelectronic devices, very often there are also equivalent alternatives in the physical model, which can have a different impact on the performance of the algorithm.

From this point of view, an important choice that has to be made in the physical model is the domain in which the calculation has to be performed. For example, in the study of nanoelectronic structures it is possible to operate in the direct domain (the spatial domain) or in the Fourier-transformed one, i.e. the reciprocal domain (the domain of the wave vectors).

An enlightening case is the transport study of graphene nanoribbons in the presence of a generic external potential, that we are performing using an envelope-function approach. In this case, a solution in the direct domain is very problematic. The reason is that the Dirac equation, if solved with a standard finite-difference approach, leads either to spurious eigenvalues (as we show in Fig. 3 for the case of a null external potential) or to the problem of fermion-doubling, which corresponds to the appearance of an unrealistic degeneracy in the wave vectors and discontinuities in the corresponding enve-
lope functions. Even though an alternative approach in the direct space exists, based on the use of two shifted grids for the points where the equations are evaluated and for the nodes where the values of the unknowns are obtained, it however has a very low numerical efficiency. On the contrary, if the problem is numerically transformed into one with periodic boundary conditions [24] and an analysis in the Fourier-transformed domain is performed, numerical problems disappear and, due to the exact treatment of derivatives, to the use of optimized FFT routines, and to the limited number of Fourier components that are generally needed in the case of slow-varying potentials, the speed of the code increases at least by two orders of magnitude with respect to a solution in the direct space.

The transition from the real space to the reciprocal space can be seen as a change [25] from a basis made up of discrete delta functions centered on the nodes of the discretization grid in the real space, to the basis of the plane waves corresponding to the different frequencies. Therefore this alternative can be seen as a particular case of the more general problem of the choice of a basis on which transport calculation are made. In general, the more similar the basis set on which transport calculation will be a proper combination of transversal modes, we have generally used a representation over a set of low-energy transverse eigenmodes (instead of over a discretization grid) in the transverse direction, in order to reduce the numerical complexity of the problem.

7 Examples of other possible choices: lattice unit vectors and potential vectors

The observation that equivalent physical representations can determine a different computational complexity in the simulation of nanoelectronic devices is, however, much more general. Here we describe a couple of significant examples deriving from the research experience of the author.

When a solid-state physics description of nanoelectronic devices is performed and we have to deal with a lattice of atoms, sometimes it can be necessary to individuate the point equivalent to a known one inside a particular region of the direct or reciprocal space. In this case, the choice of a particular set of unit vectors (among the infinite ones that can describe the lattice) can strongly reduce the times for this search. For example, nanoelectronic devices based on carbon nanotubes [26] have been proposed. Since a carbon nanotube is a graphene sheet rolled into a cylindrical shape, its energy bands can be derived cross-sectioning those of graphene and from the numerical point of view this is better obtained considering a rectangular region of the reciprocal space [27]. We have proved that that the position of the maxima and minima of the energy bands (those of main interest for electronic applications) are found much more quickly using a particular choice of the graphene unit vectors, in this way reducing the computational times [28].

Another example is the choice of the potential vector in the simulation of nanoelectronic devices in the presence of a magnetic field [3–5, 15]. In particular, we have examined this issue in the case of heterostructure-based electronic devices threaded by a uniform orthogonal magnetic field $\vec{B}$. If we define $x$, $y$ and $z$ as the longitudinal, transverse and orthogonal directions, respectively, we have that $\vec{B} = [0\ 0\ B]^T$. Since in the Schrödinger equation the potential vector $A$ appears, which is related to $\vec{B}$ by the correspondence $\vec{B} = \nabla \times A$, infinite choices for the potential vector exist for a single value of the magnetic field. Among them, we have considered $\bar{A} = [0\ Bx\ 0]^T$ and $\bar{A} = [-By\ 0\ 0]^T$. The simulation approach we have used has been to divide the device into a number of transversal sections inside which both the potential and the potential vector can be approximately considered longitudinally constant, solve the Schrödinger equation inside each of these slices, compute the scattering matrix of the transversal regions which go from the middle of each slice to the middle of the next one,
and compose these matrices to finally obtain the overall transmission of the device. Since, with the first choice, $\vec{A}$ depends on $x$, while this does not happen with the second choice, in the first case both the value of the magnetic field and the longitudinal variations of the potential limit the length of each slice, while in the second case the only constraint is given by the potential. On the other hand, considering the Schrödinger equation inside each section, in the first case its solution is immediate (once the solutions in the absence of magnetic field are known) [29], while in the second case it requires the solution of a computationally expensive eigenvalue problem [30]. Therefore, for potentials with fast variations in the longitudinal direction and low magnetic fields the first approach is preferable, due to the smaller amount of calculations for each section, while for potentials with slow variations in the longitudinal direction and high magnetic fields the second choice is the most convenient, since it strongly reduces the number of sections to be considered. In Fig. 4 we show the transmission of a chaotic cavity as a function of the electron energy for an orthogonal magnetic field $B = 3 \ T$. In this case, the second technique is clearly preferable, due to the presence of a large magnetic field and of a very small longitudinal variation in the potential profile. This is a clear example in which a proper choice at the modelling level can dramatically impact the efficiency of the simulation code.

8 Conclusion

We have described how proper choices in the physical model can allow to optimize the performance of computer-aided simulation codes for nanoelectronic devices, reporting examples from the personal research activity of the author. Together with an optimized writing of the code, with the use of efficient numerical routines, and with a smart approach to parallelization, this can be a key factor to obtain fast simulators while keeping the required degree of solution detail.

References:


