

Bottom-up device simulations: modeling electrical currents on the atomic scale

Anders Blom* and Kurt Stokbro

QuantumWise A/S, Lersø Parkallé 107, 2100 Copenhagen, Denmark
*anders.blom@quantumwise.com

ABSTRACT

On the nanoscale, electrical currents behave radically different compared to on the microscale. As the active regions become comparable to or smaller than the mean-free path of the material, it becomes necessary to describe the electron transport by quantum-mechanical methods instead of using classical relations like Ohm's law. Over the past decade, methods for computing electron tunneling currents in nanosized junctions have evolved steadily, and are now approaching a position where they can provide real assistance in the development of novel semiconductor materials and devices. At the same time, the industry's demand for such solutions is rising rapidly to meet the challenges both above and under the 16 nm node. In this paper we provide an overview of the current state-of-the-art of the field of how to model electrical currents on the nanoscale, using atomic-scale simulations.

Keywords: nanoelectronics, atomic-scale simulations, multi-scale, transport, phonons

1 INTRODUCTION

Electronic components have become smaller and smaller for more than four decades. Soon – if not already – the feature sizes of transistors are approaching length scales where the effects of individual atoms, impurities and vacancies start to have a real influence on the performance and reliability of the devices.

There are two separate, but closely related, factors to consider. First of all, atomic-scale defects are discrete in nature, as opposed to the continuum models traditionally used to describe the physical properties of semiconductor materials and device structures in TCAD models.

Secondly, as we approach length-scales where the active region under consideration (for instance the gate oxide barrier, which already is only a few atomic layers thick even when high- κ materials are used) is shorter than the mean-free path of the material, electrons will not travel diffusively across the barrier. Instead, the charge transport through the barrier becomes a ballistic tunneling process. The electrons no longer behave as particles, but must be described as coherent wave functions. The corresponding transmission probability and current can only be computed quantum-mechanically; Ohm's law does not hold any more, and we can not define uniquely the conductivity of the material, since it becomes bias-dependent (non-linear current–voltage relation) and is strongly influenced by the

presence of atomic-scale defects, which act as elastic scattering centers.

Inelastic scattering may of course also be present in the quantum-mechanical picture, especially when the mean-free path is comparable to the tunneling barrier thickness.

Nothing of this is new, and there has long been a growing awareness of the need for atomic-scale simulations of novel semiconductor materials. This is not least evident in the International Technology Roadmap for Semiconductors, which mentions “atomistic” or “atomic-scale” modeling at least 15 times in the latest section on Modeling & Simulation [1]. What is less clear, perhaps, is the precise way in which quantum-mechanical, atomic-scale models can be integrated with conventional EDA tools. This is a complex topic, which deserves dedicated attention and collaborative efforts between the TCAD and atomic-scale modeling communities.

In this presentation we will review the state-of-the-art in modeling of currents from a quantum-mechanical perspective, both in general and specifically as implemented in our nanoelectronics simulation software platform, ATOMISTIX TOOLKIT® (ATK), which is commercially provided to users at over 100 research institutes and corporations around the world [2]. We will place particular focus on recent developments to come closer to modeling realistic systems relevant for the semiconductor industry, both in terms of the physical effects and materials that can be considered, and the system sizes that can be simulated.

2 ATOMIC-SCALE MODELING OF ELECTRON TRANSPORT

2.1 Overview

It is interesting to note, that the two effects discussed in the Introduction, i.e. the discrete nature of the sources of scattering on the one hand, and the wave nature of electrons on the other, can be treated each by themselves with simpler models. One can use effective-mass theory to describe the electronic structure of heterojunctions with very thin materials, or semi-classical models to describe the influence of impurity scattering on electron transport. When we combine the quantum effects with the atomistic features, things get decidedly more complicated, however.

Bottom-up quantum-mechanical modeling of nanoscale structures such as nanowires, nanotubes, graphene devices, molecular junctions, and ultra-shallow dielectric junctions

have evolved over the last decade into a position where now several thousand atoms can be simulated, using methods ranging from density-functional theory (DFT) to simpler tight-binding models.

With the invention of non-equilibrium techniques, it has also become possible to compute the electronic transport properties of device-like structures under an applied finite bias [3–5]. In these models, the system geometry is generally made up of two (possibly different) electrodes, between which a semiconducting or insulating materials is inserted. The methodology is very general; the electrodes may consist of metallic nanotubes, bulk metals, nanowires, a zigzag graphene nanoribbon, or any other conducting material. The central part, or scattering region, on the other hand, can contain any atomic configuration; it may be a molecule (e.g. between metal surfaces, or adsorbed on graphene or a nanotube), another material to form a sandwich junction, a differently shaped piece of a graphene nanoribbon, etc.

This means that one can study the change in conductance of a functionalized nanotubes for use in sensor applications, field-effect transistors made of graphene nanoribbons, rectification and negative differential resistance in molecular diodes and switches, the contact resistance or Schottky barrier of a metal–semiconductor junction, and if spin is considered, the spin current or tunnel magneto-resistance, to just mention a few of many application areas.

The fundamental quantity to be computed is the transmission spectrum, which is a function of the energy of the incoming electrons from the electrode. The transmission spectrum can be integrated over energy to give the current at the given bias, and can also be decomposed and projected in different ways, resulting in supporting quantities which can assist in the analysis and understanding of the transport mechanisms.

Examples of such quantities, computable by our code, are transmission eigenvalues and eigentstate, transmission pathways (see Fig. 1), the density of states (real-space, or atom and orbital projected), the complex band structure of the electrodes, etc.

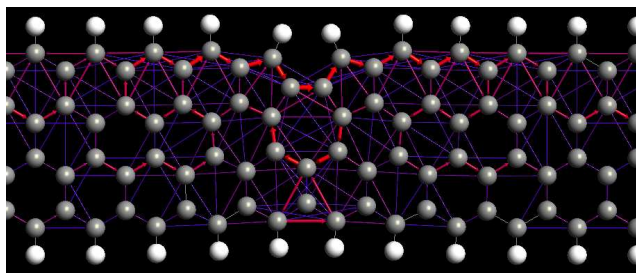


Figure 1: Transmission pathway at the Fermi level for a graphene nanoribbon with a Stone-Wales defect on one of the edges.

This understanding is one of the key advantages that can be obtained from modeling and simulation. Although one can often achieve a good qualitative agreement between experiments and computed results, the possibility to dig deeper into the mechanism is only available *in silico*. This can provide far more insight than the results themselves; insight which can be used to guide new experiments in direction of most likely success, or give rise to entirely new ideas, not tried before. Modeling does not replace experiments, but supports them.

Let us review the essential points of the models needed to compute the transmission spectrum accurately and reliably.

- *Atomistic models* are needed to describe the detailed scattering potential landscape of the central region – we already discussed this above.
- *Quantum models* are required to accurately describe the electronic structure of molecules, graphene, nanotubes, or ultrathin films. In Section 2.2 below we will discuss which particular models that are available, and their respective advantages.
- *Boundary conditions*. This is a major point that we will elaborate more on in Section 2.3.
- *Non-linear response*. In toy models or very simple systems, it may be possible to compute the transmission spectrum at zero bias and then just extrapolate to finite bias. Experience shows, however, that in the majority of systems, the transmission spectrum itself depends on the bias. Thus it is necessary to employ methods that can treat systems in the non-equilibrium situation where a finite bias is applied across the central region. This is also closely related to the boundary conditions.

An atomic-scale transport calculation can be divided into two, to a large extent separate, parts: the computation of the electronic structure, and the transfer of electrons through the system under non-equilibrium conditions when a finite bias is applied.

The transport theory used in most simulations is the well-established Landauer–Büttiker model for coherent elastic tunneling. The computation is carried out using a non-equilibrium Green’s function (NEGF) formalism, which has become the standard tool for this task.

For the electronic structure, there are however a few choices, which we will discuss in the following section.

2.2 Electronic structure methods

Ab-initio methods, to which we also count DFT, have the advantage of being able to describe any combination of elements without the need for predefined parameters. This is clearly of importance in cases where new materials are being developed, and in particular when materials are being combined in new ways, e.g. at interfaces, where the bonding character differs from bulk or molecular. It also

makes it possible to combine organic and inorganic materials.

A common concern about DFT is the difficulties related to obtaining a correct description of semiconductors. For some applications this is not really an issue (geometrical properties and total energies are often well represented by DFT) but the traditionally inaccurate estimation of the band gap does raise some questions for electronic transport properties. Recent developments of novel exchange-correlation functionals and models that allow the inclusion of on-site Hubbard terms can however cure the band gap in many situations, and alleviate the problems of interface states ending up in the conduction band, rather than in the band gap.

Tight-binding models are not in the same way transferable, meaning that parameters usually have to be tailored for each specific problem. However, they offer treatment of far larger geometries than DFT [4]. In particular, tight-binding models often only take into account interactions between adjacent atoms, and combined with routines for operating on sparse matrices, as implemented in our code, this allows for efficient handling of tens of thousands of atoms. This can not least be important for e.g. III-V materials comprised of random alloys like AlGaAs, which need to be represented by large supercells to provide statistical accuracy in terms of the material composition fluctuations.

An intermediate position is taken by semi-empirical models such as the extended Hückel method [3] and DFTB [6], where a smaller number of empirical parameters are fitted to experimental results or very accurate calculations. The computational demands are smaller than for DFT, while offering a more transferable approach than tight-binding, since one is parameterizing the atomic interactions, rather than the bonding character of the entire system.

Our code provides a uniform interface for the user to both DFT, semi-empirical and tight-binding methods, both developed in-house (ours is the only commercially engineered DFT code) and with renowned academic partners. The platform also contains a classical potential framework for extremely fast geometry optimizations, which can be carried out in the same script as the electronic transport calculation itself.

One limitation of many tight-binding and semi-empirical methods is that there is no inherent method for geometry optimizations or total energy calculations. Moreover, rather few methods include the spin degree of freedom. We are currently working on a generic framework by which spin and an inter-atomic pair potential can be added to an arbitrary – user defined – tight-binding or semi-empirical model. This will allow users to perform geometry optimizations and spin-dependent NEGF transport calculations using their own models, without the need to implement the entire transport framework for every single model separately.

2.3 Boundary conditions

There are several quantum-mechanical software packages for computing the electronic structure of atomic-scale systems. Generally, they consider one of two types of boundary conditions: periodic (bulk or otherwise infinitely extended systems) or finite (molecules). In a transport system, however, neither applies: albeit the electrodes are periodic (and are computed as such during the transport calculation), the central region is finite, and a method that can combine these two system types is required.

In addition, we must also consider the possibility to apply a finite bias across the structure, in the transport direction. This breaks any geometrical periodicity that otherwise might be present in the system.

Therefore, we need to use open boundary conditions, such that charge can flow through the system. This is where the NEGF method comes in, as it allows us to handle a non-equilibrium distribution of electrons in the central region. The entire problem must be solved self-consistently, not only for the electronic structure – as is inherent at least in DFT – but also for the charge redistribution that takes place between the electrodes, acting as reservoirs of electrons, and the central region.

The problem is highly non-linear, and convergence can be hard to achieve unless care is taken in the mixing algorithms (experience shows that mixing the Hamiltonian matrix elements is more efficient than the density matrix), and the precise way in which matrix elements are computed close the edges of the central region. This is one of the points where different implementations of the NEGF transport method differ the most. In our code, we have taken special care to include all possible contributions to the central region Hamiltonian matrix elements, also those which arise from interactions with atoms in the electrodes, so-called spill-in terms. We have also implemented a double-contour algorithm which allows for bound states to be present in the bias window, which is something that otherwise can cause big problems. This is crucial for the physical correctness of the model, and the ability to apply a realistically large bias to the structures.

For the electrostatic potential we employ Dirichlet boundary conditions on the surfaces connecting the electrodes and the central region, which allows for inclusion of the finite bias directly in the boundary conditions. This requires a multigrid method, which is considerably more time-consuming than the usual FFT methods that can be used to solve the Poisson equation under periodic boundary conditions. Moreover, in many cases the system does possess a periodic symmetry in the transverse plane, perpendicular to the transport direction.

It is therefore convenient to separate the system into two parts, and use periodic boundary conditions in the transverse plane and a multigrid method in the transport direction. This makes it easy to study interface systems using a small transverse unit cell, although for 1D or 2D systems one needs to introduce vacuum padding to avoid

spurious electrostatic and other interaction between the repeated copies of the system.

2.4 Advanced electrostatics

So far we have only discussed pure two-probe system, with source and drain electrodes. To describe transistors, however, we also need a gate. Recently, capabilities for including metallic and dielectric regions have been incorporated into quantum transport calculations (cf. Fig. 2). This enables calculations of transistor characteristics [3,4] and even makes it possible to study single electron tunneling, i.e. transport in the weak coupling or Coulomb blockade regime [7].

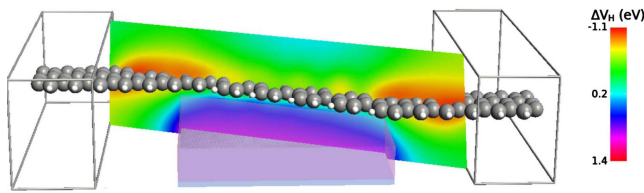


Fig. 1 The self-consistent electrostatic potential of a graphene transistor for a gate potential of 1 V. Note how the electrostatic potential from the gate is screened by the graphene device [3].

In order to mimic a realistic device geometry, it is however often necessary to place the electrostatic gates at a distance from the active device region that far exceeds the dimensions of the atomic-scale features themselves. Put shortly, the system will contain a lot of inactive space (vacuum), which becomes rather costly to simulate at the atomic-scale, not least in terms of memory. A solution here is to introduce real-space grids based on finite-element methods (FEM). This is rather straight-forward to do for methods like tight-binding where the primary real-space grid is the external potential in the Poisson equation, but decidedly harder to do for DFT where we also have to evaluate the real-space density and the exchange–correlation functional based on it. As a result of a recent PhD project [8], we plan to introduce the first ever FEM-based DFT method in the near future in our product.

There are also efforts being made for handling multi-terminal devices [9], but this is still an area for further research, as there are significantly larger challenges in obtaining stable convergence and also for the computational demands in terms of time and memory.

3 OUTLOOK

The state-of-the-art in atomic-scale quantum-mechanical calculations of electrical currents today allows for the study of systems with several hundred or several

thousand atoms, depending on the choice of electronic structure method. Properties like transistor characteristics and current–voltage relations, Schottky barriers and contact resistance can be computed for advanced systems such as high- κ interfaces or complex shapes of graphene.

There are certain elements missing from the discussion above. Partly this is because there is not room in the current format to cover all aspects; for instance, the hugely important issue of thermal transport deserves attention, and progress are being made in this area by us and others. Other areas do not yet have a satisfactory efficient solution, like how to include inelastic scattering in the general framework with DFT and NEGF, and the briefly mentioned topic of multi-terminal devices. More research is needed in these areas.

What remains to discuss is, perhaps, how accessible these methods are. The importance of having a graphical user interface (GUI) should never be underestimated, but we have also realized that it is difficult to make such an interface which fits all needs. We have therefore complemented our GUI with a Python scripting engine, which can be used not only to access low-level computational routines from an object-oriented interpreted programming language, but also allows us (or users themselves) to write plug-ins for the GUI, thus making it possible to expand its feature set to suit individual needs.

The goal and ambition of our company QuantumWise, is to support industrial users who seek to utilize the power of atomic-scale modeling to gain insight into their problems. We hope this presentation inspires further dialog on these points.

REFERENCES

- [1] International Technology Roadmap for Semiconductors (ITRS), <http://www.itrs.net/>.
- [2] Atomistix ToolKit (ATK), <http://www.quantumwise.com>
- [3] K. Stokbro, D.-E. Petersen, S. Smidstrup, M. Ipsen, A. Blom and K. Kaasbjerg, Phys. Rev. B 82, 075420 (2010).
- [4] G. Klimeck *et al.*, Comp. Model. Sci. Eng. 3, 601 (2002).
- [5] M. Brandbyge, J.-L. Mozos, P. Ordejón, J. Taylor, and K. Stokbro, Phys. Rev. B. 65, 165401 (2002).
- [6] M. Elstner *et al.*, Phys. Rev. B 58, 7260 (1998).
- [7] K. Stokbro, J. Phys. Chem. C 114, 20461 (2010).
- [8] J. Avery, Ph.D. Thesis, Copenhagen University, Feb. 2011.
- [9] K. K. Saha, W. Lu, J. Bernholc, and V. Meunier, Phys. Rev. B 81, 125420 (2010).