Current Transport in Nanoelectronic Semiconductor Devices

V. Sverdlov, H. Kosina, and S. Selberherr
Institute for Microelectronics, Technische Universität Wien
Gusshausstrasse 27–29
A-1040 Vienna, Austria

Abstract—An overview of models used for the simulation of current transport in nanoelectronic devices within the framework of TCAD applications is presented. Modern enhancements of semiclassical transport models based on microscopic theories as well as quantum mechanical methods used to describe coherent and dissipative quantum transport are specifically addressed. This comprises the incorporation of quantum correction and tunneling models up to dedicated quantum-mechanical simulators, and mixed approaches which are capable to account for both, quantum interference and scattering. Specific TCAD requirements are discussed from an engineer’s perspective and an outlook on future research directions is given.

1. INTRODUCTION

The breathtaking increase in computational power and speed of integrated circuits in the past decades has been supported by the aggressive size reduction of semiconductor devices. This trend is expected to continue in the coming decade as predicted and institutionalized by the International Technology Roadmap for Semiconductors [1]. Today, when the 90 nm technology node with physical transistor gate lengths in the range of 40 nm is in mass production, the challenge is to introduce the 65 nm technology node already in a year. A new technology node is introduced every 3 years, with a long-term projection of the 22 nm node to be in mass production by the year 2016. A possibility to build metal-on-insulator field effect transistors (MOSFETs) with even shorter gate lengths has been successfully established after the 6 nm gate length transistor has been demonstrated in research labs [2], [3]. From a theoretical viewpoint even a few nm gate length device has been predicted to be functional [4], [5]. Nevertheless, emerging outstanding technological challenges related to different aspects of MOSFET fabrication and reliability in mass production, as well as the rapidly increasing power dissipation may slow down the so far exponential scaling of Complimentary MOSFETs (CMOS). Besides, with the ongoing search for new technological solutions vital for CMOS downsizing, developing conceptually new devices and architectures is becoming increasingly important. New nanoelectronic structures, such as carbon nanotubes, nanowires, and molecules, are considered to be the most prominent candidates for the post-CMOS era. Since conventional MOSFETs are already operating in the sub-100 nm range, new nanoelectronic devices are expected to complement and substitute some of the current CMOS functions after being integrated into CMOS technology.

Technology CAD (TCAD) tools are designed to assist in development and engineering at all stages ranging from process simulation to device and circuit optimization. The main purpose of TCAD is the technology-development related cost reduction which currently amounts to 35% and is expected to rise to 40%, according to ITRS [1]. Due to the aggressive downscaling of CMOS device feature sizes and newly emerging nanoelectronic devices, various shortcomings of presently applied TCAD tools appear. These tools are frequently based on semiclassical macroscopic transport models. From an engineering point of view, classical models like the drift-diffusion model, have enjoyed an amazing success due to their relative simplicity, numerical robustness, and the ability to perform two- and three-dimensional simulations on large unstructured grids [6]. Hot-carrier effects have motivated the development of higher-order transport models such as the hydrodynamic, energy-transport and six-moments models [7]. However, inaccuracies originate from the non-local nature of carrier propagation in ultra-scaled devices [8].

Non-local effects may be of classical or quantum-mechanical nature, depending on the underlying physics relevant to the transport process. Classical non-localities appear when the mean-free path is comparable to the device feature size. Quantum mechanical non-local effects start to determine the transport properties when the device size is significantly smaller than the De-Broglie electron wave length. Size quantization of carrier motion in inversion layers of MOSFETs and in ultra-scaled multi-gate devices as well as the tunneling current, including the gate leakage current, are the most important examples of quantum effects in MOSFETs.

Fig. 1 shows the hierarchy and mutual interrelation of models currently used for the description of current transport. Semiclassical transport models are based on the Boltzmann equation which includes scattering integrals describing realistic microscopic processes. These semiclassical models, augmented with quantum corrections, are still of great importance due to their relative computational simplicity, numerical stability, and an ability to provide reasonable quantitative results within seconds even for devices with gate length as short as 50 nm. A brief overview of the currently developed semiclassical transport models will be presented in Section 2.

Quantum ballistic transport models describe a coherent propagation of carriers. They are based on the solution of the Schrödinger equation for the wave function, supplemented with the corresponding boundary conditions. This approach is efficient and provides accurate results when carrier scattering is irrelevant and can be neglected. The method will be illus-
Finally, dissipative quantum transport theory represents the most complete description of transport, which combines the coherent carrier motion between the scattering events with coherence (or phase) breaking due to carrier scattering. Different formalisms are currently used, based on the Dyson equation for the non-equilibrium Green’s functions, the Liouville/Von-Neumann equation for the density matrix, or the Wigner transport equation. Section 4 deals with quantum transport characterized by both scattering and quantization. A conclusion will summarize the main findings and give directions for future research.

II. SEMICLASSICAL TRANSPORT
After the ground-breaking work of Scharfetter and Gummel [10], who first proposed a robust discretization scheme for the drift-diffusion equation, computer programs like MINIMOS [11] and PISCES [12] played a pioneering role in numerical simulation of current transport properties of semiconductor devices. Since then, numerous transport models of increasing complexity have been introduced. The semiclassical transport description is based on the Boltzmann equation for the distribution of carriers $f(r,k,t)$ in the phase space. The Boltzmann equation includes carriers’ scattering with phonons, impurities, interfaces, and other scattering sources through the corresponding collision integrals. Although the solution of the Boltzmann equation can be found numerically by means of Monte Carlo (MC) methods, TCAD models based on moments of the distribution function are highly desirable. Being computationally significantly less expensive than the MC method, these higher-order moments’ methods provide a reasonable quantitative answer for devices as short as 50 nm within seconds. The fairly new six moments model [7] based on non-Maxwellian distribution takes naturally into account the hot-carrier effects such as avalanche generation, hot carrier induced gate currents, or hot-carrier diffusion, which typically take place in Silicon-On-Insulator (SOI) floating body MOSFETs. For the purpose of calibration the full-band MC method is often accepted, since it can precisely account for the various scattering processes [13].

Another important development of transport models is related to the MC methods for solving the Boltzmann equation. After the pioneering work of Kurosawa in 1966 [14], who was the first to apply the MC method to simulate carrier transport in semiconductors, the significantly improved MC method was successfully applied to transport description in a variety of semiconductors [15]. For electrons in silicon, the most thoroughly investigated case, it is believed that a satisfactory understanding of the band structure and of the basic scattering mechanisms has been achieved giving rise to a “standard model” [16]. Nowadays, an accurate MC evaluation of carrier transport properties in inversion layers is of primary importance for predicting performance of modern CMOS bulk devices. Due to the strong confinement of carriers in the inversion layer of bulk MOSFETs or due to the geometric confinement in multi-gate FETs the carrier motion is quantized in one or two confinement directions giving rise to the formation of subbands. One possibility to address the effect of quantum confinement on the electron concentration is to use an effective potential. This can be achieved by a convolution of the electrostatic potential with a Gaussian function, which leads to a smoothing of the original potential [17], [18], [19]. Another option is to use the self-consistent Poisson-Schrödinger-based quantum corrected potential [20], [21], which suppresses the carrier concentration close to the interface, mimicking the real quantum-mechanical behavior. These approaches combine advantages of full-band structure and flexibility of scattering processes of three-dimensional classical MC simulations with the generality of material composition and transport peculiarities due to quantum confinement and may also address the strain effects.

The MC approach may incorporate the quantized carrier motion in the direction orthogonal to the current exactly. The quantum-mechanical motion of carriers in the confined direction is addressed by the self-consistent solution of the corresponding Schrödinger and Poisson equation, leading to
the formation of subbands. The carrier motion within each subband may still be considered semiclassical and therefore can be well described by the corresponding Boltzmann equation written for the subband distribution function $f_n(r, k, t)$. Because of possible carrier transitions between different subbands due to scattering, the collision integrals on the right-hand-side of the Boltzmann equation should include the terms responsible for the intersubband scattering processes. The transport in the inversion layer of a MOSFET is finally described by a set of Boltzmann equations for every subband, coupled to each other via the intersubband scattering integrals. The set of the subband Boltzmann equations for $f_n(r, k, t)$ is conveniently solved by a MC method. This approach therefore combines the advantages of a quantum description in confinement direction with a semiclassical description in transport direction and represents a transition between semiclassical and quantum-mechanical pictures. An example of the simulation of the low-field surface mobility in inversion layers of silicon, when the transport in the current direction may be treated semiclassically is shown in Fig. 2, together with the experimental "universal mobility" curve [22]. In order to reproduce the universal mobility curve, up to 40 unprimed and 20 primed subbands formed at a (100) silicon interface were taken into account, with realistic electron-phonon and surface roughness scattering included [23].

III. QUANTUM-BALLISTIC TRANSPORT

With the aggressive downscaling of MOSFET dimensions continuing, the classical description of carrier motion in transport direction is gradually loosing its validity. When the characteristic scale of the potential variation along the channel is comparable to the De-Broglie wave length of a carrier, a TCAD transport model must include the quantum effects in transport direction. If scattering processes can be ignored and particle propagation in the device is coherent, the carrier motion is determined by the solution of the Schrödinger equation, supplemented with open boundary conditions. In order to determine the current density $J$, it is enough to know the transmission coefficient $TC(E_x)$ as well as the supply function $N(E_x)$ from the electrodes [24]:

$$J = \frac{4\pi m_{eff} q}{\hbar^3} \int_{E_{min}}^{E_{max}} TC(E_x) N(E_x) dE_x.$$  (1)

A similar approach can also be used to determine the gate leakage current [25]. The solution of the Schrödinger equation with open boundary conditions can be achieved by means of the quantum transmitting boundary method [26], [27]. An established alternative framework for these calculations is the non-equilibrium Green’s Function method [28] in its reduced coherent version. It is conveniently used for one-dimensional studies of resonant tunneling diodes or carbon nanotubes. Simulators accounting for a full two-dimensional solution of the open-boundary Schrödinger equation have been reported and applied to the simulation of 10 nm double-gate MOSFETs [29], [30].

It may appear that in the quantum-ballistic case the determination of the full wave function as a solution of the Schrödinger equation is not necessary and the knowledge of the transmission coefficient is enough for the current calculations. In the contact block reduction method [31] the transmission function is fully determined by the reduced contact part of the full Green’s function. However, the carrier concentration alters the electrostatic potential in the device via the Poisson equation. The carrier concentration is proportional to the square of the wave function, implying that the accurate determination of the transmission coefficient and therefore the
current requires a self-consistent solution of the Schrödinger and Poisson equation simultaneously. For quasi-one-dimensional transport this can be achieved straightforwardly [4]. An example of the output characteristics simulated for an ultra-thin body double-gate MOSFET with a gate length \( L \) as short as 2.5 nm is shown in Fig. 3. Surprisingly, even such a small device possesses an \( I_{on}/I_{off} \) ratio sufficient for logic applications and displays a reasonable short-channel effect and acceptable DIBL, a conclusion recently reached from more detailed atomistic calculations [5]. It should be noted that the sensitivity to small MOSFET dimension variations, control of doping as well as the whole manufacturing process development represent significant challenges for multi-gate MOSFETs with gate length below 10 nm.

Self-consistent solution of the two- or three-dimensional Schrödinger equation together with the Poisson equation represents a significant computational challenge [29]. Two- and three-dimensional quantum ballistic simulations can be performed by means of an approximate separation of the quantum motion in the confinement direction \( y \) from the motion along the current direction \( x \) by means of the following ansatz for the wave function \( \Psi_n(y, x) \):

\[
\Psi(x, y) = \sum_n \xi_n(x) \psi_n(y, x). \tag{2}
\]

This method allows the independent solution of the Schrödinger equation for the subband wave function \( \psi_n(y, x) \) at position \( x \). Transport in the current direction is characterized by a system of one-dimensional Schrödinger equations with open boundary conditions for the wave functions \( \xi_n(x) \). Each Schrödinger equation describes the transport inside the particular subband. Transport in each subband is independent from the one in other subbands, if the subband wave functions \( \psi_n(y) \) do not depend on the position \( x \) in transport direction. The Schrödinger equations describing the transport in each subband are decoupled from each other, when the potential \( U(x, y) \) in the device is the sum of two contributions, each depending either on \( y \) or \( x \) coordinate alone. In a general case when the subband wave functions depend on the position \( x \) in transport direction, the transport in the subbands \( n \) and \( m \) is coupled, with the coupling described by the Hamiltonian \( \delta H_{nm}(x) \). However, when the intersubband coupling Hamiltonian \( \delta H_{nm}(x) \) is small and may be neglected, transport in the subbands can still be considered as independent from each other. This approximation simplifies the calculations and reduces the computational effort significantly [32]–[35]. The coupling Hamiltonian is expected to be small if the dependence of the subband wave function on \( x \) is weak. An example where the subband decomposition turns out to be an excellent approximation is the quantum transport in ultra-scaled SOI MOSFETs [35]. In the opposite limit of abrupt junctions between contact reservoirs and the channel, the intersubband coupling is expected to be the strongest. However, even in this case the current value calculated self-consistently was found to be only 10% lower as compared to the calculations with neglected intersubband coupling [36]. More study is needed to clarify the situations.

The coherent quantum transport description is justified if the size of the channel region is shorter that the phase-coherence length. In carbon nanotubes, where elastic scattering can be ignored and inelastic scattering has little effect on current [39], transport can be considered to be coherent and is therefore well described within the quantum-ballistic approach [9]. Similar methods can be applied to describe the output characteristics of FinFETs in the ballistic approximation [40]. In silicon MOSFETs, however, the mean-free path in the area close to the potential maximum at 300 K is only a few nm [41], and the full quantum description which includes dissipative processes must be adopted to simulate MOSFETs with gate length of around 10 nm. A consistent introduction of realistic scattering into simulators based on the coherent description alone creates outstanding computational difficulties ranging from a necessity to invert huge matrices in NEGF formalism [28] to calculations of nonlocal scattering rates in Pauli master equation approaches [42]. Besides the difficulties of introducing scattering into the simulators based on the coherent description, these simulators are often limited to specific geometries, grids and short length scales, which makes their integration into modern engineering TCAD tools problematic. Nevertheless, these simulation approaches are necessary for the estimation of upper bounds of current transport at the quantum limit.

IV. DISSIPATIVE QUANTUM TRANSPORT

The methods described so far are either based on the assumption of semiclassical or pure quantum ballistic transport. Nevertheless, in modern microelectronic devices quantum effects are usually dominant in a small active region connected to large, heavily doped contact areas where the carrier dynamics is essentially classical. Therefore, modern TCAD simulators should be able to incorporate both classical and quantum-mechanical modeling approaches on equal footing.
To a certain extent, various quantum corrections can account for the quantum effects, as already discussed.

The non-equilibrium Green’s functions method addresses the problem in the most consistent and complete way. Due to its completeness, the method is computationally complex and is usually applied to one-dimensional problems [28] and for a restricted set of scattering mechanisms [43] only. The carbon nanotube (CNT) FET which is widely considered to be a potential alternative to the conventional MOSFETs, represents a good example where the nonequilibrium Green’s function method provides accurate results and is successfully used. Simulated output characteristics of a CNT-FET with ohmic contacts [37] are compared in Fig. 4 to experimental data [38], showing good agreement.

An alternative approach which handles both quantum-mechanical and dissipative scattering effects is based on the Wigner function formalism. Realistic scattering processes can be easily embedded into the Wigner equation via Boltzmann-like scattering integrals. The Wigner function is given by the density matrix in mixed representation [44], [45]. The kinetic equation for the Wigner function is similar to the Boltzmann equation, with the classical force term replaced by a non-local potential operator. The Wigner function formalism treats scattering and quantum mechanical effects on equal footing through the corresponding scattering integrals. It is attractive to use the well established scattering models used in classical MC simulations and solve the quantum Wigner transport equations by means of the MC technique. Such programs were recently realized in [35], [46], [47]. However, since the kernel of the quantum scattering operator is not positively defined, the numerical weight of a particle trajectory increases rapidly, and the numerical stability of a trajectory-based MC algorithm becomes a critical issue. A multiple trajectories method was recently suggested [47] in order to overcome the difficulty. In this algorithm the problem of a growing statistical weight of a single trajectory is addressed by creating an increasing number of trajectories with constant weights, which may assume positive and negative values. Being formally equivalent to the former method, the algorithm allows the annihilation of trajectories with similar statistical properties, introducing a possibility to control the number of trajectories. This approach has the advantage that a seamless transition between classical and quantum-mechanical regions in a device is possible [47]. Following [48], one can introduce a spectral decomposition of the potential profile \( V(x) \) into a slowly varying, classical component and a rapidly varying, quantum mechanical component. This decomposition is conveniently carried out by applying a low-pass filter with a cut-off wave number \( q_c \ll \pi/\Delta x \), where \( \Delta x \) is a grid step size. This separation of the total potential into a classical and a quantum mechanical contribution significantly improves the Wigner Monte Carlo convergence. The method can be applied to the simulation of resonant tunneling diodes [47], and it was recently used for the simulation of double-gate MOSFETs [35]. An example of the carrier concentration simulated with the Wigner function method is shown in Fig. 5, in comparison with the classical result.

![Fig. 5. Comparison of Wigner and classical Monte Carlo results for electron concentration in double-gate MOSFETs, for different channel lengths [35].](image)

V. CONCLUSIONS
Well established classical TCAD tools are gradually losing their ability to predict accurately the characteristics of nanoscale devices, prompting for enhancement to meet the engineering demands. Classical models using higher moments are able to include the hot-carrier effects and can reproduce results of the full-band MC, while relevant quantum corrections may be incorporated into different MC schemes.

Full quantum description is required for nanoscale devices. Contrary to the carbon nanotubes, where the transport properties can be well predicted within the coherent picture, a dissipative quantum description may be required for transport calculations in ultra-scaled MOSFETs with the gate lengths ranging around 10 nm. One option is the Wigner function approach which naturally combines the advantages of quantum description with the accurate scattering models relevant for devices in the nanoscale range. All quantum-mechanical models must be adapted for engineering applications for which timely results are more valuable than accurate analyses [49]. Device simulators must allow a coupling with process simulators, since a detailed, physics-based transport model may be misleading if geometry and doping are not described correctly. Support of unstructured grids is necessary in order to optimize the simulation time. Furthermore, the simulators should be general-purpose and not limited to specific geometries or simulation models. It is still not clear which of the outlined quantum transport approaches will finally be integrated into TCAD environments. Its further success will depend on the ability to model quantum effects efficiently with reasonable accuracy.

ACKNOWLEDGMENTS
We gratefully acknowledge financial support from the Austrian Science Fund FWF, project P17285-N02, and the European Commission, project SINANO IST-506844.