Current Transport Models for Nano-Scale Semiconductor Devices

Andreas Gehring and Siegfried Selberherr
Institute for Microelectronics, TU Vienna
Gussenhofenstr. 27–29, A-1040 Vienna, Austria
E-mail: Selberherr@TUWien.ac.at

ABSTRACT

An overview of models for the simulation of current transport in micro- and nanoelectronic devices within the framework of TCAD applications is presented. Starting from macroscopic transport models, currently discussed enhancements are specifically addressed. This comprises the inclusion of higher-order moments into the transport models, the incorporation of quantum correction and tunneling models up to dedicated quantum-mechanical simulators, and mixed approaches which are able 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.

Keywords: Device Simulation, TCAD, Transport Models

1. INTRODUCTION

The continuous minimum feature size reduction of microelectronic devices, institutionalized by the ITRS roadmap [1], has been partly enabled by the support of sophisticated Technology CAD (TCAD) tools. These tools promise to assist process and device engineers during all stages of development, ranging from process simulation to device and circuit simulation. Today, device engineers face the challenge to move from the microelectronic feature scale in the mid-90’s, with typical MOSFET gate lengths just entering the sub-micron region, to the realm of nanoelectronics with 90 nm gate length devices in volume production and 6 nm gate length transistors fabricated in research labs [2]. The continuum approximation, already questioned in the mid-1990’s, has to be abandoned in this regime, and different approaches for the simulation of devices in the nanometer regime have been proposed.

In general, the inaccuracies of presently applied semiclassical macroscopic transport models are due to non-local effects [3], either caused by classical or quantum-mechanical non-localities. Classical non-localities arise because the distribution of electrons in very small devices does not depend on local quantities alone. Quantum-mechanical non-localities occur due to the wave nature of electrons and the occurrence of quantization, either due to high electric fields as in the inversion layer of a MOSFET, or due to the geometry as in ultrasmall double-gate or FinFET devices. Fig. 1 depicts the hierarchy of models which are currently used for the description of current transport. Semiclassical transport models rely on classical states characterized by a distribution function which is governed by the Boltzmann transport equation. In Section 2 we will give a review of the evolution of current semiclassical transport models, and present recent results achieved with higher-order transport models. Quantum ballistic transport is based on pure states described by a wave function, the evolution of which follows Schrödinger’s equation. These approaches are mainly used for the simulation of closed systems, such as quantum corrections in the inversion layer of MOSFETs. In Section 3, these quantum-ballistic transport approaches will be described. Finally, quantum transport theory deals with mixed states. There exist different formulations, which can be based on the Dyson equation, the Liouville/von Neumann equation, or the Wigner transport equation. Section 4 deals with these approaches, which are characterized by both scattering and quantization. A conclusion will summarize the main findings and give directions for future research.

2. SEMICLASSICAL TRANSPORT

In the early days of semiconductor technology, the electrical characteristics of semiconductor devices could be estimated based on simple analytic compact models, employing a variety of simplifying approximations but capturing the basic physical principles of carrier transport. These models were based on the drift-diffusion (DD) formalism, where the current in the device is governed by the electric field and the concentration gradients alone. Based on the ground-breaking work of Scharfetter and Gummel [4], who first proposed a robust discretization scheme for the drift-diffusion equations, the numerical simulation of semiconductor devices was enabled. Computer programs such as Minimos [5] and Pisces [6] have been developed and played a pioneering role in the deeper understanding of current transport for engineering purposes and in the development of miniaturized devices. For the first time, it was possible to provide insight into the functioning of semiconductor devices by means of the distribution of internal device quantities, instead of global quantities such as current-voltage characteristics. Since then, numerous transport models of increasing complexity have been proposed. All models are coupled to the Poisson equation

$$\nabla \cdot (\kappa \nabla \phi) = \rho(\phi), \quad \rho(\phi) = q(n - p - C) \quad (1)$$

where $\phi$ denotes the electrostatic potential and $\kappa$ the dielectric permittivity. The question of current transport ba-
sically reduces to the self-consistent modeling of the non-linear charge density \( \rho(\phi) \) in (1), which includes the electron and hole concentration, the net concentration of impurities, and other charges such as ionized traps.

Neglecting the quantum-mechanical nature of electrons, carrier transport in a device is described by Boltzmann’s transport equation, a seven-dimensional integro-differential equation in phase space [7]

\[
\frac{\partial f}{\partial t} + v \cdot \nabla_r f + \frac{s_n q E}{\hbar} \cdot \nabla_k f = \left( \frac{\partial f}{\partial t} \right)_{\text{coll}} .
\]

(2)

Here, \( f(\mathbf{r}, \mathbf{k}, t) \) is the distribution of carriers in space (\( \mathbf{r} \)), momentum (\( h\mathbf{k} \)), and time. The charge sign \( s_n (\nu = n, p) \) distinguishes between electrons and holes. On the right-hand side stands the collision operator which describes scattering of particles due to phonons, impurities, interfaces, or other scattering sources. However, for realistic structures, the direct solution of this equation is computationally prohibitive. It is rather solved by approximate means applying the method of moments or using Monte Carlo methods.

### 2.1. The Method of Moments

In the method of moments each term of (2) is multiplied with a weight function and integrated over \( k \)-space. This yields a set of differential equations in the (\( r, t \))-space. The moments of the distribution function are defined as

\[
\langle \Phi \rangle = \frac{1}{4\pi^2} \int \Phi f(\mathbf{r}, \mathbf{k}, t) \, d^3k .
\]

(3)

This generates an infinite set of equations which must be closed by a suitably chosen ansatz [8]. Closure after the second moment and assuming a cold Maxwellian distribution leads to the drift-diffusion equations, which for electrons read

\[
\nabla \cdot \mathbf{J}_n = qR + q \frac{\partial n}{\partial t} ,
\]

(4)

\[
\mathbf{J}_n = qn \mu_n \mathbf{E} + qD_n \nabla n .
\]

(5)

In these equations \( \mathbf{J}_n \) denotes the current density, \( R \) the net recombination rate, \( \mu_n \) the mobility, \( \mathbf{E} \) the electric field, and \( D_n \) the diffusion coefficient. Together with (1), a coupled equation system is formed which is solved numerically by means of the box integration method.

From an engineering point of view, the drift-diffusion model has proven amazingly successful due to its efficiency, numerical robustness, and the feasibility to perform two- and three-dimensional studies on fairly large unstructured grids. However, several shortcomings of this model are critical for miniaturized devices. Especially hot-carrier effects such as impact ionization or velocity overshoot motivated the development of higher-order transport models such as the hydrodynamic, energy-transport, and six-moments model [9]. These models allow the electron energy distribution function to be described beyond the Maxwellian approximation, and they are used routinely in commercial and academic device simulators.

### 2.2. The Monte Carlo Method

The Monte Carlo method is well established for studying semiconductor devices and exploring semiconductor properties. The method simulates the motion of charge carriers in the six-dimensional phase space formed by position and momentum. Subjected to the action of an external force field, the point-like carriers follow trajectories governed by Newton’s law and the carrier’s dispersion relation. These drift processes are interrupted by scattering events which are assumed local in space and instantaneous in time. The duration of a drift process, the type of scattering mechanism, and the state after scattering are selected randomly from given probability distributions which are characteristic to the microscopic scattering process. As a calibration tool, the full-band Monte Carlo method has gained widespread acceptance, since it precisely accounts for the band structure of the considered semiconductor [10]. Fig. 2 shows a comparison of different macroscopic simulation approaches with full-band Monte Carlo results for a 250 nm and a 50 nm double-gate MOSFET [11]. It can be seen that transport models based on two, four, and six moments deliver similar results for the long-channel device, while only the six moments model is able to reproduce the full-band Monte Carlo results for the short-channel device. Applying the energy-transport model in this case yields almost no improvement over the drift-diffusion model.

The Monte Carlo method has been enhanced to account for quantum effects using an effective potential instead of the self-consistent potential determined by the Poisson equa-
Double Gate MOSFETs

Figure 2: Comparison of macroscopic transport models with full-band Monte Carlo. While all models yield similar results at large gate lengths, only the six-moments model reproduces the short-channel Monte Carlo results.

The effective potential can be obtained by a convolution of the electrostatic potential with a Gaussian function which leads to a smoothing of the original potential. A quantum correction based on the Schrödinger equation applied to a full-band Monte Carlo simulator is reported in [14].

3. QUANTUM-BALLISTIC TRANSPORT

Within macroscopic transport models presented above, quantum-mechanical effects are usually accounted for by means of quantum corrections. The fabrication of structures in the nanometer regime makes this approach questionable and motivated the development of quantum-mechanical modeling tools which calculate the carrier concentration by purely quantum-mechanical principles. They became especially important for the evaluation of gate dielectrics, which represent the smallest feature scale in microelectronics. Neglecting quantum confinement in this regime leads to results which are not just slightly inaccurate, but systematically wrong. As an example, the CV-characteristics of an 1.5 nm dielectric layer is shown in Fig. 3 for different poly doping concentrations calculated classically and quantum-mechanically, showing a large discrepancy. One-dimensional closed-boundary quantum simulators are today established tools for the characterization of gate dielectric layers [15]–[17]. Such one-dimensional solutions of the Schrödinger equation are also frequently used to derive correction factors for the carrier concentration calculated by macroscopic transport models [18]–[20]. They can be used to yield a quick estimate of quantum-confinement related effects without degrading the efficiency of the device simulator used. However, since they are based on the closed-boundary Schrödinger equation charge transport is neglected. Thus, they are not applicable to open systems characterized by thin or low energy barriers, which give rise to complex eigenvalues and, as a consequence, charge transport by tunneling.

3.1. Tunneling Models

Quantum-ballistic tunneling models are still predominantly applied for the simulation of gate leakage in CMOS devices. Here, the central quantity is the quantum-mechanical transmission coefficient $T_C(\varepsilon)$ which is used in the so-called Tsu-Esaki equation

$$J = \frac{4\pi n_{\text{eff}} q}{h^3} \int_{E_{\text{min}}}^{E_{\text{max}}} T_C(\varepsilon_x) N(\varepsilon_x) \, d\varepsilon_x$$

Figure 3: Comparison of CV characteristics of a 1.5 nm dielectric layer with different polysilicon doping applying classical and quantum-mechanical simulations.
### 3.2. Adiabatic Decomposition

The idea of adiabatic decomposition is demonstrated in the following for a two-dimensional structure with an electrostatic potential \( V(x, y) \) [27]. Here it is assumed that \( y \) denotes the direction normal to a channel and that current is predominantly flowing in \( x \) direction. The starting point is the two-dimensional Schrödinger equation in the \((x, y)\) simulation domain

\[
-\frac{\hbar^2}{2} \left( \frac{1}{m_x} \frac{\partial^2}{\partial x^2} + \frac{1}{m_y} \frac{\partial^2}{\partial y^2} \right) \psi(x, y) + V(x, y) \psi(x, y) = E \psi(x, y)
\]

With respect to the \( y \)-coordinate the wave function \( \psi(x, y) \) is now expanded in a series

\[
\psi(x, y) = \sum_n \phi_n(x) \zeta_n(y; x)
\]

The basis \( \zeta_n \) is obtained from a solution of the one-dimensional Schrödinger equation in transverse direction at some given lateral position \( \bar{x} \)

\[
-\frac{\hbar^2}{2m_y} \frac{\partial^2}{\partial y^2} \zeta_n(y; \bar{x}) + V(\bar{x}, y) \zeta_n(y; \bar{x}) = \epsilon_n(\bar{x}) \zeta_n(y; \bar{x})
\]

The eigenvalue \( \epsilon_n(x) \) gives the position-dependent energy of the \( n \)-th subband. The expansion coefficients \( \phi_n(x) \) are determined by a coupled system of Schrödinger equations

\[
-\frac{\hbar^2}{2m_x} \frac{d^2}{dx^2} \phi_n(x) + (\epsilon_n(x) + A_{nn}) \phi_n(x) + \sum_{m \neq n} \left( A_{nm}(x) \phi_m(x) + B_{nm}(x) \frac{d \phi_m}{dx} \right) = E_n \phi_n(x)
\]

with coupling coefficients \( A_{nm} \) and \( B_{nm} \)

\[
A_{nm}(x) = -\frac{\hbar^2}{2m_y} \int \zeta_n(y; x) \frac{\partial^2 \zeta_m(y; x)}{\partial y^2} \, dy
\]

\[
B_{nm}(x) = -\frac{\hbar^2}{2m_y} \int \zeta_n(y; x) \frac{\partial \zeta_m(y; x)}{\partial y} \, dy
\]

In the adiabatic approximation the coupling terms are neglected. The problem simplifies to a solution of decoupled one-dimensional Schrödinger equations for each subband.

The adiabatic decomposition can typically be applied to the channel of an FET, where one-dimensional confinement is due to the high electric field below the gate. The method can also be applied for situations in which two-dimensional confinement takes place, as in the channel of ultrathin FinFET structures. In this case, a set of decoupled two-dimensional Schrödinger equations is solved. Fig. 4 depicts a cross-section through the channel of different multi-gate silicon-on-insulator devices, namely a FinFET (top) and a IIG-FET (bottom) [28]. Three-dimensional device simulations have been performed for turned-off devices \( (V_{DS}=1.0 \, V, V_{GS}=0.0 \, V) \) by means of coupling a two-dimensional Schrödinger-Poisson solver to the device simulator MINIMOS-NT [29]. In this simulation, the adiabatic decomposition approach was used to calculate a quantum-mechanical correction potential which was inserted into the drift-diffusion model. The resulting carrier concentration shows typical quantum-mechanical features such as the peak in the middle of the channel and a reduced concentration at the insulator interfaces. In the turned-off case, the carrier concentration in the channel should be totally suppressed. The simulations show that the IIG-gate FET efficiently shields the channel from the drain bias as compared to the standard FinFET structure [30].

### 3.3. Multi-Dimensional Schrödinger Solvers

Recently, simulators accounting for a full two-dimensional solution of the open-boundary Schrödinger equation have been reported and applied to the simulation of double-gate MOSFETs [31], [32]. Besides the requirement for a fine and sometimes even equidistant mesh, a main obstacle in these approaches is that the treatment of scattering is not straightforwardly possible. Furthermore, these simulators are usually limited to specific geometries, restrictive grids, or small length scales, which makes their usability for engineering applications questionable. Nevertheless, these simulation approaches are necessary for the estimation of upper bounds of current transport at the quantum limit.
4. QUANTUM TRANSPORT

The methods described so far are either based on the assumption of pure classical or pure quantum transport. Modern microelectronic devices, however, are characterized by the transition between large reservoirs with strong scattering, and small regions where quantum effects may dominate. A rigorous approach to account for both effects is based on the Wigner function, which is given by a transformation of the density matrix [33], [34].

\[ f_w(r, k, t) = \int \rho \left( r + \frac{s}{2}, r - \frac{s}{2}, t \right) \exp(-ik \cdot s) \, ds. \]

The kinetic equation for the Wigner function is the Wigner transport equation which is similar to the Boltzmann equation except the Wigner potential at the right-hand side

\[ \left( \frac{\partial}{\partial t} + v \cdot \nabla_r + \frac{qE}{h} \cdot \nabla_k \right) f_w = \int V_w(r, k - k') f_w(k', r, t) \, dk' + \left( \frac{\partial f_w}{\partial t} \right)_{\text{coll}}. \]

The Wigner potential is defined by

\[ V_w(r, k) = \frac{1}{i\hbar (2\pi)^3} \int \left( V \left( r + \frac{s}{2} \right) - V \left( r - \frac{s}{2} \right) \right) \exp(-ik \cdot s) \, ds. \]

From this equation the quantum drift-diffusion or quantum hydrodynamic models can be derived applying the method of moments [35]. It is therefore more suitable for the implementation in device simulators than a Schrödinger-Poisson solver which introduces strong non-localities. However, it was reported that, while the carrier concentration in the inversion layer of a MOSFET can be modeled correctly, the method fails to reproduce tunneling currents [36]. Therefore, strong efforts have been undertaken to apply the most accurate classical device simulation approach, the Monte Carlo technique, to the Wigner transport equation.

Implementations of Monte Carlo methods for solving (13) have been reported [37], [38]. Monte Carlo methods allow scattering processes to be included on a more detailed level, as compared to the finite-difference method [39] which is practically limited to a one-dimensional momentum space and the relaxation time approximation. Unlike classical distribution functions, however, the Wigner function permits positive and negative values. Therefore, it cannot be interpreted as a probability distribution function, a peculiarity known as the negative sign problem. Instead, the Wigner function can be modeled as the difference of two positive functions which describe in-scattering and out-scattering of particles. This approach has the advantage that it allows for a seamless transition between classical and quantum-mechanical regions in a device [38]. This method has been applied to the simulation of resonant tunneling diodes as shown in Fig. 5 and it was recently used for the simulation of 10 nm double-gate MOSFETs [40].

\[ \text{Electron concentration \left[cm}^{-3}\right] \]
\[ \text{Mean energy} \]
\[ \text{Distance [nm]} \]

Figure 5: Wigner Monte Carlo results of electron concentration and mean energy for a resonant tunneling diode.

5. CONCLUSIONS

Semiconductor physics is a vast field and simulation approaches abound. Physicists are often tempted to use overly complicated approaches, in an understandable effort not to lose the important physics. However, some constraints for engineering application should be kept in mind. Models must be efficient: Timely results are often more valuable than accurate analyses [41]. There is a need for three-dimensional simulations, even if they are only rarely applied to check for spurious effects. Device simulators must allow a coupling with process simulators, since a detailed, physics-based transport model is of no use if geometry and doping are not described correctly. Therefore, support of unstructured grids is necessary. Furthermore, the simulators should be general-purpose and not limited to specific geometries or simulation modes. It is still not clear which of the outlined quantum transport approaches will find its way into integrated TCAD environments, but its further success depends on efficient and accurate modeling of these new effects.

ACKNOWLEDGEMENTS

We gratefully acknowledge support from Hans Kosina and Tibor Grasser, as well as discussions with Gerhard Klimeck and Peter Vogl.

REFERENCES
