Current transport models for nanoscale semiconductor devices — **Source link** ⧉

Viktor Sverdlov, E. Ungersboeck, Hans Kosina, Siegfried Selberherr

Related papers:

- Transport modeling for nanoscale semiconductor devices

- CMOS design near the limit of scaling

- Ab initio modeling of quantum transport properties of molecular electronic devices

- Linear combination of bulk bands method for investigating the low-dimensional electron gas in nanostructured devices

- Methods for manufacturing photolithography masks utilizing interfering beams of radiation

Current transport models for nanoscale semiconductor devices

V. Sverdlov [*], E. Ungersboeck, H. Kosina, S. Selberherr

Institute for Microelectronics, TU Vienna, Gusshausstr. 27–29, A-1040 Vienna, Austria

Abstract

Due to the rapid decrease in device dimensions the well-established TCAD tools are pushed to the limits of their applicability. Since conventional MOSFETs are already operating in the sub-100 nm range, new physical effects and principles begin to determine the transport characteristics, and the validity of conventional current transport models is in question. The drift-diffusion model, which has enjoyed a remarkable success due to its relative simplicity, numerical robustness, and the ability to perform two- and three-dimensional simulations on large unstructured meshes, must be generalized to include hot-carrier and classical non-local effects. This motivated the development of higher order moments transport models such as the hydrodynamic, the energy-transport, and the six-moments models. After the introduction of stress for device performance enhancement the demand for accurate carrier mobility calculations based on full-band Monte Carlo algorithms has significantly increased, since they allow calibration of phenomenological mobility models and thus justify closure relations for higher order moments equations.

The transport models based on the semi-classical Boltzmann transport equation already contain information which can only be obtained from quantum-mechanical consideration. These are the band structure, expressions for the scattering rates, and the Pauli exclusion principle reflecting the Fermi statistics of carriers. With scaling continuing, other quantum-mechanical effects begin to affect transport properties. Quantum confinement in the direction orthogonal to transport in inversion layers makes the energy spectrum discrete. For sufficiently long channels, however, the carrier motion in transport direction can still be treated semi-classically, and development of transport models based on a set of subband Boltzmann equations is possible.

A useful approximation to mimic the quantum-mechanical carrier concentration profile is to introduce an effective potential into otherwise classical transport models. Transport calculations can then be carried out using conventional TCAD tools providing accurate and timely results. However, when modeling transport in ultra-scaled structures with only a few subbands occupied the full subband method must be applied.

Parallel to the search for new technological solutions for MOSFET scaling, the development of conceptually new devices and architectures is becoming increasingly important. New nanoelectronic structures, such as carbon nanotubes, nanowires, and even molecules, are considered to be prominent candidates for the post-CMOS era. At this small device size the geometrical spread of the carrier wave packet in transport direction can no longer be ignored. When the device size becomes shorter than the phase coherence length, the complete information about carrier dynamics inside the device including the phase of the wave function is needed and one has to resort to a full quantum-mechanical description including scattering. Transport in advanced nanodevices is determined by the interplay between coherent propagation and scattering. Numerical methods for dissipative quantum transport based on the non-equilibrium Green's function formalism, the Liouville/von-Neumann equation for the density matrix, and the kinetic equation for the Wigner function are attaining relevance.

* Corresponding author.

E-mail addresses: sverdlov@iue.tuwien.ac.at (V. Sverdlov), ungersboeck@iue.tuwien.ac.at (E. Ungersboeck), kosina@iue.tuwien.ac.at (H. Kosina), selberherr@iue.tuwien.ac.at (S. Selberherr).

Contents

1. Introduction

1.1. TCAD tools: technological motivation and general outlook

Integrated circuits (IC) play a key role in modern information society. Rising demand for superior computer performance, increased number of integrated components per area, and reduced power consumption lead to a pressing need to scale down the channel length of MOSFETs. The minimum feature size of integrated circuits has been continuously reduced in the past decades. This trend is expected to continue in the next decade, as predicted and institutionalized by the International Technology Roadmap for Semiconductors [1] and supported by demonstration of a MOSFET with the gate length as short as 6 nm [2].

The success of microelectronics technology has been partly enabled and supported by sophisticated Technology Computer-Aided Design (TCAD) tools which are used to assist in IC development and engineering at practically all stages from process definition to circuit optimization. At this moment, the TCAD related research and development cost reduction amounts to 35%, and is expected to increase to 40% in the near future [1].

Most TCAD tools are based on semi-classical macroscopic transport models. From an engineering point of view, semi-classical models, such as the drift-diffusion transport 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 meshes [3]. However, with device size dramatically reduced the TCAD tools based on a semi-classical transport description begin to show shortcomings.

The problem is two-fold. First, with the downscaling the driving field and its gradient increase dramatically in the short channel. As a result the carrier distribution along the channel can no longer be described by the shifted and heated Maxwellian distribution. In order to properly account for hot-carrier and non-local effects, the drift-diffusion and even the energy-transport model have to be improved to incorporate the substantial modifications in the distribution function.

The second, more fundamental reason for semi-classical modeling tools to gradually loose their validity lies in the particle-wave duality of carriers. The carrier motion can be described with the classical Newton law, when the length of the device is much larger than the corresponding carrier wavelength. When the device dimensions are comparable to the carrier wavelength, the carriers can no longer be treated as classical point-like particles, and effects originating from the quantum-mechanical nature of propagation begin to determine transport.

A typical example of quantum effects in MOSFETs is the quantization of carrier motion in the potential well of the inversion layer at the Si/SiO$_2$ interface. The thickness of the channel is usually in the order of a few nanometers, which is also the size of the electron de-Broglie wavelength. In this confining potential only levels with certain, quantized energies are allowed. This results in the formation of subbands, and, therefore, transport can no longer be accurately described by the classical equations in three dimensions. In consequence, a new description based on two-dimensional subbands must be adopted [4].

Another problem of quantum-mechanical effects in modern MOSFETs is the tunneling of carriers through classically forbidden regions. Tunneling leads to increasing leakage through thin gate dielectrics. This gate leakage is responsible for an increased power consumption in modern MOSFETs. In the 65 nm technology node the gate dielectrics are already so thin that the leakage current leads to critically high values of generated power. With scaling continuing, the conventional thinning down of gate dielectrics seems no longer possible, and new technological solutions must be employed to keep the heat production within reasonable limits [5]. Several options are considered for the upcoming technology nodes. One option is the use of alternative gate dielectric materials with higher permittivity and/or dielectric stacks combined with metal gates which help reducing the gate leakage current. A second option is to increase the drive current and the performance by additionally stressing the MOSFET channel. Finally, multi-gate device architectures with improved channel control and reduced short-channel effects can eventually be employed [6].

Tunneling is not only affecting the gate leakage currents. In MOSFETs with gate lengths of 10 nm and smaller quantum-mechanical tunneling under the barrier in the transport direction from source to drain becomes important [7]. When carrier scattering is negligible, a coherent transport approach based on the solution of the Schrödinger equation provides an accurate description for quantum-mechanical transport. This method is fast and efficient and can be implemented into commercial TCAD tools. However, since all devices operate at room temperature, carrier scattering in Si-based FETs is still important, and transport is significantly affected by scattering even in ultra-scaled FETs [8]. Recent studies indicate that the crossover from diffusive to ballistic transport in nanowire transistors may occur at a much shorter distance than previously anticipated [9]. An adequate transport model for ultra-scaled MOSFETs must, therefore, account for quantum-mechanical coherent and dissipative scattering effects simultaneously.

Modern TCAD tools have to be flexible enough to address challenges due to upcoming technological changes resulting from the use of new materials and structures. They must be prepared to adequately describe the quantum-mechanical phenomena which will determine transport in ultra-scaled CMOS and post-CMOS devices.

1.2. Brief history of TCAD transport modeling

The first fully numerical transport description was already suggested in 1964 by Gummel [10] for the one-dimensional bipolar transistor. The approach was further developed and applied to pn junctions [11] and to avalanche transit-time diodes by Scharfetter and Gummel [12]. The first application of a solution of the two-dimensional Poisson equation to metal–oxide–semiconductor (MOS) structures was performed by Loeb [13] and Schroeder and Muller [14]. The first simultaneous solutions of the coupled continuity and Poisson equations applied to junction field effect transistors (FET) [15] and to bipolar transistors [16] date back to 1969.

Since these pioneering works on device modeling many different approaches have been applied to practically all important devices, and the number of papers in the field has grown exponentially. Today the modeling of integrated devices has matured into a well-established field with active research, intensive software development, and vast commercial applications. Many textbooks, monographs, and reviews devoted to theoretical and computational aspects of device modeling have been published. The first monograph comprehensively covering aspects from modeling and discretization to applications is [3]. Not pretending to cover all the literature we mention only the most recent monographs relevant to the review. Various transport models were summarized in [17], while new approaches to transport are well described in [18].

As the costs of development and maintenance of today's TCAD software have significantly increased, only few large semiconductor companies can afford to support their own TCAD development team. There is a fairly large number of commercial TCAD software products available on the market [19–21] which serve most of the industrial demands. Numerous TCAD tools developed at universities have the advantage that they are freely distributed [23,22,24]. These tools regain popularity due to the concept of a complex approach to simulations based on different

levels of precision and complexity. It makes these tools valuable not only for pure educational and research purposes. Semiconductor manufacturing companies which need more refined simulations of complex phenomena often consult researchers from universities to verify their data obtained experimentally.

1.3. Formulation of the problem and the structure of the review

In order to analyze a semiconductor device under general operating conditions, an adequate mathematical model has to be formulated first. Regardless of the complexity of carrier dynamics in the semiconductor, two equations are an important part of any model and always have to be included. The Poisson equation relates the density $\rho(\mathbf{r}, t)$ of charged carriers in the device to the electrostatic potential $\phi(\mathbf{r}, t)$:

$$\mathrm{div}(\varepsilon\,\mathrm{grad}\,\phi(\mathbf{r}, t)) = -\rho(\mathbf{r}, t). \tag{1}$$

The Poisson equation must be supplemented by the proper boundary conditions, which usually are of the Dirichlet or Neumann type. The Dirichlet boundary condition fixes the potential at the domain boundary and is conveniently applied to model the interface between an external electrical contact and the device. The Neumann boundary condition sets the normal component of the electric field at the interface $E_n = -\mathbf{n} \cdot \mathrm{grad}\phi(\mathbf{r})$ to zero in order to isolate the simulation domain from the environment.

The second equation is the continuity equation, relating the current density $\mathbf{j}(\mathbf{r}, t)$ to the time derivative of the charge density:

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} = -\mathrm{div}\mathbf{j}(\mathbf{r}, t). \tag{2}$$

In semiconductors, two types of particles carrying positive (holes) and negative (electrons) charges are present. By introducing the electron $n(\mathbf{r}, t)$ and hole $p(\mathbf{r}, t)$ concentrations, the continuity equation can be written for each carrier type:

$$
\begin{aligned}
q\frac{\partial n(\mathbf{r}, t)}{\partial t} &= \mathrm{div}\mathbf{j}_e(\mathbf{r}, t) - qR(\mathbf{r}, t), \\
-q\frac{\partial p(\mathbf{r}, t)}{\partial t} &= \mathrm{div}\mathbf{j}_p(\mathbf{r}, t) + qR(\mathbf{r}, t),
\end{aligned}
\tag{3}
$$

where $R(\mathbf{r}, t)$ is the net electron–hole generation rate.

In order to form a closed set of equations and complete the formulation of the mathematical model, the system of Eqs. (1) and (3) must be supplemented by the material properties relating the current density $\mathbf{j}(\mathbf{r}, t)$ to the electron and hole concentrations and the electric field, which requires an accurate consideration of the physical mechanisms responsible for carrier motion inside the device.

Although we do not have the intention to completely cover this enormously large field of research here, we present several important examples and outline some difficulties and challenges to describe transport in modern MOSFETs. In the next section, we review the semi-classical description of carrier dynamics inside the device, which is justified when the characteristic device size is much larger than the corresponding electron wavelength. In ultra-scaled devices quantum effects start playing a role. Different types of quantum potential and density gradient corrections are reviewed in Section 3. Special attention is paid to the subband description of carriers in single- and double-gate structures, which is quantum-mechanical in the confinement direction and semi-classical in the current direction. Transport descriptions based on a fully quantum-mechanical approach are reviewed in Section 4. Beginning with the description of ballistic coherent transport, the importance of scattering in modern devices is highlighted. Methods to include scattering based on the non-equilibrium Green's function (NEGF), on the density matrix and the Wigner function are described. We conclude with a summary and an outline of future trends for TCAD transport models.

2. Semi-classical transport

An ensemble of interacting classical particles is conveniently described by the single-particle distribution function $f(\mathbf{r}, \mathbf{k}, t)$ in phase space, formed by position $\mathbf{r}$ and momentum $\hbar\mathbf{k}$. It can be shown [25] that by introducing the

self-consistent potential and by neglecting the exchange and higher order correlations the distribution function satisfies the Boltzmann equation [3], if the potential is a smooth function both in space and time:

$$\frac{\partial f_\nu}{\partial t} + \mathbf{u} \cdot \nabla_r f_\nu + \frac{s_\nu \mathsf{q}}{\hbar} \mathbf{E} \cdot \nabla_\mathbf{k} f_\nu = \left(\frac{\partial f_\nu}{\partial t} \right)_{\text{coll}} \tag{4}$$

The charge sign s_ν distinguishes between electrons, $s_n = -1$, and holes, $s_p = 1$. The right-hand side represents the collision operator which describes scattering of particles due to phonons, impurities, rough interfaces, and other scattering sources.

For realistic structures, a direct numerical solution of this equation by discretization of the phase space is computationally too expensive. TCAD tools, however, do not usually solve the Boltzmann equation and are based on simplified transport models. Approximate solutions can be obtained by the method of moments. Defining the moments of the distribution function $f(\mathbf{r}, \mathbf{k}, t)$, one consecutively obtains the drift-diffusion model [10], the hydrodynamic model [26], the energy-transport models [27], or the six-moments model [28]. Transport models based on the moments of the Boltzmann equation are well accepted in TCAD. Although these semi-classical transport models have been used for a long time and are well understood, research on this subject is still required for several reasons. With down-scaling of device feature sizes the semi-classical transport description is gradually losing its validity. Large electric fields rapidly changing within the scaled devices give rise to hot-carrier and non-local effects, and therefore, an extension to physically more sophisticated transport model is necessary. The six-moments model which is the most advanced moment-based transport model, is briefly discussed in the first subsection.

All models depend on several material parameters. For the drift-diffusion model the most relevant parameter is the mobility [3] relating the drift velocity to the driving force. Mobility dependence on materials, structure, and physical effects, such as remote Coulomb scattering or soft-phonon scattering at the interface of high-k materials, must be quantified. Demands on mobility modeling also arise from currently investigated effects to increase transistor on-current. Such effects are, for instance, mobility enhancement in strained Si and Ge channels, possibly in combination with a favorable substrate orientation and channel orientation. An example of accurate physics-based modeling of stress-dependent low-field mobility is discussed in the second subsection.

TCAD tools are using reduced transport models, because solving the Boltzmann equation with a Monte Carlo technique is computationally very expensive. Nevertheless, Monte Carlo solutions are conveniently used for calibration of TCAD models. A Monte Carlo solution of the Boltzmann equation is necessary to obtain accurate closure relations, expressing higher moments via the moments of lower order, to introduce and control scattering mechanisms at the microscopic level and, most importantly, to incorporate the peculiarities of the semiconductor band structure. An example of the stress-induced modification of the conduction band structure and its impact on electron transport in Si is discussed in the third subsection.

2.1. From drift-diffusion to higher moments equations

From an engineering point of view, the advantages of the drift-diffusion model has been its efficiency and numerical robustness. These properties make feasible two- and three-dimensional numerical studies of fairly complex device structures. The robustness comes from the fact that in this approach the current density is given by a potential flow with the gradient of the quasi-Fermi level as the driving force. However, several shortcomings of this model are critical for miniaturized devices. Hot-carrier effects such as impact ionization are difficult to estimate correctly and non-local effects such as velocity overshoot are completely neglected. Higher order transport models such as the hydrodynamic transport [26] and the energy transport [27] models are designed to overcome some of the shortcomings of the drift-diffusion model. The energy-transport model additionally takes into account the carrier energy balance. However, problems with the energy-transport model for TCAD applications are manifold. It typically tends to overestimate the non-local effects and thus the on-current of a device. With the heated Maxwellian assumption [28] implicit in the model the high energy tail of the carrier distribution is often considerably overestimated. This may result in unacceptable errors, for example, in the estimation of the hot-carrier-induced gate tunneling current [29]. Because of an overestimation of hot-carrier diffusion into the floating body of the device the energy-transport model may even fail completely in predicting the device characteristics for the specific situation of a partially depleted SOI MOSFET [30–32]. These facts motivate the development of transport models including higher order moments.

Recently a six-moments transport model has been proposed [33]. Such a model, while computationally more efficient than the Monte Carlo method, provides additional information on the shape of the distribution function. In the following, we outline the derivation of the six-moments model, closely following Refs. [28,31,34]. We highlight the differences between the drift-diffusion, energy transport, and the six-moments transport models and discuss the validity conditions for each of them.

The derivation is based on equations for statistical averages defined as

$$\langle \Phi \rangle = \frac{1}{4\pi^3} \int \Phi(\mathbf{k}) f(\mathbf{r}, \mathbf{k}, t) \mathrm{d}^3 k, \tag{5}$$

where $\Phi(\mathbf{k})$ is a weight function in $\mathbf{k}$-space. For simplicity we assume that the energy band is isotropic and parabolic, although generalization to a non-parabolic band is possible [35]. In order to derive the six-moments model, the following weight functions are chosen:

$$\begin{array}{lll} \Phi_0 = 1, & \Phi_2 = \mathcal{E}, & \Phi_4 = \mathcal{E}^2, \\ \Phi_1 = \hbar \mathbf{k}, & \Phi_3 = \mathbf{u}\,\mathcal{E}, & \Phi_5 = \mathbf{u}\mathcal{E}^2. \end{array} \tag{6}$$

Here, $\mathcal{E} = \mathcal{E}(k) = \hbar^2 k^2/(2m_v)$ is the dispersion relation, m_v the effective mass, and $\mathbf{u} = \nabla_{\mathbf{k}}\mathcal{E}(k)$ the carrier group velocity. Taking the moment of the Boltzmann Eq. (4) gives the following general moment equation,

$$\frac{\partial \langle \Phi \rangle}{\partial t} + \nabla_r \cdot \langle \mathbf{u} \otimes \Phi_j \rangle - s_v q\,\mathbf{E}\langle \nabla_p \otimes \Phi_j \rangle = \int \mathrm{d}^3 k\, \Phi_j \left(\frac{\partial f}{\partial t} \right)_{\mathrm{coll}}, \tag{7}$$

where $\otimes$ denotes the tensor product of two vectors One defines the relaxation times τ_Φ as

$$\int \mathrm{d}^3 k \Phi \left(\frac{\partial f}{\partial t} \right)_{\mathrm{coll}} = -\frac{\langle \Phi \rangle - \langle \Phi \rangle_0}{\tau_\Phi}. \tag{8}$$

The relaxation times τ_Φ depend on the distribution function and can in principle only be determined when the solution of the Boltzmann equation is known. Therefore, approximations have to be introduced in order to obtain a closed set of equations for the moments.

The distribution function can be separated into a symmetric part and an antisymmetric part:

$$f(\mathbf{k}) = f_{\mathrm{S}}(\mathbf{k}) + f_{\mathrm{A}}(\mathbf{k}). \tag{9}$$

In the so-called diffusion approximation, it is assumed that the antisymmetric part is small compared to the symmetric part. It implies that the displacement in $\mathbf{k}$-space of the distribution function and correspondingly the drift velocity is small. The symmetric part of the distribution function is assumed to depend only on the absolute value of $\mathbf{k}$, $f_{\mathrm{S}}(\mathbf{k}) = f_{\mathrm{S}}(|\mathbf{k}|)$. With these assumptions all tensor-valued averages reduce to scalar quantities. As a result one obtains the following balance equations for the even moments:

$$\frac{\partial \langle 1 \rangle}{\partial t} + \nabla \cdot \langle \mathbf{u} \rangle = 0, \tag{10}$$

$$\frac{\partial \langle \mathcal{E} \rangle}{\partial t} + \nabla \cdot \langle \mathbf{u}\mathcal{E} \rangle - s_v\, q\,\mathbf{E} \cdot \langle \mathbf{u} \rangle = -\frac{\langle \mathcal{E} \rangle - \langle \mathcal{E} \rangle_0}{\tau_{\mathcal{E}}}, \tag{11}$$

$$\frac{\partial \langle \mathcal{E}^2 \rangle}{\partial t} + \nabla \cdot \langle \mathbf{u}\mathcal{E}^2 \rangle - s_v 2q\mathbf{E} \cdot \langle \mathbf{u}\mathcal{E} \rangle = -\frac{\langle \mathcal{E}^2 \rangle - \langle \mathcal{E}^2 \rangle_0}{\tau_{\mathcal{E}^2}}, \tag{12}$$

For the odd moments representing fluxes the equations are:

$$\frac{2}{3}\nabla\langle \mathcal{E} \rangle - s_v q\mathbf{E}\langle 1 \rangle = -m_\mu \frac{\langle \mathbf{u} \rangle}{\tau_m}, \tag{13}$$

$$\frac{2}{3}\nabla\langle \mathcal{E}^2 \rangle - s_v \frac{5}{3}q\mathbf{E}, \langle \mathcal{E} \rangle = -m_\mu \frac{\langle \mathbf{u}\mathcal{E} \rangle}{\tau_{\mathrm{S}}}, \tag{14}$$

$$\frac{2}{3}\nabla\langle \mathcal{E}^3 \rangle - s_v \frac{7}{3}q\mathbf{E}\langle \mathcal{E}^2 \rangle = -m_\mu \frac{\langle \mathbf{u}\mathcal{E}^2 \rangle}{\tau_K}. \tag{15}$$

In these equations, we have omitted the time derivatives of the fluxes. We are, therefore, restricted to describe processes changing in time with frequencies smaller than the corresponding inverse relaxation time. However, this is not a practical limitation, since the typical value of the inverse momentum relaxation time τ_m in Si is of the order of $10^{12} - 10^{14}$ s^{-1}. It will be demonstrated below that the values of the relaxation times τ_S of the energy flux $\langle \mathbf{u}\mathcal{E} \rangle$ and τ_K of the kurtosis flux $\langle \mathbf{u}\mathcal{E}^2 \rangle$ are similar to τ_m.

In order to obtain the six-moments model it is convenient to define the following quantities in analogy to the energy-transport models:

$$\langle 1 \rangle = \nu, \qquad \langle \mathcal{E} \rangle = \frac{3}{2}\mathbf{k}_B \nu T_\nu, \qquad \langle \mathcal{E}^2 \rangle = \frac{15}{4}\mathbf{k}_B^2 \nu T_\nu \Theta_\nu, \qquad \langle \mathcal{E}^3 \rangle = \frac{105}{8}\mathbf{k}_B^3 \nu M_6, \qquad \langle \mathbf{u} \rangle = \frac{\mathbf{J}_\nu}{s_\nu \mathrm{q}},$$

$$\langle \mathbf{u}\,\mathcal{E} \rangle = \mathbf{S}_\nu, \qquad \langle \mathbf{u}\,\mathcal{E}^2 \rangle = \mathbf{K}_\nu.$$

Here, ν is the carrier concentration, T_ν denotes the carrier temperature, $\mathbf{J}_\nu$ the electrical current density, and $\mathbf{S}_\nu$ the energy flux density. As compared to the energy-transport model, for the six-moments model the new variables are a second order temperature Θ_ν which is a measure for the kurtosis of the distribution function, the moment of sixth order M_6, and a flux $\mathbf{K}_\nu$ related to the kurtosis of the distribution function. With generation-recombination terms added, the balance equations for the new variables become

$$\nabla \cdot \mathbf{J}_\nu = -s_\nu \mathrm{q}\left(\frac{\partial \nu}{\partial t} + R_\nu \right), \tag{16}$$

$$\nabla \cdot \mathbf{S}_\nu = -C_4 \frac{\partial(\nu T_\nu)}{\partial t} + \mathbf{E} \cdot \mathbf{J}_\nu - C_4 \nu \frac{T_\nu - T_L}{\tau_\mathcal{E}} + G_{\mathcal{E}\nu}, \tag{17}$$

$$\nabla \cdot \mathbf{K}_\nu = -C_5 \frac{\partial(\nu T_\nu \Theta_\nu)}{\partial t} + 2s_\nu \mathrm{q}\mathbf{E} \cdot \mathbf{S}_\nu - C_5 \nu \frac{T_\nu \Theta_\nu - T_L^2}{\tau_{\mathcal{E}^2}} + G_{\Theta_\nu}, \tag{18}$$

$$C_4 = \frac{3}{2}\mathbf{k}_B, \qquad C_5 = \frac{15}{4}\mathbf{k}_B^2. \tag{19}$$

We note that the generation-recombination terms may depend on both electron and hole distribution functions in an integral, non-local manner [3], which makes the task of solving the corresponding equations extremely difficult. Therefore, generation-recombination terms have to be modeled carefully using knowledge from semiconductors physics and may represent a significant challenge [37].

The system of equations for the densities is completed with the following flux equations:

$$\mathbf{J}_\nu = -C_1 \left(\nabla(\nu\, T_\nu) - s_\nu \frac{\mathrm{q}}{\mathbf{k}_B} \mathbf{E}\nu \right), \qquad C_1 = s_\nu \mathbf{k}_B \mu_\nu, \tag{20}$$

$$\mathbf{S}_\nu = -C_2 \left(\nabla(\nu T_\nu \Theta_\nu) - s_\nu \frac{\mathrm{q}}{\mathbf{k}_B} \mathbf{E}\,\nu\, T_\nu \right), \quad C_2 = \frac{5}{2}\frac{\mathbf{k}_B^2}{\mathrm{q}}\frac{\tau_S}{\tau_m}\mu_\nu, \tag{21}$$

$$\mathbf{K}_\nu = -C_3 \left(\nabla(\nu M_6) - s_\nu \frac{\mathrm{q}}{\mathbf{k}_B} \mathbf{E}\,\nu\, T_\nu\, \Theta_\nu \right), \quad C_3 = \frac{35}{4}\frac{\mathbf{k}_B^3}{\mathrm{q}}\frac{\tau_K}{\tau_m}\mu_\nu, \tag{22}$$

where the mobility

$$\mu_\nu = \frac{\mathrm{q}\tau_m}{m_\nu} \tag{23}$$

is introduced.

The drift-diffusion transport model consists of the continuity Eq. (16) and the current relation (20). The latter is decoupled from the higher order equation by introducing a closure assumption for the second order moment, $T_\nu = T_L$. The physical meaning of this assumption is that the carrier gas is in equilibrium with the lattice. The parameters of the model are the mobilities and generation-recombination rates of electrons and holes. The mobilities can either be measured or calculated from an independent solution of the Boltzmann equation.

We briefly outline the frequently used numerical ways of solving the Boltzmann transport equations in the next section.

In bulk devices the mobilities and the generation rates are functions of the doping concentration, electric field, stress, and other external parameters of interest. In practical device simulations, analytical dependences of the mobilities and generation rates on the system parameters are used. A generic device simulator should be able to account for different semiconductor materials to model their properties for a wide range of material compositions, temperature, and doping. Several analytical models calibrated to different semiconductors are usually available in TCAD device simulation tools. Below we present an example of the mobility dependence on the driving field used in the drift-diffusion model to describe the current saturation and outline the way to include a stress dependence into the low-field mobility model for Si. More examples of mobility models can be found in [23]. The drift-diffusion model can be adapted to deal with new materials with high potential for usage in novel devices. The mobility in wide-gap GaN [38] and AlGaN/GaN [39] was recently calibrated. A narrow-band InSb has also attracted attention because of its high mobility [40,41].

The drift-diffusion model is the simplest macroscopic transport model widely used in industrial TCAD. The partial differential equations of the model can be discretized on an arbitrary unstructured mesh and the iterative solution is stable and robust. The model allows a generalization to account for an anisotropic mobility. Due to its relative simplicity the drift-diffusion model is also used for two- and three-dimensional device simulations. These simulations are needed to account for a complex geometrical device structure, doping profiles, and material compositions. However, due to the high computational costs the three-dimensional simulations are employed in rare cases, when the structure cannot be reduced to a set of two-dimensional cuts.

Since the carrier temperature is constant, the drift-diffusion model does not account for any non-local effects and cannot capture the non-local transport inside short-channel devices. This is the reason why the accuracy of the drift-diffusion model decreases for gate lengths shorter than 100 nm [42], where the restriction of constant carrier temperature must be relaxed.

Due to the temperature gradient, heat flow and thermal diffusion appear. The drift-diffusion transport model must be augmented with the energy flow, or the third moment equation. In Bløtekjaer's approach [26], the energy flow is expressed phenomenologically via the carrier current and temperature gradient. As it was pointed out in [42] the closure relations for the odd moments depend on all the fluxes in sub-100 nm devices, and the truncation of the closure relations at a fixed number of fluxes introduces an additional uncertainty into the model and should be avoided. The energy-transport model takes into account the energy flux Eq. (21) in addition to the carrier energy balance Eq. (17). To close the system of equations, an approximate dependence of the fourth order moment on the lower moments has to be introduced. The assumption of a heated Maxwellian distribution for the symmetric part of the distribution function gives the closure relation $\Theta_\nu = T_\nu$, thus decoupling the model from the equations for the higher moments. We note that a slightly different approach adopted by Stratton uses the weight functions multiplied by the microscopic relaxation time in the definition of the moments, which results in a similar set of equations of the four-moments transport model [27].

The energy-transport model requires for each carrier type the modeling of two mobilities for the current density and the energy flux, one relaxation time, and the non-parabilicity factor for non-parabolic bands [35]. Since these models are not always available, a simplified model with the energy flux mobility being proportional or equal to the carrier mobility has been proposed [23,43]. This assumption introduces an error into the energy flux even for homogeneous conditions but does not influence the current density. The assumption of a constant energy-relaxation time is more severe, because it also affects the current density [42].

The simplified four-moments model is implemented in standard TCAD simulation tools and can be applied to a large set of semiconductor materials [21,23]. We note again that the accurate calibration of the parameters of the macroscopic transport model is required, before the model can be extended to new materials. Recently, an InSb narrow-gap semiconductor as a potential material for the future low-power, high-speed planar MOSFETs was modeled with the energy-transport model [40]. Also, the models for the hydrodynamic mobility and the relaxation time were improved [39] to achieve a better agreement with Monte Carlo simulation data and measured DC and AC characteristics of AlGaN/GaN high electron mobility transistors.

Since the energy-transport model overestimates the drive current in a MOSFET as short as 25 nm [42], the next moments should be included into consideration.

Going one step further in the model hierarchy one obtains a transport model of sixth order. A balance equation for the average squared energy (18) and the related flux Eq. (22) are added. To close the equation system, the moment of

sixth order M_6 has to be approximated using the lower order moments. For a Maxwellian distribution function and parabolic energy bands one would obtain $M_6 = T_v^3$. However, as the six-moments model does not require the heated Maxwellian approximation by treating the kurtosis of the distribution function as an unknown, a more general closure relation is desirable. Sonoda et al. [36] proposed a similar six-moments transport model, however, with a very restrictive closure relation. In Ref. [44], an empirical closure relation has been proposed taking into account also the second order temperature Θ_v.

$$M_6 = T_v^3 \left(\frac{\Theta_v}{T_v} \right)^c . \tag{24}$$

From Monte Carlo simulations serving as an accurate reference, a value of $c = 2.7$ has been estimated [42].

Compared to the energy-transport models, the six-moments model requires two additional relaxation times for the second order temperature $\tau_{\varepsilon 2}$ and the kurtosis flux τ_K. In the simplest approximation, the relaxation times for Si are approximated as constants $\tau_\varepsilon = 0.33$ ps and $\tau_{\varepsilon 2} = 0.2$ ps, while the mobility ratios are taken as $\tau_S / \tau_m = 0.8$ and $\tau_k / \tau_m = 0.7$ [44].

Having too many adjustable parameters is a particular inconvenience of the six-moments model. A solution to this problem is based on tabulating some parameters of the model using Monte Carlo simulations [42]. The parameter dependences on temperature, doping, and driving field are determined from the condition that the six-moments transport model reproduces exactly all the six-moments obtained from the Monte Carlo simulator under homogeneous condition. This results in models with "no knobs to turn". However, since it requires extensive simulations, the tables are available only for Si [23], and calibration of the six-moments model for other important materials has yet to be done.

The partial differential equations of the macroscopic transport models describing the spatial distributions of continuous quantities have to be solved numerically. This is performed by discretizing the equations on a suitable mesh. The question of whether a mesh is suitable for the equations under consideration is a complicated issue [45]. Improper meshes are a common source of errors [46]. The well-established Scharfetter–Gummel discretization scheme [12] can be extended for the higher moments models, with several variants existing [47–51]. In general, the convergence properties of higher moments models degrade significantly when compared to the drift-diffusion model, and also a proper choice of a mesh is a critical issue for convergence [51].

Fig. 1 shows a comparison of different macroscopic simulation approaches with full-band Monte Carlo simulation results for a 250 nm and a 50 nm double-gate MOSFET [42]. It can be seen that transport models based on two, four,
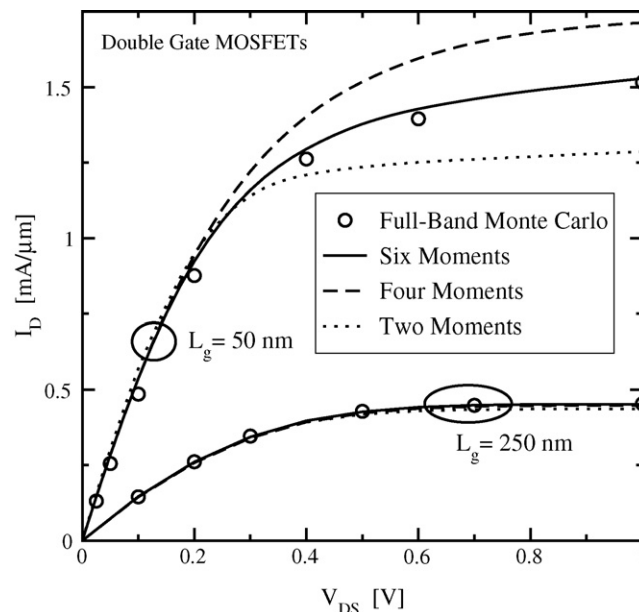

Fig. 1. Comparison of macroscopic transport models with full-band Monte Carlo calculations. While all models yield similar results at large gate lengths, only the six-moments model reproduces the short-channel Monte Carlo simulation results.

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 simulation results for the short-channel device.

2.2. Mobility compact modeling

The transport models discussed in the previous subsection contain several parameters which have to be accurately modeled for the purpose of predictive simulations. The drift-diffusion model described by (16) and the current relation (20) requires the carrier mobility (23) as a parameter. The mobility relates the drift velocity to the electric field and depends on scattering mechanisms cast into the momentum relaxation time τ as well as on the band structure via the effective mass m. Knowledge of the mobility dependence on temperature, doping, stress, substrate orientation, and driving electric field allows investigation of device performance in different regimes for arbitrary parameters.

In some cases, the physics-based consideration helps to generalize the model and to include additional effects. For example, the saturation of current at high drive voltage allows extension of the low-field mobility expression μ_L to include the dependence on the driving force E [23]:

$$\mu(E) = \frac{2\mu_L}{\left[1 + \left(1 + (2\mu_L E/v_s)^{1/b}\right)\right]^b}.$$ (25)

This expression contains two parameters determined usually from measurements or Monte Carlo simulations: b and the saturation velocity v_s at high driving forces E. Mobility dependences on E for electrons and holes in Si are illustrated in Fig. 2. We note that accurate modeling of the driving force E in inhomogeneous systems requires approximations [3]. Often, the field component in the direction of the current flow is adopted as a simple model of the driving force. We also note that transport models with higher moments use different expression for mobility which depends on carrier temperature [52,53] and not the driving force.

To develop a model for stress-dependent mobility, the peculiarity of the band structure modifications has to be taken into account. In Si, the conduction band consists of three pairs of equivalent valleys with their minima located close to the X-points in the Brillouin zone. For definitiveness we consider the biaxial stress applied in the (001) plane. This type of stress is practically realized in Si grown on (001) SiGe substrate [54]. In this case, the degeneracy between the three pairs of valleys is lifted, and a relative shift $\Delta E_C^{(i)}$ between the in-plane and out-of-plane valleys appear. The shift is given as [55]

$$\Delta E_C^{(i)} = \varXi_d(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}) + \varXi_u \varepsilon_{ii}, \quad i = x, y, z,$$ (26)
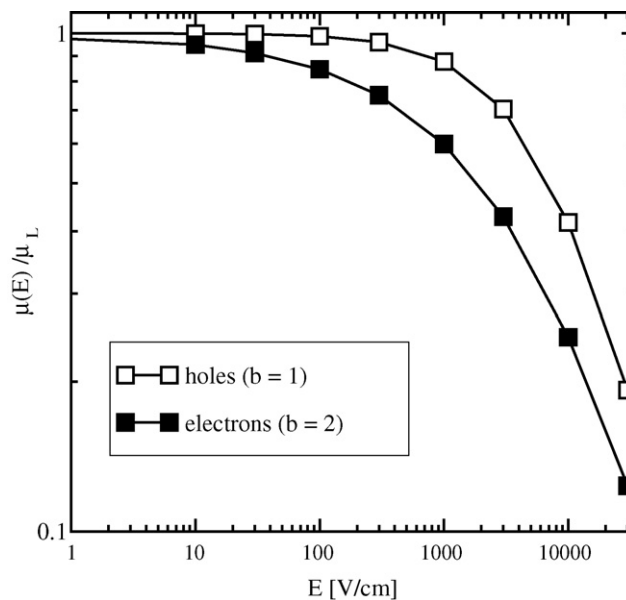

Fig. 2. Mobility (25) as function of driving force E, for electrons (filled symbols) and holes (open symbols).
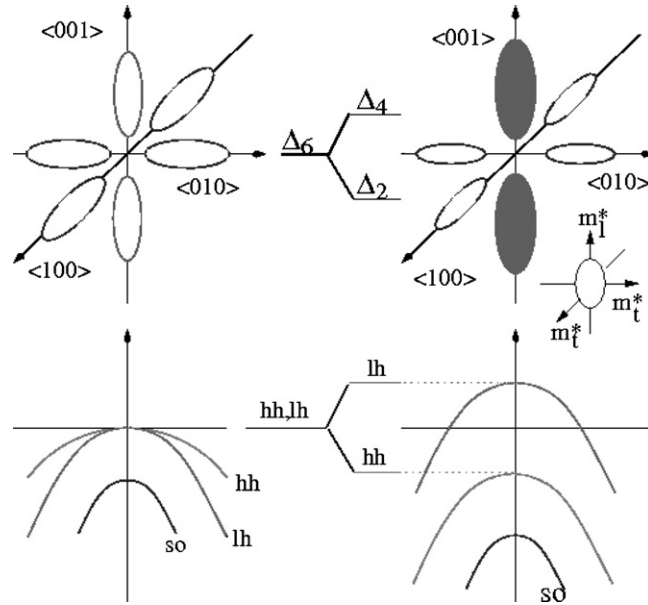
Fig. 3. Modification of the conduction band (upper part) and valence band (lower part) due to biaxial stress. Band structure of unstrained Si is shown on the left side of the figure. Stress induces the splitting between the equivalent valleys in the conduction band and between the heavy and light hole bands in the valence band.

where $\varXi_d$ denote the dilation and $\varXi_u$ the uniaxial deformation potentials for the conduction band and the ε_{ii} denote the diagonal components of the strain tensor expressed in the principal coordinate system. Changes in both the conduction and valence band structure of Si induced by biaxial stress are illustrated in Fig. 3.

The electron occupation in the valleys depends on their relative energy shifts [56,57]. As suggested in [56], the electron mobility can be computed by taking the weighted average of the electron mobility tensors $\hat{\mu}_{n,uns}^{(i)}$ of the ith pair of valleys in Si with the corresponding electron population $p^{(i)}$:

$$\hat{\mu}_n^{tot} = \sum_{i=1}^{3} p^{(i)} \cdot \hat{\mu}_{n,uns}^{(i)}, \quad p^{(i)} = \frac{n_{str}^{(i)}}{\sum_{i=1}^{3} n_{str}^{(i)}}, \tag{27}$$

$$n_{str}^{(i)} = N_C^{(i)} \cdot \exp\left[\frac{\Delta E_C^{(i)}}{k_B T}\right]. \tag{28}$$

Here $n_{str}^{(i)}$ is calculated using Boltzmann statistics with $N_C^{(i)}$ as the effective density of states, and $\Delta E_C^{(i)}$ the energy shift for the ith valley, k_B and T denote the Boltzmann's constant and lattice temperature, respectively. One advances the model (27) representing the mobility tensor as a product of a scalar mobility μ_L and the scaled inverse mass tensor and making the inverse mass tensor stress dependent [58]:

$$\hat{\mu}_{n,str}^{(i)} = \mu_L \cdot \hat{m}_{(i)}^{-1}, \quad i = x, y, z, \tag{29}$$

$$\hat{m}_x^{-1} = \begin{pmatrix} \dfrac{m_c}{m_l} & 0 & 0 \\ 0 & \dfrac{m_c}{m_t} & 0 \\ 0 & 0 & \dfrac{m_c}{m_t} \end{pmatrix}, \quad \hat{m}_y^{-1} = \begin{pmatrix} \dfrac{m_c}{m_t} & 0 & 0 \\ 0 & \dfrac{m_c}{m_l} & 0 \\ 0 & 0 & \dfrac{m_c}{m_t} \end{pmatrix}, \quad \hat{m}_z^{-1} = \begin{pmatrix} \dfrac{m_c}{m_t} & 0 & 0 \\ 0 & \dfrac{m_c}{m_t} & 0 \\ 0 & 0 & \dfrac{m_c}{m_l} \end{pmatrix}. \tag{30}$$

The tensors in (30) are the inverse effective mass tensors with m_t and m_l denoting the transversal and longitudinal masses for the ellipsoidal X-valleys in Si, respectively. The tensor is scaled to a dimensionless form by the

conductivity mass, m_c:

$$m_c = \frac{3}{(2/m_t) + (1/m_l)}.$$ (31)

Electrons in Si are scattered by phonons and impurities. Due to the later scattering mechanism the electron mobility depends strongly on the doping concentration. After scattering with phonons an electron may stay within the same valley or scatter into an equivalent valley (g-type). The after-scattering state can be also located in a non-equivalent valley (f-type scattering) [59]. The scalar electron mobility in Si is modeled by Matthiessen's rule including the effect of g-type and f-type phonon scattering, and the dependence on the doping concentration N_I via the corresponding relaxation times $\tau_{\text{equiv}}, \tau_{\text{neq}}(\Delta E_C^{(i)})$, and $\tau_I(N_I)$ as [58]:

$$\mu(N_I, \Delta E_C^{(i)}) = \frac{e}{m_c((1/\tau_{\text{equiv}}) + (1/\tau_{\text{neq}}(\Delta E_C^{(i)})) + (1/\tau_I(N_I)))}.$$ (32)

Strictly speaking even for independent scattering mechanisms Matthiessen's rule is exact only, if their relaxation times have identical dependences on energy or momentum. Deviations from Matthiessen's rule are known [4] and may be large for peculiar scattering mechanisms [60]. However, since the energy dependences for electron–phonon interaction and the impurity scattering relaxation times are similar smooth functions of energy without sharp resonances, (32) is a good approximation.

The electron mobility for the ith valley in strained Si can be written as [58]

$$\hat{\mu}_{\text{n,str}}^{(i)}(N_I, \Delta E_C^{(i)}) = \frac{\beta \cdot \mu^L}{1 + (\beta - 1) \cdot h^{(i)} + \beta \cdot ((\mu^L)/(\mu^{LI}) - 1)} \cdot \hat{m}_{(i)}^{-1}.$$ (33)

where $\hat{m}_{(i)}^{-1}$ denotes the scaled effective mass tensor for the ith valley given by (30), and h is a known function of T and $\Delta E_C^{(i)}$ [58]. The parameter $\beta = f \cdot m_t/m_c$, where f is the mobility enhancement factor, defined as the ratio of the saturation value of the transversal mobility in the valley i at high values of strain to the unstrained mobility μ^L. Eq. (33) is plugged into (27) to give the total mobility tensor for electrons in strained Si as a function of doping concentration and strain.

Fig. 4 a and b shows the doping dependence of the in-plane component of the minority and majority electron mobility components in strained Si for different Ge content in the underlying (001) SiGe substrate. The solid lines depict the results as obtained from the analytical model (33), while the symbols indicate the Monte Carlo simulation results. The model reproduces the slight increase in minority electron mobility for high doping concentrations for all strain levels. The model was recently generalized to describe the electron low-field mobility in Si under arbitrary stress conditions, including the dependences of the effective masses on shear strain [61].
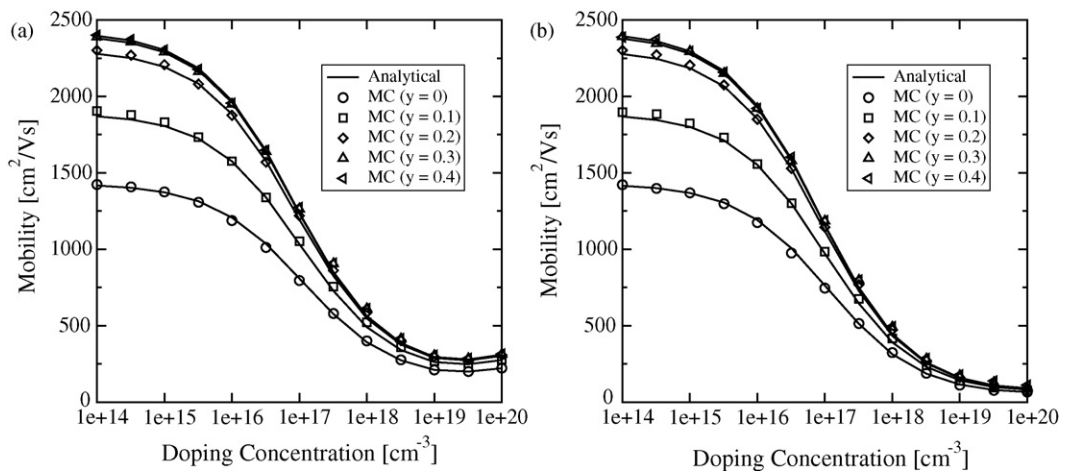

Fig. 4. Doping dependence of the in-plane electron mobility in strained Si calculated using (33) for different Ge content in the SiGe (001) substrate: (a) minority mobility and (b) majority mobility.

In order to validate compact mobility models and calibrate their parameters, results obtained by more precise methods must be used, which, however, require considerably more computational resources. A feasible approach to obtain an accurate solution of the Boltzmann transport equation is the Monte Carlo techniques discussed below.

A compact expression for the hole mobility in strained Si can also be obtained. The valence band consists of three bands: the heavy and light hole bands and the split-off band. Close to the minimum the band structure is well described by a six-band $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian [62]. Due to the interband interaction the dispersion relations are warped [63] and may not be described by the energy-dependent effective mass tensor, which complicates finding a compact expression for the mobility. At present the low-field hole mobility can be obtained and tabulated from the numerical solution of the Boltzmann equation obtained by the Monte Carlo technique.

2.3. Monte Carlo methods for transport calculations

Methods based on Monte Carlo techniques are well established for studying transport in semiconductors [64]. The motion of charge carriers is simulated in the phase space formed by position and momentum. In the presence of external fields, the carriers which are considered as point-like particles with well-defined momenta and positions move according to Newton law on classical trajectories. A dispersion relation expressing the carrier energy dependence on the crystal momentum is determined by the semiconductor's energy band structure. The free flight of carriers along the trajectory is interrupted by scattering events which are assumed local in space and instantaneous in time. Scattering is modeled as a random process. The duration of a free flight, the type of scattering mechanism, and the state after scattering are selected randomly from given probability distributions characteristic to the microscopic scattering process. The method of generating sequences of free flights and scattering events appears to be so intuitively transparent, that it is frequently interpreted as a direct emulation of the physical transport process rather than a numerical method. The first Monte Carlo algorithms used in device simulations were originally derived from merely physical considerations, viewing a Monte Carlo simulation as a computer experiment. These algorithms are called Ensemble single-particle Monte Carlo [65,66] and single-particle Monte Carlo [59,67].

Alternatively, one can formulate the transport equation first and then develop a Monte Carlo algorithm for its solution [68,69]. Interestingly, when the Boltzmann equation is transformed to an integral equation which is then iteratively substituted into itself, the iteration series results in the technique called *Monte Carlo Backward*, because the trajectories are followed in reverse direction back in time [68]. The algorithm is useful, if rare events have to be simulated or the distribution function is needed only in a small phase space domain [70]. A discussion of the relation of physics-based Monte Carlo methods to numerical Monte Carlo implementations for solving the Boltzmann equation can be found in [71].

The work of Kurosawa in 1966 [72] is considered to be the first account of an application of the Monte Carlo method to high-field transport in semiconductors. The following decade has seen considerable improvement of the method and application to a variety of materials [64]. Early papers deal with GaAs [73] and Ge [74]. In the mid-1970s, a physical model of Si has been developed, capable of explaining major macroscopic transport characteristics [75,76]. The used band structure models were represented by simple analytical expressions accounting for non-parabolicity and anisotropicity. With the increase of the carriers' energy the need for accurate, numerical energy band structure models arose [77–80]. For electrons in Si, the most thoroughly investigated case, a satisfactory understanding of the basic scattering mechanisms gives rise to a new "standard model" [81]. With the introduction of strain to enhance the performance of MOSFETs, however, the need for accurate full-band transport analysis has regained considerable interest [83,63,84].

We demonstrate the importance of full-band consideration using as an example an analysis of the electron low-field mobility in strained Si. We apply the simulator VMC [85], which offers simulation algorithms for both bulk semiconductors and one-dimensional devices with models based either on analytical band structure or on full-band structure tables. VMC includes a comprehensive set of scattering models for phonons, ionized impurities, alloy scattering, as well as impact ionization.

A phonon scattering model of Jacoboni and Reggiani [59] is used for the Monte Carlo simulation with the analytical band structure. The model takes into account long-wavelength acoustic phonons causing intravalley transitions and three f- and g-type phonons for intervalley transitions. Intravalley scattering from acoustic phonons is treated as elastic process.

Table 1
Modes, coupling constants, phonon energies, and selection rule of inelastic phonon scattering [88]

Mode	Δ (MeV/cm)	$\hbar\omega$ (meV)	Selection rule
Transversal acoustic	47.2	12.1	g
Longitudinal acoustic	75.7	18.5	g
Longitudinal optical	1042.0	62.0	g
Transversal acoustic	34.8	19.0	f
Longitudinal acoustic	232.0	47.4	f
Transversal optical	232.0	58.6	f

In the full-band approach, the scattering rates from the state **k** to the final state **k**′ are calculated from Fermi's golden rule Eq. (3.40) of [59]. The scattering rate is proportional to the square of the coupling constant of the electron–phonon interaction and to the overlap integral of the Bloch functions. The overlap integrals can in principle be determined from pseudo-potential calculations. The coupling coefficients for electron–phonon interaction depending on both initial and final electron state were calculated using a model of rigid pseudo-ions and Harris potentials [82]. It was demonstrated [86] that the total scattering rate computed with the *ab initio* model in [82] is in good agreement with the rate obtained within a simplified theory based on a so-called density-of-states Ansatz. The latter approach uses an approximation of constant matrix elements of electron–phonon interaction and a constant isotropic deformation potential. In this approximation the scattering rates are proportional to the density of the after-scattering states. While retaining the low-field transport properties of the analytical band structure model, the full-band approach accounts for the correct band structure at higher energies and improves the description of hot-electron properties [84].

In the following, "full-band Monte Carlo" refers to the density-of-states Ansatz. Several sets of parameters may be found in the literature [59,78,58] fitting equally well to the available low-field experimental data. The choice of the parameter values is not unique and, as it was pointed out in [87], depends on the model used to describe lattice dynamics and electron–phonon coupling. We favor the phonon scattering model adopted in [84]. Following [88] we recommend to slightly adjust the original values for the coupling constants for intervalley phonon scattering [59,84] to achieve a bulk mobility enhancement factor of 70% in biaxially strained Si. The coupling constants for acoustic and optical intervalley phonons as well as the phonon energies and the selection rules of the important phonon modes are listed in Table 1.

Comparing Monte Carlo simulation results using analytical band and full-band models allows the determination of the limits of validity of the analytical band model. In Fig. 5, we show simulation results for the electron mobility of strained Si for the stress directions (100) and (110) as well as predictions from a model based on the linear
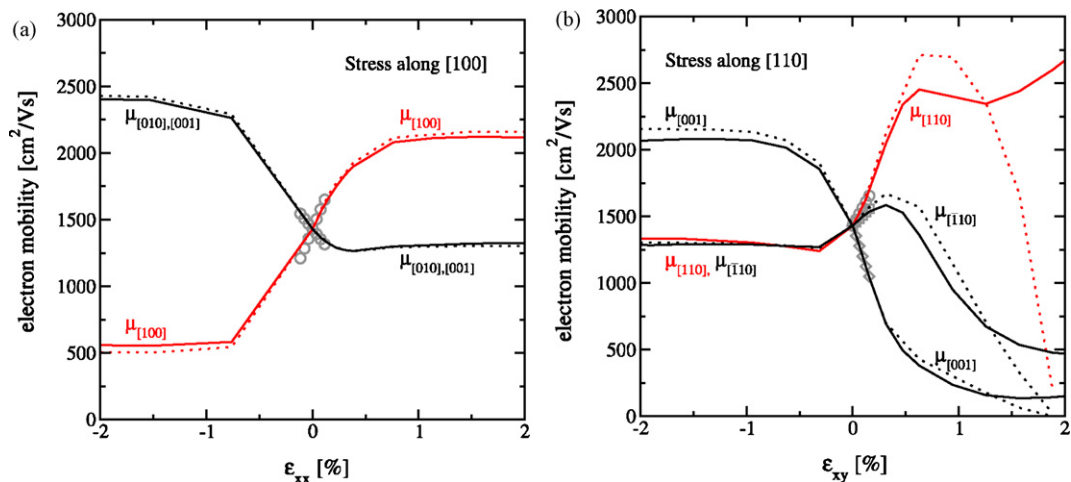

Fig. 5. Simulated bulk mobility of intrinsic Si as a function of strain for stress direction [100] (a) and [110] (b). Mobility is plotted along the stress direction and along two orthogonal directions from full-band Monte Carlo (solid lines) and analytical band Monte Carlo (dotted lines). Symbols indicate the change of mobility calculated using the piezoresistance coefficients [89].

piezoresistance coefficients [89]. Mobility is plotted in three directions, one being parallel and two being perpendicular to stress. In Fig. 5(a), the simulation results from analytical band and full-band Monte Carlo for stress along [100] are compared and good agreement is obtained. The resulting mobility is anisotropic in the (001) plane ($\mu_{[100]} \neq \mu_{[010]}$), and can be explained by strain-induced X-valley shifts. Mobility saturates at approximately 1% strain, regardless of the sign of strain. The saturated mobility values are larger for compressive strain, since in this case four X-valleys with unfavorable conductivity masses are depopulated.

In Fig. 5(b), simulation results are shown for stress along [110]. For tensile stress along [110], the mobility behavior is remarkably different from the previous case. First, mobilities along the directions [110], [$\bar{1}$ 1 0], and [001] are different from each other, with the largest mobility enhancement observed in [110] direction. Furthermore, no clear in-plane mobility saturation is observed as stress increases. The mobility enhancement for tensile stress is determined by the effective mass change induced by the shear strain component in the primarily populated valleys along [001] [90,91].

It can be seen that the results from analytical band and full-band Monte Carlo agree well up to 0.5% shear strain. At larger strain levels the band deformation is so pronounced that the energy band description in terms of an effective mass is no longer accurate, and full-band Monte Carlo simulations must be used even for the calculation of the low-field mobility.

At present the Monte Carlo method is computationally too expensive to be routinely used in TCAD. Indeed, the Boltzmann transport equation for N particles has to be solved self-consistently with the Poisson equation on a mesh in position space. An additional mesh in momentum space is required for numerical integration of the equations of motion for particles, if the Boltzmann equation is solved by a Monte Carlo technique. The mesh choice is also more restrictive than for moment-based transport models: while the use of a tetrahedron-based mesh in the momentum space is efficient [92], the problem of compensating the self-forces caused by the change of the dielectric permittivity at the Si/SiO₂interface is not resolved for unstructured meshes in the position space [93]. It deprives the particle-based device simulations from using unstructured meshes, which has been crucial for the success of the moment-based approach.

3. Mixed quantum semi-classical description and quantum corrections in current transport models

The characteristics of modern semiconductor devices are strongly influenced by quantum-mechanical effects. Therefore, purely classical device simulation is not sufficient to accurately reproduce transport. Size quantization of carrier motion in the confining potential of an inversion layer is a well understood quantum-mechanical effect in modern MOSFETs. Because of size quantization, the energy spectrum becomes discrete in the confinement direction, while it is still continuous in other directions. Thus the three-dimensional energy band structure is partitioned into a set of two-dimensional quantum subbands. In order to find the subband energy levels and the corresponding wave functions, the Schrödinger equation has to be solved. Then semi-classical transport is described by a set of Boltzmann equations for the subband distribution functions. The Boltzmann equations for different subbands are coupled because of intersubband scattering. This complicates the computation of the mobility. Even more, in inversion layers and especially in FETs with ultra-thin body (UTB) the subband occupation may be large at high effective fields, and degeneracy effects due to Pauli exclusion start playing a role.

The Pauli exclusion principle forbids the double occupancy of the same quantum state. Therefore, scattering to an already occupied state is prohibited. In order to describe transport correctly, degeneracy effects must be carefully taken into account. We discuss an efficient Monte Carlo algorithm based on the linearized Boltzmann equation including often ignored degeneracy effects. We demonstrate that in double-gate (DG) UTB FETs degeneracy effects lead to enhanced intersubband scattering which results in a carrier mobility quantatively different from the non-degenerate case.

The potential entering the Schrödinger equation depends on the carrier concentration in the inversion layer. By knowing the wave functions and occupations of the subbands, the carrier concentration in the inversion layer can be obtained. Therefore, the potential has to be found self-consistently by solving the Schrödinger and the Poisson equations simultaneously. This procedure is time consuming and should be avoided whenever timely results must be obtained. One option is to exploit the well-established semi-classical transport models, correcting them in such a way that they mimic at least the charge distribution of the quantum-mechanical system. An example of efficient quantum correction to the local density of states recently developed for double-gate FETs is discussed in details.

The quantum correction to mimic the local density of states close to the interface can also be understood as an additional quantum potential which has to be added to the classical self-consistent potential in order to describe the decrease of carrier concentration at the interface correctly. Because this additional quantum potential enters into the current relations, it opens an opportunity to introduce quantum-mechanical effects into the drift-diffusion and higher moments transport models.

3.1. Subband Monte Carlo and degeneracy effects

Excellent electrostatic channel control in DG FETs [94] makes these devices perfect candidates for the far-end of ITRS scaling [1]. Besides size quantization, degeneracy effects are more pronounced in UTB FETs, and their proper incorporation becomes an important issue for accurate transport calculations. Degeneracy leads to higher occupation of upper subbands and substantially increases the intersubband scattering.

Different approaches are known to include degeneracy effects into Monte Carlo algorithms. One method is to compute the occupation numbers self-consistently [95,96]. This approach is applicable not only to low-field mobility simulations but also for higher driving fields [97]. When the distribution function is close to the equilibrium solution, the blocking factor can be approximated with the Fermi–Dirac distribution function [78]. A similar technique to account for degeneracy effects was reported in [98].

We have recently generalized a Monte Carlo algorithm originally developed for three-dimensional simulations [99] to a quasi-two-dimensional electron gas [100]. A similar method was later used to compute transport in the presence of a magnetic field [101]. This method incorporates degeneracy effects *exactly* in the limit of vanishing driving fields and is valid for arbitrary scattering mechanisms and for general band structure. We demonstrate that in UTB DG FETs degeneracy effects lead to a qualitatively different mobility behavior than in the classical simulations. Degeneracy results in higher occupation of upper subbands which substantially increases intersubband scattering in (100) UTB DG FETs, resulting in a mobility decrease.

In order to obtain the low-field mobility, we compute the response to a small electric field $\mathbf{E}(t)$. Generated by this field a small perturbation $\delta f_n(\mathbf{k}, t)$ of the equilibrium Fermi–Dirac distribution $f_0(E_n(\mathbf{k}))$ in each quantum subband satisfies the system of coupled linearized subband equations:

$$\frac{\partial \delta f_n(\mathbf{k}, t)}{\partial t} = -e\mathbf{E}(t)\nabla_{\mathbf{k}} f_0(E_n(\mathbf{k})) + Q_n[\delta f], \tag{34}$$

where Q_n is the scattering operator of the linearized Boltzmann equation

$$Q_n[\delta f] = \sum_m \int \frac{d\mathbf{k}'}{(2\pi)^2} (\Lambda_{nm}(\mathbf{k}, \mathbf{k}')\delta f_m(\mathbf{k}, t) - \Lambda_{mn}(\mathbf{k}', \mathbf{k})\delta f_n(\mathbf{k}, t)). \tag{35}$$

The scattering rates $\Lambda_{mn}(\mathbf{k}, \mathbf{k}')$ in (35) are related to the rates $S_{mn}(\mathbf{k}, \mathbf{k}')$ of the original Boltzmann equation via

$$\Lambda_{mn}(\mathbf{k}', \mathbf{k}) = (1 - f_0(E_m(\mathbf{k}')))S_{mn}(\mathbf{k}', \mathbf{k}) + f_0(E_m(\mathbf{k}'))S_{mn}(\mathbf{k}, \mathbf{k}'), \tag{36}$$

where $E_n(\mathbf{k})$ is the total energy in the nth subband. The equation for the perturbation has a form similar to the Boltzmann equation, with two important differences: (i) the presence of a source term which depends on the small electric field and is proportional to the derivative of the equilibrium function, and (ii) renormalized scattering rates which enforce the equilibrium solution of the homogeneous Eq. (34) to be $f_0(E_n(\mathbf{k}))(1 - f_0(E_n(\mathbf{k})))$, and not $f_0(E_n(\mathbf{k}))$.

In order to calculate the mobility, a subband Monte Carlo method is used to solve the system (34). Following the procedure outlined in [99], we assume the time dependence of the driving field to be a set of instantaneous delta-like pulses:

$$\mathbf{E}(t) = \mathbf{E}_0 \tau \sum_i \delta(t - t_i). \tag{37}$$

In Eq. (37), τ is the average period between the delta-pulses and $\mathbf{E}_0$ is the value of the field averaged over a long simulation time T:

$$\mathbf{E}_0 = \frac{1}{T} \int_0^T dt \mathbf{E}(t).$$

We compute the current response $\mathbf{I}_i(t)$ produced by an electric field pulse at the moment t_i as

$$\mathbf{I}_i = eH(t - t_i) \sum_n \int \frac{d\mathbf{k}}{(2\pi)^2} \mathbf{v}_n \delta f_n(t - t_i), \quad (38)$$

where $\mathbf{v}_n$ is the velocity in the n th subband, and $H(t)$ is the Heaviside function. The instantaneous current density $\mathbf{J}(t) = \sum_i \mathbf{I}_i$ is calculated as the sum over current densities $\mathbf{I}_i$ produced by all pulses i. The current density value averaged over some long time T is then expressed as

$$\mathbf{J} = \frac{1}{T} \sum_i \int_0^T dt \mathbf{I}_i(t).$$

The low field mobility is defined as $\mu_{\alpha\beta} = J_\alpha/(enE_\beta)$, where the direction of the β-axis coincides with the direction of $\mathbf{E}_0$, and n is the carrier concentration. Now the mobility can be easily computed using a single-particle Monte Carlo technique.

The method can be illustrated as follows. The diffusion tensor $D_{\alpha\beta}$ is calculated as an integral of the velocity auto-correlation function [102]

$$D_{\alpha\beta} = \int_0^\infty d\tau \langle v_\alpha(t) v_\beta(t + \tau) \rangle, \quad (39)$$

where angular brackets denote the time averaging over the stochastic dynamics determined by the rates $\Lambda_{mn}(\mathbf{k}, \mathbf{k}')$ of the *linearized* multi-subband Boltzmann scattering integral in case of degenerate statistics.

The mobility tensor $\tilde{\mu}_{\alpha\beta}$ is related to the diffusion tensor via the Einstein relation for degenerate statistics

$$\tilde{\mu}_{\alpha\beta} = eD_{\alpha\beta} \frac{1}{n} \frac{dn}{dE_F}, \quad (40)$$

where E_F is the Fermi level.

In order to compute the mobility, we accumulate three temporary estimators t, w_β, and $v_{\alpha\beta}$ during the Monte Carlo simulations:

(i) initialize $t = 0$, $w_\beta = 0$, $v_{\alpha\beta} = 0$, and start the particle trajectory with the stochastic dynamics determined by the scattering rates $\Lambda_{mn}(\mathbf{k}, \mathbf{k}')$ from (36) of the *linearized* multi-subband Boltzmann equations;
(ii) before each scattering event update $v_{\alpha\beta}, w_\beta$, and t:

$$t = t + \frac{\tau(j)}{1 - f(E(j))},$$
$$w_\beta = w_\beta + v_\beta(j)\tau(j),$$
$$v_{\alpha\beta} = v_{\alpha\beta} + \tau(j)v_\alpha(j)w_\beta(j);$$

(iii) when t is sufficiently large, compute the mobility tensor as

$$\tilde{\mu}_{\alpha\beta} = \frac{e}{k_B T} \frac{v_{\alpha\beta}}{t},$$

where $v_\alpha(j)$ denotes the α-component of the velocity, $E(j)$ is the particle energy, $f(E)$ is the Fermi–Dirac function, and $\tau(j)$ is the time of jth free flight. The convergence of the method is improved by resetting $w_\beta = 0$ each time a velocity randomizing scattering event occurs.

We demonstrate the importance of degeneracy effects by evaluating the low-field mobility in inversion layers and in UTB FETs. The phonon-limited mobility in inversion layers shows a different behavior, if the Pauli exclusion principle is taken into account [98]. However, if surface roughness scattering is included, the relative difference decreases, and the universal mobility curve can be reproduced equally well using both degenerate and non-degenerate statistics, as shown in Fig. 6.

In UTB FETs degeneracy effects are expected to be more pronounced. We consider a 3 nm thick (100) UTB FET. The non-degenerate statistics is assured by using the rates $S_{mn}(\mathbf{k}', \mathbf{k})$ of the original Boltzmann equation in the Monte
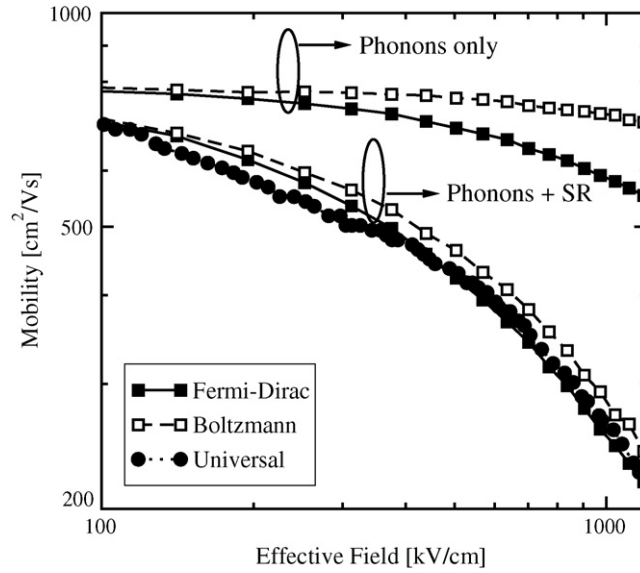
Fig. 6. The effective mobility of a Si inversion layer at (100) interface computed with Boltzmann (open symbols) and Fermi–Dirac (filled symbols) statistics reproduces well the universal mobility curve [103](circles). The phonon-limited mobility for degenerate and non-degenerate statistics is also shown.

Carlo algorithm described above. Results of mobility calculations for single-gate (SG) and DG structures, with and without degeneracy effects taken into account in the Monte Carlo simulations are summarized in Fig. 7. Mobility in a DG FET is plotted as function of the carrier concentration per channel, or $N_{DG}/2$. When degeneracy effects are neglected, the DG mobility is superior to the SG mobility. This is in agreement with the volume inversion hypothesis [104]. According to the volume inversion concept, the inversion layers located at the opposite interfaces of a thick DG structure start intersecting, when the Si body thickness shrinks. A charge maximum is formed in the middle of the UTB DG structure, yielding a larger distance between carriers and Si–SiO$_2$ interfaces. This should eventually result in a
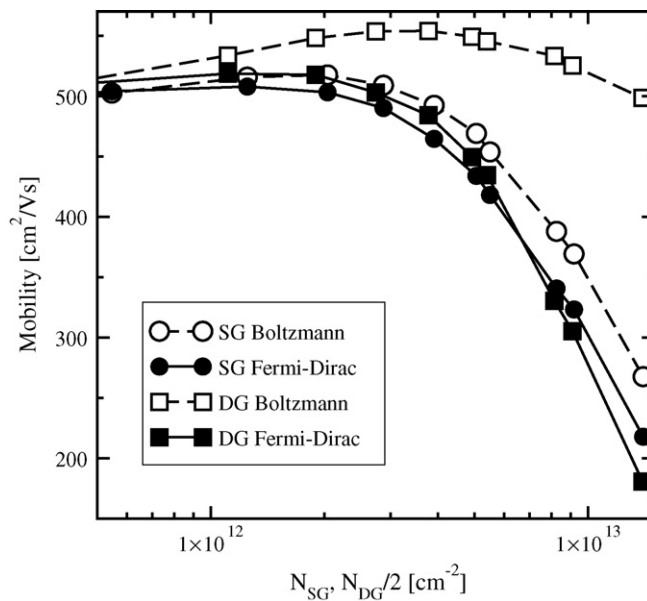

Fig. 7. Mobility in 3 nm thick (100) SG (circles) and DG (squares) structures computed with Boltzmann (open symbols) and Fermi–Dirac (filled symbols) statistics.

substantial decrease of surface roughness scattering in UTB DG structures. Therefore, a higher mobility is expected in DG structures compared to their SG analogs for similar body thicknesses. Mobility behavior consistent with this concept was recently reported experimentally for (110) UTB FETs [105]. With degeneracy effects included, Fig. 7 predicts a qualitatively different mobility behavior in the (100) DG structure, which becomes lower than in the SG structure at high carrier concentrations, in agreement with experimental data [106,105].

The difference between the mobility values for degenerate and non-degenerate statistics shown in Fig. 7 looks surprising. Indeed, at high carrier concentrations the principal scattering mechanism limiting the low-field mobility is elastic surface roughness scattering. For elastic scattering the forward and inverse scattering rates are equal: $S_{mn}^{el}(\mathbf{k}', \mathbf{k}) = S_{nm}^{el}(\mathbf{k}, \mathbf{k}')$, so that the Pauli blocking factor cancels out from the equations for the elastic scattering rates (36), and degeneracy effects seem to be irrelevant. This is not correct, however, since the Pauli blocking factor is also present in the inelastic electron–phonon part of the total scattering integral and ensures the equilibrium solution to be the Fermi–Dirac distribution function.

In case of Fermi–Dirac statistics, the Fermi level in a DG FET is higher than in a SG FET, due to a twice as high carrier concentration for the same gate voltage [100]. This results in a higher occupation of upper subbands. To study the influence of the occupation of *primed* subbands on the mobility lowering in (100) DG FETs we apply a biaxial stress of 1.6 GPa. This level of stress provides an additional splitting between the primed and unprimed subbands, which is high enough to depopulate the primed ladder completely. Results of the mobility simulation in 3 nm DG and SG structures, with biaxial stress applied, are shown in Fig. 8 together with the results for the unstrained structure. Both mobilities in strained and unstrained structures are similar in the whole range of concentrations. The inset displays the population of primed subbands in a 3 nm DG structure, showing that the primed ladder in a strained FET is practically depopulated. Since the mobilities of strained and unstrained UTB FETs are almost equivalent for both SG and DG structures, it then follows that the higher occupation of primed subbands is not the reason for the mobility lowering in (100) DG UTB structures.

Another consequence of the twice as high carrier concentration in a DG UTB FET is the higher occupation of upper *unprimed* subbands. When the carrier energy is above the bottom E_1 of the next unprimed subband, and intensive elastic intersubband scattering occurs. A step-like increase in the density of after-scattering states results in a sharp increase of scattering rates.

To demonstrate the importance of intersubband scattering for mobility calculations, we artificially switch off any scattering between unprimed subbands. We consider degenerate statistics. Results of the mobility calculations for a
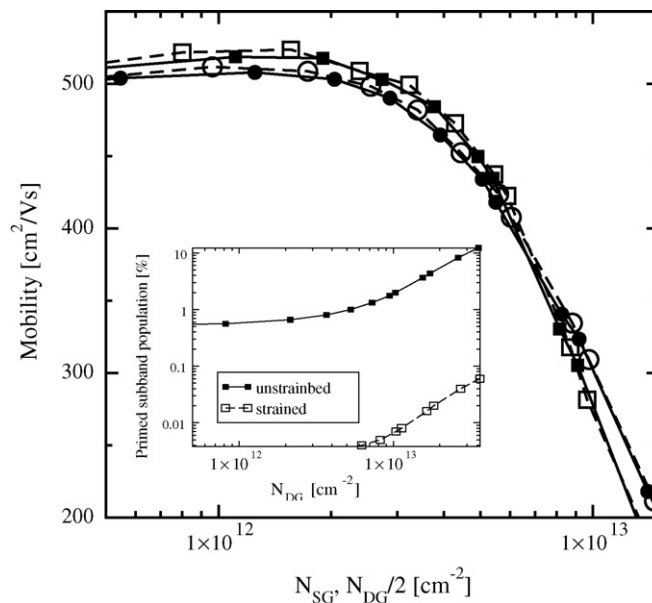

Fig. 8. Mobility in (100) 3 nm thick DG (squares) and SG (circles) structures computed with (open symbols) and without (filled symbols) in-plane biaxial stress of 1.6 GPa. *Inset*: Occupation of primed subbands in relaxed (filled symbols) and biaxially stressed (open symbols) DG structure.
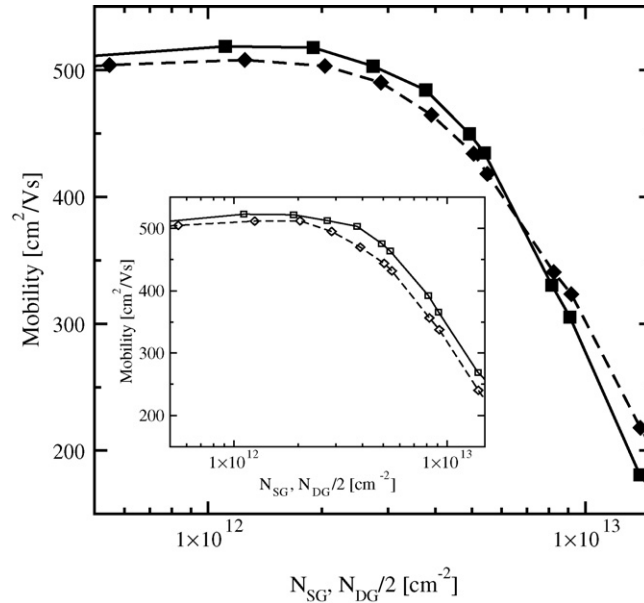
Fig. 9. Mobility in 3 nm thick (100) structures, computed with and without (inset) intersubband scattering. The higher carrier concentration in a UTB DG structure (squares) at the same gate voltage pushes the Fermi-level up and opens additional intersubband scattering channels between unprimed subbands. It decreases the mobility in (100) UTB DG FETs below its SG values (diamonds) at high carrier concentrations.

3 nm thick UTB structure with and without intersubband scattering are shown in Fig. 9. Without intersubband scattering the DG FET mobility is higher than the corresponding SG mobility, in analogy to non-degenerate results. It confirms our finding that as soon as the additional intersubband scattering channel becomes activated, the mobility in the DG FET sinks below the mobility in the SG FET. Degeneracy effects, therefore, play a significant role in compensating the volume inversion-induced mobility enhancement in (1 0 0) DG structures. They lead to a significant occupation of higher subbands in the unprimed ladder, which results in increased intersubband scattering and mobility lowering.

3.2. Quantum correction to the density of states

Although the transport description within the subband approach is based on the Boltzmann equation and is, therefore, semi-classical, finding the correct subband structure requires a self-consistent solution of the Poisson and the Schrödinger equations. Schrödinger–Poisson solvers, which deliver a self-consistent solution of a quantum-mechanically calculated carrier concentration and the electrostatic potential, provide accurate results for the carrier concentration within one-dimensional slices perpendicular to the interface. However, since the evaluation of the quantum-mechanical electron density is computationally very demanding, the application of Schrödinger–Poisson solvers may not be always practical.

Due to quantum confinement, which affects the local density of states, the carrier concentration near the gate oxide decreases. Purely classical device simulation without adequate quantum corrections is no longer sufficient to provide proper results, for example, the correct capacitance–voltage characteristics of MOS structures, since it predicts an exponential increase of the carrier concentration towards the gate oxide interface. To avoid the self-consistent solution of the Schrödinger equation, the classical semiconductor equations must be adapted to account for the quantum-mechanical decrease of the carrier concentration near the interface.

Various quantum correction models are available [107–111]. Some of these are based on empirical fits with numerous parameters [108,109]. In other models, the convergence [107] is strongly affected by the dependence on the electrical field. Some models [110] have to be re-calibrated for each particular device. A comprehensive comparison of these models was recently performed in [112].

The models discussed in literature are designed to describe a single inversion layer, and they are only of limited use for DG MOSFETs with UTB. Recently an accurate quantum correction model which is suitable for extremely scaled DG

MOSFET devices was developed [113]. This quantum correction model is intrinsically using the eigenstates that are formed in the Si film. A model confinement potential makes it possible to tabulate eigenenergies and expansion coefficients of the wave functions. It enables efficient simulation of DG MOSFET devices scaled to the decananometer regime.

The classical carrier concentration for a three-dimensional electron gas rises towards the semiconductor–oxide interface, while the physically correct quantum-mechanical carrier concentration strongly decreases towards the interface. The classically derived concentration can be adjusted to be equal to the quantum-mechanical one [114] by introducing the quantum correction potential φ_{corr} as

$$
\begin{aligned}
n_{\mathrm{cl,corr}} &= N_C \exp\left(-\frac{E_c - q\varphi_{\mathrm{corr}} - E_f}{\mathbf{k}_B T}\right), \\
n_{\mathrm{qm}} &= N_{C1} \sum_n |\Psi_n(x)|^2 \exp\left(-\frac{E_n - E_f}{\mathbf{k}_B T}\right), \\
n_{\mathrm{cl,corr}} &= n_{\mathrm{qm}}.
\end{aligned}
\tag{41}
$$

Here, N_C and N_{C1} denote the effective density of states for the classical and the quantum-mechanical carrier concentration, respectively, φ_{corr} describes the quantum correction potential, E_c the conduction band edge energy, and E_f the Fermi energy.

This approach requires the knowledge of the energy levels E_n and the wave functions $\Psi_n(x)$ of the quantized states. In order to find φ_{corr}, the above equations must be solved self-consistently with the Poisson equation. The quantum correction allows to use the drift-diffusion transport model and find the potential profile, concentration, and currents in biased devices. An example of the potential profile in a DG structure with 25 nm gate length [115] is shown in Fig. 10. To avoid the computationally expensive solution of the Schrödinger equation, one can tabulate the solutions for a parabolic shaped approximation of the conduction band edge [113]

$$
E_c(x) = E_{\max} - a\left(\frac{d}{2} - x\right)^2.
$$

Input parameters are the film thickness d and the curvature a. The wave functions are expanded as

$$
\Psi_n(x) = \sum_k \xi_{n,k} \sqrt{\frac{2}{d}} \sin\left(\frac{\pi}{d} kx\right).
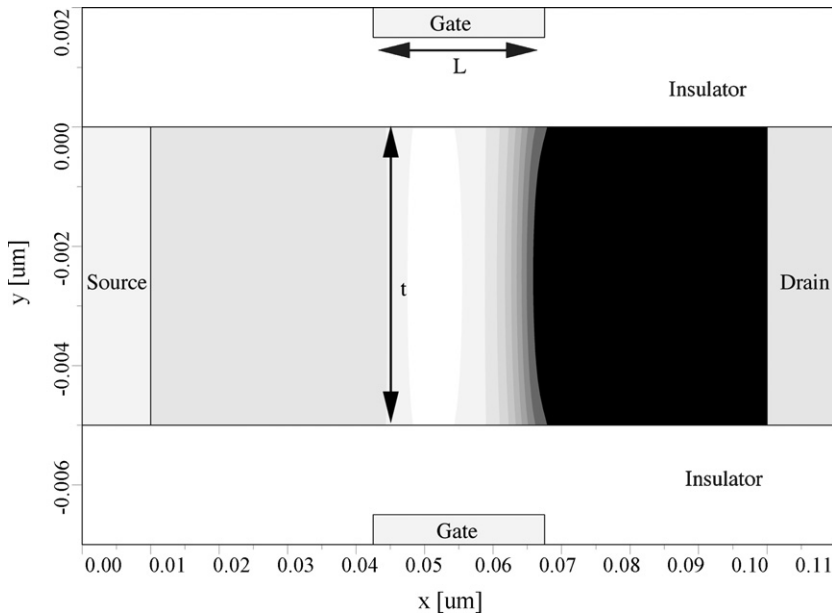$$

Fig. 10. Sketch of a 25 nm gate length double-gate MOSFET structure simulated by MINIMOS-NT. The contour plot shows the potential profile of the conduction band in the channel at the drain-source bias of 0.3 V and gate voltage 0.0 V.

Hence, the offset of the energy levels ϵ_n and the expansion coefficients of the wave functions $\xi_{n,k}$ can be found by interpolation of tabulated values. This allows an estimation of the correction potential φ_{corr} so that the corrected classical carrier concentration is consistent with the rigorous Schrödinger–Poisson solution

$$\exp\left(-\frac{q\varphi_{\text{corr}}}{\mathbf{k}_B T}\right) = \exp\left(-\frac{a(d/2-x)^2}{\mathbf{k}_B T}\right)\sum_m \frac{N_{\text{C1},m}}{N_\text{C}}\sum_n |\Psi_{m,n}(x)|^2 \exp\left(-\frac{\epsilon_{m,n} - E_f}{\mathbf{k}_B T}\right). \tag{42}$$

Here, m denotes the summation index over the different valley sorts (three for Si) [116].

The model was implemented in the general purpose device simulator MINIMOS-NT [23]. The Schrödinger–Poisson simulator VSP [117] was used to derive the reference quantum-mechanical (QM) curves. The applied iteration scheme of an initial classical simulation followed by a single quantum correction step and a final classical simulation with corrected bandedges delivers results close to a complete self-consistent simulation.

Fig. 11 shows the electron concentration at different bias points for DG MOSFETs with 5 nm film thickness. Good agreement between the QM and the corrected classical curves (DGTab) is achieved.

The inversion charge and the gate capacitance shown in Fig. 12 demonstrate excellent agreement between DGTab and QM curves for a wide range of gate voltages and relevant film thicknesses. Since the derived inversion charge is based on the accurate carrier concentration, no further fitting parameters are needed.

This model accurately reproduces both the carrier concentration distribution and gate capacitance characteristics as well as the total inversion charge even for extremely scaled DG MOSFET devices. Due to its computational efficiency, the model is well suited for TCAD.

An example of the calculated current–voltage characteristics of a double-gate MOSFET with a Si film thickness of 20 nm, with and without the quantum correction to the density of states, is shown in Fig. 13. Purely classical simulations overestimate the current in the entire range of the drain voltage.

3.3. Quantum correction potential, density gradient, and quantum hydrodynamics

The quantum correction to the local density of states can also be interpreted as an addition of a quantum correction potential [118] to the classical potential. Eq. (42) defines such a potential for double-gate structures. In fact, the
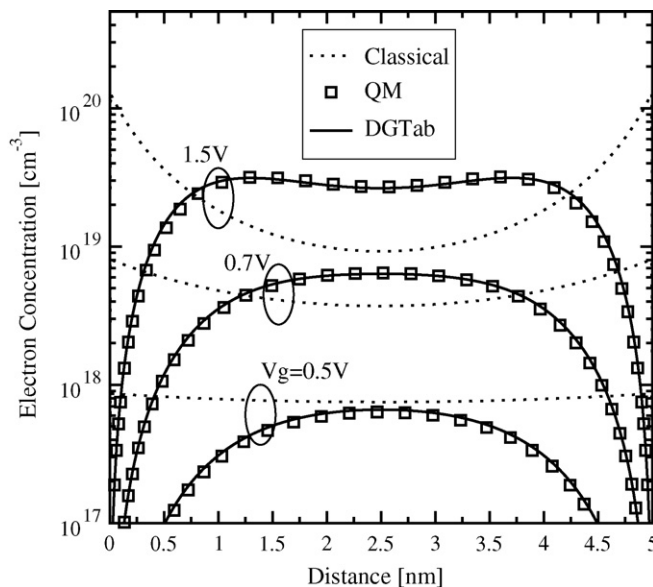

Fig. 11. The classical, the quantum-mechanical, and the corrected classical electron concentration in a double gate MOSFET structure with 5 nm Si film thickness. Gate voltages of 0.5, 0.7, and 1.5 V were applied. The DGTab quantum-corrected curves show outstanding agreement with quantum-mechanically derived curves for all applied gate voltages.
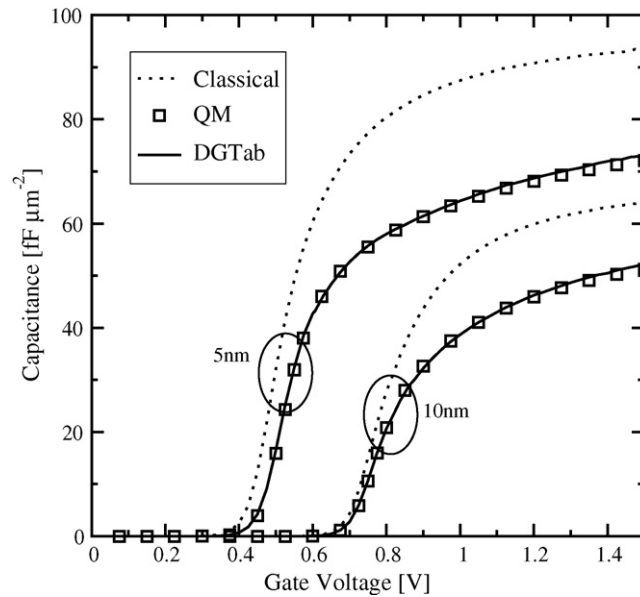
Fig. 12. The resulting gate capacitance as a function of the gate voltage for DG MOSFETs with film thicknesses of 5 and 10 nm, respectively. The DGTab quantum-corrected curves show very good agreement with quantum-mechanically derived curves for both film thicknesses.

appearance of the quantum potential can be easily illustrated, if one just substitutes the wave function in complex polar representation

$$\psi(\mathbf{r}, t) = A(\mathbf{r}, t)\exp\left(i\phi(\mathbf{r}, t)\right). \tag{43}$$

into the Schrödinger equation. Introducing the density $n(\mathbf{r}, t) = A^2(\mathbf{r}, t)$, the velocity $\mathbf{v} = \hbar\nabla\phi(\mathbf{r}, t)/m$, and the current $\mathbf{j} = N\mathbf{v}$, one obtains the following system of equations [119]:

$$\frac{\partial n(\mathbf{r}, t)}{\partial t} + \nabla[n(\mathbf{r}, t)\mathbf{v}] = 0, \tag{44}$$
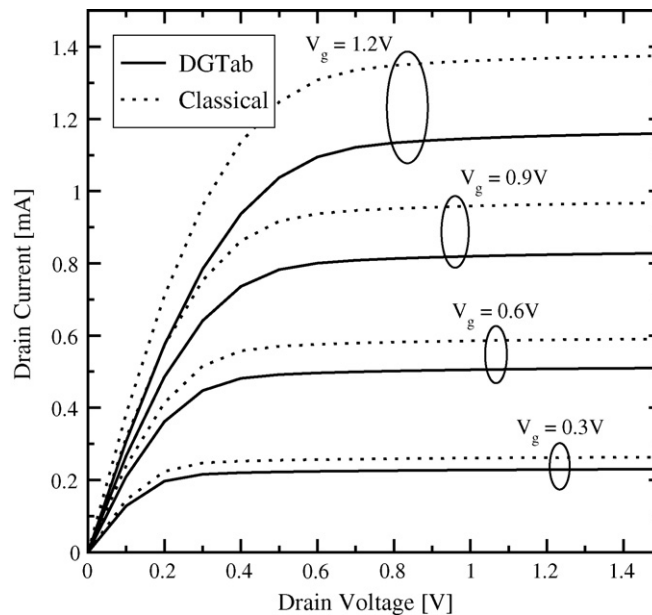

Fig. 13. Output characteristics of a double-gate MOSFET with a Si film thickness of 20 nm. The current is overestimated by a purely classical simulation because of the overestimated gate capacitance.

$$\frac{\partial \mathbf{v}}{\partial t} = \frac{v^2}{2m} + V(\mathbf{r}, t) + V_{QC}(\mathbf{r}, t). \tag{45}$$

Here, the additional quantum correction potential $V_{QC}(\mathbf{r}, t)$ is given by the following expression:

$$V_{QC}(\mathbf{r}, t) = -\frac{\hbar^2}{2m} \frac{\Delta \sqrt{n}}{\sqrt{n}}. \tag{46}$$

A similar expression for the effective quantum correction potential can be obtained from the quantum-mechanical theory based on the evolution of the moments for the Wigner function [120]:

$$V_{QC}(\mathbf{r}, t) = -\frac{\hbar^2}{3mr} \Delta \ln n, \tag{47}$$

where r is the parameter, taking values between $r = 1$ in the quantum limit with only a single subband filled and $r = 3$ in case of many subbands filled.

The concept of quantum correction potentials opens the way for using particle-based Monte Carlo methods to investigate transport in systems with strong size quantization. Therefore, the full power of three-dimensional full-band Monte Carlo methods which include accurate band-structure and scattering processes to obtain transport characteristics in inversion layers and SOI structures can be applied [8,121]. Recently, mobility of stressed Si/SiGe inversion layers was investigated [122–124]. Interestingly, due to a concentration decrease at the interface, the surface roughness scattering is underestimated, which requires special corrections [125]. Due to statistical noise in the second derivative of the carrier concentration, smooth approximations for the quantum correction potential are used in practice. The popular form of the effective potential introduced by Ferry [118] relates it to the self-consistent potential via an integral smoothing relation with a Gaussian kernel. Using this quantum correction potential, transport characteristics in ultra-thin body structures with short channels were recently investigated and compared against analytical predictions [126].

The form of the quantum potential (47) is commonly referred to as the density gradient correction and is extensively used in quantum hydrodynamic calculations [127–130]. The expression for the current density including the quantum correction reads [131]:

$$\mathbf{J} = q\mu \left[n\mathbf{E} + \frac{kT}{q} \nabla n + n\mu \frac{\hbar^2}{2r} \nabla \left(\frac{\Delta \sqrt{n}}{\sqrt{n}} \right) \right]. \tag{48}$$

Substitution of the current relation (48) into the continuity Eq. (2) results in a differential equation of the fourth order for the carrier concentration n. Such an equation needs two boundary conditions. If one considers the interface between the semiconductor and the dielectric, the first boundary condition is on the normal derivative of the concentration requiring the normal current component to be zero at the interface. The second boundary condition is obtained by setting the carrier concentration to zero at the interface. Thus, the quantum drift-diffusion theory based on (48) supplemented with the corresponding boundary conditions automatically reproduces the concentration decrease at the interface, mimicking the quantum-mechanical behavior.

For numerical transport calculations it is more convenient to avoid the discretization of the fourth order equation and to include the quantum potential correction term into a generalized electro-chemical potential [131]. After a careful calibration of the resulting density gradient model the transport calculations including the source-drain tunneling in ultra-scaled MOSFETs are possible [131]. For detailed analyses of quantum hydrodynamic models see [132]. The density-gradient theory has become popular as a physically justified approximate treatment of quantum effects well suited for engineering applications. It has also been applied to tunneling in metal–insulator–metal structures [128]. The application of the density-gradient theory to investigate the tunneling currents is less studied, probably due to the fact that direct quantum-mechanical methods provide a more accurate description of tunneling currents. Recently, a generalization of the density-gradient treatment to describe elastic multi-dimensional tunneling was reported [133]. Although these results are promising, there are indications that the density-gradient method possesses some limitations [134]. Additional studies are required in order to check whether the density-gradient theory describes tunneling correctly in all practically relevant situations.

4. Transport at nanoscale

As device sizes become comparable to the electron wavelength, quantum effects start playing a significant role. At the same time scattering with phonons, impurities, and surface roughness is still strong and cannot be neglected. At this point full quantum transport formalisms, which consistently describe quantum-coherent propagation and dissipative scattering processes, must be employed.

Established techniques used to address dissipative quantum transport can be classified according to the state functions they are based upon: non-equilibrium Green's functions, the density matrix, and the Wigner function. All three approaches are based on fundamental equations of motion and are equivalent at the most general level of formal description of a dissipative quantum system. The resulting system of integral–differential equations for the Green's function $G(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2)$, or the density matrix $\rho(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2)$, or the Wigner function $f(\mathbf{r}, \mathbf{p}, \omega, t)$ are in many cases too complex to allow for a direct numerical solution. Each function depends on two vector and two scalar arguments. For example, the Green's function $G(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2)$ in the coordinate representation depends on two position arguments $\mathbf{r}_1, \mathbf{r}_2$ and two time arguments t_1, t_2. For a numerical solution, each argument of the Green's function needs to be discretized. In the case of a three-dimensional system ($d = 3$), the total number of unknowns to be evaluated would be $N_{\text{tot}} = (N_x \cdot N_y \cdot N_z \cdot N_t)^2$. Assuming 100 grid points for each argument this results in the astronomical number $N_{\text{tot}} = 10^{16}$. Even in the two-dimensional case, $d = 2$, the number of unknowns is still very large, $N_{\text{tot}} = 10^{10}$, resulting in prohibitively large memory requirements. Approximations and simplifications must obviously be incorporated in order to make the problem numerically tractable. It is mainly these simplifying assumptions that make the differences between the approaches.

Although there are evidences that transport in MOSFETs with sub-10 nm channel length is still governed by scattering, we begin our consideration with a purely coherent carrier propagation in order to highlight the importance of quantum-mechanical effects in transport direction in short channel devices.

4.1. Ballistic transport and tunneling

A MOSFET utilizing ballistic transport was first suggested and studied by Natori [135]. It consists of a thin Si film connected to two reservoirs. One or two gate electrodes are adjusted close to the side film interfaces (Fig. 10). The gate electrodes are electrically separated from the Si film by a dielectric material. In order to reduce scattering in the channel the Si film is not doped.

The occupation of left- or right-propagating transversal modes is determined by the chemical potential of the left and the right reservoir, respectively. In equilibrium, the chemical potentials of both reservoirs are equal, the currents through the left- and right-propagating modes are equal in value and opposite in sign. Due to this compensation, the total current in equilibrium is zero. The source-drain voltage shifts the chemical potential and creates a disbalance in occupation between the left- and right-propagating modes. The number of propagating modes is controlled by the potential in the Si film, which is determined by the gate voltage. This picture is valid as long as there is no (back)scattering in the Si channel. We note that, except for the filling of propagating modes in accordance with the reservoirs' Fermi–Dirac distribution, one does not need quantum mechanics to describe the ballistic transport [136]. If it were possible to fabricate long channels without scattering, the ballistic MOSFET could be perfectly described within the semi-classical theory.

For ballistic transport the channel must be shorter than the mean free path. In Si at room temperature, the mean free path is only a few nanometers. With the Si channel scaled down, the wavelength of electrons in the channel gets comparable with the channel length. Due to gate electrode scaling the conduction band profile along the channel is no longer flat and starts bending considerably, thus forming a potential barrier between the gates. This potential barrier $V(x)$ along the channel can be well approximated by a parabola $V_{(x)} = V_m - m_l \omega^2 (x - x_0)^2/2$ close to its maximum at $x = x_0$. The characteristic curvature of the potential ω increases with the channel length L_c decreased approximately as [94]

$$\omega^2 \propto \frac{V_m}{m_l L_c^2}. \tag{49}$$

For large curvatures, quantum-mechanical tunneling of electrons under the barrier maximum becomes possible. The total current through the channel is determined by the sum of currents from all propagating modes, and the contribution from the modes tunneling under the barrier may still be negligible.

In order to estimate the curvature ω beyond which transport is becoming determined by quantum-mechanical tunneling, we use the Kemble formula for the transmission coefficient close to the maximum of the potential (49), following [94]:

$$T(E) = \left[1 + \exp\left(2\pi \frac{V_m - E}{\hbar\omega} \right) \right]^{-1}. \tag{50}$$

When electrons are injected according to the Fermi–Dirac distribution

$$f_0(E) = \left[1 + \exp\left(\frac{V_m - E}{k_B T} \right) \right]^{-1}, \tag{51}$$

the current is determined by carriers with $E \geq V_m$ flying above the potential maximum, if $\hbar\omega/2\pi \leq k_B T$. In this case, transport can be computed semi-classically. At the opposite limit $\hbar\omega/2\pi \geq k_B T$, the current must be computed quantum-mechanically, since the main contribution is due to carriers with $E \leq V_m$ tunneling under the barrier. For typical parameters of Si and $V_m = 50$ meV the channel length beyond which tunneling becomes important is around 10 nm.

Due to the absence of scattering in the channel and excellent channel control, the double-gate ballistic MOSFET can be considered as an ultimate MOSFET. In order to describe transport accurately, we assume (100) orientation of the Si channel at the Si/SiO$_2$ interface. As already discussed, a self-consistent solution of the two- or three-dimensional Schrödinger equation together with the Poisson equation represents a significant computational challenge, cf. [137]. Because of the strong lateral confinement in z direction due to the channel thickness $t = 2$ nm, only the first subband with the heavy mass m_i is populated with carriers which participate in transport for moderately high drain-source voltages. The wave function can then be written in the following approximate form:

$$\Psi(x, y, z) = \psi(x) e^{iky} \sqrt{\frac{2}{t}} \cos\left(\frac{\pi z}{t} \right). \tag{52}$$

This approximation simplifies the calculations and reduces the computational effort significantly [115,138–140]. The effective Schrödinger equation for the function $\psi(x)$ along the channel has the usual form $H\psi = E\psi$, with the Hamiltonian H

$$H = -\frac{\hbar^2 d^2}{d^2 x} + \frac{2}{t} \int_0^t dz \cos\left(\frac{\pi z}{t} \right) eV(x, z), \tag{53}$$

where the potential $V(x, z)$ is found from the Poisson equation. The Schrödinger equation must be supplemented with the boundary conditions. To describe an open system, one uses the standard boundary conditions corresponding to plane propagating waves in electrodes far from the channel. The Schrödinger equation can be solved numerically [141] or using the WKB approximation [7]. The solution of the Schrödinger equation with open boundary conditions can also be achieved by means of the quantum transmitting boundary method [142,143]. An established alternative framework for these calculations is the non-equilibrium Green's function method [144] 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 [137,145].

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 [146], for example, the transmission function is fully determined by the reduced contact part of the full Green's function. However, the carrier concentration determines the electrostatic potential in the device via the Poisson equation. Since the carrier concentration is proportional to the square of the wave function, the accurate determination of the transmission coefficient and the current requires a self-consistent solution of the Schrödinger and Poisson equation simultaneously within the whole device.

A Schrödinger–Poisson solver [117] with open boundary conditions can be used to find the wave functions self-consistently. As a result, the transmission function through the channel for each propagating mode is found. The total current is obtained by summing up the contributions from all modes propagating in forward direction from source to
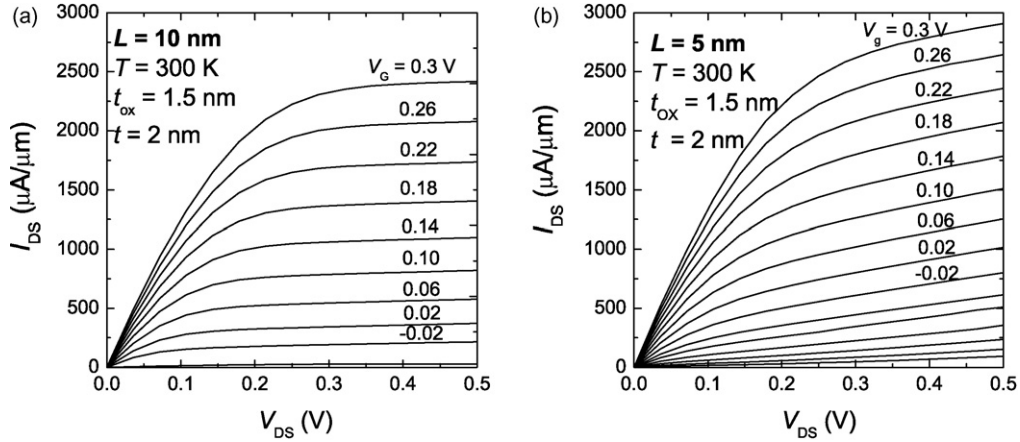
Fig. 14. Output characteristics of a DG FET with 2 nm Si body and 1.5 nm Si dioxide thickness. (a) Gate length 10 nm and (b) gate length 5 nm. With the gate length decreased, the current saturation is less pronounced.

drain and in backward direction from drain to source. In an UTB FET when only a single subband is occupied, the current as function of a source-drain voltage V is [135]

$$J(V) = e \frac{\sqrt{2mT}}{\pi^{3/2}\hbar^2} \int dE\, T(E) \left(F_{-1/2}\left(\frac{\mu_s - E}{k_B T} \right) - F_{-1/2}\left(\frac{\mu_s - E - eV}{k_B T} \right) \right), \tag{54}$$

where $F_{-1/2}(x)$ is the Fermi–Dirac integral of the order $-1/2$. The structure of (54) is quite transparent. The two terms in the brackets correspond to supply functions from the source and from the drain electrode. Their difference, multiplied by the transmission function $T(E)$, gives the current through the device.

Fig. 14 shows output characteristics of a transistor with $t_{ox} = 1.5$ nm, $t = 2$ nm, for several values of gate voltage [7]. Two sets of *IV* s are presented corresponding to the gate lengths of 5 and 10 nm. For 10 nm channel length *IV* characteristics are very similar to the characteristics of the ideal transistor, see Fig. 14(a), with a perfect saturation, suggesting an almost vanishing drain-induced barrier lowering effect (DIBL). The 10 nm transistor is, therefore, very close to the ideal double-gate ballistic transistor, described in [135]. With the decreased gate length the current saturation becomes less pronounced. The reason for this degradation is two-fold: tunneling under the barrier formed by the potential in the channel and drain-induced barrier lowering. The characteristics are acceptable for $L = 5$ nm, see Fig. 14(b), while for $L = 2.5$ nm the saturation practically vanishes. The output characteristics simulated for an ultra-thin body double-gate MOSFET with a gate length L as short as 5 nm shows that 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 [147]. It should be noted that the sensitivity to small MOSFET dimension variations, the control of doping as well as the whole manufacturing process development represent significant challenges for multi-gate MOSFETs with a gate length below 10 nm.

We note that the above consideration based on carriers' coherent tunneling through the dielectric potential barrier can be applied to study leakage currents [29]. The current expression has exactly the form of (54), if the difference of Fermi–Dirac integrals is replaced by the corresponding Tsu–Esaki supply function [148]. In case of computing leakage currents from the inversion layers the boundary condition to the Schrödinger equation must describe slowly decaying quasi-bound channel states. It can conveniently be achieved by introducing a complex coordinate stretching [149]. This technique allows the artificial absorbing layer to be applied at the interface. This perfectly matched layer method is elegant, numerically stable, and efficient for calculating lifetimes of quasi-bound states.

In the discussion above, the contacts to the device were modeled only via the supply functions and were assumed ideal. The ideal contact is assumed to be in equilibrium characterized by a value of the chemical potential. The role of the contact is two-fold. First, an ideal contact should supply enough incoming carriers with a Fermi–Dirac equilibrium distribution into the device region. The distribution should be shifted at a velocity corresponding to the current in order to avoid a pile-up of charged carriers at the contacts [150,151]. Second, an ideal contact must absorb all the carriers coming out of the device.

The concept of an ideal contact allows to consider a device isolated from the surrounding. The definition of an ideal contact may, therefore, depend on the device under consideration. A three-dimensional reservoir of electrons is a good model of an ideal contact to a Si-on-insulator FET [136]: the large mismatch of the phase spaces of a three-dimensional contact and a quasi-two-dimensional Si film makes the probability of outcoming carrier to scatter back to the active region extremely small. At the same time, the role of reservoirs is to supply enough electrons into transversal modes propagating through the constriction [152]. The model was recently shown to be valid even in the case of an abrupt connection between a Si film and a three-dimensional semi-space of the contact. By means of direct numerical simulations [153] it was demonstrated that the transport characteristics of a Si film abruptly connected to large reservoirs and of the same Si film with the boundary conditions corresponding to ideal contacts (injection with an equilibrium distribution and ideal absorption) are equivalent at room temperature.

In the light of this discussion, elevated contacts to SOI FETs can be considered as ideal. At the same time, the heavily doped extensions of the same thickness as the Si film cannot be treated as ideal contacts because of the substantial reflection of carriers back to the channel [154]. In the case of ballistic transport in the channel and elastic scattering in the extensions the total resistance of the system consists, up to a small correction, of the two Ohmic resistances of the source and drain extensions and of the Landauer resistance of the ballistic channel connected to the extensions in series [154]. However, the Landauer resistance of the channel connected to the extensions is *not* equal to the resistance of the channel between the two ideal contacts, because the resistance of the channel *depends* on the type of the contact used. Indeed, in the ideal ballistic MOSFET [135] the gate voltage fixes the carrier concentration in the channel, which is the sum of the concentrations of injected and reflected electrons. The current through the device is, however, the difference between the currents of injected and reflected carriers. Therefore, for the same gate voltage and the same source-drain voltage the contacts with different reflection properties would lead to different currents and contact-dependent resistances of the channel.

One should stress that only elastic scattering was assumed in the consideration above. Since the total energy (potential plus kinetic) of the carriers is conserved in this case, the resistance of the channel depends on the number of reflected hot-carriers with energies much higher than the characteristic energy of equilibrium carriers in the drain. Inelastic scattering with optical phonons and electron–electron scattering in the heavily doped drain will thermalize the hot-carriers in the drain and reduce their reflection into the channel. In Si, it suffices for a carrier to experience a single inelastic scattering event to lose an energy of $2k_bT$ (see Table 1). Since the characteristic energy in the channel is of the order of k_bT, this carrier will not have enough energy to return back to the channel [155]. Because of "Brownian-like" motion due to frequent elastic scattering, the characteristic distance between the two inelastic scatterings is proportional to the geometric average $l = \sqrt{l_{el}l_{in}}$ of the elastic l_{el} and inelastic l_{in} mean free paths in the contacts, which would be about 10 nm in heavily doped Si. If the active region of the device is shorter than l, a part of the contact regions of length l must be considered together with the device. The trade-off is, of course, the opportunity to set up simplified completely absorbing boundary conditions corresponding to ideal contacts at the expense of an increased simulation domain.

The device properties are independent from the contact properties only if the carriers in the simulation domain have no correlation with those in the contact reservoirs. A purely classical correlation length l between two inelastic scattering events in the contacts was identified above. The second important length is the coherence length l_{coh} in the reservoirs. Interestingly, l_{coh} coincides with l when elastic scattering is dominated by scattering with ionized impurities in heavily doped Si, because scattering with impurities preserves the coherence, while electron–phonon scattering destroys it. At finite temperature at equilibrium, however, the coherence length obtained from the thermal equilibrium density matrix [156] is proportional to $l_{coh} = \hbar\sqrt{k_BT/m}$, where m is the effective mass. It is usually assumed that the contacts represent waveguides with a flat bottom, so that plane waves are good asymptotic eigenstates [157]. If the ideal boundary conditions are set at a distance $L = \max(l, l_{coh})$ inside the contact regions form the device, the role of the contacts is to inject a (shifted) equilibrium distribution into the plane wave states incident to the device and to absorb the outcoming waves. This is exactly the idea behind the boundary conditions for the Wigner function [156] or the so-called U-scheme [158]. Although the boundary conditions for the Pauli master equation [150,151,159] discussed in the end of this section look differently, it was argued [151] that, if they are set at a distance further than the coherence length inside the contacts, injections into the plane wave states or the scattering states should be equivalent. Formulation of the boundary conditions too close to the active device region may result in a non-physical behavior of the solution [159,160]. Finally, the contact self-energy [161] discussed in the next section describes a decay of the Green's function of an isolated system because of the coupling of the system to the modes propagating in equilibrium reservoirs.

The transport models described so far are either based on the assumption of semi-classical or pure quantum-mechanical ballistic transport. The former modeling approach has proven to be adequate to describe transport in previous generations of microelectronic devices. The latter one may be used for transport description, when the carrier coherence length is larger than the device size. Recent studies show that even for devices with a channel length as short as 15 nm scattering will still play a significant role [8] and, therefore, determine the current, in accordance with estimations of the mean-free path in MOSFET structures [162]. The crossover from diffusive to ballistic transport in Si nanowire transistors occurs at approximately 2 nm [9], a much shorter distance than previously anticipated. An adequate transport model for ultra-scaled MOSFETs must, therefore, account for quantum-mechanical and dissipative effects simultaneously.

4.2. Non-equilibrium Green's function approach to scattering

The non-equilibrium Green's function method addresses the quantum transport problem in the most consistent and complete way. However, the method is computationally complex and, with a few exceptions [163–165], usually applied to one-dimensional problems [144] and for a restricted set of scattering mechanisms [166] only. Scattering requires the knowledge of the corresponding self-energies and thus complicates computations significantly [166]. The carbon nanotube (CNT) FET which is widely considered to be a potential alternative to the conventional MOSFET in the far future, represents a good example where the non-equilibrium Green's function method provides accurate results and is successfully used. CNTs can be considered as a graphite sheet which has been wrapped into a tube. The way the graphite sheet is wrapped is represented by a pair of indices (n, m) called the chiral vector. The integers n and m denote the number of basis vectors along two directions in the honeycomb crystal lattice of graphite (Fig. 15).

CNTs with $n - m = 3$ are metals, otherwise they are semiconductors [167]. Semiconducting CNTs can be used as channels for transistors [168], whereas metallic CNTs can serve as interconnect wires [169]. The fabrication of CNT based devices with Schottky type [170] and Ohmic [171] contacts for holes has been reported.

Due to the relatively small number of atoms in short CNTs, the non-equilibrium Green's functions can even be solved at an atomistic level by using p_z orbitals of carbon atoms as the basis. Calculations of current through electronic devices based on the non-equilibrium Green's function formalism (NEGF) was first described in a series of papers in the early 1970s [172–175]. The NEGF technique has widely been used to study quantum transport of electrons and holes in a variety of materials and devices, such as III–V resonant tunneling diodes [144,176–183], electron waveguides [184], Si tunneling diodes [185,186], ultra-scaled Si MOSFETs [163,166,187], CNTs [188–203], metal wires [204,205], and organic molecules [206–212].

Green's function methods provide a powerful technique to evaluate the properties of a many-body system both in thermodynamic equilibrium and non-equilibrium situations. The single-particle Green's function of a system allows evaluation of carrier density and current. The many-particle information about interactions is cast into self-energies. The perturbative solution of the equations for the Green's functions and self-energies is the key technique to evaluate the properties of an interacting system in external fields.
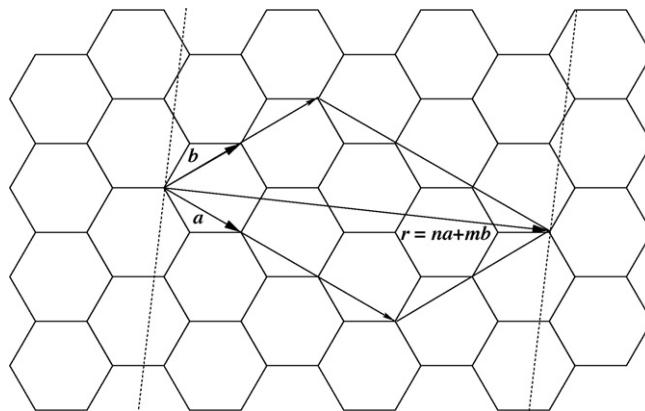

Fig. 15. Indices (n, m) defining the chiral vector **r**. A graphite sheet is cut along the dotted lines and wrapped along **r** to form a (3,2) CNT.

Four types of Green's functions can be defined as the non-equilibrium statistical ensemble averages of the single particle correlation operator [213]:

$$
\begin{aligned}
G^>(\mathbf{r}, t; \mathbf{r}', t') &= -i\hbar^{-1}\langle \hat{\psi}(\mathbf{r}, t)\hat{\psi}^\dagger(\mathbf{r}', t')\rangle, \\
G^<(\mathbf{r}, t; \mathbf{r}', t') &= +i\hbar^{-1}\langle \hat{\psi}^\dagger(\mathbf{r}', t')\hat{\psi}(\mathbf{r}, t)\rangle, \\
G^R(\mathbf{r}, t; \mathbf{r}', t') &= \theta(t - t')[G^>(\mathbf{r}, t; \mathbf{r}', t') - G^<(\mathbf{r}, t; \mathbf{r}', t')], \\
G^A(\mathbf{r}, t; \mathbf{r}', t') &= \theta(t' - t)[G^<(\mathbf{r}, t; \mathbf{r}', t') - G^>(\mathbf{r}, t; \mathbf{r}', t')],
\end{aligned}
\tag{55}
$$

where $\theta(t)$ is the unit step function, and $\hat{\psi}^\dagger(\mathbf{r}, t)$ and $\hat{\psi}(\mathbf{r}, t)$ are the field operators creating or destroying a particle at point $(\mathbf{r}, t)$ in space–time, respectively. Only three of the Green's functions are independent.

Under steady-state conditions the equation of motion for the Green's functions can be written as [161]:

$$
[E - H(\mathbf{r})]G^R(\mathbf{r}, \mathbf{r}'; E) - \int d\mathbf{r_1} \Sigma^R(\mathbf{r}, \mathbf{r_1}; E)G^R(\mathbf{r_1}, \mathbf{r}'; E) = \delta(\mathbf{r} - \mathbf{r}'),
\tag{56}
$$

$$
G^{\lessgtr}(\mathbf{r}, \mathbf{r}'; E) = \int d\mathbf{r}_1 \int d\mathbf{r}_2 G^R(\mathbf{r}, \mathbf{r}_1; E)\Sigma^{\lessgtr}(\mathbf{r}_1, \mathbf{r}_2; E)[G^R(\mathbf{r}_2, \mathbf{r}'; E)]^\dagger,
\tag{57}
$$

where H is the one-particle Hamiltonian, and Σ^R, $\Sigma^<$, and $\Sigma^>$ are the retarded, lesser, and greater self-energies, respectively. The total self-energy contains contributions due to the coupling of the device to the contacts and other interactions, such as electron–phonon interaction, $\Sigma = \Sigma_c + \Sigma_{\text{el-ph}}$. The self-energies due to the coupling of the device to the contacts are only non-zero at the boundaries [161]

$$
\Sigma_c^R(E) = \tau g_c \tau^\dagger,
\tag{58}
$$

$$
\Sigma_c^<(E) = -2i\Im m[\Sigma_c^R]f_c(E)
\tag{59}
$$

$$
\Sigma_c^>(E) = +2i\Im m[\Sigma_c^R](1 - f_c(E)),
\tag{60}
$$

where τ is the coupling matrix between the device and the contact, g_c is the surface Green's function [161], and $f_c(E)$ is the Fermi–Dirac distribution function at the contact c. Within the self-consistent Born approximation, the self-energies for the electron–phonon interaction are

$$
\Sigma_{\text{el-ph}}^<(\mathbf{r}, \mathbf{r}'; E) = \sum_j \int \frac{d\mathbf{q}}{(2\pi)^3} e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')} D_{\mathbf{q},j}\left(n_{\mathbf{q},j} + \frac{1}{2} \pm \frac{1}{2}\right)G^<(\mathbf{r}, \mathbf{r}'; E \pm \hbar\omega_{\mathbf{q},j}),
\tag{61}
$$

$$
\Sigma_{\text{el-ph}}^>(\mathbf{r}, \mathbf{r}'; E) = \sum_j \int \frac{d\mathbf{q}}{(2\pi)^3} e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')} D_{\mathbf{q},j}\left(n_{\mathbf{q},j} + \frac{1}{2} \pm \frac{1}{2}\right)G^>(\mathbf{r}, \mathbf{r}'; E \mp \hbar\omega_{\mathbf{q},j}),
\tag{62}
$$

$$
\Sigma_{\text{el-ph}}^R(\mathbf{r}, \mathbf{r}'; E) = -\frac{i}{2}\Gamma_{\text{el-ph}}(\mathbf{r}, \mathbf{r}'; E) + P\int \frac{dE'}{2\pi}\frac{\Gamma_{\text{el-ph}}(\mathbf{r}, \mathbf{r}'; E')}{E - E'}.
\tag{63}
$$

$\hbar\omega_{\mathbf{q},j}$ denotes the phonon energy of branch j at the wave-vector $\mathbf{q}$, $n_{\mathbf{q},j}$ is the average phonon occupation number, $D_{\mathbf{q},j}$ is the electron–phonon interaction strength, $\Gamma_{\text{el-ph}} \equiv i(\Sigma_{\text{el-ph}}^> - \Sigma_{\text{el-ph}}^<)$ defines the broadening, and $P\int$ represents the principal part of the integration. The imaginary part of the retarded self-energy broadens the density of states, whereas the real part shifts it. The plus and minus signs in Eq. (61) and Eq. (62) denote the phonon emission and absorption processes, respectively. Assuming that the environment stays in thermal equilibrium, $n_{\mathbf{q},j}$ is given by

$$
n_{\mathbf{q},j} = \frac{1}{\exp(\hbar\omega_{\mathbf{q},j}/k_B T) - 1}.
\tag{64}
$$

These equations must be solved self-consistently with the Poisson equation [214]. If the radius of a CNT is larger than the distance between two carbon atoms, the charge density can be assumed to be distributed uniformly over the surface of the CNT [215]. The self-consistent Born approximation for the self-energy is an extremely time consuming but necessary step, because it guarantees the current continuity. The convergence of the self-consistent iteration is a

critical issue where fine resonances at some energies have to be resolved accurately [215,216]. For that purpose an adaptive method for selecting the energy grid is essential [216].

For numerical solutions it is convenient to transform the equations for the Green's function in the eigen-mode representation [217]. The introduction of CNT subbands makes the equations for the Green's function one-dimensional, which results in a significant reduction of computational time [199,202]. The one-dimensional equation of motion for the Green's functions, Eqs. (56) and (57), are solved for $G^{R,\nu}$ and $G^{\lessgtr,\nu}$ for each subband ν using a recursive Green's function algorithm [163]. Coupling to reservoirs [161,199,203,207] is described by the contact self-energies Eq. (58).

The carrier concentration at a node l of the spatial grid and the current density at the edge between the nodes l and $l+1$ are given by

$$n_l = -4i \sum_\nu \int \frac{dE}{2\pi} G^{<,\nu}_{l,l}(E), \tag{65}$$

$$j_{l,l+1} = \frac{4q}{\hbar} \sum_\nu \int \frac{dE}{2\pi} 2\Re e\{G^{<,\nu}_{l,l+1}(E) t^\nu_{l+1,l}\} \tag{66}$$

where the factor 4 is due to the spin and band degeneracy. In Fig. 16 simulated output characteristics of a CNT-FET with Ohmic contacts [218] are compared to experimental data [219], showing good agreement.

In the case of ballistic transport the current is independent of the device length, but in the presence of scattering it decreases with the CNT length increased. The self-energy due to electron–phonon interaction comprises the contributions of elastic and inelastic scattering mechanisms, $\Sigma^\nu_{\text{e-ph}} = \Sigma^\nu_{\text{el}} + \Sigma^\nu_{\text{inel}}$. By definition the particle energy is conserved for elastic scattering. For inelastic scattering, the after-scattering energy is different from its initial value. The energy difference is carried away by a scattering mediator, e.g. by a phonon. For optical phonons the frequency dependence $\omega_{\mathbf{q}}$ on the wave vector q is weak and it has a finite value ω_0 as $\mathbf{q} \to 0$, therefore, carrier scattering with optical phonons is inelastic. Inelastic interaction of carriers with optical phonons is usually considered local [179]. In this case, the self-energies can be written as

$$\Sigma^{<,\nu}_{\text{inel}}(E) = \sum_j D_{\text{inel},j}\left(n_B(\hbar\omega_j) + \frac{1}{2} \pm \frac{1}{2}\right) G^{<,\nu}(E \pm \hbar\omega_j), \tag{67}$$

$$\Sigma^{>,\nu}_{\text{inel}}(E) = \sum_j D_{\text{inel},j}\left(n_B(\hbar\omega_j) + \frac{1}{2} \pm \frac{1}{2}\right) G^{>,\nu}(E \mp \hbar\omega_j). \tag{68}$$
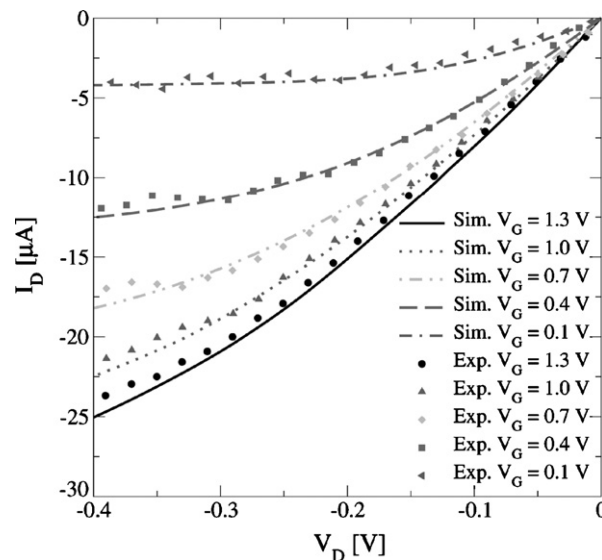

Fig. 16. Output characteristics of CNT-FET compared to experimental data.

The electron–phonon interaction strength is given by

$$D_{\text{inel},j} = \frac{\hbar |M_j^{\text{OP}}|^2}{2nm_c\omega_j}, \tag{69}$$

where m_c is the mass of a carbon atom.

For acoustic phonons $\omega_{\mathbf{q},j} = v_j q$ is proportional to q. Therefore, the acoustic phonon energy at small q is negligible compared to the carrier energy, and carrier scattering with acoustic phonons is frequently considered elastic. The corresponding self-energies for acoustic phonon interaction are

$$\Sigma_{\text{el}}^{\lessgtr,\nu}(E) = D_{\text{el}}^{\nu} G^{\lessgtr,\nu}(E), \tag{70}$$

$$D_{\text{el},j} = \frac{k_B T |M_j^{\text{AP}}|^2}{2nm_c v_j}. \tag{71}$$

The electron–phonon coupling strength and the phonon energy depend on the chirality and the diameter of the CNT. The calculation of these parameters is found in [220,221].

Fig. 17 shows the ballisticity as a function of the electron–phonon coupling strength [222]. The ballisticity is defined as $I_{\text{Sc}}/I_{\text{Bl}}$, the ratio of the on-current in the presence of electron–phonon interaction to the current in the ballistic case [223]. With increasing electron–phonon coupling strength the self-energy increases. Elastic scattering conserves the energy of carriers, but the current decreases due to elastic back-scattering of carriers.

4.3. Dissipative transport: density matrix and Wigner function approach

Alternative approaches which can handle both quantum-mechanical and dissipative scattering effects are based on the reduced density matrix [150,151,224] and on the Wigner function [156,225,226]. The evolution for the reduced electron density matrix is described by a quantum Liouville equation. It is obtained from the equation of motion for the total density matrix by tracing out the degrees of freedom of the environment. To simplify the equation, the limit of weak interaction between the device and the environment is applied. Memory effects can be neglected, if one is interested in system evolution for times larger than the reservoir correlation time [227]. This coarse-graining in time
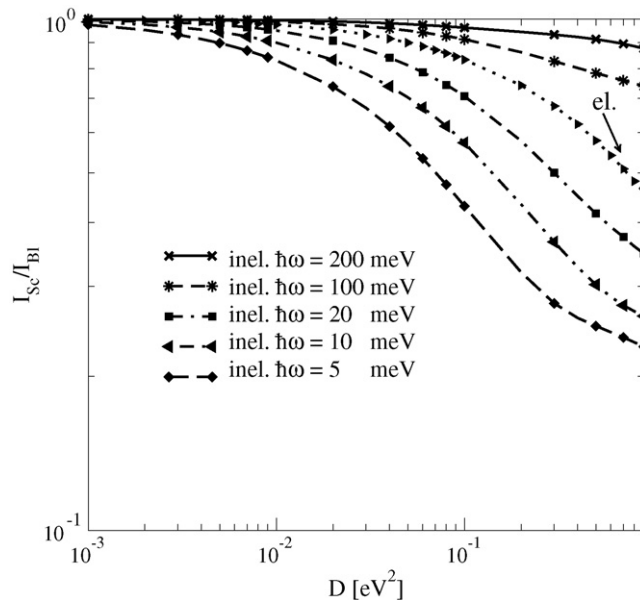

Fig. 17. Ballisticity parameter as function of electron–phonon coupling strength for a CNT of 50 nm length. The ballisticity is defined as the ratio of the on-current in the presence of electron–phonon interaction to the current in the ballistic case $I_{\text{Sc}}/I_{\text{Bl}}$. Results for both elastic and inelastic scattering with different phonon energies are shown. $V_{\text{G}} = V_{\text{D}} = 1$ V.

results in a simplified, Markovian description, when the future dynamics of the system is determined by its current state and is independent from the past. The resulting quantum Liouville equation for the reduced density matrix is of Lindblad form and describes the quantum evolution of a system with loss [228]. The Lindblad form guarantees positive definiteness, normalization, and hermiticity of the reduced density matrix.

Coupling to contacts introduces carrier exchange between device and reservoirs. The difference between the electrochemical potentials of the reservoirs causes current flow through the device. In the approach developed by Fischetti, based on the Pauli master equation, this coupling is introduced in a phenomenological manner [150,151]. Assuming a flat potential profile within each reservoir, electron wave functions are plane waves with well-defined momentum. All electron states may be separated in two distinct groups, representing electrons traveling towards and away from the device, respectively. The role of each reservoir is to supply the incident wave functions with occupation determined by the equilibrium distribution in the reservoir. Therefore, the basis to analyze the dynamics of the reduced density matrix is formed by the scattering states incoming from each reservoir. These scattering states are obtained by a numerical solution of the Schrödinger equation subject to open boundary conditions. Without dissipation in the system, i.e., when the evolution is governed by the system Hamiltonian, the density matrix is diagonal in the basis of the scattering states. Occupation of each scattering state is determined by the equilibrium distribution within the reservoirs. Once the occupations of scattering states are known, the quantum-ballistic current is readily determined. In the presence of weak dissipation, when the device length is shorter than the phase coherence length, the state occupations are found from a solution of the Pauli master equation, which contains only diagonal elements of the density matrix [150,151]. Carrier density and current density are calculated from the occupations of the scattering states. The Pauli master equation and the Poisson equation are solved repeatedly in a self-consistent iteration loop. In this way, subband quantization and rapid potential variations in the transport direction are taken into account self-consistently. The eigenfunctions are used to compute transition rates among the corresponding states. Application of the Pauli master equation is restricted to stationary systems, since in the non-stationary case current continuity would be violated [156].

An interesting solution free from the above mentioned shortcoming, which also allows to circumvent the issue of phenomenological coupling of the device to the reservoirs, was recently suggested by Gebauer and Car [229,230]. They suggest to impose periodic boundary conditions upon the non-perturbed system. A constant electric field is introduced into the system through a vector-potential term, which linearly depends on time. The master equation is solved in two steps. First, the Hamiltonian dynamics of the density matrix is evaluated for a certain time-step t. During this time evolution the vector potential grows linearly with time. It can be scaled out from the Hamiltonian by means of a gauge transformation [231] at the expense of an additional phase factor in the wave function. The time-step at which the gauge transformation is performed is chosen such that the phase added to the wave function satisfies the periodic boundary condition. In the second step, the change of the density matrix due to the Lindblad scattering operator is evaluated. This procedure is repeated until a steady state is achieved. Since the eigenfunctions of the periodic system do not carry any current, the current in the Gebauer and Car formulation is due to off-diagonal elements. This approach can also be used to describe transients. The only limitation is that the time scale is much larger than the reservoir coherence time, a condition arising from coarse graining. It has been shown that in addition to the usual Hamiltonian current component a dissipative current component due to interaction with the environment appears. The total current is then conserved also for transients.

Another approach capable of handling both quantum-coherent propagation and dissipative scattering effects is based on the Wigner function. The Wigner function is defined as the density matrix in a mixed coordinate/momentum representation [225,226]. A practically used approximation to incorporate realistic scattering processes into the Wigner equation is to utilize a properly adapted Boltzmann scattering operator [232]. In this way, well-established scattering models already calibrated within semi-classical transport approaches can be employed in quantum transport calculations. The inclusion of dissipation through the Boltzmann scattering operator, although intuitively appealing, raises some concerns about the validity of such a procedure. The Boltzmann scattering operator is semi-classical by its nature, and represents a good approximation for sufficiently smooth device potentials.

To account for scattering more rigorously, spectral information has to be included into the Wigner function, resulting in an energy-dependence in addition to the momentum dependence [213]. An alternative approach to construct quantum-mechanical extensions of the semi-classical electron–phonon scattering operator has been reported in [233,234]. These quantum collision operators satisfy a quantum H-theorem and relax systems towards quantum equilibria.

The kinetic equation for the Wigner function is similar to the semi-classical Boltzmann equation, except for a non-local potential term. In the case of a slowly varying potential this non-local term reduces to the local classical force term, and the semi-classical description given by the Boltzmann equation is obtained from the Wigner equation. This semi-classical limit of the Wigner transport equation allows to link seamlessly a semi-classical description of the extended contact regions with the quantum description of the active region of a device using the same formalism [232].

Reports on finite difference solutions of the one-particle Wigner equation for device applications were given by Ravaioli [235], Kluksdahl [236], and coworkers, and date back to the mid-1980s. Frensley [237–239] was the first who introduced boundary conditions on the Wigner function to model open quantum systems. Later, self-consistency was added to the Wigner equation solvers [240,241]. Main and Haddad included a reduced Boltzmann scattering operator in transient Wigner function-based simulations [242]. Research on finite difference methods for the Wigner equation culminated in 1990 when review articles of Frensley [156] and Buot and Jensen [243] appeared. In 2002, implementations of Monte Carlo methods for solving the Wigner device equation have been reported [244,245]. With the finite difference method scattering was restricted to the relaxation time approximation and the momentum space to one dimension.

The Monte Carlo method allows scattering processes to be included on a more detailed level, e.g., through the Boltzmann scattering operator, assuming a three-dimensional momentum-space [246,247]. Realistic scattering processes can be easily embedded into the Wigner equation via Boltzmann-like scattering integrals, which turns out to be a good approximation. The Wigner function approach reduces to a semi-classical transport description in contacts providing an important advantage of a seamless treatment between classical and quantum-mechanical regions in device simulations [232].

The Wigner function is given by the density matrix in mixed representation [225,226] defined by the Wigner–Weyl transform

$$f_{\mathrm{w}}(\mathbf{r}, \mathbf{k}, t) = \int \rho\left(\mathbf{r} + \frac{\mathbf{s}}{2}, \mathbf{r} - \frac{\mathbf{s}}{2}, t\right) \exp\left(-\imath \mathbf{k} \cdot \mathbf{s}\right) \mathrm{d}\mathbf{s}.$$

The kinetic equation for the Wigner function is similar to the Boltzmann equation:

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla_r\right) f_{\mathrm{w}} = \int V_{\mathrm{w}}(\mathbf{r}, \mathbf{k}' - \mathbf{k}) f_{\mathrm{w}}(\mathbf{k}', \mathbf{r}, t) \mathrm{d}\mathbf{k}' + \left(\frac{\partial f_{\mathrm{w}}}{\partial t}\right)_{\mathrm{coll}}. \tag{72}$$

The Wigner potential entering into the non-local operator in the right-hand side is defined as

$$V_{\mathrm{w}}(\mathbf{r}, \mathbf{k}) = \frac{1}{\imath \hbar (2\pi)^3} \int \left(V\left(\mathbf{r} - \frac{\mathbf{s}}{2}\right) - V\left(\mathbf{r} + \frac{\mathbf{s}}{2}\right)\right) \exp\left(-i\mathbf{k} \cdot \mathbf{s}\right) \mathrm{d}\mathbf{s}. \tag{73}$$

In case of slowly varying potentials the difference term in the right-hand side of (73) can be developed into the series. Keeping the first non-vanishing terms, one rewrites (73) as

$$V_{\mathrm{w}}(\mathbf{r}, \mathbf{k}' - \mathbf{k}) = -\frac{1}{\hbar (2\pi)^3} \nabla_{\mathbf{r}} V(\mathbf{r}) \frac{\partial}{\partial \mathbf{k}} \int \exp\left(-i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{s}\right) \mathrm{d}\mathbf{s}. \tag{74}$$

After substituting (74) into (72) the term with the non-local potential reduces to the classical force term. Following [248], one can introduce a spectral decomposition of the potential profile $V(x)$ into a slowly varying, classical component and a rapidly changing component treated quantum-mechanically.

$$V(\mathbf{r}) = V_{\mathrm{cl}}(\mathbf{r}) + V_{\mathrm{qm}}(\mathbf{r}). \tag{75}$$

This separation of the total potential into a smooth classical and a small quantum-mechanical contribution can improve the stability of a numerical solution method. The quantum-mechanical contribution may be moved into the right-hand side of the transport equation. Considered as a perturbation, the quantum-mechanical term can be interpreted as a quantum scattering integral. It allows to treat quantum effects on equal footing with classical scattering:

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla_r - \frac{q \nabla_r V_{\mathrm{cl}}(\mathbf{r})}{\hbar} \cdot \nabla_k\right) f_{\mathrm{w}} = \int V_{\mathrm{qm}}(\mathbf{r}, \mathbf{k}' - \mathbf{k}) f_{\mathrm{w}}(\mathbf{k}', \mathbf{r}, t) \mathrm{d}\mathbf{k}' + \left(\frac{\partial f_{\mathrm{w}}}{\partial t}\right)_{\mathrm{coll}}. \tag{76}$$

By applying the method of moments to Eq. (76), the quantum drift-diffusion or quantum hydrodynamic models can be derived [249]. These models are more convenient for the implementation in TCAD device simulators than a Schrödinger–Poisson solver which strongly depends on non-local quantities. However, it was reported that, while the carrier concentration in the inversion layer of a MOSFET is reproduced correctly, the method fails to account properly for tunneling currents [134].

The Wigner function formalism treats scattering and quantum-mechanical effects on equal footing through the corresponding scattering integrals. By analogy to the Monte Carlo methods used for the Boltzmann transport equation, it is tempting to solve the quantum Wigner transport Eq. (76) by means of the Monte Carlo technique. Such a program was recently realized in [115,250–252]. However, since the kernel of the quantum scattering operator is not positively defined, the numerical weight of the particle trajectory increases rapidly, and the numerical stability of the trajectory-based Monte Carlo algorithm becomes a critical issue. A multiple trajectories method was suggested [232] to overcome this difficulty. In the algorithm developed 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 particles with similar statistical properties, introducing a possibility to control the number of trajectories.

This method was recently applied to double-gate MOSFETs [115]. In the coherent mode, where scattering is turned off, a comparison to conventional Schrödinger solvers can be performed. In order to estimate the tunneling component of the current the Wigner Monte Carlo simulations were carried out for a MOSFET with a gate length of 10 nm, and a good agreement between the two approaches was found.

The carrier concentration must be used to update the potential in the device by solving the Poisson equation. A superimposed iteration loop makes the Wigner–Poisson solver self-consistent. An example of self-consistent potentials for n–i–n Si structures with an intrinsic region of length W ranging from 20 to 2.5 nm, as calculated with the Wigner Monte Carlo method and the classical Monte Carlo method is shown in Fig. 18. The doping profile is assumed to increase gradually from the intrinsic channel to the highly doped contacts over the same distance W. Phonon and Coulomb scattering were included. As expected, for long W the classical and quantum-mechanical calculations yield similar results for the self-consistent potential. For $W = 2.5$ nm an extra space charge due to electrons tunneling under the barrier becomes important, which results in a potential barrier increase. Despite of the potential barrier increase, the current in self-consistent Wigner simulations is approximately 20% higher compared to its classical value found by a self-consistent solution of the Boltzmann and the Poisson equations.
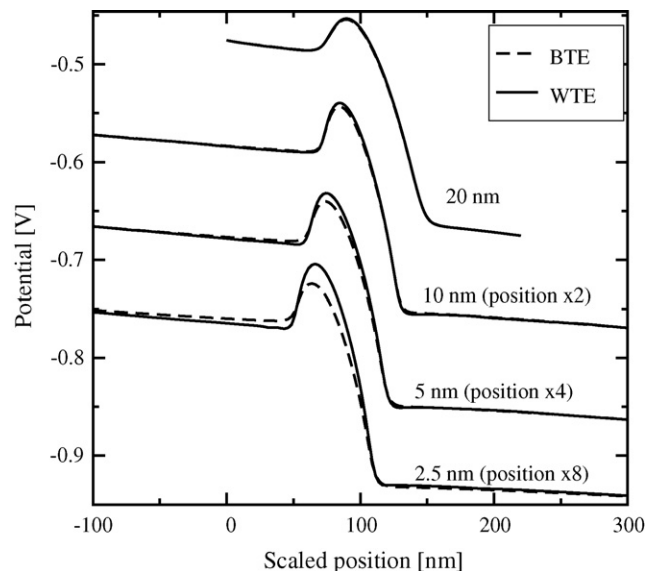

Fig. 18. Self-consistent potential profiles calculated for n–i–n structures with Wigner (solid lines) and Boltzmann (dashed lines) transport equations. For long n–i–n structures the results are similar. For short n–i–n structures the additional charge due to tunneling electrons results in a higher potential barrier.
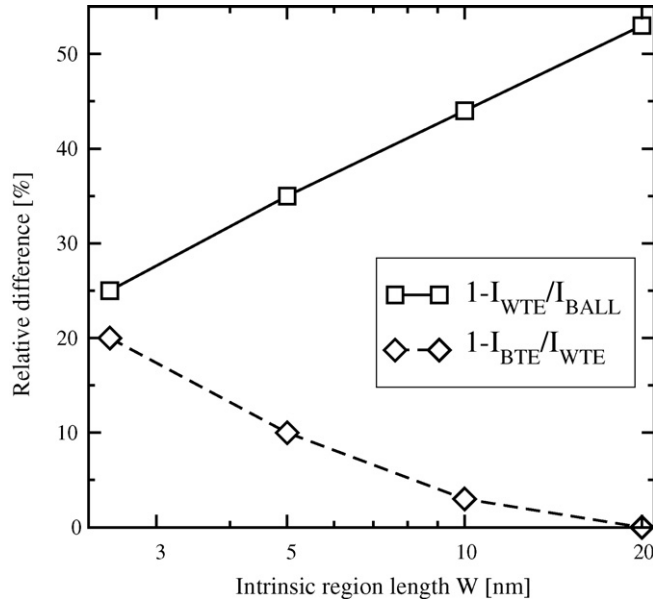
Fig. 19. Relative difference between currents calculated with the Wigner and Boltzmann Monte Carlo methods (diamonds) and calculated with the Wigner Monte Carlo method for n–i–n structures, with and without scattering in the intrinsic region.

Relative differences between I_{WIG} and the current I_{BALL} computed for a "ballistic" device with scattering inside the intrinsic and transition regions turned off is shown in Fig. 19. For $W = 2.5$ nm the relative differences in current due to quantum effects and scattering in the barrier are still of the order of 25% and cannot be neglected.

The Wigner Monte Carlo method gives accurate results not only for single-barrier devices, but it can also be applied to purely quantum-mechanical systems such as resonant tunneling diodes [232]. A typical output characteristic of a GaAs resonant tunneling diode is shown in Fig. 20. Scattering with polar optical phonons as well as Coulomb
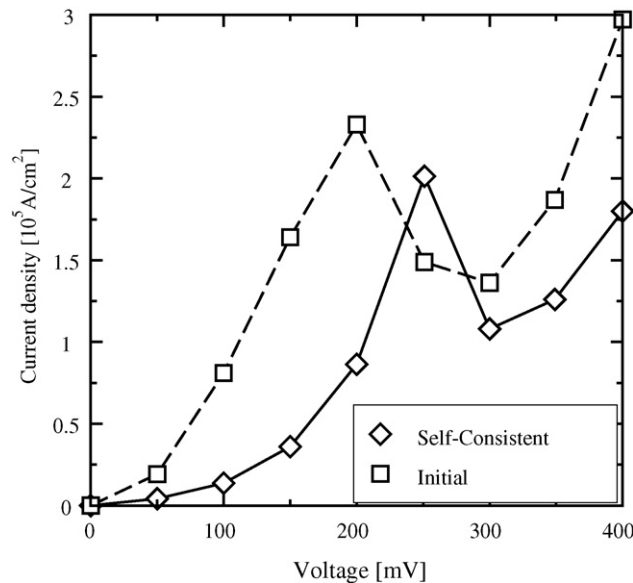

Fig. 20. Typical *IV* curve of a resonant tunneling diode, calculated self-consistently (solid line), contrasted against a not self-consistent characteristics. Charge accumulation before the potential barrier as well as a higher charge density inside the RTD between the potential barriers are responsible for the significant shift of the *IV* resonance peak, thus demonstrating the importance of space–charge effects.
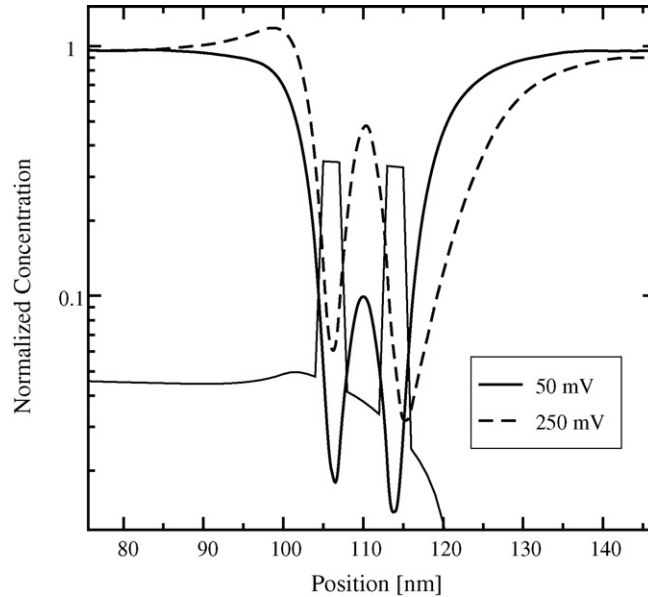
Fig. 21. Normalized electron concentration off-resonance (solid line) and at resonance (dashed line) in RTD.

scattering in the contacts is considered. A region of negative differential resistance common to transport via a resonant level is clearly visible after the resonance peak at 250 mV applied voltage. A self-consistent solution of the Wigner transport and Poisson equation is mandatory for the correct determination of the resonance position due to charge accumulation at the cathode side of the resonant tunneling diode. Before the barrier, an accumulation layer forms, depending on the applied voltage, as seen in Fig. 21. This results in a voltage shift of the resonance peak of the *IV* characteristics shown in Fig. 20. A typical distribution of the concentration in resonance condition and off-resonance is presented in Fig. 21. The amount of charge localized in the potential well is much higher at resonance. This leads to a potential barrier increase and also contributes to the shift of the resonance peak. This example demonstrates the importance of quantum-mechanical effects for the simulation of properties of ultra-scaled devices. It also shows that space charge effects are of crucial importance for the accurate prediction of output characteristics of single- and double-barrier devices.

5. Conclusion and trends

Transport modeling for TCAD applications has grown into a mature field of research, software development, and applications. Models of different complexity, precision and accuracy are offered and implemented in various commercial and academic TCAD tools. Starting with the drift-diffusion model and higher moments models, a complete hierarchy of transport models for semiconductor device simulation has been gradually constructed. Depending on parameter values and device scales, either a semi-classical or a quantum-mechanical transport description has to be adopted. Monte Carlo techniques are used to obtain solutions of the Boltzmann transport equation with arbitrary scattering mechanisms and general band structure. These methods require significant CPU resources and are relatively rarely used for industrial TCAD applications, when timely, but perhaps less accurate, results are of primary importance. Monte Carlo solvers are, however, indispensable for calibration of the parameters of higher order moment models. With quantum corrections carefully added, full-band Monte Carlo methods can provide accurate results in ultra-scaled devices with strong size quantizations, although the use of the subband Monte Carlo in ultra-thin body FETs is a must to include intersubband scattering correctly.

With the channel length of MOSFETs in the decananometer region, the development of conceptually new devices and architectures is becoming increasingly important. New nanoelectronic structures, utilizing carbon nanotubes, nanowires, and molecules are considered to be the most prominent candidates for the post-CMOS era. Nanoelectronic devices are expected to complement and substitute some of the current CMOS functions. New physical principles for

carrier propagation must be considered by transport models for nanodevices. Indeed, at this small device size the geometrical spread of the carrier wave packet cannot be ignored, as this spread approaches the device size. Thus, the complete information about carrier dynamics inside the device including the phase of the wave function is needed. Although many particles with different energies and phases are injected into a device, the phases of an individual carrier may not average out completely if the potential in transport direction is changing rapidly. A full quantum-mechanical description is needed, when the curvature of the potential close to its maximum is so big that tunneling under the potential barrier contributes significantly to the current.

When the channel length is smaller than the coherence length current transport inside the device becomes quasi-ballistic. The coherence length is limited by scattering with phonons and is only a few nanometers in modern MOSFETs operating at room temperature. Therefore, transport in 10 nm long devices is still not ballistic. However, if the potential changes rapidly enough on the scale of the carrier wavelength, a quantum-mechanical description of electron motion between two scattering events must be adopted. The interplay between coherent propagation and scattering determines transport in advanced nanodevices. Methods for dissipative quantum transport are based on the non-equilibrium Green's function formalism, the Liouville/von-Neumann equation for the density matrix, and the kinetic equation for the Wigner function. All the numerical methods for quantum transport are time consuming, and highly accurate simulations of emerging nanoscale devices represent an outstanding modeling and computational challenge. Development of new efficient numerical algorithms as well as a comprehensive comparison between different quantum-mechanical models describing dissipative transport in open systems is mandatory and currently on the research agenda.

In modern microelectronic devices quantum effects are usually dominant in a small active region connected to relatively large, heavily doped contact areas where the carrier dynamics is essentially classical. Modern TCAD simulators must be able to incorporate both semi-classical and (dissipative) quantum-mechanical modeling approaches within the same body. It is, therefore, necessary to have an accurate, multiscale simulation technique which is also able to bridge the gap between semi-classical and quantum-mechanical simulations.

With shrinking device dimensions, the demand for full three-dimensional accurate solvers for the coupled transport/Poisson equations [24,165,253] and atomistic based simulations [254,255] has grown significantly. With the advances in computer architectures, increased computational power and memory capabilities, state-of-the art software, development of fast numerical algorithms and conceptually new generic simulation platforms a fundamental breakthrough in speed, reliability, and accuracy of multiscale three-dimensional TCAD simulation tools is anticipated. This allows careful design of the macroscopic transport properties of emerging nanoelectronic and molecular devices based on their atomistic electronic structure.

Acknowledgments

References

[1] International technology roadmap for semiconductors: 2005 edition, http://www.itrs.net/Common/2005ITRS/Home2005.htm, 2005.
[2] B. Doris, M. Ieong, T. Kanarsky, Y. Zhang, R.A. Roy, O. Documaci, Z. Ren, F.-F. Jamin, L. Shi, W. Natzle, H.-J. Huang, J. Mezzapelle, A. Mocuta, S. Womack, M. Gribelyuk, E.C. Jones, R.J. Miller, H.-S.P. Wong, W. Haensch, IEDM Techn. Dig. (2002) 267–270.
[3] S. Selberherr, Analysis and Simulation of Semiconductor Devices, Springer, Wien, New York, 1984.
[4] T. Ando, A.B. Fowler, F. Stern, Rev. Modern Phys, 54 (2) (1982) 437–672.
[5] W. Haensch, E.J. Novak, R.H. Dennard, P.M. Solomon, A. Bryant, O.H. Documaci, A. Kumar, X. Wang, J.B. Johnson, M.V. Fischetti, IBM J. Res. Dev. 50 (4/5) (2006) 339–361.
[6] J.G. Fossum, Solid-State Electron. (2007) 188–194.
[7] V.A. Sverdlov, T.J. Walls, K.K. Likharev, IEEE Trans. Electron Devices 50 (9) (2003) 1926–1933.
[8] P. Palestri, D. Esseni, S. Eminente, C. Fiegna, E. Sangiorgi, L. Selmi, IEEE Trans. Electron Devices 52 (12) (2005) 2727–2735.
[9] M. Gilbert, R. Akis, D. Ferry, Phonon-assisted ballistic to diffusive crossover in silicon nanowire transistors, J. Appl. Phys. 98 (9) (2005), p. 094303-1-8.
[10] H. Gummel, IEEE Trans. Electron Devices ED-11 (1964) 455–465.
[11] A. DeMari, Solid-State Electron. 11 (1968) 33–58.

[12] D. Scharfetter, H. Gummel, IEEE Trans. Electron Devices 16 (1) (1969) 64–77.

[13] H. Loeb, R. Andrew, W. Love, Electron. Lett. 4 (1968) 352–354.

[14] J. Schroeder, R. Muller, IEEE Trans. Electron Devices ED-15 (12) (1968) 954–961.

[15] D. Kennedy, Proc. IEEE 54 (6) (1969) 1202–1203.

[16] J. Slotboom, Electron. Lett. 5 (1969) 677–678.

[17] M. Lundstrom, Fundamentals of Carrier Transport, Cambridge University Press, Cambridge, 2000.

[18] S. Datta, Quantum Transport: Atom to Transistor, Cambridge University Press, Cambridge, 2005.

[19] ATLAS User's Manual, Silvaco, 2007.

[20] TAURUS MEDICI User's Manual, Synopsys, 2007.

[21] DESSIS 9.0 User's Manual, Synopsys, 2007.

[22] Online simulations and more, http://www.nanohub.org, 2007.

[23] MINIMOS-NT 2.1 User's Guide, Institut für Mikroelektronik, Technische Universität Wien, Austria, 2004.

[24] A. Trellakis, T. Zibold, T. Andalauer, S.B.A.K. Smith, R. Morschal, P. Vogl, Proc. IWCE, Wien, (2006), pp. 173–174.

[25] L.P. Kadanoff, G. Baym, Quantum Statistical Mechanics, W.A. Benjamin, New York, 1962.

[26] K. Bløtekjaer, IEEE Trans. Electron Devices 17 (1) (1970) 38–47.

[27] R. Stratton, Phys. Rev. 126 (6) (1962) 2002–2014.

[28] T. Grasser, H. Kosina, C. Heitzinger, S. Selberherr, J. Appl. Phys. 91 (6) (2002) 3869–3879.

[29] A. Gehring, T. Grasser, H. Kosina, S. Selberherr, J. Appl. Phys. 92 (10) (2002) 6019–6027.

[30] M. Gritsch, H. Kosina, T. Grasser, S. Selberherr, Solid-State Electron. 45 (4) (2001) 621–627.

[31] M. Gritsch, Numerical modeling of SOI MOSFETs, Dissertation, Technische Universität Wien, http://www.iue.tuwien.ac.at/phd/gritsch, 2002.

[32] M. Gritsch, H. Kosina, T. Grasser, S. Selberherr, IEEE Trans. Electron Devices 49 (10) (2002) 1814–1820.

[33] T. Grasser, H. Kosina, S. Selberherr, Intl. J. High Speed Electron. 13 (3) (2003) 873–901.

[34] H. Kosina, S. Selberherr, Intl. J. High Speed Electron. 16 (1) (2006) 115–136.

[35] T. Grasser, R. Kosik, C. Jungemann, H. Kosina, S. Selberherr, J. Appl. Phys. 97 (9) (2005) 0937101–09371012.

[36] K.-I. Sonoda, M. Yamaji, K. Taniguchi, C. Hamaguchi, S.T. Dunham, J. Appl. Phys. 80 (9) (1996) 5444–5448.

[37] T. Grasser, H. Kosina, S. Selberherr, J. Appl. Phys 90 (12) (2001) 6165–6171.

[38] I.M. Abdel-Motaleb, R.Y. korotkov, J. Appl. Phys. 97 (2005), 03715-1-6.

[39] S. Vitanov, V. Palankovski, S. Murad, T. Rödle, R. Quay, S. Selberherr, Proc. Intl. Conf. on Simulation of Semiconductor Processes and Devices, 2007, 273–276.

[40] L. Ma, Y.-W. Jin, C. Zeng, D.W. Barlage, Proc. Intl. Semiconductor Device Research Symposium, 2005, 338–339.

[41] E. Sijercic, B. Pejcinovic, Proc. Intl. Conference Applied Electromagnetics and Communications, 2005, 1–4.

[42] T. Grasser, C. Jungemann, H. Kosina, B. Meinerzhagen, S. Selberherr, Proc. Intl. Conf. on Simulation of Semiconductor Processes and Devices, 2004, 1–8.

[43] T. Tang, R. Ramaswamy, J. Nam, IEEE Trans. Electron devices 40 (8) (1993) 1469–1477.

[44] T. Grasser, H. Kosina, M. Gritsch, S. Selberherr, J. Appl. Phys. 90 (5) (2001) 2389–2396.

[45] S.E. Laux, R.J. Lomax, IEEE Trans. Electron Devices (1981) 120–122.

[46] M. Duane, IEICE Trans. Electron. 82-C (6) (1999) 976–982.

[47] T.-W. Tang, IEEE Trans. Electron Devices 31 (1984) 1912–1914.

[48] C.C. McAndrew, K. Singhal, E.L. Haesell, IEEE Electron Device Lett. 6 (1985) 446–447.

[49] A. Forgieri, R. Guerrieri, P. Ciamolini, A. Gnudi, M. Rudan, G. Baccarani, IEEE Trans. Computer-Aided Design 7 (1988) 231–242.

[50] W.-S. Choi, J.-G. Ahn, Y.-J. Park, H.-S. Min, G.-G. Hwang, IEEE Trans. Computer-Aided Design 13 (1994) 899–908.

[51] T.-W. Tang, M.-K. Ieong, IEEE Trans. Computer-Aided Design 14 (1995) 1309–1315.

[52] W. Hänsch, The Drift Diffusion Equation and its Applications in MOSFET Modeling, Springer, Wien, New York, 1991.

[53] T. Grasser, Physica A 349 (2005) 221–258.

[54] J. Welser, J. Hoyt, J. Gibbons, IEDM Techn. Dig. (1992) 1000–1002.

[55] I. Balslev, Phys. Rev. 143 (1966) 636–647.

[56] T. Manku, A. Nathan, IEEE Trans. Electron Devices 39 (9) (1992) 2082–2089.

[57] J. Egley, D. Chidambarao, Solid-State Electron. 36 (12) (1993) 1653–1664.

[58] S. Dhar, H. Kosina, V. Palankovski, E. Ungersboeck, S. Selberherr, IEEE Trans. Electron Devices 52 (4) (2005) 527–533.

[59] C. Jacoboni, L. Reggiani, Rev. Modern Phys. 55 (3) (1983) 645–705.

[60] H. Smith, J.W. Wilkins, Phys. Rev. Lett. 24 (5) (1970) 221–225.

[61] S. Dhar, E. Ungersboeck, H. Kosina, T. Grasser, S. Selberherr, IEEE Trans. Nanotechnol. 6 (1) (2007) 97–100.

[62] J.M. Luttinger, W. Kohn, Phys. Rev. 97 (4) (1955) 869–883.

[63] M.V. Fischetti, Z. Ren, P.M. Solomon, M. Yang, K. Rim, J. Appl. Phys. 94 (2) (2003) 1079–1095.

[64] C. Jacoboni, P. Lugli, The Monte Carlo Method for Semiconductor Device Simulation, Springer, Wien, New York, 1989.

[65] R. Hockney, J.W. Eastwood, Computer Simulation Using Particles, Adam Hilger, Bristol and Philadelphia, 1988.

[66] C. Moglestue, Comput. Methods Appl. Mech. Eng (1982) 173–208.

[67] P.J. Price, Semicond. Semimet. 14 (1979) 249–308.

[68] C. Jacoboni, P. Poli, L. Rota, Solid-State Electron. 31 (3/4) (1988) 523–526.

[69] M. Nedjalkov, P. Vitanov, Solid-State Electron. 32 (10) (1989) 893–896.

[70] H. Kosina, M. Nedjalkov, S. Selberherr, A stable backward Monte Carlo method for the solution of the Boltzmann equation, in: Lecture Notes in Computer Science 2907: Large-Scale Scientific Computing, Springer Verlag, 2003, pp. 170–177.

[71] H. Kosina, M. Nedjalkov, S. Selberherr, IEEE Trans. Electron Devices 47 (10) (2000) 1899–1908.
[72] T. Kurosawa, Proc. Intl. Conf. on Physics of Semiconductors, 1966, 424–426.
[73] W. Fawcett, A. Boardman, S. Swain, J. Phys. Chem. Solids 31 (1970) 1963–1990.
[74] W. Fawcett, E. Paige, J. Phys. C: Solid State Physics 4 (1971) 1801–1821.
[75] C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, A. Quaranta, Phys. Rev. B 12 (4) (1975) 2265–2284.
[76] C. Jacoboni, R. Minder, G. Majni, J. Phys. Chem. Solids 36 (1975) 1129–1133.
[77] H. Shichijo, K. Hess, Phys. Rev. B 23 (8) (1981) 4197–4207.
[78] M.V. Fischetti, S.E. Laux, Phys. Rev. B 38 (14) (1988) 9721–9745.
[79] T. Kunikiyo, M. Takenaka, Y. Kamakura, M. Yamaji, H. Mizuno, M. Morifuji, K. Taniguchi, C. Hamaguchi, J. Appl. Phys. 75 (1) (1994) 297–312.
[80] P. Yoder, J. Higman, J. Bude, K. Hess, Semicond. Sci. Technol. 7 (3B) (1992) 357–359.
[81] M. Fischetti, S. Laux, Proc. European Solid-State Device Research Conf., 1996, 813–820.
[82] P.D. Yoder, V.D. Natoli, R.M. Martin, J. Appl. Phys. 73 (9) (1993) 4378–4383.
[83] M.V. Fischetti, F. Gámiz, W. Hänsch, J. Appl. Phys. 92 (12) (2002) 7320–7324.
[84] C. Jungemann, B. Meinerzhagen, Hierarchical Device Simulation, The Monte Carlo Perspective, Springer, 2003.
[85] Vienna Monte Carlo 2.0 User's Guide, Institut für Mikroelektronik, http://www.iue.tuwien.ac.at/software, Technische Universität Wien, Austria, 2006.
[86] M.V. Fischetti, S.E. Laux, J. Appl. Phys. 80 (4) (1996) 2234–2252.
[87] M.V. Fischetti, S.E. Laux, Proc. 26th European Solid State Device Research Conference, 1996, 813–820.
[88] S. Dhar, H. Kosina, G. Karlowatz, E. Ungersboeck, T. Grasser, S. Selberherr, IEEE Trans. Electron Devices 53 (12) (2006) 3054–3062.
[89] C.S. Smith, Phys. Rev. 94 (1) (1954) 42–49.
[90] E. Ungersboeck, S. Dhar, G. Karlowatz, V. Sverdlov, H. Kosina, S. Selberherr, IEEE Trans. Electron Devices.
[91] V. Sverdlov, E. Ungersboeck, H. Kosina, S. Selberherr, Proc. EUROSOI, 2007, 39–40.
[92] G. Karlowatz, W. Wessner, H. Kosina, 5th Mathmod Vienna Proceedings (ISBN: 3–901608-30–3), 2006, pp. 4–1–46–46.
[93] S.E. Laux, IBM Research Report, RC 20101 Enginnering and Technology), 1995, pp. 1–37.
[94] K.K. Likharev, Advanced Semiconductor and Organic Nano-Techniques, Academic Press, New York, 2003, pp. 239–302.
[95] S. Bosi, C. Jacoboni, J. Phys. C:Solid State Phys. 9 (1976) 315–319.
[96] P. Lugli, D.K. Ferry, IEEE Trans. Electron Devices 32 (11) (1985) 2431–2437.
[97] L. Lucci, P. Palestri, D. Esseni, L. Selmi, IEDM Techn. Dig. (2005) 531–534.
[98] E. Ungersboeck, H. Kosina, Proc. Intl. Conf. on Simulation of Semiconductor Processes and Devices, 2005, 311–314.
[99] S. Smirnov, H. Kosina, M. Nedjalkov, S. Selberherr, J. Appl. Phys. 94 (9) (2003) 5791–5799.
[100] V. Sverdlov, E. Ungersboeck, H. Kosina, S. Selberherr, Solid-State Electron. 51 (2) (2007) 299–305.
[101] C. Jungemann, A.T. Pham, B. Meinerzhagen, Proc. IWCE, Wien, (2006), pp. 13–14.
[102] L. Reggiani, P. Lugli, S. Gantsevich, V. Gurevich, R. Katilius, Phys. Rev. B 40 (18) (1989) 12209–12214.
[103] S.I. Takagi, A. Toriumi, M. Iwase, H. Tango, IEEE Trans. Electron Devices 41 (12) (1994) 2357–2362.
[104] F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini, T. Elewa, IEEE Electron Device Lett. 8 (9) (1987) 410–412.
[105] G. Tsutsui, M. Saitoh, T. Saraya, T. Nagumo, T. Hiramoto, IEDM Techn. Dig. (2005) 747–750.
[106] K. Uchida, J. Koga, S. Takagi, IEDM Techn. Dig. (2003) 805–808.
[107] M.J. van Dort, P.H. Woerlee, A.J. Walker, Solid-State Electron. 37 (3) (1994) 411–414.
[108] C. Jungemann, C.D. Nguyen, B. Neinhüs, S. Decker, B. Meinerzhagen, Proc. Intl. Conf. Modeling and Simulation of Microsystems, 2001, 458–461.
[109] C.D. Nguyen, C. Jungemann, B. Meinerzhagen, Proc. Nanotechnol. 3 (2005) 33–36.
[110] W. Hänsch, T. Vogelsang, R. Kircher, M. Orlowski, Solid-State Electron. 32 (10) (1989) 839–849.
[111] G. Paasch, H. Übensee, Phys. Stat. Sol. (b) 118 (1) (1983) 255–266.
[112] M. Wagner, M. Karner, T. Grasser, Proc. of the XIII Intl. Workshop on Semiconductor Devices, 2005, 458–459.
[113] M. Wagner, M. Karner, T. Grasser, H. Kosina, Proc. IWCE, Wien, (2006), pp. 87–88.
[114] M. Karner, M. Wagner, T. Grasser, H. Kosina, Proc. MRS Spring Meeting, San Francisco, (2006), pp. 104–105.
[115] V. Sverdlov, A. Gehring, H. Kosina, S. Selberherr, Solid-State Electron. 49 (9) (2005) 1510–1515.
[116] F. Stern, Phys. Rev. B 5 (12) (1972) 4891–4899.
[117] Vienna Schrödinger–Poisson Solver 1.0 User's Guide, Institut für Mikroelektronik, http://www.iue.tuwien.ac.at/software, Technische Universität Wien, Austria, 2007.
[118] D. Ferry, R. Akis, D. Vasileska, IEDM Tech. Dig. (2000) 287–290.
[119] D. Vasileska, D. Ferry, S. Goodnick, Handbook of Theoretical and Computational Nanotechnology, Vol. 10, American Scientific Publishers, Los Angeles, 2006, pp. 1–135 (Chapter 1. Computational nanoelectronics).
[120] G.J. Iafrate, H.L. Grubin, D.K. Ferry, J. Phys. 42 (1981) 307–312.
[121] G. Kathawala, B. Winstead, U. Ravaioli, IEEE Trans. Electron Devices 50 (12) (2003) 2467–2473.
[122] V. Aubry-Fortuna, P. Dollfus, S. Galdin-Retailleau, Solid-State Electron. 49 (8) (2005) 1320–1329.
[123] T. Krishnamohan, C. Jungemann, D. Kim, E. Ungersboeck, S. Selberherr, P. Wong, Y. Nishi, K. Saraswat, IEDM Tech. Dig. (2006) 937–940.
[124] X.F. Fan, L.F. Register, B. Winstead, M.C. Foisy, W.Q. Chen, X. Zheng, B. Ghosh, S.K. Banerjee, IEEE Trans. Electron Devices 54 (2) (2007) 291–296.
[125] F.M. Bufler, R. Hudé, A. Erlebach, Proc. IWCE, Wien, (2006), pp. 101–102.
[126] P. Palestri, S. Eminente, D. Esseni, C. Fiegna, E. Sangiorgi, L. Selmi, Solid-State Electron. 49 (2005) 727–732.

[127] M.G. Ancona, H.F. Tiersten, Phys. Rev. B 39 (13) (1989) 9536–9540.
[128] M.G. Ancona, Phys. Rev. B 42 (2) (1990) 1222–1233.
[129] M.G. Ancona, Z. Yu, R.W. Dutton, P.J.V. Voorde, M. Cao, D. Vook, Proc. Intl. Conf. on Simulation of Semiconductor Processes and Devices, 1999, 235–238.
[130] M.G. Ancona, Z. Yu, R.W. Dutton, P.J.V. Voorde, M. Cao, D. Vook, IEEE Trans. Electron Devices 47 (12) (2000) 2310–2319.
[131] A. Asenov, A.R. Brown, J.R. Watling, Solid-State Electron. 47 (7) (2003) 1141–1145.
[132] A. Jungel, Quasi-hydrodynamic semiconductor equations, Vol. 41 of Progress in Nonlinear Differential Equations and Their Applications, Birkhauser, 2001.
[133] M.G. Ancona, Proc. IWCE, Wien, (2006), pp. 9–10.
[134] T. Hoehr, A. Schenk, A. Wettstein, W. Fichtner, Proc. Simulation of Semiconductor Processes and Devices, 2002, 275–278.
[135] K. Natori, J. Appl. Phys. 76 (8) (1994) 4879–4890.
[136] Y. Naveh, K.K. Likharev, IEEE Electron Device Lett. 21 (5) (2000) 242–244.
[137] S. Laux, A. Kumar, M. Fischetti, IEEE Trans. Nanotechnol. 1 (4) (2002) 255–259.
[138] F. Heinz, A. Schenk, A. Scholze, W. Fichtner, J. Comput. Electron. 1 (1) (2002) 161–164.
[139] G. Curatola, G. Fiori, G. Iannaccone, Solid-State Electron. 48 (4) (2004) 581–587.
[140] J. Wang, E. Polizzi, A. Ghosh, S. Datta, M. Lundstrom, J. Appl. Phys. (2005) 0431011–0431013.
[141] T.J. Walls, V.A. Sverdlov, K.K. Likharev, Solid-State Electron. 48 (2004) 857–865.
[142] C. Lent, D. Kirkner, J. Appl. Phys. 67 (10) (1990) 6353–6359.
[143] W. Frensley, Superlattices Microstruct. 11 (3) (1992) 347–350.
[144] R. Lake, G. Klimeck, R.C. Bowen, D. Jovanovic, J. Appl. Phys. 81 (12) (1997) 7845–7869.
[145] M. Sabathil, S. Hackenbuchner, J.A. Majewski, G. Zandler, P. Vogl, J. Comput. Electron. 1 (2002) 81–85.
[146] D. Mamaluy, M. Sabathil, P. Vogl, J. Appl. Phys. 93 (8) (2003) 4628–4633.
[147] L. Risch, Proc. European Solid-State Device Research Conf. (2005) 63–68.
[148] R. Tsu, L. Esaki, Appl. Phys. Lett. 22 (11) (1973) 562–564.
[149] M. Karner, A. Gehring, S. Holzer, M. Pourfath, M. Wagner, H. Kosina, T. Grasser, S. Selberherr, Proc. IWCE, Wien, (2006), pp. 255–256.
[150] M.V. Fischetti, J. Appl. Phys. 83 (1) (1998) 270–291.
[151] M.V. Fischetti, Phys. Rev. B 59 (7) (1999) 4901–4917.
[152] A. Szafer, A.D. Stone, Phys. Rev. Lett. 62 (3) (1989) 300–303.
[153] R. Venugopal, S. Goasguen, S. Datta, M.S. Lundstrom, J. Appl. Phys. 95 (1) (2004) 292–305.
[154] V.A. Sverdlov, X. Oriols, K.K. Likharev, IEEE Trans. Nanotechnol. 2 (1) (2003) 59–63.
[155] R. Zhibin, R. Venugopal, S. Goasguen, S. Datta, M.S. Lundstrom, IEEE Trans. Electron Devices 50 (9) (2003) 1914–1925.
[156] W. Frensley, Rev. Modern Phys. 62 (3) (1990) 745–791.
[157] S.E. Laux, A. Kumar, M.V. Fischetti, J. Appl. Phys. 95 (10) (2004) 5546–5582.
[158] F. Rossi, A. Di Carlo, P. Lugli, Phys. Rev. Lett. 80 (15) (1998) 3348–3351.
[159] R.P. Zaccaria, F. Rossi, Phys. Rev. B 67 (2003) 113311-1-4.
[160] G. Ferrari, P. Bordone, C. Jacoboni, Phys. Lett. A 356 (2006) 371–375.
[161] S. Datta, Electronic Transport in Mesoscopic Systems, Cambridge University Press, Cambridge, 1995.
[162] C. Jungemann, N. Subba, J.-S. Goo, C. Riccobene, Q. Xiang, B. Meinerzhagen, Solid-State Electron. 48 (8) (2004) 1417–1422.
[163] A. Svizhenko, M.P. Anantram, T.R. Govindan, B. Biegel, R. Venugopal, J. Appl. Phys. 91 (2002) 2343–2354.
[164] A. Martinez, A. Svizhenko, M.P. Anantram, J.R. Barker, A.R. Brown, A. Asenov, IEDM Techn. Dig. (2005) 627–630.
[165] A. Martinez, J.R. Barker, M.P. Anantram, A. Svizhenko, A. Asenov, Proc. IWCE, Wien, (2006), pp. 275–276.
[166] A. Svizhenko, M.P. Anantram, IEEE Trans. Electron Devices 50 (2003) 1459–1466.
[167] R. Saito, G. Dresselhaus, M. Dresselhaus, Physical Properties of Carbon Nanotubes, Imperial College Press, London, 1998.
[168] R. Martel, T. Schmidt, H.R. Shea, T. Hertel, P. Avouris, Appl. Phys. Lett. 73 (17) (1998) 2447–2449.
[169] W. Hoenlein, F. Kreupl, G. Duesberg, A. Graham, M. Liebau, R. Seidel, E. Unger, IEEE Trans. Components Packag. Technol. 27 (4) (2004) 629–634.
[170] J. Appenzeller, M. Radosavljevic, J. Knoch, P. Avouris, Phys. Rev. Lett. 92 (2004) 048301.
[171] A. Javey, J. Guo, Q. Wang, M. Lundstrom, H. Dai, Lett. Nat. 424 (6949) (2003) 654–657.
[172] C. Caroli, R. Combescot, P. Nozieres, D. Saint-James, J. Phys. C:Solid State Phys. 4 (8) (1971) 916–929.
[173] C. Caroli, R. Combescot, D. Lederer, P. Nozieres, D. Saint-James, J. Phys. C:Solid State Phys. 4 (16) (1971) 2598–2610.
[174] R. Combescot, J. Phys. C:Solid State Phys. 4 (16) (1971) 2611–2622.
[175] C. Caroli, R. Combescot, P. Nozieres, D. Saint-James, J. Phys. C:Solid State Phys. 5 (1) (1972) 21–42.
[176] G. Kim, G.B. Arnold, Phys. Rev. B 38 (5) (1988) 3252–3262.
[177] E.V. Anda, F. Flores, J. Phys.: Cond. Matter 3 (46) (1991) 9087–9101.
[178] L.Y. Chen, C.S. Ting, Phys. Rev. B 43 (5) (1991) 4534–4537.
[179] R. Lake, S. Datta, Phys. Rev. B 45 (12) (1992) 6670–6685.
[180] C.H. Grein, E. Runge, H. Ehrenreich, Phys. Rev. B 47 (19) (1993) 12590–12597.
[181] G. Kim, H. Suh, E. Lee, Phys. Rev. B 52 (4) (1995) 2632–2639.
[182] G. Klimeck, R. Lake, C. Bowen, W.R. Frensley, T.S. Moise, Appl. Phys. Lett. 67 (17) (1995) 2539–2541.
[183] C. Bowen, C.L. Fernando, G. Klimeck, A. Chatterjee, D. Blanks, R. Lake, J. Hu, J. Davis, M. Kulkarni, S. Hattangady, I.-C. Chen, IEDM Techn. Dig. (1997) 35.1.1–3514.3514.3514.
[184] M.J. McLennan, Y. Lee, S. Datta, Phys. Rev. B 43 (17) (1991) 13846–13884.

[185] C. Rivas, R. Lake, G. Klimeck, W.R. Frensley, M.V. Fischetti, P.E. Thompson, S.L. Rommel, P.R. Berger, Appl. Phys. Lett. 78 (8) (2001) 814–916.
[186] C. Rivas, R. Lake, W.R. Frensley, G. Klimeck, P.E. Thomson, K.D. Hobart, S.L. Rommel, P.R. Berger, J. Appl. Phys. 94 (8) (2003) 5005–5013.
[187] R. Venugopal, M. Paulsson, S. Goasguen, S. Datta, M.S. Lundstrom, J. Appl. Phys. 93 (9) (2003) 5613–5625.
[188] M.B. Nardelli, Phys. Rev. B 60 (11) (1999) 7828–7833.
[189] J. Taylor, H. Guo, J. Wang, Phys. Rev. B 63 (2001) 245407.
[190] H. Mehrez, J. Taylor, H. Guo, J. Wang, C. Roland, Phys. Rev. Lett. 84 (12) (2000) 2682–2685.
[191] D. Orlikowski, H. Mehrez, J. Taylor, H. Guo, J. Wang, C. Roland, Phys. Rev. B 63 (2001) 155412.
[192] M.P. Anantram, Appl. Phys. Lett. 78 (14) (2001) 2055–2057.
[193] A. Maiti, A. Svizhenko, M.P. Anantram, Phys. Rev. Lett. 88 (2002) 126805.
[194] C.-C. Kaun, B. Larade, H. Mehrez, J. Taylor, H. Guo, Phys. Rev. B 65 (2002) 205416.
[195] G. Cuniberti, R. Gutiérrez, G. Fagas, F. Grossmann, K. Richter, R. Schmidt, Physica E 12 (1-4) (2002) 749–752.
[196] J.J. Palacios, A.J.P. Jimenez, E. Louis, E. SanFabioa, J. Verges, Phys. Rev. Lett. 90 (2003) 106801.
[197] Y. Xue, M.A. Ratner, Appl. Phys. Lett. 83 (12) (2003) 2429–2431.
[198] T.-S. Xia, L.F. Register, S.K. Banerjee, J. Appl. Phys. 95 (3) (2004) 1597–1599.
[199] J. Guo, S. Datta, M. Lundstrom, M. Anantram, Int. J. Multiscale Comput. Eng. 2 (2004) 257.
[200] J. Guo, J. Appl. Phys. 98 (2005) 063519.
[201] J. Guo, S. Hassan, A. Javey, G. Bosman, M. Lundstrom, IEEE Trans. Nanotechnol. 4 (6) (2005) 715–721.
[202] A. Svizhenko, M. Anantram, T. Govindan, IEEE Trans. Nanotechnol. 4 (5) (2005) 557–562.
[203] A. Svizhenko, M. Anantram, Phys. Rev. B 72 (2005) 085430–085440.
[204] M. Brandbyge, J.-L. Mozos, P. Ordejón, J. Taylor, K. Stokbro, Phys. Rev. B 65 (2002) 165401.
[205] E. Louis, J.A. Vergés, J.J. Palacios, A.J. Pérez-Jiménez, E. SanFabián, Phys. Rev. B 67 (2003) 155321.
[206] W. Tian, S. Datta, S. Hong, R. Reifenberger, J.I. Henderson, C.P. Kubiak, J. Chem. Phys. conf 109 (7) (1998) 2874–2882.
[207] P.S. Damle, A.W. Ghosh, S. Datta, Phys. Rev. B 64 (2001) 201403.
[208] J.M. Seminario, L.E. Cordova, P.A. Derosa, Proc. IEEE 91 (11) (2003) 1958–1975.
[209] M. Galperin, A. Nitzan, S. Sek, M. Majda, J. Electroanalytical Chem. 550/551 (2003) 337–350.
[210] P.A. Derosa, S. Guda, J.M. Seminario, J. Am. Chem. Soc. 125 (47) (2003) 14240–14241.
[211] Y. Xue, M.A. Ratner, Phys. Rev. B 69 (2004) 085403.
[212] A.W. Ghosh, T. Rakshit, S. Datta, Nano Lett. 4 (4) (2004) 565–568.
[213] G.D. Mahan, Many-Particle Physics, Premium Press, New York, 1990.
[214] F. Zahid, A. Ghosh, M. Paulsson, E. Polizzi, S. Datta, Phys. Rev. B 70 (2004) 245317.
[215] D. John, L. Castro, P. Pereira, D. Pulfrey, Proc. NSTI Nanotech. 3 (2004) 65–68.
[216] M. Pourfath, H. Kosina, Fast convergent Schrödinger–Poisson solver for the static and dynamic analysis of carbon nanotube field effect transistors, in: Lecture Notes in Computer Science 3743: Large-Scale Scientific Computing, Springer Verlag, 2006, pp. 578–585.
[217] R. Venugopal, Z. Ren, S. Datta, M.S. Lundstrom, D. Jovanovic, J. Appl. Phys. 92 (7) (2002) 3730–3739.
[218] M. Pourfath, H. Kosina, B. Cheong, W. Park, S. Selberherr, Proc. European Solid-State Device Research Conf., 2005, 541–544.
[219] A. Javey, J. Guo, D. Farmer, Q. Wang, E. Yenilmez, R. Gordon, M. Lundstrom, H. Dai, Nano Lett. 4 (7) (2004) 1319–1322.
[220] G. Mahan, Phys. Rev. B 68 (2003) 125409.
[221] V.N. Popov, P. Lambin, Phys. Rev. B 74 (2006) 075415.
[222] M. Pourfath, H. Kosina, S. Selberherr, J. Comp. Electron. 5 (2/3) (2006) 155–159.
[223] S. Koswatta, S. Hasan, M. Lundstrom, M. Anantram, D. Nikonov, Ballisticity of nanotube FETs: Role of phonon energy and gate bias, Appl. Phys. Lett 89 (2) (2006), pp. 23125–1-3.
[224] C. Jacoboni, Semicond. Sci. Technol. 7 (3B) (1992) B6–B11.
[225] E. Wigner, Phys. Rev. 40 (1932) 749–759.
[226] H. Kosina, M. Nedjalkov, Handbook of Theoretical and Computational Nanotechnology, Vol. 10, American Scientific Publishers, Los Angeles, 2006, Ch. 14. Wigner function based device modeling, pp. 731–763.
[227] W.H. Louisell, Quantum Statistical Properties of Radiation, J. Willey & Sons., New York, 1973.
[228] G. Lindblad, On the generators of quantum dynamical semigroups, Comm. Math. Phys. 48 (1976) 119–130.
[229] R. Gebauer, R. Car, Phys. Rev. Lett. 93 (16) (2004) 160404.
[230] R. Gebauer, R. Car, Phys. Rev. B 70 (12) (2004) 125324.
[231] J.D. Jackson, Classical Electrodynamics, 3rd Ed., J. Wiley & Sons., New York, 1998.
[232] H. Kosina, M. Nedjalkov, S. Selberherr, J. Comp. Electron. 2 (2-4) (2003) 147–151.
[233] P. Degond, C. Ringhofer, J. Stat. Phys. 112 (3) (2003) 587–628.
[234] J.-P. Bourgade, P. Degond, F. Mehats, C. Ringhofer, J. Math. Phys. 47 (4) (2006) 043302.
[235] U. Ravaioli, M. Osman, W. Pötz, N. Kluksdahl, D. Ferry, Physica B 134 (1985) 36–40.
[236] N. Kluksdahl, W. Pötz, U. Ravaioli, D. Ferry, Superlattices Microstruct. 3 (1) (1987) 41–45.
[237] W. Frensley, Phys. Rev. Lett. 57 (22) (1986) 2853–2856.
[238] W. Frensley, IEDM Techn. Dig. (1986) 571–574.
[239] W. Frensley, Phys. Rev. B 36 (3) (1987) 1570–1580.
[240] N. Kluksdahl, A. Kriman, D. Ferry, C. Ringhofer, Phys. Rev. B 39 (11) (1989) 7720–7735.
[241] W. Frensley, Solid-State Electron. 32 (12) (1989) 1235–1239.

[242] R.K. Mains, G.I. Haddad, J. Appl. Phys. 64 (7) (1988) 3564–3569.
[243] F. Buot, K. Jensen, Phys. Rev. B 42 (15) (1990) 9429–9457.
[244] L. Shifren, D.K. Ferry, J. Comp. Electron. 1 (2002) 55–58.
[245] M. Nedjalkov, R. Kosik, H. Kosina, S. Selberherr, Proc. Intl. Conf. on Simulation of Semiconductor Processes and Devices, 2002, 187–190.
[246] L. Shifren, C. Ringhofer, D. Ferry, Phys. Lett. A 306 (2003) 332–336.
[247] H. Kosina, M. Nedjalkov, S. Selberherr, M. Laudon, B. Romanowicz, Quantum Monte Carlo simulation of a resonant tunneling diode including phonon scattering, Nanotech, Computational Publications, San Francisco, 2003, pp. 190–193.
[248] A. Gehring, H. Kosina, J. Comp. Electron. 4 (1/2) (2005) 67–70.
[249] A. Wettstein, A. Schenk, W. Fichtner, IEEE Trans. Electron Devices 48 (2) (2001) 279–284.
[250] L. Shifren, C. Ringhofer, D. Ferry, IEEE Trans. Electron Devices 50 (3) (2003) 769–773.
[251] D. Querlioz, J. Saint-Martin, V.N. Do, IEEE Trans. Nanotechnol. 5 (6) (1998) 737–744.
[252] H. Kosina, V. Sverdlov, T. Grasser, Proc. Intl. Conf. on Simulation of Semiconductor Processes and Devices, 2006, 357–360.
[253] J. Wang, E. Polizzi, M. Lundstrom, J. Appl. Phys. 96 (4) (2004) 2192–2203.
[254] J. Wang, A. Rahman, A. Ghosh, G. Klimeck, M. Lundstrom, IEEE Trans. Electron Devices 52 (7) (2005) 1589–1595.
[255] M. Luisier, A. Schenk, W. Fichtner, G. Klimeck, Phys. Rev. B 74 (2006), pp. 205323-1-12.