Acta Numerica (1997), *pp.* 485–521

Computational methods for semiclassical and quantum transport in semiconductor devices

Christian Ringhofer
Department of Mathematics, Arizona State University
Tempe, AZ 85287–1804, USA
E-mail: ringhofer@asu.edu

The progressive miniaturization of semiconductor devices, and the use of bulk materials other than silicon, necessitates the use of a wide variety of models in semiconductor device simulation. These include classical and semiclassical models, such as the Boltzmann equation and the hydrodynamic system, as well as quantum transport models such as the quantum Boltzmann equation and the quantum hydrodynamic system. This paper gives an overview of recently developed numerical methods for these systems. The focus is on Galerkin methods for the semiclassical and quantum kinetic systems and on difference methods for the classical and quantum hydrodynamic systems. The stability and convergence properties of these methods and their relation to the analytical properties of the continuous systems are discussed.

CONTENTS

1. Introduction

The goal of numerical semiconductor device simulation is to model the flow of electrons in a crystal in order to predict macroscopically measurable quantities, such as currents and heat fluxes, in given operating and environmental conditions, such as the bias applied to a given device and ambient temperature. Other than in process simulation, it is always the same physical process that is considered, namely the transport of charged particles in a solid state medium. Different mathematical models are used

only because of the wide range of device dimensions and operating conditions. Since one and the same set of equations can be used to model a wide variety of devices, it is reasonable to develop customized numerical methods for the governing equations.

The key parameters influencing the choice of model equations are the mean free path of electrons (the average length of free flight before the electron undergoes a scattering event), the number of free electrons in a given device, the size of the Planck constant in relation to the dimensions of the simulation domain, and the ambient temperature. These parameters determine whether the electrons can be modelled as a continuum, as classical particles or via quantum mechanical descriptions.

The resulting model equations range from the Schrödinger equation for the evolution of the electron wave function to the drift–diffusion system for the evolution of an 'electron gas' which is close to a Maxwellian equilibrium. Because of the progressive miniaturization of semiconductor devices and the use of materials whose mean free path is considerably longer than that of silicon, the trend in device simulation is certainly towards a more and more microscopic description. Since the field is now so wide, one necessarily has to limit the scope of an overview of numerical techniques in device simulation. There are two types of models and simulation techniques which are extremely well developed and documented at this point. One consists of finite difference and finite element techniques for the drift–diffusion system, and the other of Monte Carlo methods for the Boltzmann equation. Since it would be impossible to do all this work justice in the space provided, we have instead decided to focus on more recent developments, and refer the reader to excellent reference works such as Kersch and Morokoff (1995) and Selberherr (1981) for these topics. The first category of methods presented in this paper deals with the intermediate regime between the Boltzmann equation and the drift–diffusion system. This category comprises methods based on series expansion of the Boltzmann equation and various forms of moment closure hierarchies, including the so-called hydrodynamic models. The second category includes methods for quantum kinetic equations and their moment closure hierarchies, such as the so-called quantum hydrodynamic model.

This paper is organized as follows. Section 2 presents a brief overview of the various models, pointing out some of the features relevant to numerical simulations. Section 3 deals with methods for semiclassical transport descriptions, based on the semiclassical Boltzmann equation. Series expansion methods around a Maxwell distribution are discussed in Section 3.1, numerical methods for the hydrodynamic model are discussed in Section 3.2, and extensions of hydrodynamic models are presented in Section 3.3. Section 4 is devoted to numerical methods for quantum transport models. In Section 4.1 numerical methods for the quantum Boltzmann equation are discussed.

As is the case for the classical Boltzmann equation, many of the interesting effects can be studied using much simpler macroscopic models based on moment hierarchies, leading to the quantum hydrodynamic model. Section 4.2 deals with numerical methods for the quantum hydrodynamic system.

2. Model equations

In this section a brief overview of the underlying model equation is presented. Models for semiconductor device simulations generally fall into two categories, namely semiclassical models, based on the semiclassical Boltzmann equation, and quantum mechanical models, derived from the Schrödinger equation. In Section 2.1 we discuss the semiclassical Boltzmann equation together with some of its features, such as conservation properties. In Section 2.2 its quantum mechanical equivalent, namely the quantum Boltzmann equation, is presented.

2.1. The semiclassical Boltzmann equation

The basis for the semiclassical description of electron transport is the Boltzmann equation in the form

$$\partial_t f + \operatorname{div}_x(v(k)f) - \frac{1}{\hbar}\operatorname{div}_k(E(x,t)f) = Q(f). \tag{2.1}$$

Here $f(x,k,t)$ denotes the density of electrons, x stands for position, k denotes the three-dimensional wave vector, and t time. If we let $\varepsilon(k)$ denote the energy of an electron with wave vector k in a certain band, the corresponding velocity in (2.1) is given by $v(k) = \hbar^{-1}\nabla_k\varepsilon$. In a vacuum, the classical Hamiltonian yields the energy-wave vector relationship

$$\varepsilon(k) = \frac{\hbar^2}{2m}|k|^2, \quad v(k) = \frac{\hbar}{m}k. \tag{2.2}$$

Thus the velocity v and the wave vector k are identical up to the constant $\hbar/m$ and the classical Boltzmann transport equation is obtained. In a crystal, the relationship between the wave vector and the energy is given by the parametrization of the eigenfunctions of the Schrödinger equation with a potential that is periodic on the crystal lattice, and the energy band function $\varepsilon(k)$ has to be computed. However, for small wave vectors, and consequently for small velocities, the energy band function is often approximated locally by a parabolic function via the effective mass approximation, for analytical purposes. The collision integral $Q(f)$ on the right-hand side of (2.1) is given by

$$Q(f)(x,k,t) = \int S(k,k')f'(1-f) - S(k',k)f(1-f')\,\mathrm{d}k', \tag{2.3}$$

where the notation

$$f = f(x,k,t), \quad f' = f(x,k',t) \tag{2.4}$$

is used. The collision integral $Q(f)$ models the interaction of electrons with the crystal lattice. These interactions include scattering with crystal impurities and acoustic and polar optical phonons (the vibrations of the lattice). A more complicated collision operator Q_{ee} is used to model the interaction of electrons with each other. The electron–electron collision operator is of the form

$$\begin{aligned} Q_{ee}(f)(x,k,t) = \\ \int S_{ee}(x,k,k_1,k',k_1')f'f_1'(1-f)(1-f_1) - \\ S_{ee}(x,k',k_1',k,k_1)ff_1(1-f')(1-f_1')\,\mathrm{d}k_1\,\mathrm{d}k'\,\mathrm{d}k_1', \end{aligned} \tag{2.5}$$

where $f = f(x,k,t)$, $f' = f(x,k',t)$, $f_1 = f(x,k_1,t)$, and $f_1' = f(x,k_1',t)$. However, other than in gas dynamics, particle–particle scattering is a rather rare event in most semiconductor devices and the operator Q_{ee} is rarely used.

Conservation and equilibrium

There are two important features of the collision operator Q that need to be reflected by any numerical method, namely conservation and the existence of a thermal equilibrium. A quantity $g(k)$ is said to be conserved if

$$\int g(k)Q(f)(x,k,t)\,\mathrm{d}k = 0 \tag{2.6}$$

holds for any density function f. For reasons of symmetry $g(k) = 1$, the number of particles, is obviously conserved by all collision operators. The second property is the existence of a thermal equilibrium, namely a density function f_e such that, because of the principle of detailed balance (Markowich, Ringhofer and Schmeiser 1990),

$$S(k,k')f_e'(1-f_e) = S(k',k)f_e(1-f_e') \tag{2.7}$$

holds. The thermal equilibrium density f_e is given by the Fermi–Dirac density function

$$f_e = F_D\left(\frac{\varepsilon(k)-\varepsilon_F}{k_BT}\right), \quad F_D(z) = \frac{1}{1+e^z}, \tag{2.8}$$

where ε_F is the Fermi energy, k_B is the Boltzmann constant, and T denotes the lattice temperature of the crystal. The principle of detailed balance (2.7) implies the relation

$$S(k,k') = M(k)s(k,k'), \quad M(k) = \exp\left(\frac{-\varepsilon(k)}{k_BT}\right) \tag{2.9}$$

for the scattering rate S, where M is called the Maxwellian distribution and s is symmetric in the variables k and k', so $s(k,k') = s(k',k)$ holds.

Low density and relaxation time approximations

In order to derive simplified models from the Boltzmann equation (2.1), it is often necessary to make simplifying approximations to the collision operator Q. The first approximation is to assume the density function f to be small, and therefore to drop the quadratic terms in the collision operator Q in (2.3), giving the linear operator

$$Q(f)(x,k,t) = \int s(k,k')f' - s(k',k)f \,\mathrm{d}k'. \tag{2.10}$$

Next, it is assumed that the density function f is close to a Maxwellian distribution of the form

$$f(x,k,t) \approx \frac{n(x,t)}{n_0} \exp\left(\frac{-\varepsilon(k)}{k_B T}\right), \quad n_0 = \int \exp\left(\frac{-\varepsilon(k)}{k_B T}\right) \mathrm{d}k. \tag{2.11}$$

Replacing f' in the linear collision operator (2.10) by the expression (2.11) gives

$$Q(f)(x,k,t) = \frac{1}{\tau(x,k)}\Big(n(x,t)M(k) - f\Big), \tag{2.12}$$

with

$$\frac{1}{\tau(x,k)} = \int s(k',k)\,\mathrm{d}k', \quad n(x,t) = \int f(x,k,t)\,\mathrm{d}k, \tag{2.13}$$

$$\text{and} \quad M(k) = \frac{1}{n_0} \exp\left(\frac{-\varepsilon(k)}{k_B T}\right).$$

The term $\tau(x,k)$ is called the relaxation time.

Collision frequency, mean free path and scaling

One of the most important quantitative parameters determining which model to choose is the mean free path. The mean free path is given by the shape of the scattering rate $s(k,k')$ and the the energy band function ε. If we define the collision frequency ω by

$$\omega(k) = \int M(k')s(k',k)\,\mathrm{d}k', \tag{2.14}$$

then ω^{-1} is the average time an electron travels freely before undergoing a collision event. Scaling the velocity wave vector relationship

$$v(k) = \frac{1}{\hbar}\nabla_k \varepsilon(k) = v_0 \tilde{v}\left(\frac{k}{k_0}\right), \tag{2.15}$$

where v_0 and k_0 are chosen such that $\tilde{v}(k)$ is an $O(1)$ function, the expression

$$\lambda_0 = \frac{v_0}{\omega} \tag{2.16}$$

gives the average distance an electron travels between collision events. If we now scale the position variable x by the device length L, the wave vector k by k_0 and the time by $\gamma = \omega_0 L^2/v_0^2$, we obtain the scaled Boltzmann equation

$$\lambda^2 \partial_t f_s + \lambda \operatorname{div}_x\Big(\tilde{v}(k) f_s\Big) - \lambda \operatorname{div}_k\Big(E_s f_s\Big) = Q(f_s) \tag{2.17a}$$

$$Q(f_s) = \int_{\mathbb{R}^3} \mathrm{d}k\; s_s\Big(M_s f_s'(1-f_s) - M_s' f_s(1-f_s')\Big), \tag{2.17b}$$

where x, k and t are now dimensionless and the scaled field and scattering rate E_s and s_s are given by

$$\begin{aligned} E(x,t) &= \frac{v_0 \hbar k_0}{L} E_s\left(\frac{x}{L}, \frac{t}{\gamma}\right), \\ s(k,k') &= \frac{\omega_0}{k_0^3} s_s\left(\frac{k}{k_0}, \frac{k}{k_0}\right), \\ M(k) &= M_s\left(\frac{k}{k_0}\right), \end{aligned} \tag{2.18}$$

and $\lambda = \lambda_0/L$ is the Knudsen number, the ratio of the mean free path and the size of the simulation domain. To what extent macroscopic models provide an accurate transport picture depends mainly on the size of the Knudsen number λ. In the limit for $\lambda \to 0$ one obtains from the Hilbert expansion (see Markowich et al. (1990)) that $f_s = n(x,t)M_s(k) + O(\lambda)$ holds, where the macroscopic electron density n satisfies the drift–diffusion equation

$$\partial_t n - \operatorname{div}_x\left(D\nabla_x n - mE_s n\right) = 0. \tag{2.19}$$

(We will drop the subscript s from now on.) Through miniaturization, the device length L decreases, and through the use of materials such as gallium arsenide, the mean free path λ_0 increases, making the drift–diffusion system less and less valid. Current state-of-the-art technology for devices like MOSFETs works with values of $\lambda = O(0.1)$, which makes simulations based on alternative models necessary.

Boundary conditions for the semiclassical Boltzmann equation

For the simulation of an actual device, the simulation domain will consist of a bounded region Ω, whose boundary $\partial\Omega$ is made up of segments $\partial\Omega_c, \partial\Omega_i, \partial\Omega_a$, corresponding to contacts, insulating surfaces and artificial boundaries, introduced to limit the size of the simulation domain. At contacts, the inflow of electrons according to a certain given distribution f_c is

prescribed. At insulating surfaces we usually prescribe a specular reflection condition, and at artificial surface segments zero influx is required. The situation will be somewhat more complicated in the quantum case (see Section 4.1). So altogether we have

$$\begin{aligned}
\partial\Omega &= \partial\Omega_c \cup \partial\Omega_a \cup \partial\Omega_i && && && \text{(2.20a)}\\
f(x,k,t) &= b(x,t) f_c(k) && \text{for} && x \in \partial\Omega_c, \quad k\cdot\nu < 0 && \text{(2.20b)}\\
f(x,k,t) &= 0 && \text{for} && x \in \partial\Omega_a, \quad k\cdot\nu < 0 && \text{(2.20c)}\\
f(x,k,t) &= f(x,-k,t) && \text{for} && x \in \partial\Omega_i, \quad k\cdot\nu < 0, && \text{(2.20d)}
\end{aligned}$$

where ν denotes the outward normal vector on the boundary $\partial\Omega$. The function $b(x,t)$ in (2.20b) gives the amount of electrons injected. Assuming that the device is part of a circuit, b is given by Ohm's law.

The Poisson equation

The electric field E is, in general, derived from Maxwell's equations. However, since we operate in a regime where the speed of light can safely be set to infinity, Maxwell's equations become

$$\operatorname{div}_x(\varepsilon_D E) = \rho, \qquad \nabla_x \times E = 0, \tag{2.21}$$

where ε_D denotes the dielectric constant of the material. The charge density ρ is given by $\rho = e(N_D - N_A - n)$, where e is the unit charge and N_D and N_A are the pre-concentrations of donor and acceptor atoms in the crystal, due to doping. Here n is the spatial density of free electrons, given by the zeroth order moment of the density function f in the Boltzmann equation. Introducing the potential V by $E = -\nabla_x V$, we obtain the Poisson equation

$$\operatorname{div}_x(\varepsilon_D E) = e(N_D - N_A - n), \quad E = -\nabla_x V, \quad n = \int_{\mathbb{R}^3} \mathrm{d}k\; n. \tag{2.22}$$

The Poisson equation (2.22) is coupled to the Boltzmann equation (2.1) via the charge density n in (2.12), and therefore the two equations have to be solved simultaneously. A bias is applied to the contacts of the device by prescribing a potential difference between the contacts, that is, by setting V at the boundary segments corresponding to contacts.

2.2. Quantum transport models

As device dimensions decrease, quantum mechanical transport phenomena play an increasing role in the function of devices. It is therefore necessary to develop simulation models that are capable of describing these effects. These models are a generalization of the classical models in the sense that they reduce to the Boltzmann transport picture in the classical limit, that is, when an appropriately scaled form of the Planck constant tends to zero. The quantum mechanical description of the motion of an electron in a vacuum

under the influence of a potential field is given by the Schrödinger equation

$$i\hbar\partial_t\psi = H\psi, \quad H\psi = -\frac{\hbar^2}{2m}\Delta\psi - eV\psi, \tag{2.23}$$

where ψ denotes the wave function and V denotes the potential. The operator H is called the quantum Hamiltonian. The electron density n and the electron current density J are then given by

$$n = |\psi|^2 \quad \text{and} \quad eJ = \frac{\hbar}{m} Im(\psi\nabla\psi). \tag{2.24}$$

In order to describe transport in an actual device, several features have to be added to the above transport picture.

- An ensemble of Schrödinger equations must be considered in order to model the mixed state of an electron.
- The electron is moving in a crystal and not in a vacuum.
- Collisions, representing the interaction of the electron with the crystal lattice, have to be modelled.
- The system has to interact with the outside world via boundary conditions at device contacts and insulating surfaces.

Several steps are taken to achieve these goals. Some are mathematically precise, whereas some are purely phenomenological. First, the density matrix for mixed states of the form

$$\rho(r,s,t) = \sum_j \sigma(\omega_j)\psi(r,t)^*\psi(s,t) \tag{2.25}$$

is introduced, where each of the wave functions ψ_j satisfies the Schrödinger equation (2.23). The density matrix ρ then satisfies the quantum Liouville equation

$$i\hbar\partial_t\rho = (H_s - H_r)\rho = \frac{\hbar^2}{2m}(\Delta_r - \Delta_s)\rho + e\Big(V(r) - V(s)\Big)\rho, \tag{2.26}$$

and the electron and current densities n and J are given by

$$n(x,t) = \rho(x,x,t), \quad eJ(x,t) = \frac{i\hbar}{2m}(\nabla_r\rho - \nabla_s\rho)(r=x, s=x, t). \tag{2.27}$$

In order to relate the quantum picture to the classical picture it is convenient to introduce the Wigner function (Wigner 1932)

$$w(x,k,t) = (2\pi)^{-3}\int_{\mathbb{R}^3} \mathrm{d}\eta\ \rho\left(x + \frac{1}{2}\eta, x - \frac{1}{2}\eta, t\right) e^{i\eta\cdot k}, \tag{2.28}$$

which then satisfies the Fourier transformed version of the quantum Liouville equation, often referred to as the Wigner equation

$$\partial_t w + \frac{\hbar}{m}k\cdot\nabla_x w + \frac{ie}{\hbar}\delta V\left(x, \frac{1}{2i}\nabla_k\right) w = 0, \tag{2.29}$$

and the electron and current densities are given by

$$n(x,t) = \int_{\mathbb{R}^3} \mathrm{d}k\; w, \quad J(x,t) = \frac{\hbar}{m} \int_{\mathbb{R}^3} \mathrm{d}k\; kw, \tag{2.30}$$

and the operator $\delta V\,(x, (1/2i)\nabla_k)$ in (2.29) is defined in the sense of pseudo-differential operators (Taylor 1981) as

$$\begin{aligned} &\delta V\left(x, \frac{1}{2i}\nabla_k\right) w(x,k,t) = \\ &\qquad (2\pi)^{-3} \int_{\mathbb{R}^3} \mathrm{d}\eta \int_{\mathbb{R}^3} \mathrm{d}k'\; \delta V\left(x, \frac{1}{2}\eta\right) w(x,k',t) e^{i\eta\cdot(k-k')}, \\ \text{where} \qquad &\delta V\left(x, \frac{1}{2}\eta\right) = V\left(x+\frac{1}{2}\eta\right) - V\left(x+\frac{1}{2}\eta\right). \end{aligned} \tag{2.31}$$

The advantage of the Wigner formulation lies in the fact that it relates the quantum mechanical picture to the classical picture. For quadratic potentials V, the Wigner equation (2.29) reduces to the Boltzmann equation without collision terms. It can be shown (Markowich and Ringhofer 1989) that the Wigner function converges to the solution of the collisionless Boltzmann equation in the limit of large time and spatial scales. However, from the point of view of device simulation, we are interested in quantum transport equations in regimes which are quite far away from the classical picture. Here, the advantage of the Wigner equation lies in the fact that it allows for a more phenomenological treatment of collision terms and boundary conditions. Clearly the Wigner equation (2.29) is the quantum equivalent of the Boltzmann equation with a parabolic band structure (2.2), since the starting point was the quantum Hamiltonian for a vacuum. In order to describe the motion of the electron in a crystal, a modified Hamiltonian of the form

$$H = -\frac{\hbar^2}{2m}\Delta_x - e(V_L + V) \tag{2.32}$$

has to be considered, where V_L denotes the potential due to a periodic crystal lattice (Ashcroft and Mermin 1976). So

$$V_L(x+\gamma z_j) = V_L(x), \quad j = 1,2,3 \tag{2.33}$$

holds, where the z_j are the lattice directions and γ is the length scale of the lattice, chosen such that $\det(Z) = 1$, where $Z = (z_1, z_2, z_3)$. It can be shown by using a Bloch wave decomposition (Arnold, Degond, Markowich and Steinrück 1989, Poupaud and Ringhofer 1995, Markowich, Mauser and Poupaud 1994) that the projection of the wave function onto the mth energy band satisfies the Schrödinger equation

$$i\hbar\partial_t\psi = \left(\frac{\gamma}{2\pi}\right)^3 \int_{\mathbf{B}} \mathrm{d}k \sum_m \psi(x+\gamma Zm)\varepsilon_m(k) e^{-i\gamma k^T Zm} - eV\psi, \tag{2.34}$$

where $\varepsilon_m(k)$ is the mth eigenvalue of the Hamiltonian $H_L = -(\hbar^2/2m)\Delta + V_L$ together with quasi-periodic boundary conditions. Here $\mathbf{B}$ denotes the Brillouin zone, the unit cell of the dual crystal lattice defined as $\mathbf{B} = Z^{-T}[-\pi/\gamma, \pi/\gamma]^3$. Performing the Wigner transformation for the mixed state, as before, now gives the Wigner equation in a crystal of the form

$$\partial_t w + \operatorname{div}_x \left(\int_{-1/2}^{1/2} \mathrm{d}s \sum_m \hat{v}(m) w(x + \gamma s Z m, k, t) \exp\left(i\gamma k^T Z m\right) \right) + \frac{ie}{\hbar} \delta V \left(x, \frac{1}{2i} \nabla_k \right) w = 0, \tag{2.35}$$

where all functions are now periodic in the wave vector k on the Brillouin zone $\mathbf{B}$ and Fourier transforms and pseudo-differential operators are appropriately reformulated as

$$\begin{aligned} v(k) &= \sum_m \hat{v}(m) \exp\left(i\gamma k^T Z m\right), \\ \delta V \left(x, \frac{1}{2i} \nabla_k \right) w(x,k,t) &= \\ \left(\frac{\gamma}{2\pi} \right)^3 \sum_n \int_{\mathbf{B}} \mathrm{d}k' \; \delta V \left(x, \frac{\gamma}{2} Z n \right) w(x,k',t) \exp\left(i\gamma (k-k')^T Z n\right). \end{aligned} \tag{2.36}$$

Since the length scale γ of the crystal lattice will be small even for quantum mechanical simulations, the formal limit $\gamma \to 0$ is used in equation (2.35) for actual simulations giving

$$\partial_t w + \operatorname{div}_x (v(k) w) + \frac{ie}{\hbar} \delta V \left(x, \frac{1}{2i} \nabla_k \right) w = 0. \tag{2.37}$$

In the limit $\gamma \to 0$, the Brillouin zone becomes infinite and the definition of Fourier transforms and pseudo-differential operators reverts to (2.31).

Modelling the scattering processes of electrons with phonons quantum mechanically is a much more complicated task. Most models lead to equations which are too high-dimensional to be actually used in the simulation of devices (Ferry and Grubin 1995). Therefore, we are in practice reduced to two approaches to formulating the quantum Boltzmann equation

$$\partial_t w + \operatorname{div}_x (v(k) w) + \theta[V] w = Q(w), \quad \theta[V] w = \frac{ie}{\hbar} \delta V \left(x, \frac{1}{2i} \nabla_k \right) w, \tag{2.38}$$

namely the relaxation time model and the Fokker–Planck term. The relaxation time model is, as in the classical case, given by

$$\begin{aligned} Q(w) &= \frac{1}{\tau} \left(\frac{n}{n_0} w_0 - w \right), \\ n(x,t) &= \int_{\mathbb{R}^3} \mathrm{d}k \; w(x,k,t), \end{aligned}$$

$$n_0(x) = \int_{\mathbb{R}^3} \mathrm{d}k \; w_0(x,k), \tag{2.39}$$

where $w_0(x,k)$ denotes the quantum mechanical thermal equilibrium density, the quantum equivalent to the Maxwellian. The Fokker–Planck term model is given by

$$Q(w) = \frac{1}{\tau} \operatorname{div}_k \left(\frac{mT_0}{\hbar^2} \nabla_k w + kw \right), \tag{2.40}$$

where T_0 denotes the lattice temperature.

Thermal equilibrium

To carry out actual simulations it is necessary to compute a quantum mechanical thermal equilibrium solution. This is necessary for two reasons. First, the thermal equilibrium solution w_0 is used in the relaxation time approximation (2.39). Second, transient simulations are started by using the thermal equilibrium as initial datum. For a mixed state, the thermal equilibrium density matrix is defined by

$$\rho_{TE}(r,s) = \sum_j \sigma(\omega_j)\psi_j(r)\psi_j(s), \tag{2.41}$$

where the ψ_j and l_j are the eigenfunctions and eigenvalues of the quantum Hamiltonian and $\sigma(\omega)$ is the statistical distribution. For Boltzmann statistics, σ is of the form $\sigma(\omega) = \exp(-\omega/T_0)$.

3. Numerical methods for semiclassical and classical transport

In this section we describe two types of approaches to simulating semiconductor devices based on the semiclassical Boltzmann equation (2.17). Both approaches are more or less restricted to the case of parabolic band structures, so (2.2) is assumed. A relatively easy generalization is to assume a more general quadratic band structure of the form $\varepsilon(k) = (\hbar^2/2m)k^T Z k$ with a positive definite symmetric matrix Z. This corresponds to a Taylor expansion of the band energy ε for small wave vectors k and leads to the so-called effective mass approximation. Since generalizing the presented numerical methods to this case is straightforward, it will not be considered separately here. The discussed methods cannot be expected to represent the physical transport picture as accurately as a complete Monte Carlo simulation in all possible cases. They have, however, the great advantage of dealing with deterministic computational models that possess a well defined steady state. At the same time, they seem to give a reasonably accurate transport description for current device dimensions, as verified by comparisons with experiments and Monte Carlo calculations.

3.1. Series expansion methods

Series expansion methods for the Boltzmann transport equation have the advantage that they give, in some sense, a direct extension of macroscopic models such as the drift–diffusion system and the hydrodynamic system, to be discussed in the next section. Usually a spectral Galerkin approach is used in the wave vector direction, while some other standard finite difference or finite element discretization is employed in the spatial and time directions. Most series expansion methods assume a parabolic band structure (2.2). So, after an appropriate scaling, the wave vector can be identified with the velocity vector. We will discuss series expansion methods on the scaled Boltzmann transport equation

$$\lambda^2 \partial_t f + \lambda v \cdot \nabla_x f - \lambda E \cdot \nabla_v f = Q(f), \tag{3.1}$$

where λ denotes the Knudsen number, the ratio of the mean free path to the length scale of the simulation domain. Most expansion methods for the Boltzmann transport equation use the assumption of isotropic scattering, that is, that the scattering rates $s(v, v')$ in (2.3) as well as the Maxwellian depend only on the energy ε. So, after scaling, and assuming a parabolic band structure, the collision integral $Q(f)$ in (3.1) and the Maxwellian are of the form

$$\begin{aligned} Q(f)(x,v,t) &= \int_{\mathbb{R}^3} \mathrm{d}v \, \Big\{ s(|v|,|v'|)\Big(M f'(1-f) - M' f(1-f') \Big) \Big\}, \\ M(v) &= \exp\left(\frac{-|v^2|}{2} \right). \end{aligned} \tag{3.2}$$

One of the drawbacks of series expansion methods is that the evaluation of the collision integral is quite complicated and expensive if no Monte Carlo approach is used. The dependence of the integral kernel on the energy alone, reduces the complexity of this problem considerably once polar coordinates

$$\begin{aligned} v &= (r\cos\theta, r\sin\theta\cos\phi, r\sin\theta\sin\phi)^T, \\ &\quad \text{where} \quad \theta \in [0,\pi], \quad \phi \in [-\pi,\pi], \quad r \in [0,\infty), \end{aligned} \tag{3.3}$$

are used. In polar coordinates, the collision integral Q then becomes

$$\begin{aligned} &Q(f)(x,r,\theta,\phi,t) = \\ &\int_0^\infty \mathrm{d}r' \int_0^\pi \mathrm{d}\theta' \int_{-\pi}^{\pi} \mathrm{d}\phi' \; r'^2 \sin(\theta) s(r,r') \Big(M f'(1-f) - M' f(1-f') \Big), \end{aligned} \tag{3.4}$$

and the integration over the angular variables can be carried out explicitly, giving

$$Q(f)(x,r,\theta,\phi,t) = \int_0^\infty \mathrm{d}r' \; r'^2 s(r,r') \Big(M F'(1-f) - M' f(4\pi - F') \Big), \tag{3.5}$$

where $F(x,r,t)$ denotes the average of the density function over spheres of

equal energy

$$F(x,r,t) = \int_0^{\pi} \mathrm{d}\theta \int_{-\pi}^{\pi} \mathrm{d}\phi \ \sin(\theta) f(x,r,\theta,\phi,t), \tag{3.6}$$

and again the terms f' and F' mean that the corresponding functions are evaluated at (r',θ',ϕ') rather than at (r,θ,ϕ). The advantage of the use of polar coordinates lies in the fact that the collision integral is now one-dimensional and the collision terms for scattering of acoustic and polar optical phonons, whose scattering rates are of the form

$$s(r,r') = \sum_{j=-1}^{1} \gamma(|j|)\delta(r^2 - r'^2 - j\omega_{ph}), \tag{3.7}$$

with ω_{ph} the energy of emission/absorption of a polar optical phonon, amount to pointwise evaluation of F. The Boltzmann transport equation in polar coordinates takes the form

$$\lambda^2 \partial_t f + \lambda r a \cdot \nabla_x f - \lambda E \cdot (a\partial_r f + b\partial_\theta f + c\partial_\phi f) = Q(f), \tag{3.8}$$

where

$$a = \frac{1}{r}v = \begin{pmatrix} \cos\theta \\ \sin\theta\cos\phi \\ \sin\theta\sin\phi \end{pmatrix}, \quad b = \frac{1}{r}\begin{pmatrix} -\sin\theta \\ \cos\theta\cos\phi \\ \cos\theta\sin\phi \end{pmatrix}, \quad c = \frac{1}{r\sin\theta}\begin{pmatrix} 0 \\ -\sin\phi \\ \cos\phi \end{pmatrix}.$$

Starting with Odeh, Gnudi, Baccarani, Rudan and Ventura (Ventura, Gnudi and Baccarani 1991, Ventura, Gnudi, Baccarani and Odeh 1992), and continuing with the work of Goldsman and Frey (Goldsman, Wu and Frey 1990, Goldsman, Henrickson and Frey 1991), spherical harmonic expansions of the Boltzmann transport equation in polar coordinates have been used with great success, meaning that good agreement with Monte Carlo simulations has been achieved for realistic devices using only a relatively small number of terms. We recall that spherical harmonic functions take the form

$$S_n(\theta,\phi) = L_n(\cos\theta)(\sin\theta)^{n_2}\exp(in_2\phi), \quad n = (n_1,n_2), \tag{3.9}$$

where L_n is the associated Legendre polynomial of degree (n_1,n_2). Thus L_n is a polynomial of degree n_1 satisfying the orthogonality relation

$$\int_{-1}^{1} L_{n_1,n_2}(y)L_{\nu_1,n_2}(y)(1-y^2)^{n_2}\,\mathrm{d}y = \frac{1}{2\pi}\delta(n_1-\nu_1), \tag{3.10}$$

and consequently the spherical harmonics satisfy

$$\int_{-\pi}^{\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta \sin(\theta) S_n^*(\theta,\phi)S_\nu(\theta,\phi) = \delta(n-\nu). \tag{3.11}$$

Expanding the density function f in spherical harmonics gives

$$f \approx f^N = \sum_{n\in N} f_n(x,r,t)S_n(\theta,\phi), \tag{3.12}$$

where N denotes some suitable index set, and using the standard Galerkin approach, we find

$$\lambda^2 \partial_t f_n + \lambda r A_{inm} \partial_{x_i} f_m - E_i \left(A_{inm} \partial_r f_m + B_{inm} f_m \right) = q_n, \tag{3.13}$$

where the summation convention is used in (3.13) and the coefficients A_{inm} and $B_{inm}(r)$ are given by

$$A_{inm} = \int_0^\pi \mathrm{d}\theta \int_{-\pi}^\pi \mathrm{d}\phi \ \sin(\theta) S_n a_i S_m, \tag{3.14a}$$

$$B_{inm} = \int_0^\pi \mathrm{d}\theta \int_{-\pi}^\pi \mathrm{d}\phi \ \sin(\theta) S_n (b_i \partial_\theta S_m + c_i \partial_\phi S_m) \tag{3.14b}$$

$$q_n = \int_0^\infty \mathrm{d}r' \ r'^2 s(r, r') \left(M f_0' \left(4\pi \delta(n) - \sqrt{4\pi} f_n \right) - M' f_n \left(4\pi - \sqrt{4\pi} f_0' \right) \right). \tag{3.14c}$$

Stability and discretization

Equation (3.13) represents a hyperbolic first-order system in the spatial and time variables (x, t) and the energy variable $r = |v|$. In principle, the system (3.13) could be discretized by any number of methods suitable for hyperbolic systems. In Ventura et al. (1991) and (1992), standard finite differences are used in all variables. However, in addition to exhibiting the usual stiffness of PDE discretizations, (3.13) is extremely stiff in time close to the drift–diffusion regime, for small values of the Knudsen number λ, and in space for large values of the electric field E. It therefore pays to investigate the stability properties of the system (3.13) before writing down any approximation scheme. We will now give a simple linear stability estimate, first derived by Poupaud (1991), which indicates how to discretize the system in the spatial, time and energy variables. The Galerkin approach implies the equation

$$2 \int_0^\pi \mathrm{d}\theta \int_{-\pi}^\pi \mathrm{d}\phi \ \sin(\theta) f^N \left(\lambda^2 \partial_t f^N + \lambda v \cdot \nabla_x f^N - \lambda E \cdot \left(a \partial_r f^N + b \partial_\theta f^N + c \partial_\phi f^N \right) - Q(f^N) \right) = 0. \tag{3.15}$$

Integrating (3.15) by parts yields

$$\begin{aligned} &\lambda^2 \partial_t G \\ &\quad + \int_0^\pi \mathrm{d}\theta \int_{-\pi}^\pi \mathrm{d}\phi \ \sin(\theta) \lambda \left(\mathrm{div}_x \left(v f^{N2} \right) - E \cdot a \partial_r \left(f^{N2} \right) \right) \\ &\quad + f^{N2} E \cdot \left(\partial_\theta (b \sin\theta) + \partial_\phi (\sin\theta c) \right) - 2 f^N Q \left(f^N \right) = 0, \end{aligned} \tag{3.16}$$

where

$$G(x,r,t) = \int_0^\pi \mathrm{d}\theta \int_{-\pi}^\pi \mathrm{d}\phi \ \sin(\theta) f^{N2} = \sum_{n\in N} f_n(x,r,t)^2 \tag{3.17}$$

is the norm of the coefficient vector (f_n).

Because of the properties of polar coordinates, we have

$$\partial_\theta(b\sin\theta) + \partial_\phi(\sin\theta c) = -2\frac{\sin(\theta)}{r}a, \tag{3.18}$$

and (3.16) can be rewritten as

$$\begin{aligned}&\lambda^2\partial_t G\\ &\quad + \int_0^\pi \mathrm{d}\theta \int_{-\pi}^\pi \mathrm{d}\phi \ \sin(\theta)\lambda\left(\mathrm{div}_x\left(vf^{N2}\right) - E\cdot a\frac{1}{r^2}\partial_r\left(r^2 f^{N2}\right)\right)\\ &\quad -2f^N Q\left(f^N\right) = 0.\end{aligned} \tag{3.19}$$

Using the scaled Maxwellian $M(r) = \exp(-r^2/2)$, the term $r^{-2}\partial_r(r^2 f^{N2})$ can be rewritten as $r^{-2}\partial_r(r^2 f^{N2}) = (M/r^2)\partial_r((r^2/M)f^{N2}) - rf^{N2}$ and, since $v = ar$ and $E = -\nabla_x V$ holds, we have

$$E\cdot a\frac{1}{r^2}\partial_r\left(r^2 f^{N2}\right) = E\cdot a\frac{M}{r^2}\partial_r\left(\frac{r^2}{M}f^{N2}\right) + \nabla_x V\cdot v f^{N2}. \tag{3.20}$$

Thus equation (3.16) can be written in conservation form as

$$\begin{aligned}&\lambda^2\partial_t G\\ &\quad + \int_0^\pi \mathrm{d}\theta \int_{-\pi}^\pi \mathrm{d}\phi \ \sin(\theta)\lambda\left(e^{-V}\,\mathrm{div}_x\left(e^V v f^{N2}\right) - E\cdot a\frac{M}{r^2}\partial_r\left(\frac{r^2}{M}f^{N2}\right)\right)\\ &\quad -2f^N Q(f^N) = 0.\end{aligned} \tag{3.21}$$

Next, we note the following basic identity for the linearized collision operator $Q(f) = \int_{\mathbb{R}^3} \mathrm{d}v s[Mf' - M'f]$:

$$\begin{aligned}&2\int_0^\pi \mathrm{d}\theta \int_{-\pi}^\pi \mathrm{d}\phi \ \sin(\theta) r^2 f Q(f)\\ &\quad = \ -\int_0^\pi \mathrm{d}\theta \int_{-\pi}^\pi \mathrm{d}\phi \int_0^\pi \mathrm{d}\theta' \int_{-\pi}^\pi \mathrm{d}\phi' \ \sin(\theta) r^2 \sin(\theta') r'^2 s M M' \left(\frac{f}{M} - \frac{f'}{M'}\right)^2\\ &\quad < \ 0,\end{aligned} \tag{3.22}$$

which can be verified by direct calculation. Therefore, multiplying (3.21) by $e^V r^2/M$ and integrating with respect to x and r gives

$$\begin{aligned}&\partial_t \int_\Omega \mathrm{d}x \int_0^\infty \mathrm{d}r \ e^{-V}\frac{r^2}{M}G(x,r,t) <\\ &\quad -\int_\Omega \mathrm{d}x \int_0^\infty \mathrm{d}r \ (\partial_t V)e^{-V}\frac{r^2}{M}G(x,r,t) - \frac{1}{\lambda}\int_{\partial\Omega}\int_0^\infty \mathrm{d}r\frac{r^2}{M}e^{-V}(\nu\cdot v)f^{N2},\end{aligned} \tag{3.23}$$

where ν denotes the unit outward normal vector of the domain Ω. The second term on the right-hand side of (3.23) has to be controlled by the boundary fluxes and (3.23) yields a Gronwall inequality for the term

$$\int_\Omega \mathrm{d}x \int_0^\infty \mathrm{d}r\ \frac{r^2}{M} e^{-V} G(x,r,t) = \sum_{n\in N} \int_\Omega \mathrm{d}x \int_0^\infty \mathrm{d}r\ \frac{r^2}{M} e^{-V} f_n(x,r,t)^2. \tag{3.24}$$

The significance of the estimate (3.24) lies in the fact that it is independent of the Knudsen number λ, and therefore is a valid stability result even in the stiff case close to the drift–diffusion regime. Therefore, the discretization of the hyperbolic system (3.13) in the x,t and r variables should reflect this estimate. Without going into the explicit details, we will now show how to construct the discretization for system (3.13). Because of the equality (3.18) the matrices A_i and B_i, made up of the coefficients A_{inm} and B_{inm} in (3.13), satisfy

$$B_i + B_i^T = \frac{2}{r} A_i, \quad i = 1,2,3, \tag{3.25}$$

and, consequently, the matrices $B_i - (1/r)A_i$ are skew symmetric. In the linear case the terms q_n on the right-hand side of (3.13) are given by a matrix integral operator of the form

$$q_n = \sum_{n\in N} C_{nm} f_m, \tag{3.26}$$

$$\text{where} \quad C_{nm} g(x,r,t) = 4\pi \int_0^\infty \mathrm{d}r'\ r'^2 s\Big(\delta(n)\delta(m) M g' - \delta(n-m) M' g\Big),$$

and, because of (3.22), the matrix operator C satisfies

$$\int_0^\infty \mathrm{d}r\ \frac{r^2}{M} \mathbf{f}^T C \mathbf{f} < 0, \tag{3.27}$$

where $\mathbf{f}$ denotes the vector of coefficients (f_n). The stability estimate for the semi-discretized Boltzmann transport equation suggests rewriting the system as

$$\lambda^2 \left(\partial_t \left(e^{-V/2}\mathbf{f}\right) + \frac{1}{2}(\partial_t V) e^{-V/2}\mathbf{f}\right) + \lambda r A_i \partial_{x_i}\left(e^{-V/2}\mathbf{f}\right) + \tag{3.28}$$
$$\frac{\lambda}{r}\left(M e^{-V}\right)^{1/2} E_i A_i \partial_r \left(r M^{-1/2}\mathbf{f}\right) + \lambda e^{-V/2} E_i \left(B_i - \frac{1}{r} A_i\right)\mathbf{f} = e^{-V/2} C\mathbf{f}.$$

Multiplying (3.28) from the left with $(r^2/M)e^{-V/2}\mathbf{f}^T$ and integrating with respect to x and r gives the stability estimate

$$\partial_t \int_\Omega \mathrm{d}x \int_0^\infty \mathrm{d}r\ \frac{r^2}{M} e^{-V} |\mathbf{f}|^2 = -\int_\Omega \mathrm{d}x \int_0^\infty \mathrm{d}r\ \frac{r^2}{M} (\partial_t V) e^{-V} |\mathbf{f}|^2. \tag{3.29}$$

Therefore, any difference discretization of the system (3.13) should build on differencing the terms $e^{-V/2}\mathbf{f}$ in the spatial and $rM^{-1/2}\mathbf{f}$ in the energy direction. A completely time-implicit scheme for (3.13) would immediately reproduce the stability estimate (3.29). Unfortunately, completely time-implicit schemes are usually too computationally expensive. From (3.28) it is clear, however, that at least the collision term on the right-hand side should be discretized implicitly in time in order to reduce the $O(\lambda^2)$ stiffness of the system to $O(\lambda)$. So, system (3.13) should be discretized as

$$\lambda^2\partial_t f_n - q_n = \lambda r \exp\left(\frac{V}{2}\right) A_{inm}\partial_{x_i}\left(\exp\left(-\frac{V}{2}\right) f_m\right) - \lambda E_i\left(\frac{1}{r}M^{1/2}A_{inm}\partial_r\left(rM^{-1/2}f_m\right) + \left(B_{inm} - \frac{1}{r}A_{inm}\right) f_m\right) \tag{3.30a}$$

$$q_n = \sum_{j=-1}^{1}\gamma(|j|)\int_0^\infty \mathrm{d}r'\; r'^2\delta(r^2 - r'^2 - j\omega_{ph}) \left(Mf_0'\left(4\pi\delta(n) - \sqrt{4\pi}f_n\right) - M'f_n\left(4\pi - \sqrt{4\pi}f_0'\right)\right), \tag{3.30b}$$

where the terms on the left-hand side are taken implicitly, and a standard hyperbolic scheme, such as Lax–Wendroff, is used on the right-hand side, yielding a Courant–Friedrich–Lewy (CFL) condition of the form $\Delta t/(\lambda\Delta x) <$ const. In the case of a linear collision operator ($1 - f \approx 1$ in (3.30b)), if the mesh size in the energy direction is taken as an integer fraction of the energy ω_{ph} of the emission of a polar optical phonon, the collision operator C becomes a sparse matrix whose LU decomposition can be computed once and reused for every grid point in the x-direction (Ventura et al. 1991, 1992). An alternative to this approach is to use a spectral discretization in the energy direction as well. In order to preserve the stability estimate (3.23), it is necessary to use the function r^2/M as a weight function for the scalar product. This approach has, in principle, been used in Schmeiser and Zwirchmayr (1995) and (1997), although there, Cartesian coordinates in v and Laguerre polynomial basis functions are used. Consequently, the matrix collision operator matrix C becomes a full matrix and the scheme is restricted to relatively few terms in the expansion.

3.2. The hydrodynamic model

The series expansion methods described in the previous section are centred around an almost spherically symmetric density function. Although they present a non-perturbative theory and are always convergent, we can expect slow convergence far from equilibrium, that is, in the case of large group velocities. As device dimensions decrease, the value of the Knudsen number λ increases, and the transport picture is not dominated by the collision term any longer. In order to study ballistic transport in short channels of

a transistor, an alternative model is used. This model corresponds to the compressible Euler equations for a gas driven by an external force. This model, not very aptly named the 'hydrodynamic model for semiconductors', is obtained by taking moments of the Boltzmann equation with respect to the wave vector k, corresponding to the macroscopic particle density, the group velocity and the energy. The hydrodynamic model usually assumes a parabolic band structure as given in (2.2). Multiplying the Boltzmann transport equation (2.17) by $1, k$ and $|k|^2$ and integrating with respect to the wave vector k gives the moment equations

$$\lambda^2 \partial_t \langle 1 \rangle + \lambda \partial_{x_l} \langle k_l \rangle = 0, \tag{3.31a}$$

$$\lambda^2 \partial_t \langle k_j \rangle + \lambda \partial_{x_l} \langle k_l k_j \rangle + \lambda E_j \langle 1 \rangle = \left\langle \frac{k_j}{f} Q(f) \right\rangle, \tag{3.31b}$$

$$\lambda^2 \partial_t \left\langle \frac{1}{2}|k|^2 \right\rangle + \lambda \partial_{x_l} \left\langle \frac{1}{2} k_l |k|^2 \right\rangle + \lambda E_l \langle k_l \rangle = \left\langle \frac{|k|^2}{2f} Q(f) \right\rangle, \tag{3.31c}$$

where the symbol $\langle \cdot \rangle$ denotes the expectation of a quantity with respect to the density f. So

$$< g >= \int_{\mathbb{R}^3} \mathrm{d}k \; (gf) \tag{3.32}$$

holds and the summation convention is again used in (3.31). (3.31) is regarded as a system for the particle density $n = \langle 1 \rangle$, the moment $\langle k \rangle = nv$ and the total energy $W = \langle (1/2)|k|^2 \rangle$. The so-called closure problem is then given by expressing the higher order moments $k_j k_l$, $k_l |k|^2$ and the moments of the collision operator Q in terms of the primary variables n, v and W. This is achieved by the assumption that the density function f is approximately equal to a displaced Maxwellian of the form

$$f(x,k,t) \approx \mu(x,t) \exp\left(-\frac{|k - \lambda v(x,t)|^2}{2T(x,t)} \right), \tag{3.33}$$

where v denotes the macroscopic velocity and T the electron temperature. Under assumption (3.33), the higher order moments are of the form

$$\langle k \rangle = \lambda n v, \quad \langle k_j k_l \rangle = n(T\delta_{jl} + \lambda^2 v_j v_l), \tag{3.34}$$

$$W = \frac{1}{2}\left\langle |k|^2 \right\rangle = \frac{n}{2}(3T + \lambda^2 |v|^2), \quad \left\langle k_l |k|^2 \right\rangle = \lambda v_l n \frac{1}{2}(5T + \lambda^2 |v|^2),$$

and (3.31) can be written as

$$\partial_t \langle n \rangle + \partial_{x_l} \langle v_l n \rangle = 0 \tag{3.35a}$$

$$\lambda^2 \partial_t (n v_j) + \lambda^2 \partial_{x_l} (n v_l v_j) + \partial_{x_j} (nT) + E_j n = \left\langle \frac{k_j}{\lambda f} Q(f) \right\rangle \tag{3.35b}$$

$$\partial_t W + \partial_{x_l}(v_l W) + \partial_{x_l}(n v_l T) \qquad (3.35c)$$
$$+\lambda E_l v_l n - \operatorname{div}_x(\kappa \nabla_x T) = \left\langle \frac{|k|^2}{2f\lambda^2} Q(f) \right\rangle.$$

Here the term $\operatorname{div}_x(\kappa \nabla_x T)$ denotes the heat flux which is derived from a higher order perturbation theory, and κ denotes the heat conductivity. The moments of the collision terms are modelled phenomenologically as

$$\left\langle \frac{k_j}{f} Q(f) \right\rangle = -\frac{\lambda}{\tau_p} n v_j, \quad \left\langle \frac{|k|^2}{2f} Q(f) \right\rangle = -\frac{1}{2\tau_W}\left(3(T - T_0) + \lambda^2 |v|^2\right) \qquad (3.36)$$

(see Baccarani and Wordeman (1985)).

Note that, for $\lambda \to 0$, the transport terms in (3.35) vanish and $T = T_0$ holds. Thus the drift–diffusion system is recovered in the limit. However, the hydrodynamic model is used in regimes where the active region of the device is of the same order as the mean free path and $\lambda = O(1)$ usually holds. For $\kappa = 0$, neglecting the heat flux, the hydrodynamic model represents a nonlinear hyperbolic system with sound speed $c = (1/\lambda)\sqrt{5T/3}$ which will usually exhibit shocks for $\lambda = O(1)$. For finite values of the heat conductivity κ, these shocks will have a finite width of order $O(\kappa)$ (Gardner 1991*b*). For short and relatively small active regions (of the order of 0.1–0.5 μm) the hydrodynamic model equations (3.35) usually lead to a sufficiently good agreement with Monte Carlo simulations (Gardner 1993*b*).

Steady state calculations

The most common approach to discretizing the hydrodynamic model equations in steady state is upwind box integration (Gardner 1992, Gardner 1991*a*). To this end, the steady state version of the model equations (3.35) is put in conservation form as

$$\partial_{x_l}(v_l G_{lj}) + H_j = 0, \quad j = 0, \ldots, 5, \qquad (3.37)$$

where the G_{lj} and H_j are given by

$$G_{l0} = n \quad \text{and} \quad G_{lj} = \lambda^2 n v_j, \quad j = 1, 2, 3, \qquad (3.38)$$

$$G_{l4} = (n/2)(5T + \lambda^2 |v|^2) - n\partial_{x_l} V, \quad G_{l5} = 0,$$

$$H_0 = 0, \quad H_j = \partial_{x_j}(nT) - n\partial_{x_j} V - \frac{1}{\tau_p} n v_j, \quad j = 1, 2, 3,$$

$$H_4 = -\nabla_x \cdot (\kappa \nabla_x T) + \frac{1}{2\lambda^2 \tau_W}\left(3(T - T_0) + \lambda^2 |v|^2\right),$$

$$\text{and} \quad H_5 = -\operatorname{div}_x(e \nabla_x V) = n - D,$$

where we have already included the Poisson equation for the potential V. Using the upwind box integration method, the terms $\partial_{x_l}(v_l G_{lj})$ in (3.37) are discretized by

$$\partial_{x_l}(v_l G_{lj})(x) = \frac{1}{(\mu_l \Delta x_l)} \delta_l \left((\mu_l v_l)(\mu_l G_{lj}) - \frac{1}{2} |\mu_l v_l| (\delta_l G_{lj}) \right), \tag{3.39}$$

where the discrete difference and averaging operators δ_l and μ_l are defined by

$$\begin{aligned} \delta_l z(x) &= z\left(x + \frac{1}{2}\Delta x_l e_l\right) - z\left(x - \frac{1}{2}\Delta x_l e_l\right), \\ \mu_l z(x) &= \frac{1}{2}\left(z\left(x + \frac{1}{2}\Delta x_l e_l\right) + z\left(x - \frac{1}{2}\Delta x_l e_l\right)\right), \end{aligned} \tag{3.40}$$

where Δx_l denotes the (variable) stepsize and e_l denotes the unit vector in the lth coordinate direction. The derivatives in the terms H_j in (3.37) are usually discretized by standard centred finite differences (Gardner 1991*b*, Gardner 1992). In order to deal with locally large electric fields, *a modification of the Scharfetter–Gummel scheme (Selberherr 1981) may be* used for $H_j, j = 1, 2, 3$. In this modification the derivative

$$H_j = \partial_{x_j}(nT) - n\partial_{x_j} V - \frac{n}{\tau_p} v_j \tag{3.41}$$

is written as

$$H_j = -\partial_{x_j}\left(\frac{V}{T}\right) \frac{\partial_{x_j}\left(nT \exp(-\frac{V}{T})\right)}{\partial_{x_j}\left(\exp(-\frac{V}{T})\right)} - nV\partial_{x_j}\left(\log(|T|)\right) - \frac{n}{\tau_p} v_j. \tag{3.42}$$

The derivatives in (3.42) are then discretized by using standard differences. For constant temperature T the discretization of (3.42) then reduces to the classical Scharfetter–Gummel scheme, which has the advantage of correctly performing the right upwinding in the direction of the electric field $E = -\nabla_x V$.

After carrying out the discretization, a large sparse nonlinear system of algebraic equations has to be solved. After linearization this leads to the solution of the linear system

$$J\Delta F = -F(z), \quad z = (n, v, T, V)^T, \tag{3.43}$$

at each step. Here the vector F denotes the discretization of (3.37) on the

mesh and the Jacobian J has the block structure

$$J = \begin{pmatrix} \frac{\partial F_0}{\partial n} & \frac{\partial F_0}{\partial v} & 0 & 0 \\ \frac{\partial F_j}{\partial n} & \frac{\partial F_j}{\partial v} & \frac{\partial F_j}{\partial T} & \frac{\partial F_j}{\partial V} \\ \frac{\partial F_4}{\partial n} & \frac{\partial F_4}{\partial v} & \frac{\partial F_4}{\partial T} & \frac{\partial F_4}{\partial V} \\ \frac{\partial F_5}{\partial n} & 0 & 0 & \frac{\partial F_0}{\partial V} \end{pmatrix}. \tag{3.44}$$

In Lanzkron, Gardner and Rose (1991) and Gardner, Jerome and Rose (1989), detailed investigations of the convergence of block iterative methods for the solution of the Newton equations (3.43) have been carried out. The basic result is that block under-relaxation methods, in conjunction with conjugate gradient methods for the individual blocks, perform well as long as the equations for the density n and the velocities v are treated as one block. In this case chaotic under-relaxation methods on parallel architectures also give good results.

3.3. Generalizations of the hydrodynamic model – Grad systems

The relative simplicity of the hydrodynamic model equations and their certain shortcomings, such as the overestimation of velocities close to P-N junctions (see Ringhofer (1997)), suggests a generalization of the underlying principle to higher order moment methods. In the ballistic regime, that is, in the presence of large electron velocities, series expansion methods based on a perturbation of a spherical symmetric density function will in general not perform well. This suggests the introduction of a modified series expansion approach based not on a centred Maxwellian distribution function but on a wave vector displaced Maxwellian instead. This idea was first introduced by Grad (1949, 1958) for the study of the fine structure of shock waves in fluid dynamics. Since the assumption underlying the hydrodynamic model is that the density function is approximately of the form (3.33), it is natural to expand the Boltzmann equation around a Maxwellian distribution function in a stretched variable coordinate system in wave vector space with the macroscopic velocity u as the origin and the square root of the macroscopic temperature T as the stretching factor. Introducing the coordinate transformation

$$(x, v, t) \to (x, w, t), \quad v = \alpha(x, t)w + u(x, t) \tag{3.45}$$

gives the transformed Boltzmann equation

$$\lambda^2 \partial_t f + \lambda v \cdot \nabla_x f - \frac{\lambda}{\alpha} H \cdot \nabla_w f = Q(f) \tag{3.46}$$

$$H = E + \lambda((\partial_t \alpha)w + \partial_t u) + \left(w(\nabla_x \alpha)^T + \frac{\partial u}{\partial x} \right) v, \quad v = \alpha w + u.$$

Equation (3.46) is now approximated by a Galerkin method in the microscopic velocity variable v where α and u are still kept as free parameters. There are two conditions which we have to place on the choice of basis functions and the scalar product in order to obtain a generalization of the hydrodynamic model.

- The lowest order basis function is the displaced Maxwellian $e^{-|w|^2/2}$.
- The Galerkin approach should correspond to taking the moments of the Boltzmann equation (3.46).

This is achieved by choosing basis functions of the form

$$\psi_m = M(w)p_m(w), \quad M(w) = \exp\left(-\frac{|w|^2}{2}\right), \tag{3.47}$$

where the p_m in (3.47) are vector basis polynomials containing the polynomials $1, w$ and $|w|^2$. So

$$\{1, w, |w|^2\} \subseteq \text{span}\{p_0, \ldots, p_N\} \tag{3.48}$$

holds. Secondly, the scalar product is taken to be of the form

$$< f, g >= \int_{\mathbb{R}^3} \mathrm{d}w \; \frac{1}{M(w)} f^T g. \tag{3.49}$$

Thus, taking the scalar product of the Boltzmann equation with the basis function ψ_m corresponds to integrating against the polynomial p_m, and the moments leading to the hydrodynamic model are reproduced. Expanding the density function

$$f(x, w, t) \approx \alpha^{-3} \sum_n f_n(x, t)\psi_n(w) \tag{3.50}$$

and using the Galerkin procedure yields the system

$$\lambda^2 \partial_t f_m + \lambda \partial_{x_l}[A_{lmn} f_n] + \frac{\lambda}{\alpha} B_{mn} f_n = C_{mn} f_n, \tag{3.51}$$

with

$$\begin{aligned} A_{lmn} &= \langle \psi_m, v_l \psi_n \rangle, \\ B_{mn} &= -\langle \psi_m, \operatorname{div}_w(H\psi_n) \rangle = \int_{\mathbb{R}^3} \mathrm{d}w \; \psi_n H \cdot \nabla_w p_m, \\ C_{mn} &= \langle \psi_m, Q\psi_n \rangle. \end{aligned} \tag{3.52}$$

It remains to choose the parameters $\alpha(x, t)$ and $u(x, t)$. In the original Grad system (Grad 1949), they are chosen to correspond to the square root of the temperature and the group velocity:

$$u = \frac{\int_{\mathbb{R}^3} \mathrm{d}v \; vf}{\int_{\mathbb{R}^3} \mathrm{d}v \; f}, \quad |u|^2 + 3\alpha^2 = \frac{\int_{\mathbb{R}^3} \mathrm{d}v \; |v|^2 f}{\int_{\mathbb{R}^3} \mathrm{d}v \; f}, \tag{3.53}$$

which implies

$$\int_{\mathbb{R}^3} \mathrm{d}w\ [wf] = 0, \quad \int_{\mathbb{R}^3} \mathrm{d}w\ [|w|^2 - 3]f = 0. \tag{3.54}$$

If we denote the basis functions $M, w_1 M, w_2 M, w_3 M, |w|^2 M$ by $\psi_0, \ldots, \psi_4$, this becomes

$$f_j = 0, \quad j = 1, 2, 3 \quad \text{and} \quad f_4 = \text{const} f_0. \tag{3.55}$$

Thus $f_1, \ldots, f_4$ can be eliminated from the system and the corresponding equations determine the free parameters α and u. By virtue of construction the system reduces to the hydrodynamic model (without heat conduction) if only the basis functions $\psi_0, \ldots, \psi_4$ are used. One of the major problem of Grad systems is that they are not necessarily well-posed. Equation (3.51), together with the constraints (3.55), represents a hyperbolic differential algebraic system. If the coefficient functions $f_1, \ldots, f_4$ are eliminated from the system the resulting equations form a first-order system for the variables $(f_0, \alpha, u_1, u_2, u_3, f_5 \ldots)$ whose linearization can have complex eigenvalues, and therefore modes can grow proportionally to their frequencies. This ill-posedness occurs at quite moderate Mach numbers and has been analysed by Cordier (1994*a*, 1994*b*) for some special sets of basis functions. Several approaches to remedy this problem have been given by Ringhofer (1994*b*, 1994*a*, 1997). They involve relaxing the conditions (3.55) in some way or another, and lead to well-posed problems.

4. Numerical methods for quantum transport

The numerical methods discussed in this section are essentially mirror images of the methods for semiclassical transport from Section 3. The quantum Boltzmann equation (2.38) replaces the semiclassical Boltzmann equation (2.1) and its moment expansion gives the quantum hydrodynamic model. There are, however, several important differences which do not allow us to treat quantum transport phenomena as just a perturbation of semiclassical transport. First, the discretization of the pseudo-differential operator θ in (2.38) is not trivial. The transport term on the left-hand side of (2.38) does not possess classical characteristics. (They would be replaced by the paths in the Feynman path integrals.) Second, because of the nonlocality of the transport operator, the formulation of proper boundary conditions is more complicated than in the classical case. Finally, because of the dispersive nature of the underlying Schrödinger equation, moment models, such as the quantum hydrodynamic equations, will also be dispersive, that is, waves will be able to travel at all speeds. Therefore, the artificial diffusion introduced by a discretization scheme will play a crucial role in its accuracy.

4.1. Discretization of the quantum Boltzmann equation

We now turn to the design of numerical methods for the quantum Boltzmann equation (2.38). There is a variety of possible choices for discretization schemes in the spatial and temporal directions, which will be discussed in more detail later. The more fundamental problem is posed by the discretization of the wave vector k, in particular by the approximation of the pseudo-differential operator θ, the quantum equivalent of the operator $-\nabla_x V \cdot \nabla_k$ in the classical Boltzmann equation. Since the quantum Boltzmann equation does not possess characteristics in the classical sense, and the Wigner function w does not necessarily remain nonnegative (see Tatarski (1983)), a Monte Carlo type approach becomes too complicated to be practically feasible. On the other hand, since the operator θ is defined in terms of Fourier transforms, a spectral discretization using trigonometric basis functions seems natural.

First we note that the quantum Boltzmann equation (2.38) allows for a reduction in dimension. If the potential V is only dependent on the variables $x_1, \ldots, x_d$ with $d = 1$ or 2 (so there is no field pointing in the direction $x_{d+1}, \ldots, x_3$) the quantum Boltzmann equation with both collision terms (2.39),(2.40) allows for solutions of the form

$$w(x,k,t) = \exp\left(-\frac{1}{2}\left(k_{d+1}^2 + \cdots + k_3^2\right)\right)\tilde{w}(x_1,\ldots,x_d,k_1,\ldots,k_d,t). \quad (4.1)$$

The dimensionally reduced quantum Boltzmann equation for $\tilde{w}$ is then of the form

$$\partial_t \tilde{w} + \tilde{k}\cdot\nabla_{\tilde{x}}\tilde{w} + \theta[V]\tilde{w} = Q(\tilde{w}) \quad (4.2a)$$

$$\theta[V]w(\tilde{x},\tilde{k},t) = (2\pi)^{-d} \quad (4.2b)$$
$$\int_{\mathbb{R}^d} \mathrm{d}\tilde{k}' \int_{\mathbb{R}^d} \mathrm{d}\eta\ \frac{i}{h}\delta V\left(\tilde{x},\frac{h}{2}\eta,t\right) w(\tilde{x},\tilde{k},t)\exp\Big(i\eta\cdot(\tilde{k}-\tilde{k}')\Big),$$

where $\tilde{x} = (x_1,\ldots,x_d)^T$, $\tilde{k} = (k_1,\ldots,k_d)^T$.

Here we have already used the quantum Boltzmann equation in a scaled and dimensionless form, where h denotes the scaled Planck constant $\hbar$. Of course, the Poisson equation has to be modified accordingly to take into account the integral of the Maxwellian in the directions $k_{d+1},\ldots,k_3$. Since the reduced quantum Boltzmann equation has the same form as the three-dimensional equation, we will from now on drop the tilde symbol. Following Ringhofer (1990) and (1992), we approximate the Wigner function w by a trigonometric polynomial of the form

$$w \approx w_N(x,k,t) = \sum_{n\in\mathbf{N}} c(x,n,t)\phi_n(k), \quad \mathbf{N} = \{-N+1,\ldots,N\}^d, \quad (4.3)$$

$$\phi_n(k) = \left(\frac{\alpha}{2\pi}\right)^{d/2} \exp(i\alpha n \cdot k).$$

Thus we approximate the L^2 function w by the $(2\pi/\alpha)$-periodic function w_N and, consequently, α will have to go to zero in the limit to achieve convergence. The quantum Boltzmann equation (4.2) is simply approximated by collocation at the appropriate equally spaced nodes. So

$$\partial_t w_N + k \cdot \nabla_x w_N + \theta[V] w_N = Q_N(w_N) \quad \text{at} \quad k = k_j, \tag{4.4}$$

$$k_j = \beta j, \quad j \in \mathbf{N}, \quad \beta = \frac{\pi}{N\alpha},$$

holds. If the relaxation time approximation is used for the collision term Q then the corresponding approximation Q_N has to be modified accordingly to account for the fact that densities are now computed from periodic basis functions. So, in this case,

$$Q_N(w)(x,k,t) = \frac{1}{\tau}\left(\frac{n}{n_0} w_{0N} - w\right), \tag{4.5}$$

$$n = \int_{[-\pi/\alpha,\pi/\alpha]^d} w, \quad n_0 = \int_{[-\pi/\alpha,\pi/\alpha]^d} w_{0N}, \quad M_N = \sum_{n\in\mathbf{N}} c_0(x,n)\phi_n,$$

holds. Here w_{0N} denotes a suitable approximation of the thermal equilibrium density w_0. The advantage of this approach lies in the fact that the highly oscillatory integrals in the definition (4.2b) of the pseudo-differential operator θ can be evaluated exactly. The basis functions ϕ_n satisfy the orthogonality relations

$$\int_{[-\pi/\alpha,\pi/\alpha]^d} \phi_m^* \phi_n = \delta(m-n), \quad \beta^d \sum_{\nu\in\mathbf{N}} \phi_m^*(k_\nu)\phi_n(k_\nu) = \delta_N(m-n), \tag{4.6}$$

where δ_N denotes the Kronecker δ on $\mathbf{N}$ periodically extended over all integers. Using these orthogonality relations, a direct calculation yields

$$\begin{aligned} \theta w_N(x,k,t) &= \sum_{n\in\mathbf{N}} \frac{i}{h} \delta V(x, \frac{\alpha h}{2} n, t) c(x,n,t) \phi_n(k), \\ c(x,n,t) &= \beta^d \sum_{\nu\in\mathbf{N}} \phi_n^*(k_\nu) w_N(x,k_\nu,t). \end{aligned} \tag{4.7}$$

Collecting the function values of the trigonometric polynomial w_N at the collocation points k_ν into a vector W, one obtains the hyperbolic system

$$\partial_t W + \Lambda_j \partial_{x_j} W + B(V)W = QW, \tag{4.8}$$

where the Λ_j are diagonal matrices made up of the jth component of the

collocation points k_ν, and the tensor B is given by

$$B(\mu,\nu) = \beta^d \sum_{n\in\mathbf{N}} \frac{i}{h}\delta V\left(x, \frac{\alpha h}{2}n, t\right)\phi_N^*(\beta\mu)\phi_n(\beta\nu). \tag{4.9}$$

Multiplication with the tensor B can now be carried out using FFTs, a significant advantage in higher dimensions.

Spectral accuracy

A complete convergence proof for the semi-discretized scheme can be found in Ringhofer (1990) and (1992), and turns out to be quite tedious. We will here only sketch the consistency of the discretization in order to indicate under what conditions and with what order the scheme is convergent. Note that the discretization scheme (4.8) is somewhat nonstandard. The basis functions ϕ_n are not elements of the same space as the exact solution, since we have approximated the L^2 solution by periodic functions. Let the interpolation operator P be defined by

$$Pw(x,k,t) = \sum_{n\in\mathbf{N}} c(x,n,t)\phi_n(k), \quad Pw(x,k_\nu,t) = w(x,k_\nu,t), \quad \nu\in\mathbf{N}. \tag{4.10}$$

Then the scheme can be formally written as

$$P\Big(\partial_t w_N + k\cdot\nabla_x w_N + \theta[V]w_N - Q_N w_N\Big) = 0. \tag{4.11}$$

Defining the global discretization error as $e = w_N - Pw$ we obtain

$$P\Big(\partial_t e + k\cdot\nabla_x e + \theta[V]e - Q_N e\Big) = L, \tag{4.12}$$

where the local discretization error L is given by

$$\begin{aligned} L &= -P\Big(\partial_t Pw + k\cdot\nabla_x Pw + \theta[V]Pw - Q_N Pw\Big) \\ &= (P\theta - \theta P)w + (Q_N P - PQ)w \end{aligned} \tag{4.13}$$

The interpolation operator P has the representation

$$Pf(k) = \beta^d \sum_{n\in\mathbf{N}}\sum_{\nu\in\mathbf{N}} f(k_\nu)\phi_\nu^*(k_\nu)\phi_\nu(k) \tag{4.14}$$

and, consequently, the interpolant of any $(2\pi/\alpha)$-periodic function f is given by

$$Pf = \sum_{n\in\mathbf{N}}\sum_{s\in Z^d} \hat{f}(n+2Ns)\phi_n, \quad f = \sum_{n\in Z^d}\hat{f}(n)\phi_n. \tag{4.15}$$

The formula (4.15) represents the usual aliasing error. The exact solution w is now smoothly decomposed into a part which vanishes identically outside the interval $[-\pi/\alpha, \pi/\alpha]^d$ and a part which vanishes identically inside a

subinterval of $[-\pi/\alpha, \pi/\alpha]^d$, that is,

$$w = w_i + w_o, \quad w_i = 0 \quad \text{for} \quad k \notin [-\frac{\pi}{\alpha+\varepsilon}, \frac{\pi}{\alpha+\varepsilon}]^d,$$
$$w_o = 0 \quad \text{for} \quad k \in [-\frac{\pi}{\alpha+2\varepsilon}, \frac{\pi}{\alpha+2\varepsilon}]^d. \tag{4.16}$$

So w_o denotes the tail of the distribution and w_i equals w in the smaller domain. Clearly w_i is $(2\pi/\alpha)$-periodic and, therefore, using (4.15),

$$\begin{aligned}(\theta P - P\theta)w_i &= \sum_{n\in\mathbf{N}} \sum_{s\in Z^d-\{0\}} \\ &\hat{w}_i(n+2Ns)\Big(\delta V(x,\alpha n,t) - \delta V(x,\alpha n+2Ns,t)\Big)\phi_n, \\ &\text{where} \quad w_i = \sum_{n\in Z^d} \hat{w}_i(n)\phi_n, \end{aligned} \tag{4.17}$$

holds. Note that the sum in equation (4.17) only contains Fourier coefficients with indices larger than N, and therefore

$$\|(\theta P - P\theta)w_i\|_{L^2([-\frac{\pi}{\alpha},\frac{\pi}{\alpha}]^d)} < c_q \|\delta V\|_\infty \|w_i\|_{H_p([-\frac{\pi}{\alpha},\frac{\pi}{\alpha}]^d)} \tag{4.18}$$

holds, which gives the usual estimate for spectral accuracy of the discretization scheme.

Time discretization

After employing the spectral collocation scheme in the wave vector direction, it remains to discretize the first-order hyperbolic system (4.8) in space and time. Of course, every method for hyperbolic systems would do this job. However, the use of a standard hyperbolic scheme for (4.8) will result in a CFL condition of the form $\Delta t/(\alpha\Delta x) < \text{const}$, which will be prohibitive in practice since $\alpha \to 0$ has to hold for the spectral discretization to be convergent. The best alternative, given in Arnold and Ringhofer (1995*b*) is to employ operator splitting to the semi discretizeed equation (4.8). In the operator splitting approach, one time step of length Δt for the equation (4.8), starting from $W(x,t_n)$ is performed by

$$\partial_t W_1 + \Lambda_j \partial_{x_j} W_1 = 0, \quad W_1(x,t) = W(x,t_n) \tag{4.19a}$$
$$\partial_t W_2 + B(V)W_2 = Q(W_2), \quad W_2(x,t_n) = W_1(x,t_n+\Delta t) \tag{4.19b}$$
$$W(x,t_{n+1}) = W_2(x,t_{n+1}), \quad t_{n+1} = t_n + \Delta t. \tag{4.19c}$$

This discretization is first-order accurate in time. A second-order accurate discretization can be achieved with a slight modification using so-called Strang splitting (Arnold and Ringhofer 1995*a*) . The step (4.19b) represents the solution of a system of ordinary differential equations. This can be achieved using any ODE integrator. (Actually, in the absence of the collision term Q, this step can be carried out exactly.) In theory, the first step

(4.19a) could be carried out exactly as well, since it is given by the shift

$$W_1(x, k_j, t_n + \Delta t) = W(x - k_j \Delta t, k_j, t_n), \tag{4.20}$$

eliminating any type of CFL condition. In practice, since the vector W is given on a fixed mesh on the x-axis, the term $W(x - k_j \Delta t, k_j, t_n)$ will have to be interpolated between the nearest gridpoints. Using second-order interpolation between nearest neighbours in the x-mesh and a first-order ODE integrator for step (4.19b) gives a first-order accurate scheme (Arnold and Ringhofer 1995*b*).

Boundary conditions

One of the major problems in the application of quantum kinetic models to the simulation of actual devices is the appropriate formulation of boundary conditions. In a device, the simulation region will be a bounded domain whose boundaries will consist of contacts, insulating surfaces or artificial boundaries, which are introduced to limit the size of the simulation domain. The quantum Boltzmann equation is nonlocal in the wave vector k and the transport term on the left-hand side of (2.38) does not possess classical characteristics. Nevertheless, the quantum Boltzmann equation allows for wave solutions since, at least in the collisionless case, it is equivalent to the Schrödinger equation. Thus, if care is not taken in the formulation of boundary conditions, artificial reflections of waves at the boundaries will occur, and these spurious waves will propagate back in the interior of the simulation domain. We will first treat the case of an artificial boundary, where the boundary conditions should be such that reflection of waves at the boundary is kept to a minimum. For simplicity, let us consider a one-dimensional model $x \in \mathbb{R}^1, k \in \mathbb{R}^1$ which is obtained from the Schrödinger equation in one spatial dimension. The presented methodology is given in detail in Ringhofer, Ferry and Kluksdahl (1989) and represents a generalization of the approach of Engquist and Majda (1977) for hyperbolic systems to the infinite-dimensional case. In the one-dimensional collisionless case the Wigner equation becomes

$$\partial_t w + k \partial_x w + \theta[V] w = 0, x, k \in R^1. \tag{4.21}$$

We will assume the boundary to be located at $x = 0$ and the simulation domain to be given by the half plane $x > 0$. Generalizations to more than one boundary are straightforward. In the absence of the pseudo-differential operator θ, the absorbing boundary condition would trivially be given by

$$w(x = 0, k, t) = 0, \quad \text{for} \quad k > 0, \tag{4.22}$$

since we assume that no waves enter the domain from outside the region. The same would be true for an equation of the form

$$\partial_t w + k\partial_x w + \Gamma w = 0, \quad \Gamma w = \int_{\mathbb{R}} \mathrm{d}k'\ G(x,t,k,k')w(x,k',t), \tag{4.23}$$

if the operator Γ is block diagonal in the sense that $G(x,t,k,k') = 0$ when $kk' < 0$, since the solution $w(x,k,t)$ for $k < 0$ is completely decoupled from the solution $w(x,k,t)$ for $k > 0$. The goal of the presented approach is to achieve such a decoupling asymptotically for large wave speeds. If we assume a plane wave solution in the x,t plane of the form $w(x,k,t) = g(k)\exp[i\xi(x-\omega t)]$ with a velocity ω and a frequency ξ, $\partial_t w = -i\xi\omega w$ holds and the time derivative will be proportional to the velocity ω. Thus, we formally decouple $k < 0$ from $k > 0$ by expanding the operator in powers of ∂_t^{-1} for 'large ∂_t'. This will be made more precise later. Setting formally

$$u = w - \partial_t^{-1} A[kw], \quad A[f](x,k,t) = \int_{\mathbb{R}} \mathrm{d}k'\ a(x,t,k,k')f(x,k',t), \tag{4.24}$$

we obtain

$$k\partial_x u = (1 - \partial_t^{-1} kA)[k\partial_x w] - k\partial_t^{-1} A_x[kw], \tag{4.25}$$

where the operator A_x arises from differentiating the product, so

$$A_x[f](x,k,t) = \int_{\mathbb{R}} \mathrm{d}k'\ d_x a(x,t,k,k')f(x,k',t) \tag{4.26}$$

holds. Using the differential equation (4.21) yields

$$k\partial_x u = (1 - \partial_t^{-1} kA)(-\partial_t - \theta[V])[w - k\partial_t^{-1} A_x(kw)]. \tag{4.27}$$

Asymptotically, the inverse of the operator $1 - \partial_t^{-1} Ak$ will be given by $1 + \partial_t^{-1} Ak$ and $w = u + \partial_t^{-1} A(ku) + O(\partial_t^{-2})$ holds. So, formally, up to terms of order $O(\partial_t^{-2})$ we obtain

$$\begin{aligned} k\partial_x u &= -\partial_t u + (\partial_t^{-1} kA\partial_t - Ak - \theta)u + O(\partial_t^{-2} u) \\ &= -\partial_t u + (kA - Ak - \theta)u + O(\partial_t^{-2} u). \end{aligned} \tag{4.28}$$

Therefore, we choose the operator A such that it diagonalizes the equation (4.28). If we write the pseudo-differential operator θ in terms of its kernel

$$\begin{aligned} \theta(u)(x,k,t) &= \int_{\mathbb{R}} \mathrm{d}k'\ D(x,t,k-k')u(x,k',t), \\ D(x,t,r) &= \int_{\mathbb{R}} \mathrm{d}\eta\ \frac{i}{\hbar}\delta V\left(x, \frac{h}{2}\eta\right) e^{i\eta r}, \end{aligned} \tag{4.29}$$

then $(k-k')a(x,t,k,k') - D(x,t,k-k') = 0$ has to hold for $kk' < 0$. So, we set

$$a(x,t,k,k') = \begin{cases} \frac{D(x,t,k-k')}{k-k'} & \text{for} \quad kk' < 0, \\ 0 & \text{otherwise,} \end{cases} \tag{4.30}$$

and the absorbing boundary condition reads

$$u = w - \partial_t^{-1} \int_{\mathbb{R}} \mathrm{d}k' \; a(x,t,k,k')k'w(x,k',t) = 0 \quad \text{for} \quad x = 0, \quad k > 0. \tag{4.31}$$

Differentiating (4.31) once with respect to time gives an implementable boundary condition. The above formal manipulations can be made precise to make the whole approach more plausible. Given a solution w of the Wigner equation (4.21), we define u by

$$\partial_t u = \partial_t w - \int_{\mathbb{R}} \mathrm{d}k' \; a(x,t,k,k')k'w(x,k',t), \quad u(x,k) = w(x,k). \tag{4.32}$$

A direct calculation gives the residuals for the inverse transformations

$$R = w - u, \quad S = \partial_t(w-u) - \int_{\mathbb{R}} \mathrm{d}k' \; a(x,t,k,k')k'u(x,k',t) \tag{4.33}$$

$$\partial_t R = A(kw), \quad S = A(kR).$$

Inserting the new variable u into the transport operator and differentiation with respect to time gives

$$\begin{aligned} &\partial_t\Big(\partial_t u + k\partial_x u\Big) \\ &= \partial_t\Big(\partial_t + k\partial_x\Big)w - \Big(\partial_t + k\partial_x\Big)A(kw) \\ &= -\partial_t\theta(w) - \Big(\partial_t + k\partial_x\Big)A(kw) \\ &= -\theta(\partial_t w) - A(k\partial_t w) - kA(k\partial_x w) - \Big(\theta_t w + A_t(kw) - kA_x(kw)\Big) \\ &= -\theta(\partial_t w) - A(k\partial_t w) - kA(\partial_t w + \theta(w)) - \Big(\theta_t w + A_t(kw) - kA_x(kw)\Big) \\ &= -\Gamma(\partial_t w) - L(w), \end{aligned} \tag{4.34}$$

where the operators θ_t, A_t and A_x are the ones obtained from differentiating the kernels with respect to x and t, the block diagonal operator Γ is given by $\Gamma(f) = \theta(f) + A(kf) - kA(f)$ and L is given by $L(f) = \theta_t(f) + A_t(kf) - kA_x(f) - kA(\theta(f))$. Setting $w = u + R$ and $\partial_t w = \partial_t u + A(ku) + S$, and integrating with respect to time gives

$$\begin{aligned} \partial_t u + k\partial_x u + \Gamma(u) &= H, \\ \partial_t H &= \Gamma_t(u) - \Gamma(A(ku)) - \Gamma(A(kR)) - L(u) - L(R), \\ \partial_t R &= A(k(u+R)). \end{aligned} \tag{4.35}$$

The above system is decoupled up to the lower order term H. Imposing the boundary condition $u(x=0,k>0,t)=0$ and inserting a plane wave of the form

$$\begin{aligned} u(x,k,t) &= g_u(k,\omega,\xi)\exp\Big(i\xi(x-\omega t)\Big), \\ H(x,k,t) &= g_H(k,\omega,\xi)\exp\Big(i\xi(x-\omega t)\Big), \\ R(x,k,t) &= g_R(k,\omega,\xi)\exp\Big(i\xi(x-\omega t)\Big) \end{aligned} \tag{4.36}$$

immediately gives that the waves travelling to the right (for $k>0$) have amplitudes of order $O(\omega^{-2})$, that is, $g_u(k,\omega,\xi)=O(\omega^{-2})$ for $k>0$ holds. So, the second-order absorbing boundary condition is of the form

$$\partial_t u(0,k>0,t)=\Big(\partial_t w - A(kw)\Big)(0,k>0,t)=0. \tag{4.37}$$

In the case of an insulating surface, perfect reflection is imposed instead. So, in this case, the boundary condition $u(0,k,t)=u(0,-k,t)$, or

$$\Big(\partial_t w - A(kw)\Big)(0,k,t)=\Big(\partial_t w - A(kw)\Big)(0,-k,t), \quad \text{for} \quad k>0, \tag{4.38}$$

holds. Finally, in the case of a contact, we will impose a boundary condition modelling the injection of electrons according to a certain distribution. The corresponding boundary condition is then given such that nothing but the injected part of the distribution is reflected. Therefore, if we denote the injection distribution by $f(k)$, the absorbing boundary condition in (4.31) acts on $w-\rho(t)f$, giving

$$\begin{aligned} &\partial_t w - \int_{\mathbb{R}} \mathrm{d}k'\ a(x,t,k,k')k'w(x,k',t) = \\ &\qquad f(k)\partial_t\rho(t) - \rho(t)\int_{\mathbb{R}} \mathrm{d}k'\ a(x,t,k,k')k'f(k') \quad \text{for} \quad x=0, \quad k>0, \end{aligned} \tag{4.39}$$

where the function $\rho(t)$ is chosen such that the total charge in the device is conserved.

4.2. Quantum hydrodynamic models

The calculation of quantum transport phenomena via the quantum Boltzmann equation becomes prohibitively expensive in more than two dimensions. However, certain essential effects, such as non-monotone voltage current characteristics or negative differential resistance (Gardner 1993*a*), which are characteristic of the behaviour of quantum devices, can be simulated using much simpler macroscopic models. Like their classical counterpart, these model equations, the so-called quantum hydrodynamic equations (Gardner 1994), are derived from a moment expansion of the underlying kinetic equation. So, in the classical limit for $\hbar\to 0$, they reduce to the hydrodynamic model equations treated previously. Denoting the momentum

$\hbar k$ by p and the corresponding moments by

$$\langle 1\rangle = \int_{\mathbb{R}^3} \mathrm{d}k\ w, \quad \langle p_j\rangle = \int_{\mathbb{R}^3} \mathrm{d}k\ \hbar k_j w,$$

$$\langle p_j p_l\rangle = \int_{\mathbb{R}^3} \mathrm{d}k\ \hbar^2 k_j k_l w, \quad \left\langle |p|^2\right\rangle = \int_{\mathbb{R}^3} \mathrm{d}k\ \hbar^2 |k|^2 w,$$

$$\left\langle p_j |p|^2\right\rangle = \int_{\mathbb{R}^3} \mathrm{d}k\ \hbar^3 k_j |k|^2 w, \tag{4.40}$$

and taking the first three moments of the quantum Boltzmann equation (2.38) gives

$$\partial_t \langle 1\rangle + \frac{1}{m}\partial_{x_l}\langle p_l\rangle = 0 \tag{4.41a}$$

$$\partial_t \langle p_j\rangle + \frac{1}{m}\partial_{x_l}\langle p_l p_j\rangle - e\partial_{x_j} V \langle 1\rangle = \langle p_j Q\rangle \tag{4.41b}$$

$$\partial_t \left\langle |p|^2\right\rangle + \frac{1}{m}\partial_{x_l}\left\langle p_l |p|^2\right\rangle - 2e\partial_{x_l} V \langle p_l\rangle = \left\langle |p|^2 Q\right\rangle. \tag{4.41c}$$

(In (4.41) the summation convention is used again.) The system has to be closed again by expressing the pseudo-expectations $\langle p_l p_j\rangle, \langle p_l |p|^2\rangle, \langle p_j Q\rangle$ and $\langle |p|^2 Q\rangle$ in terms of the primary variables $\langle 1\rangle, \langle p_j\rangle$ and $\langle |p|^2\rangle$. If the Fokker–Planck term (2.40) is used as a collision operator, the moments on the right-hand side of (4.41) become

$$\langle p_j Q\rangle = -\frac{1}{\tau}\langle p_j\rangle, \quad \left\langle |p|^2 Q\right\rangle = \frac{2}{\tau}\left(3mT_0\langle 1\rangle - \left\langle |p|^2\right\rangle\right). \tag{4.42}$$

As in the classical case, closure is achieved by assuming that the Wigner function w is close to a wave vector displaced equilibrium density. Note that the first three moments of the quantum Boltzmann equation are the same as in the classical case. Therefore, quantum effects will enter solely through the closure conditions. If we assume the form of a wave vector displaced equilibrium density, so that

$$w(x,k,t) = w_e\left(x, k - \frac{m}{\hbar}u(x,t)\right) \tag{4.43}$$

holds with some group velocity vector u, we obtain

$$\langle 1\rangle = n, \quad \langle p_j\rangle = mnu_j, \quad \langle p_j p_l\rangle = m^2 n u_j u_l - mP_{jl}, \tag{4.44}$$
$$\left\langle |p|^2\right\rangle = m^2 n|u|^2 + mP =: 2mW, \quad \left\langle p_j |p|^2\right\rangle = 2m^2\left(u_j W - P_{jl}u_l\right),$$

where the P_{jl} and P denote the second moments of the equilibrium density, that is,

$$P_{jl} = -\frac{\hbar}{m}\int_{\mathbb{R}^3} \mathrm{d}k\ k_j k_l w_e, \quad P = -\frac{1}{2}\sum_j P_{jj} \tag{4.45}$$

holds. (It can always be assumed that the equilibrium density is symmetric, which implies that the odd order moments of w_e vanish.) Following Gardner (1994), the approximate equilibrium density is taken as

$$w_e(x,k) = A(x,t) f_c \left[1 + \hbar^2 \left(-\frac{1}{8mT^2}\Delta_x V + \frac{1}{24mT^3}|\nabla_x V|^2 + \frac{p_k p_l}{24m^2T^3}\partial^2_{x_k x_l} V \right) + O(\hbar^4)\right], \tag{4.46}$$

where $p = \hbar k$ holds and f_c denotes the classical equilibrium density

$$f_c = A(x,t)\exp\left(-\frac{|p|^2}{2mT} + \frac{V}{T}\right). \tag{4.47}$$

The form (4.46) is derived from an $O(\hbar^4)$ approximation of the thermal equilibrium density first given by Wigner (1932). With this form of the equilibrium density the moments P_{jl} and P in (4.44) become

$$P_{jl} = -nT\delta_{jl} - \frac{\hbar^2 n}{12m}\partial^2_{x_j x_l}\log(n) + O(\hbar^4), \quad P = \frac{3}{2}nT + \frac{\hbar^2 n}{24m}\Delta_x \log(n) \tag{4.48}$$

and the quantum hydrodynamic equations become

$$\partial_t n + \partial_{x_l}\langle n u_l\rangle = 0 \tag{4.49a}$$

$$\partial_t \langle mnu_j\rangle + \partial_{x_l}\left(mnu_l u_j - P_{jl}\right) - e\partial_{x_j} V n = -\frac{1}{\tau}mnu_j \tag{4.49b}$$

$$\partial_t\left(\frac{1}{2}mn|u|^2 + P\right) + \partial_{x_l}\left(u_l\left(\frac{1}{2}mn|u|^2 + P\right) - P_{lj}u_j + q_l\right) \tag{4.49c}$$

$$-e\,n\,\partial_{x_l} V u_l = \frac{2}{\tau}\left(3T_0 n - m|u|^2\right).$$

The structure of the quantum hydrodynamic equations is considerably more complex than that of the classical hydrodynamic model. Because of the presence of the term $\frac{\hbar^2 n}{12m}\partial^2_{x_j x_l} n$ in the correction to the stress tensor P_{jk}, the quantum hydrodynamic equations show the same dispersive behaviour as the underlying Wigner or Schrödinger equation. More precisely, an analysis of the linearized problem shows that the corresponding matrix has two hyperbolic (pure imaginary) eigenvalues, two dispersive modes (real eigenvalues which are proportional to the frequency) of order $\hbar^2$ and one parabolic eigenvalue, due to the presence of the heat conduction term $\operatorname{div}_x(\kappa\nabla_x T)$ (Gardner 1993*a*). At present, there are essentially two approaches to discretizing the quantum hydrodynamic system (4.49). The first treats the quantum hydrodynamic equations as a perturbation of the classical hydrodynamic system and uses a discretization appropriate for hyperbolic conservation laws. In this approach, the system is written as

$$\partial_t Z_j + \partial_{x_l} F_{lj}(Z) = R_j(Z), \quad Z = (n, mnu, W_c^T), \tag{4.50}$$

where W_c denotes the classical energy term $W_c = (1/2)(3nT + mn|u|^2)$, and the flux F and the right-hand side R are given by

$$F(Z) = \begin{pmatrix} nu \\ mnuu^T + nT\mathbf{I} \\ (W_c + nT)u \end{pmatrix},$$

$$R(Z) = \begin{pmatrix} 0 \\ n\nabla_x V - \frac{mnu}{\tau} - \frac{n}{3}\nabla_x Q \\ nu^T\nabla_x V - (W_c - \frac{3}{2}nT)/\tau + \mathrm{div}_x(\kappa\nabla_x T) + Q_w \end{pmatrix}, \quad (4.51a)$$

$$Q = \frac{\hbar^2}{2m}\frac{1}{\sqrt{n}}\Delta_x\sqrt{n},$$

$$Q_w = \frac{\hbar^2}{24m\tau}\Delta_x\Big(\log(n)\Big) - \frac{\hbar^2}{24m}\mathrm{div}_x\Big(n\Delta_x u\Big) - \frac{n}{3}u^T\nabla_x Q. \quad (4.51b)$$

The term Q in (4.51b) is referred to as the Bohm potential. Writing the quantum hydrodynamic system in the form (4.51), numerical methods suitable for hyperbolic conservation laws are used. In Chen, Cockburn, Gardner and Jerome (1995), a discontinuous Galerkin method is used in the spatial direction to simulate hysteresis effects in resonant tunnelling diodes. The time variable is discretized by a second-order explicit Runge–Kutta method, where each of the intermediate stages are projected orthogonally onto the manifold given by the Poisson equation.

A different approach to the discretization of the quantum hydrodynamic system is used in Gardner (1993*a*) for one-dimensional steady state simulations. Here, as in the classical case, the system is written in a form suitable for upwinding methods as

$$\partial_x(uG_j) + H_j + S_j = 0, \quad j = 0, \ldots, 3, \quad (4.52)$$

with the G_j, h_j and S_j given by

$$G_0 = n, \qquad G_1 = mnu,$$
$$G_2 = \frac{5}{2}nT + \frac{1}{2}mnu^2 - \frac{\hbar^2 n}{m}\partial_x^2\log(n) - nV, \qquad G_3 = 0 \quad (4.53a)$$

$$H_0 = 0, \qquad H_1 = \partial_x(nT) - \partial_x\left(\frac{\hbar^2 n}{12m}\partial_x^2\log(n)\right) - n\partial_x V,$$
$$H_2 = -\partial_x(\kappa\partial_x T), \qquad H_3 = e\partial_x^2 V \quad (4.53b)$$

$$S_0 = 0, \qquad S_1 = \frac{mnu}{\tau},$$
$$S_2 = \frac{3}{2}nT + \frac{1}{2}mnu^2 - \frac{\hbar^2 n}{24m}\partial_x^2\log(n) - \frac{3}{2}nT_0,$$
$$S_3 = e^2(N_D - N_A - n). \quad (4.53c)$$

Here the Poisson equation has already been included in the system. In this form the one-dimensional quantum hydrodynamic equations are discretized by conservative upwinding similar to the classical case. Thus, the term $\partial_x(uG_j)$ is discretized as

$$\partial_x(uG_j) \approx \frac{1}{\mu_x \Delta x} \delta_x \left((\mu_x u)(\mu_x G_j) - \frac{1}{2} |\mu_x u| (\delta_x G_j) \right), \tag{4.54}$$

where the averaging operator μ_x and the difference operator δ_x are defined as in (3.40). Notice, that the philosophy behind the upwind discretization differs from the one presented above since the quantum correction term $(\hbar^2 n/m)\partial_x^2 \log(n)$ is included in the transport term G. This is only possible in the one-dimensional case and results in the dispersive modes of the quantum hydrodynamic system being heavily damped out through the artificial diffusion produced by upwinding method. However, this one-dimensional scheme has proven nevertheless to be quite successful in the simulation of quantum mechanical phenomena, such as negative differential resistance in actual devices.

REFERENCES

A. Arnold and C. Ringhofer (1995*a*), 'An operator splitting method for the Wigner–Poisson problem', *SIAM J. Numer. Anal.* **32**, 1895–1921.

A. Arnold and C. Ringhofer (1995*b*), 'Operator splitting methods applied to spectral discretizations of quantum transport equations', *SIAM J. Numer. Anal.* **32**, 1876–1894.

A. Arnold, P. Degond, P. Markowich and H. Steinrück (1989), 'The Wigner–Poisson equation in a crystal', *Appl. Math. Lett.* **2**, 187–191.

N. Ashcroft and M. Mermin (1976), *Solid State Physics*, Holt-Saunders, New York.

G. Baccarani and M. Wordeman (1985), 'An investigation of steady state velocity overshoot effects in Si and GaAs devices', *Solid State Electr.* **28**, 407–416.

Z. Chen, B. Cockburn, C. Gardner and J. Jerome (1995), 'Quantum hydrodynamic simulation of hysteresis in the resonant tunneling diode', *J. Comput. Phys.* **117**, 274–280.

S. Cordier (1994*a*), 'Hyperbolicity of Grad's extension of hydrodynamic models for ionospheric plasmas I: the single species case', *Math. Mod. Meth. Appl. Sci.* **4**, 625–645.

S. Cordier (1994*b*), 'Hyperbolicity of Grad's extension of hydrodynamic models for ionospheric plasmas II: the two species case', *Math. Mod. Meth. Appl. Sci.* **4**, 647–667.

B. Engquist and A. Majda (1977), 'Absorbing boundary conditions for the numerical simulation of waves', *Math. Comput.* **31**, 629–651.

D. Ferry and H. Grubin (1995), 'Modelling of quantum transport in semiconductor devices', *Solid State Phys.* **49**, 283–448.

C. Gardner (1991*a*), 'Numerical simulation of a steady state electron shock wave in a submicrometer semiconductor device', *IEEE Trans. Electr. Dev.* **38**, 392–398.

C. Gardner (1991*b*), Shock waves in the hydrodynamic model for semiconductor devices, in *IMA Volumes in Mathematics and its Applications*, Vol. 59, pp. 123–134.

C. Gardner (1992), Upwind simulation of a steady state electron shock wave in a semiconductor device, in *Viscous Profiles and Numerical Methods for Shock Waves* (M. Shearer, ed.), pp. 21–30.

C. Gardner (1993*a*), The classical and the quantum hydrodynamic models, in *Proc. Int. Workshop on Computational Electronics, Leeds 1993* (J. Snowden, ed.), pp. 25–36.

C. Gardner (1993*b*), 'Hydrodynamic and Monte Carlo simulations of an electron shock wave in a $1\mu m$ $n^+ - n - n^-$ diode', *IEEE Trans. Electr. Dev.* **40**, 455–457.

C. Gardner (1994), 'The quantum hydrodynamic model for semiconductor devices', *SIAM J. Appl. Math.* **54**, 409–427.

C. Gardner, J. Jerome and D. Rose (1989), 'Numerical methods for the hydrodynamic device model', *IEEE Trans. CAD* **8**, 501–507.

N. Goldsman, L. Henrickson and J. Frey (1991), 'A physics based analytical–numerical solution to the Boltzmann equation for use in semiconductor device simulation', *Solid State Electr.* **34**, 389.

N. Goldsman, J. Wu and J. Frey (1990), 'Efficient calculation of ionization coefficients in silicon from the energy distribution function', *J. Appl. Phys.* **68**, 1075.

H. Grad (1949), 'On the kinetic theory of rarefied gases', *Comm. Pure Appl. Math.* **2**, 331–407.

H. Grad (1958), 'Principles of the kinetic theory of gases', *Handbooks Phys.* **12**, 205–294.

A. Kersch and W. Morokoff (1995), *Transport Simulation in Microelectronics*, Birkhäuser, Basel.

P. Lanzkron, C. Gardner and D. Rose (1991), 'A parallel block iterative method for the hydrodynamic device model', *IEEE Trans. CAD* **10**, 1187–1192.

P. Markowich and C. Ringhofer (1989), 'An analysis of the quantum Liouville equation', *ZAMM* **69**, 121–127.

P. Markowich, N. Mauser and F. Poupaud (1994), 'A Wigner function approach to semiclassical limits', *J. Math. Phys.* **35**, 1066–1094.

P. Markowich, C. Ringhofer and C. Schmeiser (1990), *Semiconductor Equations*, Springer.

F. Poupaud (1991), 'Diffusion approximation of the linear Boltzmann equation: analysis of boundary layers', *Asympt. Anal.* **4**, 293–317.

F. Poupaud and C. Ringhofer (1995), 'Quantum hydrodynamic models in semiconductor crystals', *Appl. Math. Lett.* **8**, 55–59.

C. Ringhofer (1990), 'A spectral method for the numerical solution of quantum tunneling phenomena', *SIAM J. Numer. Anal.* **27**, 32–50.

C. Ringhofer (1992), 'On the convergence of spectral methods for the Wigner–Poisson problem', *Math. Mod. Meth. Appl. Sci.* **2**, 91–111.

C. Ringhofer (1994*a*), Galerkin methods for kinetic equations in time variant coordinate systems, in *Proc. 'Mathematical Methods in Semiconductor Simulation'* (R. Natalini, ed.), pp. 32–49.

C. Ringhofer (1994*b*), A series expansion method for the Boltzmann transport equation using variable coordinate systems, in *Proc. Int. Wkshp. on Comp. Electr.* (S. Goodnick, ed.), Portland, pp. 128–132.

C. Ringhofer (1997), 'An adaptive Galerkin procedure for the Boltzmann transport equation', *Math. Mod. Meth. Appl. Sci.* To appear.

C. Ringhofer, D. Ferry and N. Kluksdahl (1989), 'Absorbing boundary condition for the simulation of quantum transport phenomena', *Transp. Theory and Stat. Phys.* **18**, 331–346.

C. Schmeiser and A. Zwirchmayr (1995), Galerkin methods for the semiconductor Boltzmann equation, in *Proc. ICIAM 95, Hamburg.*

C. Schmeiser and A. Zwirchmayr (1997), 'Convergence of moment methods for the semiconductor Boltzmann equation', *SIAM J. Numer. Anal.* To appear.

S. Selberherr (1981), *Analysis of Semiconductor Devices*, 2nd edn, Wiley, New York.

V. Tatarski (1983), 'The Wigner representation of quantum mechanics', *Soviet. Phys. Uspekhi* **26**, 311–372.

M. Taylor (1981), *Pseudodifferential Operators*, Princeton University Press, Princeton.

D. Ventura, A. Gnudi and G. Baccarani (1991), One dimensional simulation of a bipolar transistor by means of spherical harmonics expansions of the Boltzmann equation, in *Proc. SISDEP 91 Conference (Zürich)* (W. Fichtner, ed.), pp. 203–205.

D. Ventura, A. Gnudi, G. Baccarani and F. Odeh (1992), 'Multidimensional spherical harmonics expansions for the Boltzmann equation for transport in semiconductors', *Appl. Math. Lett.* **5**, 85–90.

E. Wigner (1932), 'On the quantum correction for thermodynamic equilibrium', *Phys. Rev.* **40**, 749–759.