Wigner function approach to nano device simulation

Hans Kosina
TU Vienna, Institute for Microelectronics, Gusshausstrasse 27–29, A-1040 Vienna, Austria
E-mail: kosina@iue.tuwien.ac.at

Abstract: Coherent transport in mesoscopic devices is well described by the Schrödinger equation supplemented by open boundary conditions. When electronic devices are operated at room temperature, however, a realistic transport model needs to include carrier scattering. In this work, the kinetic equation for the Wigner function is employed as a model for dissipative quantum transport. Carrier scattering is treated in an approximate manner through a Boltzmann collision operator. The development of Monte-Carlo algorithms for this quantum kinetic equation is complicated by the fact that, as opposed to the semi-classical case, the integral kernel is no longer positive. This so-called negative sign problem requires the introduction of new numerical techniques in order to obtain stable Monte-Carlo methods. A particular method for the solution of the stationary Wigner equation is presented. Applications to single barrier and double barrier structures are discussed.

Keywords: device simulation; quantum transport; Wigner function; Monte Carlo method; resonant tunnelling diode.


Biographical notes: Hans Kosina received the ‘Diplomingenieur’ degree in Electrical Engineering, a PhD in Technical Sciences and the ‘venia docendi’ in Microelectronics from the Technical University of Vienna in 1987, 1992 and 1998, respectively. Currently, he is an Associate Professor at the Institute for Microelectronics at the same University. His current research interests include modelling of semiconductor devices, nanoelectronic devices, organic semi-conductors and optoelectronic devices, development of novel Monte Carlo algorithms for classical and quantum transport problems and computer aided engineering in ULSI-technology.

1 Introduction

Modelling of electronic transport in mesoscopic systems requires a theory that describes open, quantum-statistical systems driven far from thermodynamic equilibrium. Several formulations of quantum transport have been employed practically, such as those based on the density matrix, non-equilibrium Green’s functions and the Wigner function.

A quantum-mechanical phase-space distribution was introduced by Eugene Wigner in 1932. The purpose was the formulation of a quantum correction for the thermodynamic equilibrium of a many-body system by means of a quasiprobability function. In more recent times, the definition of the Wigner function has been generalised as a Fourier transform of a many-body Green’s function (Mahan, 1983).

The Wigner function is a real-valued but not necessarily positive definite quasidistribution and represents a quantum generalisation of Boltzmann’s N-particle distribution. The Wigner function formalism is attractive as it allows the expression of quantum dynamics in a phase-space formulation, directly comparable with the classical analogue.

A phase-space approach may appear more intuitive compared with the more abstract density matrix and Green’s function approaches. The method of quasidistributions has proved especially useful in providing reductions to classical physics and kinetic regimes under suitable conditions.

To discuss the physical interpretation of a quasidistribution, let us consider the simple case of a one-particle distribution. Starting with the classical case, the distribution \( f_{cl}(p, r, t) \) is proportional to the probability density of finding a particle of momentum \( p \) and position \( r \) in the phase-space volume \( dp \times dr \). This is a purely classical interpretation, directly conflicting with the uncertainty principle. The quantum mechanical quasidistribution \( f_w(p, r, t) \), however, is not positive definite and has to be interpreted as a joint density of \( p \) and \( r \) (Tatarskii, 1983). Only the marginal distributions are positive definite, that is, integrating \( f_w(p, r, t) \) over momentum space gives the probability density in \( r \)-space and vice versa.

An excellent review of quantum-mechanical phase-space distributions in scattering theory has been given by Carruthers and Zachariasen (1983). This work deals with potential scattering, the two-body problem and the...
1.1 History and state of the art review

Reports on finite-difference solutions of the one-particle Wigner equation for device applications are due to Ravaioli et al. (1985), Kluksdahl et al. (1987) and date back to the mid 1980s. Frensky (1986a,b, 1987) 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 (Frensky, 1989; Kluksdahl et al., 1989). Main and Haddad included a reduced Boltzmann scattering operator in transient Wigner function-based simulations (Mains and Haddad, 1988). Research on finite-difference solution methods for the Wigner equation culminated in 1990 when the review papers of Frensky (1990) and Buot and Jensen (1990) appeared.

The 1990s have seen further extensions and applications of the finite-difference Wigner function method. High-frequency operation of resonant tunnelling diodes has been studied by Jensen and Buot (1990, 1991), and the transient response by Gullapalli et al. (1994) and Biegel and Plummer (1997) and later by Woolard et al. (2002). A new finite-difference discretisation scheme has been proposed by Mains and Haddad (1994).

In 2002, implementations of Monte Carlo methods for solving the Wigner device equation were reported (Nedjalkov et al., 2002; Shifren and Ferry, 2002). Although 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, assuming a three-dimensional momentum-space (Kosina et al., 2003a,b; Shifren et al., 2003). Issues such as choosing proper up-winding schemes, restrictions on matrix size and momentum space resolution are largely relaxed or do not exist when using the Monte Carlo method. Construction of new Monte Carlo algorithms is complicated by the fact that the kernel of the integral equation to solve is not positive semidefinite. As a consequence, the commonly applied Markov chain Monte Carlo method shows a variance exponentially increasing with time, prohibiting its application to realistic structures or larger evolution times (Nedjalkov et al., 1996, 2002; Rossi et al., 1994). Because of this so-called negative sign problem, the concept of Wigner paths alone (Bordone et al., 2003; Pascoli et al., 1998) is not sufficient to construct a stable Monte Carlo algorithm. Instead, additional measures have to be introduced that prevent a runaway of the particle weights and hence of the variance (Kosina et al., 2003a,b; Shifren et al., 2003). Note that in Shifren et al. (2003), the statistical weights are termed affinities.

Large basic research efforts on the Monte Carlo modelling of electron–phonon interaction based on the Wigner formalism have been reported Rossi et al. (1994) Bordone et al. (2003) Bertoni et al. (1999), Bordone et al. (1999) and Jacoboni et al. (2001).

The effect of a spatially varying effective mass in Wigner device simulations has been demonstrated by Tsuchiya et al. (1991) and Shih et al. (1994). A non-parabolic version of the Wigner equation has been derived by Butler and Schlösser (1994). Multiband models have been reported by Miller and Neikirk (1991), Demaio et al. (2002) and Unlu et al. (2004).

A Wigner equation including a magnetic field has been solved by Wu and Wu (1992). The gauge-invariant formulation of the Wigner equation has been given by Levinson (1970) and a discussion can be found in various works (Carruthers and Zachariasen, 1983; Ferry and Goodnick, 2001; Holland and Kyprianidis, 1986; Levanda and Fleurov, 2001; Sonego, 1991). Two-time and frequency-dependent Wigner functions are considered by Mahan (1983), Levanda and Fleurov (2001), Hänsch (1991) and Brunetti et al. (2000).

In various types of field-effect transistors carrier transport takes place in inversion layers, quantum wells or thin semiconductor films. To model such devices, the effect of quantum confinement needs to be included in the transport model. An expansion of the three-dimensional wave function in the basis set of transverse wave functions is employed in the mode space approach (Venugopal et al., 2002). The problem is significantly simplified if coupling of the modes or subbands, can be neglected. In this case a set of decoupled Wigner equations, one for each subband, can be used (Croitoru et al., 2003, 2004). If the cross section of the semiconductor film changes rapidly, intermode coupling becomes important (Venugopal et al., 2004). Inclusion of intermode coupling in the Wigner function formalism for confined systems is an open issue.

Finally, we note that the Wigner function formalism is often used to derive reduced transport models, such as...
the quantum hydrodynamic model (Degond and Ringhofer, 2003; Gardner, 1994; Gardner and Ringhofer, 1996; Zhou and Ferry, 1992) or to find quantum corrections to classical models, such as the ensemble Monte Carlo method (Tsuchiya and Ravaïoli, 2001) or the spherical harmonics expansion method (Goldsmann et al., 2000; Han et al., 2000).

2 The Wigner function formalism

In the Schrödinger picture, a physical system is quantum-mechanically described by a state vector \(|\Psi(t)\rangle\) as function of time \(t\). Often, the precise quantum-mechanical state of a system is not known, but rather some statistical information about the probabilities for the system being in one of a set of states. Suppose that there is a set of orthonormal states \(|\Psi_1\rangle, |\Psi_2\rangle, \ldots\) and that the probabilities that the system is in one of these states are \(p_i\). Then, the expectation value of operator \(\hat{A}\) associated with the observable \(A\) is given by

\[
\langle A \rangle = \sum_i p_i \langle \Psi_i | \hat{A} | \Psi_i \rangle
\]

(1)

which is a quantum and statistical average. Introducing the density operator \(\hat{\rho}\) as

\[
\hat{\rho} = \sum_i p_i |\Psi_i\rangle \langle \Psi_i|
\]

(2)

the expectation value becomes

\[
\langle A \rangle = \text{Tr}(\hat{\rho}\hat{A}) = \text{Tr}(\hat{A}\hat{\rho})
\]

(3)

Equations (1) and (3) require the operator \(\hat{A}\) to be self-adjoint. Equation (3) can be easily verified by expressing the trace of some operator \(\hat{X}\) in the basis \(|\Psi_i\rangle\).

\[
\text{Tr}(\hat{X}) = \sum_i \langle \Psi_i | \hat{X} | \Psi_i \rangle
\]

(4)

The fact that the probabilities sum up to unity, \(\sum_i p_i = 1\), is expressed by the fact that the trace of the density operator is also unity, \(\text{Tr}(\hat{\rho}) = 1\). If the system is in a pure state \(|\Psi_i\rangle\) it holds \(p_i = 1\) and \(p_j = 0\ \forall j \neq i\) and the density operator is idempotent, \(\hat{\rho}^2 = \hat{\rho}\). Otherwise, the system is in a mixed state and \(\hat{\rho}\) does not obey the idem-potency condition. From the Schrödinger equation for the state vector and the definition of \(\hat{\rho}\), we immediately obtain the Liouville-von Neumann equation for the evolution of the density operator:

\[
i\hbar \frac{\partial \hat{\rho}}{\partial t} = [\hat{H}, \hat{\rho}]
\]

(5)

Introducing the one-particle approximation (Rammer, 1991) implies that the electron system is modelled as consisting of many, non-interacting electrons. In the next step, one chooses the coordinate representation, where the set of basis vectors is given by the electron position states \(|r\rangle\). The eigenstates of the system are then represented by the wavefunctions \(\Psi_i(r, t) = \langle r | \Psi_i(t) \rangle\) and the density operator by the density matrix \(\rho(r_1, r_2, t)\).

\[
\rho(r_1, r_2, t) = \langle r_1 | \hat{\rho}(t) | r_2 \rangle = \sum_i p_i \Psi_i(r_1, t) \Psi_i^*(r_2, t)
\]

(6)

The Liouville-von Neumann equation in coordinate representation is found as

\[
\frac{\partial \rho(r_1, r_2, t)}{\partial t} = (H_{r_1} - H_{r_2})\rho(r_1, r_2, t)
\]

(7)

2.1 The Wigner function

The Wigner function is obtained from the density matrix by means of the Wigner–Weyl transformation. This transformation consists of a change of independent coordinates to diagonal and cross-diagonal coordinates

\[
r = \frac{1}{2}(r_1 + r_2), \quad s = r_1 - r_2
\]

(8)

followed by a Fourier transformation with respect to \(s\) (Frensley, 1990). The variables \(r_1\) and \(r_2\) may be expressed in terms of the new ones.

\[
r_1 = r + \frac{s}{2}, \quad r_2 = r - \frac{s}{2}
\]

(9)

Then, the elementary definition of the Wigner distribution is given by the following transformation of the density matrix.

\[
f_w(k, r, t) = \int \rho\left(r + \frac{s}{2}, r - \frac{s}{2}, t\right) e^{-iks} ds
\]

(10)

The Wigner function (10) is real-valued, but not positive semidefinite. In terms of the wave functions, the definition (10) becomes

\[
f_w(k, r, t) = \sum_i p_i \int \Psi_i^*\left(r + \frac{s}{2}, t\right) e^{-iks} ds
\]

(11)

The normalisation of the Wigner function results from the normalisation of the wave functions.

\[
\frac{1}{(2\pi)^3} \int dr \int dk f_w(k, r, t) = 1
\]

(12)

Here, the \(k\)-integration can be performed first, giving \(\int e^{-iks} dk = (2\pi)^3 \delta(s)\). The normalisation (12) ensures that the quantity \(N f_w\), where \(N\) is the number of electrons in the system, will approach the classical distribution function \(f_{cl}\) in the classical limit (Jacobi et al., 2001).

Sometimes it is convenient to use the inverse Fourier transform of (10).

\[
\rho\left(r + \frac{s}{2}, r - \frac{s}{2}, t\right) = \frac{1}{(2\pi)^3} \int f_w(k, r, t) e^{iks} dk
\]

(13)

Changing variables gives a transformation that inverts the Wigner–Weyl transformation.

\[
\rho(r_1, r_2, t) = \frac{1}{(2\pi)^3} \int f_w\left(k, \frac{r_1 + r_2}{2}, t\right) e^{i(kr_1 - kr_2)} dk
\]

(14)

An important feature of the phase-space approach is the possibility of expressing quantum-mechanical expectation
values in the same way as it is done in classical statistical mechanics, employing integration over the phase-space. The expectation values of operators of the form $A(\mathbf{r})$ and $B(\mathbf{k})$, where $\mathbf{k} = \mathbf{p}/\hbar$, are given as follows.

$$\langle A(\mathbf{r}) \rangle = \frac{1}{(2\pi)^2} \int f_w(\mathbf{k}, \mathbf{r}, t) A(\mathbf{r}) \, d\mathbf{k} \, d\mathbf{r}$$

$$= \sum_i p_i \int A(\mathbf{r}) |\Psi_i(\mathbf{r}, t)|^2 \, d\mathbf{r}$$

(15)

$$\langle B(\mathbf{k}) \rangle = \frac{1}{(2\pi)^3} \int f_w(\mathbf{k}, \mathbf{r}, t) B(\mathbf{k}) \, d\mathbf{k} \, d\mathbf{r}$$

$$= \sum_i p_i \int B(\mathbf{k}) |\Phi_i(\mathbf{k}, t)|^2 \, d\mathbf{k}$$

(16)

If the classical observable $C(\mathbf{k}, \mathbf{r})$ is a function of both momentum and position, the definition of a corresponding Hermitian operator $\hat{C}$ is not unique. In this case, the Weyl quantisation can be applied. Thereby, the function $C$ is expressed through its Fourier transform $c$.

$$C(\mathbf{k}, \mathbf{r}) = \int c(\mathbf{a}, \mathbf{b}) e^{(i\mathbf{k} \cdot \mathbf{r} + i\mathbf{a} \cdot \mathbf{b})} \, d\mathbf{a} \, d\mathbf{b}$$

(17)

The operator $\hat{C}$ is defined by the following rule of correspondence.

$$\hat{C} = \int c(\mathbf{a}, \mathbf{b}) e^{i(\mathbf{k} \cdot \mathbf{a} + \mathbf{b})} \, d\mathbf{a} \, d\mathbf{b}$$

(18)

Then, the expectation value of $\hat{C}$ is given by the phase-space integral.

$$\text{Tr}(\hat{C} \hat{\rho}) = \int C(\mathbf{k}, \mathbf{r}) f_w(\mathbf{k}, \mathbf{r}, t) \, d\mathbf{k} \, d\mathbf{r}$$

(19)

To proceed with (18), one may employ the Baker-Campbell-Hausdorff formula,

$$e^{\hat{A} + \hat{B}} = e^{\hat{A}} e^{\hat{B}} e^{-[\hat{A}, \hat{B}]/2}$$

(20)

which is generally valid when $[\hat{A}, [\hat{A}, \hat{B}]] = [\hat{B}, [\hat{A}, \hat{B}]] = 0$, or in particular when $[\hat{A}, \hat{B}]$ is a $c$-number.

2.2 Marginal distributions

The Wigner function (10) can assume negative values. Only the marginal distributions of $f_w(\mathbf{k}, \mathbf{r}, t)$ are positive semi-definite and have the meaning of probability distributions in real space and momentum space, respectively.

$$n(\mathbf{r}) = \frac{1}{(2\pi)^3} \int f_w(\mathbf{k}, \mathbf{r}, t) \, d\mathbf{k}$$

$$= \sum_i p_i |\Psi_i(\mathbf{r}, t)|^2$$

(21)

$$p(\mathbf{k}) = \frac{1}{(2\pi)^3} \int f_w(\mathbf{k}, \mathbf{r}, t) \, d\mathbf{r}$$

$$= \sum_i p_i |\Phi_i(\mathbf{k}, t)|^2$$

(22)

Here, $\Phi_i(\mathbf{k}, t)$ denotes the momentum representation of the state vector $|\Psi_i\rangle$. The integration in (22) can easily be carried out after changing variables, using (8).

$$\int \, d\mathbf{r} \int d\mathbf{s} \psi_i(\mathbf{r} + s \frac{2}{\sqrt{\pi}} \mathbf{1}, t) \psi_i^*(\mathbf{r} - s \frac{2}{\sqrt{\pi}} \mathbf{1}, t) e^{-i\mathbf{s} \cdot \mathbf{k} \mathbf{1}}$$

$$= \int \, d\mathbf{r} \int d\mathbf{r}_1 \psi_i(\mathbf{r}_1, t) \psi_i^*(\mathbf{r}_2, t) e^{-i\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)}$$

(23)

The marginal distributions (21) and (22) can also be expressed as the diagonal elements of the density matrix.

$$\frac{1}{(2\pi)^3} \int f_w(\mathbf{k}, \mathbf{r}, t) \, d\mathbf{k} = |\mathbf{r} \rangle \langle \mathbf{r} | = \rho(\mathbf{r}, \mathbf{r})$$

(24)

$$\frac{1}{(2\pi)^3} \int f_w(\mathbf{r}, \mathbf{r}, t) \, d\mathbf{r} = |\mathbf{k} \rangle \langle \mathbf{k} | = \sigma(\mathbf{k}, \mathbf{k})$$

(25)

Here, $|\mathbf{k}\rangle$ denotes the electron momentum eigenstate with eigenvalue $\hbar \mathbf{k}$ and $\sigma$ the density matrix in momentum representation. Note that the latter can be used for a dual definition of the Wigner function (Rossi et al., 1994; Fannjiang et al., 2002).

$$f_w(\mathbf{k}, \mathbf{r}, t) = \int \sigma(\mathbf{k} + \frac{1}{2}, \mathbf{r} - \frac{1}{2}) e^{i\mathbf{k} \cdot \mathbf{r}} \, d\mathbf{k}$$

(26)

This definition follows, for example, from (11), when the $\Psi_i$ are replaced by

$$\psi_i(\mathbf{r}, t) = (2\pi)^{-3/2} \int \phi_i(\mathbf{k}', t) e^{i\mathbf{k}' \cdot \mathbf{r}} \, d\mathbf{k}'$$

(27)

Other marginal distributions than the elementary ones, (21) and (22), have to be constructed with care. Only Hermitian operators give real marginal distributions. For the current density, this operator would be $(\mathbf{k} \hat{\rho} + \hat{\rho} \mathbf{k})/2$. Expressing $\hat{\rho}$ in terms of the wave functions, we get the elementary current definition from wave mechanics.

$$\mathbf{j}(\mathbf{r}) = \frac{\hbar}{2m^*} \mathbf{1} \cdot \hat{\rho} + \hat{\rho} \mathbf{1} \cdot \mathbf{k}$$

$$= \frac{\hbar}{2m^*} \sum_i p_i \left[ \langle \mathbf{r} | \mathbf{k} \rangle |\Psi_i\rangle \langle \Psi_i| + \langle \Psi_i| \langle \mathbf{k} | \mathbf{r} \rangle \langle \Psi_i| \mathbf{k} angle \right]$$

(28)

Choosing the momentum representation of $\hat{\rho}$,

$$\mathbf{j}(\mathbf{r}) = \frac{\hbar}{2m^*} \int d\mathbf{k}_1 \int d\mathbf{k}_2 \left[ \langle \mathbf{r} | \mathbf{k}_1 \rangle \langle \mathbf{k}_1 | \hat{\rho} | \mathbf{k}_2 \rangle \langle \mathbf{k}_2 | \mathbf{k} \rangle + \langle \mathbf{k} | \mathbf{k}_1 \rangle \langle \mathbf{k}_1 | \hat{\rho} | \mathbf{k}_2 \rangle \langle \mathbf{k}_2 | \mathbf{k}_1 \rangle \right]$$

we get the current density expressed in terms of the Wigner function.

$$\mathbf{j}(\mathbf{r}) = \frac{\hbar}{2m^*} \int d\mathbf{k}_1 \int d\mathbf{k}_2 \sigma(\mathbf{k}_1, \mathbf{k}_2) (\mathbf{k}_1 + \mathbf{k}_2) e^{i\mathbf{k}_1 \cdot \mathbf{r}}$$

(29)

Here, the Wigner function has been introduced using (26). The current density is given by the first-order moment of the Wigner function, in full analogy with the classical phase-space definition.
For the definition of the energy density we discuss several options. Starting from the trace operation for the statistical average, one would consider the symmetrised operator \((\hat{k}^2\hat{\rho} + \hat{\rho}\hat{k}^2)/2\) and derive the marginal distribution.

\[
w_1(r) = \frac{\hbar^2}{4m^*} \left( (r) \hat{k}^2\hat{\rho} + (r) \hat{\rho}\hat{k}^2 \right) = -\frac{\hbar^2}{4m^*} \sum_i p_i \left[ \Psi_i^*(r) \nabla^2 \Psi_i(r) + \Psi_i(r) \nabla^2 \Psi_i^*(r) \right]
\]

The last expression in (30) is obtained with the help of the distribution can be shown to be positive semi-definite.

\[
\sum_i p_i |E_i - V(r)|^2 \Psi_i^*(r) \Psi_i(r) \]

The last expression in (30) is obtained with the help of the stationary Schrödinger equation. Apparently, \(w_1\) describes the kinetic energy density, as the potential energy term \(V(r)\) is subtracted from the total energy term. This energy density can become negative in tunnelling regions, where for one or more states \(E_i < V(r)\) holds. In a derivation similar to (29), one finds the Wigner representation of \(w_1\).

\[
w_1(r) = \frac{\hbar^2}{4m^*} \int \frac{dk}{2\pi} \left( |k|^2 - \frac{1}{4} \nabla^2 r \right) f_w(k, r, t) \text{ } dk
\]

To ensure positiveness of the energy density, in (Kosik, 2004) the Hermitian operator \(\hat{k}^2\hat{\rho}\) is considered. Its marginal distribution can be shown to be positive semi-definite.

\[
w_2(r) = \frac{\hbar^2}{2m^*} \sum_i p_i \left( |\Psi_i(r)|^2 \right) \geq 0
\]

The Wigner representation of \(w_2\) is obtained as

\[
w_2(r) = \frac{\hbar^2}{4m^*} \int \frac{dk}{2\pi} \left( |k|^2 - \frac{1}{4} \nabla^2 r \right) f_w(k, r, t) \text{ } dk
\]

Conditions for obtaining non-negative marginal distributions are theoretically discussed by Wlodarz (1999). The Weyl correspondence (18) gives the definition of the energy density as the second-order moment of the Wigner function.

\[
w_3(r) = \frac{1}{(2\pi)^3} \int \frac{dk}{2m^*} |k|^2 f_w(k, r, t) \text{ } dk
\]

It can be seen that (35) is just the arithmetic mean of (31) and (34), \(w_3 = (w_1 + w_2)/2\). Therefore, (35) represents the marginal distribution of the symmetrised operator \((\hat{k}^2\hat{\rho} + 2\hat{k}\hat{\rho}\hat{k} + \hat{\rho}\hat{k}^2)/4\).

All three definitions of the energy density give the same statistical average \(\langle \epsilon \rangle = \text{Tr}(\epsilon \hat{k})\). The differences among the definitions are in the \(\nabla^2\) term, which vanishes after the \(r\)-integration. However, only the density \(w_1\) seems to have a clear physical interpretation as the kinetic energy density.

2.3 The Wigner equation

In this section, we consider a system consisting of one electron interacting with a potential distribution \(V_{\text{ext}}(r)\). This potential is assumed to be a superposition of some potential \(V(r)\) and a uniform electric field: \(V_{\text{ext}}(r) = V(r) - h\mathbf{F} \cdot \mathbf{r}\), with \(h\mathbf{F} = -e\mathbf{E}\). Although the existence of a field term is not physically motivated at this point, it is introduced here to demonstrate its treatment in the Wigner function formalism. The potential \(V(r)\) comprises the electrostatic potential and the band-edge profile of the semi-conductor. A uniform effective mass \(m^*\) is assumed. In the usual coordinate representation, the Hamiltonian of the system is then given by

\[
H = H_0 + V(r) - h\mathbf{F} \cdot \mathbf{r}
\]

with

\[
H_0 = -\frac{\hbar^2}{2m^*} \nabla^2 r
\]

Electron–phonon interaction is neglected here. The evolution equation for the Wigner function is found by taking the time derivative of the defining Equation (10) and substituting the Liouville-von Neumann equation (7) on the right-hand side.

\[
\frac{\partial}{\partial t} f_w(k, r, t) = 1 \text{ } \int \left( H_{\text{rel}} - H_{\text{tot}} \right) \rho \left( r + \frac{s}{2}, r - \frac{s}{2}, t \right) e^{-iks} ds
\]

In the following, the three parts of the Hamiltonian (36) will be separately transformed. Unlike in Section 2.2, where calculations were done in momentum representation, we choose below the configuration representation to carry out the transformations (Bertoni et al., 1999).

The free-electron Hamiltonian is given by \(H_0\). To calculate the Wigner transform of \(H_0\), we have to transform the gradients first. Differentiating the density matrix with respect to the new variables \(r\) and \(s\)

\[
\nabla_r \rho \left( r + \frac{s}{2}, r - \frac{s}{2}, t \right) = \nabla_{r_1} \rho + \nabla_{r_2} \rho
\]

\[
\nabla_s \rho \left( r + \frac{s}{2}, r - \frac{s}{2}, t \right) = \frac{1}{2} \nabla_{r_1} \rho - \frac{1}{2} \nabla_{r_2} \rho
\]

gives the relations

\[
\nabla_{r_1} + \nabla_{r_2} = \nabla_r
\]

\[
\nabla_{r_1} - \nabla_{r_2} = 2\nabla_s
\]

\[
\nabla_{r_1}^2 - \nabla_{s}^2 = 2\nabla_{r} \nabla_{s}
\]

Now the free-electron term transforms to a diffusion term. For the sake of brevity, we write \(\rho_{r,s} = \rho(r + s/2, r - s/2, t)\) in the following.

\[
\frac{1}{i\hbar} \int -\frac{\hbar^2}{2m^*}(\nabla_{r_1}^2 - \nabla_{r_2}^2) \rho_{r,s} e^{-iks} ds = -\frac{\hbar}{im^*} \nabla_r \int (\nabla_s \rho_{r,s}) e^{-iks} ds
\]

\[
= -\frac{\hbar k}{m^*} \nabla_r \int \rho_{r,s} e^{-iks} ds
\]

\[
= -\frac{\hbar k}{m^*} \nabla_r f_w(k, r, t)
\]
Next, we transform the potential term.

$$\frac{1}{\hbar} \int \left[ V \left( \mathbf{r} + \frac{\mathbf{s}}{2} \right) - V \left( \mathbf{r} - \frac{\mathbf{s}}{2} \right) \right] \rho_{\mathbf{r}, \mathbf{s}} e^{-i \mathbf{k} \cdot \mathbf{s}} \, d\mathbf{s}$$

$$= \int V_{\nu}(\mathbf{k} - \mathbf{k}', \mathbf{r}) \, f_{\nu}(\mathbf{k}', \mathbf{r}, t) \, d\mathbf{k}'$$

This transformation is readily found by replacing \( \rho_{\mathbf{r}, \mathbf{s}} \) on the left-hand side by the inverse Fourier transformation (13). The remaining integral over \( \mathbf{s} \) is denoted by \( V_{\nu} \) and referred to as the Wigner potential.

$$V_{\nu}(\mathbf{q}, \mathbf{r}) = \frac{1}{(2\pi)^3 \hbar} \int \left[ V \left( \mathbf{r} + \frac{\mathbf{s}}{2} \right) - V \left( \mathbf{r} - \frac{\mathbf{s}}{2} \right) \right] e^{-i \mathbf{q} \cdot \mathbf{s}} \, d\mathbf{s}$$

Using the simple relation \(-(\mathbf{Fr}_1 - \mathbf{Fr}_2) = -\mathbf{Fs} \), the constant-field term transforms as

$$\frac{1}{\hbar} \int \left( -\mathbf{hF}\rho_{\mathbf{r}, \mathbf{s}} e^{-i \mathbf{k} \cdot \mathbf{s}} \right) \, d\mathbf{s} = -\frac{1}{\hbar} \mathbf{FV}_{\nu}(\mathbf{k}, \mathbf{r}, t)$$

Collecting the above results gives the Wigner equation for the system Hamiltonian (36).

$$\left( \frac{\partial}{\partial t} + \frac{\hbar \mathbf{k}}{m} \nabla_{\mathbf{r}} + \mathbf{FV}_{\nu} \right) f_{\nu}(\mathbf{k}, \mathbf{r}, t)$$

$$= \int V_{\nu}(\mathbf{k} - \mathbf{k}', \mathbf{r}) \, f_{\nu}(\mathbf{k}', \mathbf{r}, t) \, d\mathbf{k}'$$

The terms are arranged so to form the classical Liouville operator on the left-hand side. The interaction of the electron with the potential distribution \( V(\mathbf{r}) \) is described by the potential operator on the right-hand side. As can be seen, the Wigner function in \( \mathbf{k} \) and \( \mathbf{r} \) depends in a non-local manner on the Wigner function in all other momentum points \( \mathbf{k}' \) and through \( V_{\nu} \) also on the potential at all other locations \( \mathbf{r} \pm s/2 \).

### 2.4 Electron–phonon interaction

The Wigner equation has frequently been solved using the finite-difference method (Biegel and Plummer, 1996; Frensley, 1990), assuming the phenomenological relaxation time approximation for dissipative transport. Recently developed Monte Carlo methods allowed phonon scattering to be included semi-classically in quantum device simulations (Kosina et al., 2003a,b; Shifren and Ferry, 2002). Use of a Boltzmann scattering operator acting on the Wigner distribution was originally suggested by Frensley (1990). A rigorous treatment of electron–phonon interaction in the Wigner formalism based on a one-electron, many-phonon Hamiltonian was reported by Rossi et al. (1994), Bertoni et al. (1999) and Bordone et al. (1999). However, for the purpose of numerical simulation of mesoscopic devices this formalism turned out to be rather complex. Introducing the weak scattering limit and assuming the phonon system to be in thermodynamic equilibrium, an equation for the reduced Wigner function is obtained (Brunetti et al., 1989; Nedjalkov et al., 2002; Rossi et al., 1994). Since this equation takes into account the finiteness of the interaction time, it is time reversible and includes effects such as collisional broadening, collisional retardation and the intracollisional field effect (Gurov et al., 2002; Nedjalkov et al., 2001). For a uniform electric field, the equation reduces to the Levinson equation (1970). Introducing the classical limit, the scattering operator of the equation for the reduced Wigner function reduces to the Boltzmann scattering operator. This limit implies that the system is considered on a timescale much larger than the timescale of the lattice vibrations (Ringhofer et al., 2004). The resulting Wigner equation is of the form

$$\left( \frac{\partial}{\partial t} + \frac{\hbar \mathbf{k}}{m} \nabla_{\mathbf{r}} + \mathbf{M} \mathbf{V}_{\nu} \right) f_{\nu} = \Theta_{\nu} f_{\nu} + \mathcal{Q}[f_{\nu}]$$

where \( \Theta_{\nu} \) denotes the potential operator,

$$\Theta_{\nu}[f_{\nu}](\mathbf{k}, \mathbf{r}, t) = \int V_{\nu}(\mathbf{k} - \mathbf{k}', \mathbf{r}) \, f_{\nu}(\mathbf{k}', \mathbf{r}, t) \, d\mathbf{k}'$$

and \( \mathcal{Q} \) the Boltzmann scattering operator,

$$\mathcal{Q}[f_{\nu}](\mathbf{k}, \mathbf{r}, t) = \int \left[ \mathcal{S}(\mathbf{k}, \mathbf{k}') f_{\nu}(\mathbf{k}') - \mathcal{S}(\mathbf{k}', \mathbf{k}) f_{\nu}(\mathbf{k}) \right] d\mathbf{k}'$$

The scattering rates are given by Fermi’s golden rule, with \( \mathcal{M} \) denoting the interaction matrix element and \( \hbar \omega_q \) the phonon energy.

$$\mathcal{S}(\mathbf{k}', \mathbf{k}) = \frac{V}{(2\pi)^3} \sum_{\nu = \pm 1} \frac{2\pi}{\hbar} M^2(\mathbf{q}) \left( N_q + \frac{1}{2} - \frac{\nu}{2} \right) \times 8|\epsilon(\mathbf{k}') - \epsilon(\mathbf{k}) + \hbar \omega_q|$$

Consequences of the classical limit are that scattering events are instantaneous in time, energy is conserved exactly as expressed by the \( \delta \)-function in (48) and the equation is time-irreversible.

#### 2.5 Classical force

For a one-dimensional device, the definition of the Wigner potential (44) reduces to

$$V_{\nu}(q, x) = \frac{1}{2\pi \hbar} \int_{-\infty}^{\infty} \left[ V \left( x + \frac{s}{2} \right) - V \left( x - \frac{s}{2} \right) \right] e^{-iqs} \, ds$$

This potential is assumed to be constant outside the simulation domain. When a voltage is applied to the device, the integrand of (44) becomes constant for large \( s \):

$$V(x + s/2) - V(x - s/2) \rightarrow (-e) V_{\text{app}}$$

Since the integrand does not vanish for \( s \rightarrow \infty \), the Fourier integral will diverge at \( q = 0 \), giving rise to a singularity in \( V_{\nu}(q, x) \). Therefore, one should exclude the interval \( |q| < q_c / 2 \) from the domain of integration, where \( q_c \) is some small wave number (Gehring and Kosina, 2004). The potential operator (46) is rewritten as

$$\Theta_{\nu}[f_{\nu}](k, x, t) = \int_{-q_c/2}^{q_c/2} V_{\nu}(q, x) \, f_{\nu}(k - q, x, t) \, dq$$

In the the integral over the small wave numbers the term \( f_{\nu}(k - q) \simeq -q \partial_q f_{\nu}(k) \) can be linearised, yielding a local classical force term.
A double gate MOSFET with 10 nm gate length is considered. An example for this decomposition is shown in Figure 1. Scattering is taken into account through the Boltzmann scattering operator $Q$. The Wigner potential is calculated for the 10 nm gate length device $L_{\text{g}} = 10 \text{nm}$.

Realistic devices often consist of a quantum region embedded in an extended classical region. The discrete Fourier transform requires a uniform discretisation of the quantum region with some step size $\Delta x$. To obtain a good resolution of the Wigner potential, a fairly large value for $N_k$, the number of points for the discrete Fourier transform, has to be chosen. Typical values are in the range $10^2 - 10^3$. In this case the domain of the Fourier transform, $L_{\text{FT}} = N_k \Delta x$, can become much larger than the simulation domain. For the purpose of the Fourier transform, the potential outside the device region is extrapolated by a constant. In the discrete system, the classical potential component defined by (52) becomes:

$$V_c(x_j) = \sum_{l=1}^{N_k/2} w_{jl} \frac{V(x_l)}{\sum_{j=1}^{N_k/2} w_{jl}}, \quad -\frac{N_k}{2} \leq l \leq \frac{N_k}{2}$$

$$w_{jl} = \frac{\sin((j-l)q_c \Delta x)}{(j-l)q_c \Delta x}$$

The cutoff wavenumber $q_c$ is conveniently specified by a cutoff wave length as:

$$q_c = \frac{2\pi}{\lambda_c}$$

It can be seen from (56) that $\lambda_c = 2\Delta x$ would give $V_c(x_j) = V(x_j)$ and hence the quantum component would vanish. This choice of $\lambda_c$ gives the classical limit of the discrete system. For a quantum transport calculation one has to choose $\lambda_c \gg 2\Delta x$.

2.6 Integral form of the Wigner equation

From the integro-differential form of the Wigner equation, a path-integral formulation can be derived. The equation to be transformed reads:

$$\left( \frac{\partial}{\partial t} + v(\mathbf{k}) \mathbf{\nabla} \right) f_w(\mathbf{k}, \mathbf{r}, t)$$

$$= \int \left[ S(\mathbf{k}, \mathbf{k}') + V_w(\mathbf{k} - \mathbf{k}', \mathbf{r}) + \alpha(\mathbf{k}, \mathbf{r}) \delta(\mathbf{k} - \mathbf{k}') \right] f_w(\mathbf{k}', \mathbf{r}, t)$$

where $\lambda$ denotes the total scattering rate, $\lambda(\mathbf{k}) = \int S(\mathbf{k}', \mathbf{k}) \mathbf{d}k'$ and a fictitious scattering mechanism of the form $\alpha \delta(\mathbf{k} - \mathbf{k}')$ is introduced, referred to as self-scattering (Jacoboni and Reggiani, 1983). Because of the $\delta$-function, this mechanism does not change the state of the electron and hence does not affect the solution of the equation. For the sake of brevity, we define an integral kernel $\Gamma$ and the symbols $\mu$ and $U$.

$$\mu(\mathbf{k}, \mathbf{r}) = \lambda(\mathbf{k}, \mathbf{r}) + \alpha(\mathbf{k}, \mathbf{r})$$

$$\Gamma(\mathbf{k}, \mathbf{k}', \mathbf{r}) = S(\mathbf{k}, \mathbf{k}') + V_w(\mathbf{k} - \mathbf{k}', \mathbf{r}) + \alpha(\mathbf{k}, \mathbf{r}) \delta(\mathbf{k} - \mathbf{k}')$$

$$U(\mathbf{k}, \mathbf{r}, t) = \int \Gamma(\mathbf{k}, \mathbf{k}', \mathbf{r}) \mu(\mathbf{k}', \mathbf{r}) f_w(\mathbf{k}', \mathbf{r}, t) \mathbf{d}k'$$

The Liouville operator in (58) is treated by the method of characteristics. One introduces path variables $\mathbf{K}(t)$ and $\mathbf{R}(t)$ and takes the total time derivative of $f_w$.

$$\frac{d}{dt} f_w(\mathbf{K}(t), \mathbf{R}(t), t)$$

$$= \left( \frac{\partial}{\partial t} + \frac{d\mathbf{K}(t)}{dt} \mathbf{\nabla}_k + \frac{d\mathbf{R}(t)}{dt} \mathbf{\nabla}_r \right) f_w$$
The right-hand side equals the Liouville operator if the path variables satisfy the following equations of motion.

$$\frac{d}{dt}K(t) = F(R(t)) \quad \text{and} \quad \frac{d}{dt}R(t) = v(K(t)) \quad (59)$$

Now we assume some phase-space point $k$, $r$ and some time $t$ to be given. A phase-space trajectory with the initial condition $K(t') = k$ and $R(t') = r$ is obtained by formal integration.

$$K(t') = k + \int_{t'}^{t} F(R(y)) \, dy, \quad R(t') = r + \int_{t'}^{t} v(K(y)) \, dy$$

Note that $k$, $r$, $t$ are treated as constants in the following derivation, only $t'$ is a variable. Introducing the functions

$$\tilde{f}_w(t') = f_w(K(t'), R(t'), t') \quad \text{and} \quad \tilde{\mu}(t') = \mu(K(t'), R(t'))$$

$$\tilde{U}(t') = U(K(t'), R(t'), t')$$

allows (58) to be rewritten as an ordinary differential equation of first order.

$$\frac{d}{dt'} \tilde{f}_w(t') + \tilde{\mu}(t') \tilde{f}_w(t') = \tilde{U}(t') \quad (60)$$

When multiplied by an integrating factor of the form $\exp\left[\int_{t'}^{t} \tilde{\mu}(y) \, dy\right]$, the equation can easily be integrated in time.

$$\frac{d}{dt'} \exp\left[\int_{t'}^{t} \tilde{\mu}(y) \, dy\right] \tilde{f}_w(t') = \exp\left[\int_{t'}^{t} \tilde{\mu}(y) \, dy\right] \tilde{U}(t') \quad (61)$$

The choice of the upper and lower bounds of time integration depends on whether the problem under consideration is time-dependent or stationary.

The ordinary differential equation Equation (61), which is the result of treating the Liouville operator by the method of characteristics, has the same structure as the corresponding differential equation for the Boltzmann equation. Therefore, we can refer to the work on the Boltzmann equation regarding the details of the time integration of (61) (Kosina and Nedjalkov, 2003; Kosina et al., 2000).

2.6.1 The time-dependent equation

The upper bound of the time integration should be $t' = t$ to obtain $\tilde{f}_w(t) = f_w(k, r, t)$, the value of the unknown at the given phase-space point. At $t' = 0$, an initial distribution $f_i(k, r)$ is assumed to be given. In analogy with the Boltzmann equation (Kosina and Nedjalkov, 2003), the integral form of the Wigner equation is obtained.

$$f_w(k, r, t) = \int_{t'}^{t} \int dK' \exp\left\{-\int_{t'}^{t'_{b}} \mu(K(y), R(y)) \, dy\right\}$$

$$\times \Gamma[K(t'), k', R(t')] \mu[k', R(t')] f_w[k', R(t'), t'] \quad (62)$$

$$+ \exp\left\{-\int_{t'}^{t} \mu(K(y), R(y)) \, dy\right\} f_0[K(0), R(0)]$$

This equation states that the Wigner function at time $t$ depends on the Wigner function at some previous time $t'$. Using (62) in an iterative procedure, with each iteration the time variable would move to smaller values. Therefore, another equation is desirable that describes the evolution of the system in forward time direction. Such an equation is given by the adjoint equation of (62).

$$g_w(k', r', t') = \int_{t'}^{\infty} d\tau \int \int dK' g_w[K(\tau), R(\tau), \tau]$$

$$\times \exp\left\{-\int_{t'}^{\tau} \mu(K(y), R(y)) \, dy\right\} \Gamma(k, k', r') \mu(k', r') \quad (63)$$

$$+ g_0(k', r', t')$$

The derivation of the adjoint equation (63) is discussed in detail by Kosina et al. (2000) and Kosina and Nedjalkov (2003).

2.6.2 The stationary equation

In a stationary system, the potential and all material parameters are independent of time. A phase-space trajectory is invariant under time translations. This property can be conveniently used to adjust the time reference of each trajectory (Kosina et al., 2003a,b; Nedjalkov et al., 2003).

In the stationary case we assume the phase-space point $k$, $r$ to be given at $t' = 0$. So the initial condition for the phase-space trajectory is $K(0) = k$ and $R(0) = r$. For the upper bound of time integration of (61), we choose now $t' = 0$ to obtain $f_w(0) = f_w(k, r)$. The lower time bound has to be chosen such that the functions $K(t)$ and $R(t)$ take on values at which the Wigner function is known. In the steady-state, this function is known only at the domain boundary. An appropriate lower time bound is therefore the time when the trajectory enters the simulation domain. This time is denoted by $t'^{-}$ and depends on the point $k$, $r$ under consideration. The case that the real space trajectory $R(t)$ never intersects the domain boundary can occur for a classically bound state. Then the trajectory forms a closed loop and the appropriate choice is $t'^{-} = -\infty$. Integration of (61) in the time bounds discussed above results in the integral form of the stationary Wigner equation (cf. Kosina et al., 2003a,b).

$$f(k, r) = f_0(k, r)$$

$$+ \int_{t'^{-}}^{0} \int dK' \exp\left\{-\int_{t'}^{0} \mu[K(y), R(y)] \, dy\right\}$$

$$\times \Gamma[K(t'), k', R(t')] \mu[k', R(t')] f_w[k', R(t'), t'] \quad (64)$$

$$f_0(k, r) = f_b[K(t'^{-}, k, r)], \quad R[t'^{-}, k, r)]$$

$$\times \exp\left\{-\int_{t'^{-}}^{0} \lambda[K(y), R(y)] \, dy\right\} \quad (65)$$

Here, $f_b$ denotes the boundary distribution. The integral form (64) represents a backward equation. The corresponding forward equation is given by the adjoint equation.
\[ g_w(k, r) = g_0(k, r) + \int_0^1 dt \int dK g_w[K(t), R(t)] \times \exp \left\{ -i \mu[K(y), R(y)]dy \right\} \]

\[ \Gamma(k', k, r) \mu(k, r) \Theta_D(r) \]

\( \Theta_D \) denotes the indicator function of the simulation domain \( D \). The initial condition for the phase-space trajectory is \( K(t) = k' \) and \( R(t) = r \).

### 3 The Monte Carlo method

Monte Carlo is a numerical method that can be applied to solve integral equations. Applying this method to the various integral formulations of the Wigner equation gives rise to a variety of Monte Carlo algorithms, as discussed in the following.

#### 3.1 The general scheme

This section introduces the general scheme of the Monte Carlo method and outlines its application to the solution of integrals and integral equations. To calculate some unknown value \( m \) by the Monte Carlo method, one has to find a random variable \( \xi \) whose expectation value equals \( E(\xi) = m \). The variance of \( \xi \) is designated \( \sigma^2 \), with \( \sigma \) being the standard deviation.

Now consider \( N \) independent random variables \( \xi_1, \xi_2, \ldots, \xi_N \) with distributions identical to that of \( \xi \). Consequently, their expectation values and their variance are equal.

\[ E(\xi_i) = m, \quad \text{Var}(\xi_i) = \sigma^2, \quad i = 1, 2, \ldots, N \]  

Expectation value and variance of the sum of all these random variables are given by

\[ E(\xi_1 + \xi_2 + \cdots + \xi_N) = E(\xi_1) + \cdots + E(\xi_N) = Nm \]  

\[ \text{Var}(\xi_1 + \xi_2 + \cdots + \xi_N) = \text{Var}(\xi_1) + \cdots + \text{Var}(\xi_N) = N\sigma^2 \]

Using the properties \( E(c\xi) = cE(\xi) \) and \( \text{Var}(c\xi) = c^2\text{Var}(\xi) \), one obtains from (68) and (69)

\[ E\left( \frac{1}{N}(\xi_1 + \xi_2 + \cdots + \xi_N) \right) = m \]  

\[ \text{Var}\left( \frac{1}{N}(\xi_1 + \xi_2 + \cdots + \xi_N) \right) = \frac{\sigma^2}{N} \]

Therefore, the random variable

\[ \bar{\xi} = \frac{1}{N} \sum_{i=1}^{N} \xi_i \]  

has the same expectation value as \( \xi \) and an \( N \) times reduced variance. A Monte Carlo simulation of the unknown \( m \) consists of drawing one random number \( \bar{\xi} \). Indeed, this is equivalent to drawing \( N \) values of the random variable \( \xi \) and evaluating the sample mean (72).

The Monte Carlo method gives an estimate of both the result and the error. According to the central limit theorem, the sum \( \rho_N = \xi_1 + \xi_2 + \cdots + \xi_N \) of a large number of identical random variables is approximately normal. For this reason, the following three-sigma rule holds only approximately

\[ P\left( |\rho_N - Nm| < 3\sqrt{N\sigma^2/N} \right) \approx 0.997 \]

In this equation, the expectation value and the variance of \( \rho_N \) are given by (68) and (69), respectively. Dividing the inequality by \( N \) and using \( \bar{\xi} = \rho_N/N \) we arrive at an equivalent inequality and the probability will not change:

\[ P\left( |\bar{\xi} - m| < \frac{\sigma}{\sqrt{N}} \right) \approx 0.997 \]

This formula indicates that the sample mean \( \bar{\xi} \) will be approximately equal to \( m \). The error of this approximation will most probably not exceed the value \( 3\sigma/\sqrt{N} \). This error evidently approaches zero as \( N \) increases (Sobol, 1984).

#### 3.1.1 Monte Carlo integration

We apply the Monte Carlo method to the evaluation of an integral.

\[ m = \int_{a}^{b} \phi(x) \, dx \]  

For this purpose, the integrand has to be decomposed into a product \( \phi = p \psi \), where \( p \) is a density function, which means that \( p \) is non-negative and satisfies \( \int_{a}^{b} p(x) \, dx = 1 \). Integral (75) becomes

\[ m = \int_{a}^{b} p(x)\psi(x) \, dx \]  

and denotes the expectation value \( m = E(\psi) \) of some random variable \( \Psi = \psi(X) \). Now the general scheme described in the previous section can be applied. First, a sample \( x_1, \ldots, x_N \) is generated from the density \( p \). Then the sample \( \psi_1, \ldots, \psi_N \) is obtained by evaluating the function \( \psi: \psi_i = \psi(x_i) \). The sample mean

\[ m \simeq \bar{\psi} = \frac{1}{N} \sum_{i=1}^{N} \psi_i \]  

approximates the expectation value. To employ the error estimation (74), the variance of \( \Psi \) can be approximately evaluated by the sample variance

\[ \sigma^2 \simeq \overline{\sigma^2} = \frac{1}{N - 1} \sum_{i=1}^{N} (\psi_i - \overline{\psi})^2 \]  

Because the factorisation of the integrand is not unique, different random variables can be introduced depending on the choice of the density \( p \). All of them have the same expectation value but different variance.
3.1.2 Integral equations

The kinetic equations considered in this work can be formulated as integral equations of the form

\[ f(x) = \int K(x, x') f(x') \, dx' + f_0(x) \]  

(79)

where the kernel \( K \) and the source term \( f_0 \) are given functions. Equations of this form are known as Fredholm integral equations of the second kind. In the particular cases of the Boltzmann equation and the Wigner equation the unknown function \( f \) represents the phase-space distribution function. The multidimensional variable \( x \) stands for \((k, r, t)\) in the transient case and for \((k, r)\) in the steady-state.

Substituting (79) recursively into itself gives the Neumann series by Monte Carlo integration, as pointed out within a small but finite volume around \( x \) is non-zero. However, in many cases one is interested in mean values of \( f \) rather than in a point-wise evaluation. Such a mean value can be expressed as an inner product.

\[ \langle f, A \rangle = \int f(x) A(x) \, dx \]  

(82)

It is to note that (79) is a backward equation. The corresponding forward equation is given by the adjoint equation,

\[ g(x') = \int K^\dagger(x', x) g(x) \, dx + A(x') \]  

(83)

where the kernel is defined by \( K^\dagger(x', x) = K(x, x') \). Multiplying (79) by \( g(x) \) and (83) by \( f(x') \), and integrating over \( x \) and \( x' \), respectively, results in the equality

\[ \langle f, A \rangle = \langle g, f_0 \rangle \]  

(84)

By means of (84), one can calculate a statistical mean value not only from \( f \), but also from \( g \), the solution of the adjoint equation. The given function \( A \) has to be used as the source term of the adjoint equation. The link with the numerical Monte Carlo method is established by evaluating the terms of the Neumann series by Monte Carlo integration, as pointed out in the previous section.

Note that usage of (84) precludes a point-wise evaluation of the distribution function using a forward algorithm, because \( A(x) = \delta(x) \) cannot be treated by the Monte Carlo method. The probability for a continuous random variable \( x' \) to assume a given value \( x \) is zero. Only the probability of finding \( x' \) within a small but finite volume around \( x \) is non-zero.

3.2 Particle models

Each term of the Neumann series of the adjoint equation describes a sequence of alternating free flight and scattering events. A transition consisting of a free flight with initial state \( k_i \) at time \( t_i \) and a scattering process to the final state \( k_f \) at time \( t_f \) is described by the following expression. For the sake of brevity, the \( r \)-dependence of \( \Gamma \) and \( \mu \) is omitted in the following.

\[ P(k_i, t_i, k_f, t_f) = \Gamma[k_i, K_i(t_f)] \mu[k_i(t_f)] \times \exp(-\int_{t_i}^{t_f} \mu[K_i(\tau)] \, d\tau) \]  

(85)

In a Monte Carlo simulation, the time of the next scattering event, \( t_f \), is generated from an exponential distribution, given by the terms \( \mu \exp() \) in (85). Then, a transition from the trajectory end point \( K_i(t_f) \) to the final state \( k_f \) is realised using the kernel \( \Gamma \). In contrast to the classical case, where \( P \) would represent a transition probability, such an interpretation is not possible in the case of the Wigner equation because \( P \) is not positive semi-definite. The problem originates from the Wigner potential, which assumes positive and negative values. However, because of its antisymmetry with respect to \( k \), the Wigner potential can be reformulated in terms of one positive function \( V_w^+ \) (Kosina et al., 2003a,b).

\[ V_w^+(q, r) = \max(0, V_w(q, r)) \]  

(86)

\[ V_w(q, r) = V_w^+(q, r) - V_w^-(q, r) \]  

(87)

Then, the kernel \( \Gamma \) is rewritten as a sum of transition probabilities.

\[ \Gamma(k, k') = \frac{\lambda}{\mu} s(k, k') + \frac{\alpha}{\mu} \delta(k' - k) \]  

(88)

\[ + \frac{\gamma}{\mu} [w(k, k') - w^*(k, k')] \]

The transition probabilities are given by

\[ s(k, k') = \frac{S(k', k)}{\lambda(k)} \]  

(89)

\[ w(k, k') = \frac{V_w^+(k - k')}{\gamma} \]  

(90)

\[ w^*(k, k') = w(k', k) \]  

(91)

The normalisation factor associated with the Wigner potential is defined as

\[ \gamma(r) = \int V_w^+(q, r) \, dq \]  

(92)

In the following, different variants of generating the final state \( k_f \) from the kernel \( \Gamma \) will be discussed.

3.2.1 The Markov chain method

In analogy to the simple integral (76), we have now to decompose the kernel \( P \) into a transition probability \( p \) and the remaining function \( P/p \). More details on the Markov chain method can be found by Ermakov (1975) and Hammersley...
An ensemble of particles would evolve as shown schematically in Figure 2. As the multiplier (95) is always greater than one, the absolute value of the particle weight will inevitably grow with the number of transitions on the trajectory. To solve the problem of growing particle weights, one can split particles. In this way, an increase in particle weight is transformed to an increase in particle number.

3.2.2 Pair generation method

The basic idea of splitting is refined so to avoid fractional weights. Different interpretations of the kernel are presented that conserve the magnitude of the particle weight. Choosing the initial weight to be +1, all generated particles will have weight +1 or −1. This is achieved by interpreting the potential operator in (45) as a generation term of positive and negative particles. We consider the kernel (88).

$$\Gamma(k_i, k') = \frac{\lambda}{\mu} s(k_i, k') + \frac{\alpha}{\mu} \delta(k_i - k')$$

The normalisation factor is $v = \lambda + \alpha + 2\gamma$. The free-flight time is generated from the exponential distribution appearing in (85).

$$p(s(t_i, t_f, k_i) = \mu[K_i(t_f)] \exp \left(-\int_{t_i}^{t_f} \mu[K_i(\tau)] d\tau \right)$$

For the sake of brevity, the state at the end of the free flight is labelled $k' = K_i(t_f)$ in the following. To generate the final state $k_i$, one of the four terms in (93) is selected with the associated probabilities $\lambda/v$, $\alpha/v$, $\gamma/v$, and $\gamma/v$, respectively. Apparently, these probabilities sum up to one.

If classical scattering is selected, $k_i$ is generated from $s$. If self-scattering is selected, the state does not change and $k_i = k'$ holds. If the third or fourth term is selected, the particle state is changed by scattering from the Wigner potential and $k_i$ is selected from $w$ or $w^*$, respectively. The particle weight has to be multiplied by the ratio

$$\frac{\Gamma}{p} = \pm \left(1 + \frac{2\gamma}{\lambda + \alpha} \right)$$

where the minus sign applies if $k_i$ has been generated from $w^*$. For instance, for a quantum mechanical system, where the classical scattering rate $\lambda$ is less than the Wigner scattering rate $\gamma$, the self-scattering rate $\alpha$ can be chosen in such a way that $\lambda + \alpha = \gamma$. Then, the multiplier (95) evaluates to ±3. An ensemble of particles would evolve as shown schematically in Figure 2. As the multiplier (95) is always greater than one, the absolute value of the particle weight will inevitably grow with the number of transitions on the trajectory. To solve the problem of growing particle weights, one can split particles. In this way, an increase in particle weight is transformed to an increase in particle number.

Figure 2 With the Markov chain method, the number of numerical particles is conserved. The magnitude of the particle weight increases with each event, and the sign of the weight changes randomly according to a given probability distribution

- $+1$ +3 −9
- $+1$ −3 +9
- $+1$ +3 −9

3.2.2 Pair generation method

The basic idea of splitting is refined so to avoid fractional weights. Different interpretations of the kernel are presented that conserve the magnitude of the particle weight. Choosing the initial weight to be +1, all generated particles will have weight +1 or −1. This is achieved by interpreting the potential operator in (45) as a generation term of positive and negative particles. We consider the kernel (88).

$$\Gamma(k_i, k') = \frac{\lambda}{\mu} s(k_i, k') + \frac{\alpha}{\mu} \delta(k_i - k')$$

If the Wigner scattering rate $\gamma$ is larger than the classical scattering rate $\lambda$, the self-scattering rate $\alpha$ has to be chosen large enough to satisfy the inequality $\gamma/\mu \leq 1$. Typical choices are $\mu = \text{Max}(\lambda, \gamma)$ or $\mu = \lambda + \gamma$. These expressions also hold for the less interesting case $\gamma < \lambda$, where quantum interference effects are less important than classical scattering effects. In the following, we discuss the case $\gamma > \lambda$, where quantum effects are dominant. We choose the self-scattering rate equal to $\alpha = \gamma$ and regroup the kernel as

$$\Gamma(k_i, k') = \frac{\lambda}{\mu} s(k_i, k') + \left(1 - \frac{\lambda}{\mu}\right) \left[\delta(k_i - k') + w(k_i, k') - w^*(k_i, k')\right]$$

As in the classical Monte Carlo method, the distribution of the free-flight duration is given by the exponential distribution (94). At the end of a free flight, classical scattering is selected with probability $p_s = \lambda/\mu$. With the complementary probability, $1 - p_s$, a self-scattering event and a pair generation event occur. The weight of the state generated from $w^*$ is multiplied by −1. The weights of the states generated from $w$ and from self-scattering do not change. Therefore, the magnitude of the initial particle weight is conserved, as shown in Figure 3.

Figure 3 With the pair generation method the magnitude of the particle weight is conserved, but one initial particle generates a cascade of numerical particles. At all times mass is exactly conserved

In this algorithm, classical scattering and pair generation cannot occur at the same time, as shown in Figure 4.

In the pair-generation method described, the weight of the generated particles is ±1, because the generation rate used equals $2\gamma$. If a generation rate larger than $2\gamma$ or a fixed time-step less than $1/2\gamma$ were used, the magnitude of the generated weight would be less than one. This approach has been followed by Shifren et al. (2002), where the resulting fractional weights are termed affinities. On the other hand, a generation rate less than $2\gamma$ would result in an under-sampling of the physical process. Then, the magnitude of the generated weights would be generally greater than one.
3.3 The negative sign problem

In the following, we analyse the growth rates of particle weights and particle numbers associated with the different Monte Carlo algorithms. In the Markov chain method discussed in Section 3.2.1, the weight increases at each scattering event by the multiplier (95). The growth rate of the weight can be estimated for the case of constant coefficients \( \gamma \) and \( \mu \). Because free-flight times are generated with rate \( \mu \), the mean free-flight time will be \( 1/\mu \). During a given time interval \( t \), on average \( n = \mu t \) scattering events will occur. The total weight is then estimated asymptotically for \( t \gg 1/\mu \).

\[
|W(t)| = \left( 1 + \frac{2\gamma}{\mu} \right)^n = \left( 1 + \frac{2\gamma t}{n} \right)^n \approx \exp(2\gamma t) \quad (98)
\]

This expression shows that the growth rate is determined by the Wigner scattering rate \( \gamma \) independently of the classical and the self-scattering rates. The growth rate \( 2\gamma \) is equal to the \( L_1 \) norm of the Wigner potential.

In the pair generation method, the potential operator

\[
\Theta_\text{w}[f_w](k) = \int V^+(q)[f_w(k - q) - f_w(k + q)]dq \quad (99)
\]

is interpreted as a generation term. It describes the creation of two new states, \( k - q \) and \( k + q \). The pair generation rate is equal to \( \gamma \). When generating the second state, the sign of the statistical weight is changed. It should be noted that the Wigner equation strictly conserves mass, as can be seen by taking the zeroth-order moment of (45): \( \partial n/\partial t + \text{div} \mathbf{J} = 0 \). Looking at the number of particles regardless of their statistical weights, that is, counting each particle as positive, would correspond to using the following potential operator.

\[
\Theta_\text{w}^*[f_w^*](k) = \int V_w^+(q)[f_w^*(k - q) + f_w^*(k + q)]dq \quad (100)
\]

Using (100), a continuity equation for numerical particles is obtained as \( \partial n^*/\partial t + \text{div} \mathbf{J}^* = 2\gamma \mathbf{r} n^* \). Assuming a constant \( \gamma \), the generation rate in this equation will give rise to an exponential increase in the number of numerical particles \( N^* \).

\[
N^*(t) = N^*(0) \exp(2\gamma t) \quad (101)
\]

This discussion shows that the appearance of an exponential growth rate is independent of the details of the particular Monte Carlo algorithm and must be considered to be a fundamental consequence of the non-positive kernel.

3.4 Particle annihilation

The discussed particle models are instable, because either the particle weight or the particle number grows exponentially in time. Using the Markov chain method, it has been demonstrated that tunnelling can be treated numerically by means of a particle model (Nedjalkov et al., 2002). However, because of the exponentially increasing particle weight at the very short timescale \( (2\gamma)^{-1} \), application of this algorithm turned out to be restricted to single-barrier tunnelling and small barrier heights only. This method can be useful for devices where quantum effects are weak and the potential operator is a small correction to the otherwise classical transport equation.

A stable Monte Carlo algorithm can be obtained by combining one of the particle generation methods with a method to control the particle number. One can assume that two particles of opposite weight and a sufficiently small distance in phase-space annihilate each other. The reason is that the motions of both particles are governed by the same equation. Therefore, when they come close to each other at some time instant, the two particles have approximately the same initial condition and thus a common probabilistic future. In an ensemble Monte Carlo method, a particle removal step should be performed at given time steps. During the time step, the ensemble is allowed to grow to a certain limit, then particles are removed and the initial size of the ensemble is restored. In this work, the problem has been solved for the stationary transport problem. In the algorithm, the trajectory of only one sample particle is followed, whereas other numerical particles are temporarily stored on a phase-space grid. Due to the opposite sign, particle weights annihilate to a large extent in the cells of the grid. The total residual weight in each cell has to be minimised, as it represents a measure for the numerical error of the method (Kosina et al., 2003a,b).

4 Simulation results

The Monte Carlo method for solving the Wigner equation has been applied to different nanoelectronic devices. Since quantum ballistic transport can be described by both the collision-less Wigner equation and the Schrödinger equation, this case can be used for a consistent comparison. The equivalence of both approaches is demonstrated for the case of a double gate MOSFET. Effects of scattering are then discussed for resonant tunnelling structures.

4.1 Comparison of Wigner and Schrödinger equation-based results

To demonstrate the validity of the numerical method presented, quantum ballistic transport in a double gate MOSFET is studied. The test device has a gate length of 10 nm and a silicon film thicknesses of 3 nm. Metal gates with mid-gap work function are assumed. To obtain potential profiles for the quantum transport study a classical simulation using Minimos-NT (Institut für Mikroelektronik) is performed. To account for the effect of transverse confinement due to the small silicon film thickness, the first longitudinal subband \( m_1 = 0.91 m_0 \) is calculated in a post-processing step by solving the Schrödinger equation. Due to the computational requirements of the quantum
Monte Carlo simulation, the potential is not determined self-consistently with the quantum mechanical problem in this study. Figure 5 shows the resulting subband edges along the channel of the MOSFET at a drain voltage of 0.4 V and gate voltages of 0.2 and 0.4 V. Assuming a cutoff length of $\lambda_c = 60$ nm, the spectral decomposition described in Section 2.5 is applied to obtain the classical and the quantum mechanical potential profiles.

Figure 6 compares carrier concentrations for different gate voltages resulting from the quantum Monte Carlo and a classical ballistic Monte Carlo method. In the channel the quantum mechanical concentration is significantly higher than the classical concentration. The additional concentration originates from electrons able to tunnel through the energy barrier.

Mean energies are calculated as the second moments of the Wigner function and the classical ballistic distribution function, respectively (Figure 7). In the source region both the quantum mechanical and classical mean energies are very close to the equilibrium value of $3kT/2$. The difference in mean energy is most pronounced in the region of the potential barrier. Since tunnelling electrons formally possess a negative energy, it can be expected that in a tunnelling region the quantum mechanical mean energy will be lower than the classical one. In the drain region the average energy profiles decay to the same value, which is higher than the equilibrium value. Because both the classical and quantum transport models under consideration are ballistic, electrons reaching the drain region either by surmounting the potential barrier or tunnelling, acquire considerable kinetic energy due to the applied drain-source voltage. Since scattering is neglected, the high momenta of these electrons are preserved, which leads to a significant contribution to the mean energy throughout the drain area. The opposite effect, although much weaker, can be observed in the source region. Because of a lacking back flow of electrons above the energy barrier, the negative, high-momentum tail of the Maxwellian distribution is truncated, resulting in a mean energy slightly below the equilibrium value (Figure 7).

The ballistic current-voltage characteristics, as calculated by the Wigner and the classical Monte Carlo methods, are shown in Figure 8. Current is normalised with respect to the classical on-current. For all bias points, the quantum mechanical current is significantly higher than the classical thermionic current. The current increase is caused by carriers tunnelling through the potential barrier. To justify the correctness of the Wigner Monte Carlo simulations, the current is also determined from a solution of the Schrödinger equation with open boundary conditions for the same potential profile. Results of the Wigner Monte Carlo simulations are in good agreement with the solutions of the Schrödinger equation, as shown in Figure 8.

For a quantum ballistic transport problem, the numerical solution of the Schrödinger equation will always be computationally more efficient than the solution of the Wigner equation. The important advantage of the Wigner function-based method, however, is that dissipative processes due to scattering can be included.

4.2 The effect of scattering

Virtually all published results of Wigner function–based device modelling focus on resonant tunneling diodes (Sun
et al., 1998; Mizuta and Tanoue, 1995). In this section, two different devices are discussed. Their parameter values are collected in Table 1, where RTD1 (Shifren and Ferry, 2002) is a device from the literature. The semi-classical scattering model includes polar optical, acoustic deformation potential and ionised-impurity scattering. Parameter values for GaAs have been assumed.

**Figure 8** Classical ballistic and quantum ballistic current, both obtained by Monte Carlo simulations (open symbols). Currents are normalised to the classical on-current at $V_G = 0.4V$. Solid lines show the analytically calculated thermionic emission current and the current from a numerical solution of the open boundary Schrödinger equation, respectively.

**Table 1** Parameter values of the simulated resonant tunnelling diodes. The lattice constant of GaAs is $a_0 = 0.565$ nm.

<table>
<thead>
<tr>
<th>Device name</th>
<th>Barrier height (eV)</th>
<th>Barrier width (nm)</th>
<th>Well width (nm)</th>
<th>Device length (nm)</th>
<th>Contact doping ($\text{cm}^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTD1</td>
<td>0.3</td>
<td>3.0</td>
<td>5.0</td>
<td>200.0</td>
<td>$10^{16}$</td>
</tr>
<tr>
<td>RTD2</td>
<td>0.47</td>
<td>3.0</td>
<td>4.0</td>
<td>270.0</td>
<td>$10^{18}$</td>
</tr>
</tbody>
</table>

In RTD1, the potential changes linearly in a region of 40 nm length, starting 10 nm before the emitter barrier and extending 19 nm after the collector barrier, as shown in Figure 9. The Wigner potential is discretised using $N_k = 640$ equidistant $k_x$ points and $\Delta x = 0.5$ nm spacing in $x$-direction. Assuming a cutoff length of $L_c = 80$ nm, one would require at least $N_k = L_c/\Delta x = 160$. This minimum value is often used in finite-difference simulations for the Wigner equation, but in the Monte Carlo simulation we use the considerably larger value stated above in order to get a better resolution of the energy domain. The annihilation mesh is three-dimensional. In $x$-direction, the grid covers the region where the Wigner potential is non-zero. Because of the cylindrical symmetry of the Wigner function, only two momentum coordinates have to be considered. The mesh extends to an energy of $6eV$ in both axial and radial $k$-direction.

**Figure 9** Conduction band edge of RTD1 for different voltages. A linear voltage drop is assumed.

**Figure 10** Electron concentration in RTD1 for voltages less than the resonance voltage.

For voltages above the resonance voltage, the concentration in the well drops, whereas the depletion layer continues to grow (Figure 11). The mean kinetic energy of the electrons is depicted in Figure 12. The energy density has been calculated from the second-order moment of the Wigner function (35) and divided by the electron density to get the mean energy per electron. In the zero-field regions, an energy close to the equilibrium energy is obtained, which demonstrates that the energy conservation property of the Wigner potential operator...
is also satisfied by the numerical Monte Carlo procedure. One has to keep in mind that the Wigner potential can produce a rather large momentum transfer. For the chosen value for $\Delta x$, the related energy transfer can reach values as large as 5 eV, which shows that a large degree of cancellation occurs in the estimator for the mean energy. Electrons injected from the second barrier into the collector space charge region show initially a high kinetic energy.

**Figure 11** Electron concentration in RTD1 for voltages greater than the resonance voltage

![Figure 11](image1)

**Figure 12** Mean kinetic energy in RTD1 for two different voltages

![Figure 12](image2)

Phonon scattering strongly affects the current–voltage characteristic of RTD1 (Figure 13). As compared to the coherent case, phonon scattering leads to an increase in the valley current and a resonance voltage shift. The large difference in the valley current can be explained by the electron concentration in off-resonance condition (Figure 14). With phonon scattering included, a significantly higher concentration forms in the emitter notch and injection in the double barrier is increased. This indicates that a quasi bound state forms in the emitter notch. The population of this state increases when scattering is switched on. On the other hand, in resonance condition where the applied voltage is lower, such a bound state does not form and very similar electron concentrations are observed for the coherent and noncoherent case (Figure 15).

**Figure 13** Influence of phonon scattering on the current–voltage characteristics of the RTD1

![Figure 13](image3)

**Figure 14** Electron concentration in RTD1 in off-resonance condition

![Figure 14](image4)

4.3 Inclusion of extended contact regions

Since the Wigner equation simplifies to the Boltzmann equation when the potential variation is sufficiently smooth, the proposed quantum Monte Carlo method turns into the semiclassical Monte Carlo method for vanishing Wigner potential. Therefore, one can simulate a quantum region embedded in an extended classical region with the interface between the regions correctly treated in an implicit way. By means of the Wigner generation rate $\gamma$, the simulation domain can be decomposed into quantum regions ($\gamma > 0$) and classical regions ($\gamma \approx 0$). In Figure 16, these regions within RTD1 are marked. The electron concentration and the mean energy are smooth in the extended contact regions and not affected by the strong onset of the Wigner generation rate.

In the simulation of RTD2, the Wigner potential $V_w(k_x, x)$ is discretised using $N_k = 1200$ equidistant $k_x$ points.
and $\Delta x = 0.5\,\text{nm}$ spacing in the $x$-direction. A cutoff length of $L_c = 60\,\text{nm}$ is assumed. The annihilation mesh consists of 480 points in the longitudinal and 120 points in the perpendicular momentum direction and the real space coordinate is discretised using $\Delta x = 0.5\,\text{nm}$. The electrostatic potential has been computed using the self-consistent Schrödinger-Poisson solver NANOTCAD-1D (Iannaccone et al., 2001). Figure 17 shows the electron concentration profile in the device. At the resonance voltage of $1.2\,\text{V}$, the concentration in the quantum well is considerably higher than in the off-resonance condition at $1.6\,\text{V}$. The concentration in the depletion region left of the barrier depends on the injected current and is thus correlated with the concentration in the well.

**Figure 15** Electron concentration in RTD1 in resonance condition

**Figure 16** Electron concentration and mean electron energy in RTD1 at $T = 300\,\text{K}$ and $0.1\,\text{V}$ applied voltage

**Figure 17** Electron concentration profiles in RTD2

### 5 Conclusion

The examples presented in Section 4 demonstrate that a numerical solver for the Wigner equation can provide quantitatively correct results. One requirement is that the cutoff length is chosen sufficiently large. The completeness relation of the discrete Fourier transform reflecting Heisenberg’s uncertainty principle, $\Delta k_c = \pi / L_c$, shows that a small $L_c$ will result in a coarse grid in momentum space, and resonance peaks in the transmission coefficient might not be resolved properly. In the past, the Wigner equation has been solved most frequently by finite-difference methods. Due to the non-locality of the potential operator, all points in momentum space are coupled, resulting in a very poor sparsity pattern of the matrix. Therefore, increasing the number of grid points in $k$-space, related to the cutoff length by $N_k = L_c / \Delta x$, is limited by prohibitive memory and computation time requirements. This might be one reason why quantitatively correct solutions were difficult to obtain in the past. We believe that the frequently reported accuracy problems with finite-difference Wigner function-based device simulations result from a too coarse $k$-space discretisation. As this problem occurs already for one-dimensional geometries, higher dimensional simulations using the finite-difference method are probably out of reach. It is interesting to note that Frensley, who pioneered the finite-difference method for the Wigner equation (Frensley, 1990), later abandoned this method and developed the quantum-transmitting boundary method to describe coherent transport in open systems (Frensley, 1992).

The Monte Carlo method allows the number of $k$-points to be increased. In this work, the Wigner potential has been discretised using $N_k$ of the order $10^3$. However, high-performance resonant tunnelling diodes with very high peak-to-valley current ratio pose still a problem for the Monte Carlo method. In such a device, the density can vary over several orders of magnitude, which often cannot be resolved by the Monte Carlo method. This problem is also well-known from the classical Monte Carlo method. As a solution, one could apply statistical enhancement techniques in such cases. At present, an equidistant $k$-grid is used for the discretisation of the Wigner potential. Because the transmission coefficient of double-barrier structures may show very narrow resonance peaks, using an equidistant $k$-grid may not be the optimal choice. However, because of the discrete Fourier transform of the potential involved in the computation of the Wigner potential, the use of a non-equidistant $k$-grid appears to be problematic.

In a Wigner function-based simulation of one-dimensional heterostructures, fundamental simulation parameters such as the cutoff length are closely linked to physical device parameters such as the spacing from the contacts. This property stems from the choice of plane-wave basis sets in a quantum mechanical regime of broken translational symmetry. Although analytically appealing, this basis set can cause numerical difficulties. Other approaches such as the non-equilibrium Green’s function formalism may have the advantage that other basis sets can be used more straightforwardly.
These considerations indicate that from a numerical point of view, the Wigner function formalism might not be the optimal choice for resonant tunnelling simulation. However, because the formalism describes quantum effects and scattering effects with equal accuracy, it appears well suited especially when a quasi-ballistic transport condition without energetically sharp resonances is present. One strength of the Wigner function approach is the treatment of contact regions. Non-equilibrium transport can be simulated in the whole device formed by a central quantum region embedded in extended classical regions. The presented Wigner Monte Carlo method can bridge the gap between classical device simulation and pure quantum ballistic simulations.

Development of Monte Carlo methods for the solution of the Wigner equation is still in the beginning. Research efforts are needed especially with respect to the negative sign problem. The particle generation-annihilation algorithm developed by the authors is just one solution to that problem. Improved variants of this algorithm or even new solution strategies are yet to be devised. Extension of the Monte Carlo methods to higher dimensional device geometries, however, is straightforward.

References


H. Kosina


