Effective Potentials and Quantum Fluid Models: A Thermodynamic Approach *

C. Gardner † C. Ringhofer ‡ D. Vasileska §

December 9, 2002

Abstract

We present a thermodynamic approach to introducing quantum corrections to the classical transport picture in semiconductor device simulation. This approach leads to a modified Boltzmann equation with a quantum corrected force term and to quantum corrected fluid, or quantum hydrodynamic models. We present the quantum interaction of electrons with a gate oxide barrier potential and quantum hydrodynamic simulations of a resonant tunneling diode as application examples.

KEYWORDS: Quantum thermodynamics, effective potentials, quantum hydrodynamics

1 Introduction

A quantum mechanical description of the evolution of electron ensembles in solid state materials is given by the many body Schrödinger equation for the electron ensemble. While this equation, in combination with mean field theories and the associated Poisson equation, is able to describe ballistic transport, the inclusion of scattering mechanisms on length and time scales necessary for device simulation is still an unresolved problem. Scattering of electrons with impurities and phonons leads in general to integral operators in the resulting transport

---

*This work was supported by NSF grant DECS-0218008.
†Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804 (gardner@math.la.asu.edu).
‡Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804 (ringhofer@asu.edu).
§Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287(vasileska@asu.edu).
equations which are completely non-local in space and time [2], [4], [10]. As a consequence, quantum mechanical models in device simulation usually treat quantum effects with a domain decomposition approach, modeling ballistic transport via the mean field Schrödinger equation in 'quantum regions', and linking them to a purely classical transport picture with collision effects in 'classical regions' [5], [24], [25], [26].

This paper gives an overview over a variety of alternatives to this approach, which consist of slightly modifying classical transport models to include quantum effects, thus deriving what could be called quantum corrected classical transport models. The purpose of these models is to provide a relatively inexpensive way to include quantum effects. Therefore, the derived models should in general satisfy the following criteria:

1. Their computational complexity should be comparable to their classical equivalent, i.e. drift-diffusion models, hydrodynamic models or particle based (Monte Carlo) approaches to solving the Boltzmann equation.

2. They should be corrections to this classical models in the sense that they reduce to the classical picture in the limit $\hbar \to 0$.

3. The resulting equations should be of the same, or at least a similar, type as their classical counterparts.

Satisfying the first requirement allows us to perform simulations in more complex geometries, corresponding to real devices, on the same level as is done with classical models. The purpose of the second requirement is twofold. First, it allows us to easily monitor the influence of quantum effects on solutions by comparing to the classical model equivalent. Second, it allows us to mix classical and quantum mechanical models in an engineering fashion, by, c.f. using quantum models for the electrons during free flight and semi-classical models for collisions with phonons. The third requirement guarantees that the quantum corrected models can be linked relatively easily to the outside world by using the same type of boundary conditions as for their classical counterparts.

We will in general consider two types of models in this paper, namely

**Effective Potential Approaches:**

In an effective potential approach, one replaces the quantum equation by a classical equation with a modified potential. Thus all the quantum effects in the system are modeled solely through the forces acting on the electron. For this reason the same form of numerical simulations tools can be used as for the Boltzmann transport equation, i.e. only the acceleration terms in the free flight phase of a Monte Carlo method have to be modified. The effective potential approach has been applied to modeling transport in nano-scale MOSFETs with gate length of 25 nm that are of both industrial and scientific interest. We find that
the inclusion of the barrier potential into a classical Monte Carlo particle-based simulation scheme, details of which are given in Section 8, gives rise to on-state current reduction of about 20% due to the displacement of the charge from the interface and the introduction of the inversion layer capacitance that degrades the oxide capacitance and, therefore, the MOSFET transconductance. The approach proposed here is compared to a similar approach by Ferry and coworkers [1], in which the effective potential is calculated as a convolution of a Gaussian function and the Hartree potential obtained from solving the two-dimensional or three-dimensional Poisson equation.

Quantum Fluid Models:
These models are derived in general from moment closures of quantum kinetic transport equations, like the Wigner equation. They involve roughly the same level of computational complexity as the classical drift-diffusion systems and hydrodynamic equations. On this level of modeling collisions appear only through relaxation times and diffusion coefficients, which are taken from classical theory. Quantum fluid models have been used successfully in the simulation of heterostructure devices whose function is based on quantum tunneling phenomena. There, they are able to predict negative differential resistance as well as resonant tunneling phenomena [11], [13].

This paper is organized as follows. Effective potentials as well as fluid models are derived from a quantum mechanical description, either in form of the Schrödinger equation or a quantum kinetic transport equation for the Wigner function. In Section 2 we present these equations in a setting appropriate for our purposes and in Section 3 we give the general setting for the effective potential and moment closure approach. The approach mainly considered in this paper is a thermodynamic one, and relies on an appropriate quantum representation of the thermodynamic equilibrium state. We discuss this representation and its approximations in Sections 4 and 5. Sections 6 and 7 contain the derivation of effective potentials and moment closures based on this thermodynamic approach. Section 8 gives a glimpse on simulation results for the for interaction of electrons with the Si−SiO₂ interface in a short channel MOSFET, using effective potentials and a resonant tunneling diode using a quantum fluid model.

2 Quantum Kinetic Equations

In this section we give brief overview over the kinetic equations and their relation to the quantum mechanical transport picture. The purpose of this section is mainly to provide an appropriate framework for the following derivation of macroscopic approximations and effective potentials. We start with the Schrödinger equation for a mixed state. The transient
Schrodinger equation for the wave function $\psi_\lambda$ is given by

$$i\hbar \partial_t \psi_\lambda = H\psi_\lambda := -\frac{\hbar^2}{2m_*} \Delta_x \psi_\lambda + e(V^C + V^B)(x,t)\psi_\lambda,$$  

(1)

where $\psi_(x,t)$ is the wave function corresponding to the energy level $\lambda$, $\hbar$ denotes Planck’s constant and $m_*$ is the effective mass of the electron. The effective single electron Schrodinger equation (1) is already derived from a mean field approximation for the many body Schrodinger equation. Therefore, if the forces between the electrons are Coulomb forces, the Hartree potential $V^C(x,t)$ in (1) is given by the Poisson equation

$$\nabla_x \cdot \epsilon \nabla_x V^C = e(n - D)$$  

(2)

where $\epsilon$ and $e$ denote the dielectric tensor and the electron charge respectively. $D(x)$ denotes the usual doping concentration and $n(x, t)$ is the number density of electrons given by

$$n(x, t) = \sum_\lambda F(\lambda) |\psi_\lambda(x, t)|^2.$$  

(3)

The summation in (3) ranges over the whole spectrum of the Hamiltonian $H$ in (1) and $F(\lambda)$ denotes a given statistical distribution function. The potential $V^B$ in (1) denotes the bandgap of the given material. If only one type of material is considered $V^B$ can be taken to be equal to zero. However, in device simulation it is usually necessary to consider geometries made up of several materials and $V^B$ is will in general a discontinuous function taking on different constant values in the semiconductor and the gate oxide or in the different materials making up a hetero-junction device. The Schrodinger Poisson system (1)-(2) is the most accurate description of purely ballistic transport. It does, however, bear some severe disadvantages for the actual use in device simulation. These are:

1. It does not include any scattering mechanisms important to the actual function of devices.
2. It still has to be linked to the outside world via boundary conditions which are not trivial in the case of contacts and interfaces.
3. Its steady state is given by the solution of an eigenvalue problem for the energies $\lambda$ and cannot be computed by simply setting formally $\partial_t \psi_\lambda = 0$ in (1).

The above problems are usually overcome in practice in the following way: The Schrodinger equation (1) is linked to the outside world by setting potentials constant and assuming a plain wave solution outside the domain of computation. This plain wave solution is then linked via continuity conditions to the solution inside the computational domain. Since now the steady state eigenvalue problem $H\psi_\lambda = \lambda\psi_\lambda$ has a continuous spectrum, solving the eigenvalue problem is actually replaced by solving a boundary value
Problem for the Hamiltonian for a select set of eigenvalues $\lambda$. In one dimension this approach takes the following form:

If the simulation domain (in one dimension) consists of the interval $[-L, L]$, then we assume a plain wave solution of the form

$$\psi_\lambda(x) = e^{i\xi x} + Re^{-i\xi x} \text{ for } x < -L$$

This corresponds to injecting an electron with positive momentum $\hbar\xi$ from the left into the domain. This gives the relation $\lambda = \lambda(\xi) = \frac{\hbar^2 \xi^2}{2m_\ast} + eV^-$ between the eigenvalue $\lambda$ and the wave number $\xi$, where $V^-$ denotes the constant value of the potential to the left of $x = -L$.

The corresponding plane wave solution to the right of the interval $[-L, L]$ is given by

$$\psi_\lambda(x) = Te^{i\eta(\xi)x} \text{ for } x > L, \quad \eta(\xi) = \sqrt{\frac{2m_\ast}{\hbar^2} (\lambda(\xi) - eV^+)} = \sqrt{\xi^2 + \frac{2em_\ast}{\hbar^2} (V^- - V^+)} ,$$

where $V^+$ denotes the constant value of the potential for $x > L$. It remains to compute the reflection and transmission coefficients $R$ and $T$. They can be eliminated by matching the wave function and its derivative to the outer solutions, giving the boundary conditions

$$\partial_x \psi_\lambda + i\xi \psi_\lambda |_{x=-L} = 2i\xi e^{-i\xi L}, \quad \partial_x \psi_\lambda - i\eta(\xi) \psi_\lambda |_{x=L} = 0. \quad (4)$$

Thus, for a given set of wave numbers $\xi$, a boundary value problem of the form $H \psi_{\lambda(\xi)} = \lambda(\xi) \psi_{\lambda(\xi)}$, together with the boundary conditions (4) has to be solved. A symmetric procedure has to be carried out, injecting electrons from the $x > L$ and the solution is then pieced together, using the statistic $F(\lambda)$ in (3). Starting from the so constructed equilibrium state a system of transient Schrödinger equations (1) can then be solved using the boundary conditions (4). This whole procedure becomes of course significantly more complicated in more than one spatial dimension and several approximations have to be used. We refer the reader to [26],[24],[25] for an overview.

The basic concept outlined above does, however, not yet address the issue of collisions. These are incorporated by linking the ballistic quantum transport picture to a classical transport equation (either the Boltzmann equation on a kinetic level or hydrodynamics on a macroscopic level) and by assuming that quantum effects and collisions are not relevant in the same subdomains. It should be mentioned that this linkage still poses some problems in maintaining momentum conservation [5].

Quantum corrections to the classical transport picture present an alternative to the approach outlined above. They start from formally transforming the Schrödinger - Poisson system (1)-(2) into a transport equation which is the analog of the collisionless Boltzmann equation. The tool for this is the Wigner transform [31], given by

$$f(x, p, t) = (2\pi)^{-3} \int_{\mathbb{R}^3} \rho(x + \frac{\hbar}{2} \eta, x - \frac{\hbar}{2} \eta) \exp(i\eta \cdot p) \, d\eta, \quad \rho(x, y, t) := \sum_\lambda F(\lambda) \psi_\lambda^*(x) \psi_\lambda(y) ,$$

(5)
where $\rho(x, y, t)$ denotes the density matrix of the mixed state. $f(x, p, t)$, the Wigner function, is a quasi-probability, and the equivalent of the kinetic density in the Boltzmann equation, i.e. $f$ is a function of position $x$ and momentum $p$. Direct calculus yields that positive definite density matrices $\rho$ correspond to real Wigner functions $f$ and that

$$n(x, t) = \int_{\mathbb{R}^3} f(x, p, t) \, dp = \rho(x, x, t) = \sum_{\lambda} F(\lambda) |\psi_\lambda(x, t)|^2$$

holds. Direct calculus again yields that, if the wave functions $\psi_\lambda$ satisfy the Schrödinger equation (1), the Wigner function $f$ satisfies the Wigner equation

$$\partial_t f + \frac{1}{m_*} \nabla_x \cdot (pf) - e \theta[V] f = 0 \quad V := V^C + V^B,$$

with the operator $\theta[V]$ given by

$$\theta[V] = \frac{i}{\hbar} [V(x + \frac{\hbar}{2i} \nabla_p) - V(x - \frac{\hbar}{2i} \nabla_p)].$$

The operator $\theta[V]$ has to be understood in the sense of pseudo-differential operators. So the action of $\theta[V]$ is given by

$$\theta[V] f(x, p, t) = (2\pi)^{-3} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{i}{\hbar} [V(x + \frac{\hbar}{2i} \eta) - V(x - \frac{\hbar}{2i} \eta)] f(x, q, t) \exp[i \eta \cdot (p - q)] \, dq \, d\eta.$$

At least on a purely formal level, the pseudo differential operator $\theta[V]$ tends to the classical acceleration term $\nabla_x V \cdot \nabla_p$ in the classical limit for $\hbar \to 0$, and therefore, on a formal level again, the Wigner equation (6) tends to the Vlasov equation

$$\partial_t f + \frac{1}{m_*} \nabla_x \cdot (pf) - e \nabla_p \cdot (\nabla_x V f) = 0.$$

It is this similarity which is exploited in the general approach to quantum corrections to classical kinetic and fluid dynamic modeling. So the starting point of quantum corrections to the classical transport picture will be the Wigner - Boltzmann - Poisson system

$$(a) \quad \partial_t f + \frac{1}{m_*} \nabla_x \cdot (pf) - e \theta[V] f = Q(f)$$

$$(b) \quad \nabla \cdot \varepsilon \nabla V = e(n - D), \quad n(x, t) = \int_{\mathbb{R}^3} f(x, p, t) \, dp,$$

where $Q(f)$ denotes the collision operator. In the case of fluid approximations only the knowledge of the moments of the collision operator $Q$, i.e. the relaxation times, are required and we will assume those to be given by the classical theory.
3 Moment Closures and Effective Potentials

In this section we will introduce the concept of effective potentials and give the formal derivation of the moment system for the Wigner - Boltzmann - Poisson system (7). We start with the most famous effective potential approximation, the Bohm - potential, which is exact for a single pure state.

3.1 Bohm Potentials

Writing the complex wave function $\psi_\lambda$ in (1) in terms of its amplitude and phase, i.e.

$$\psi_\lambda = a_\lambda \exp(i\phi_\lambda)$$

we obtain the Schrödinger equations for amplitude and phase of the form

$$\partial_t \ln(a_\lambda) + \frac{\hbar}{2m^*_\lambda} [\Delta \phi_\lambda + 2\nabla \ln(a_\lambda) \cdot \nabla \phi_\lambda] = 0$$

$$\hbar \partial_t \phi_\lambda + \frac{\hbar^2}{2m^*_\lambda} [|\nabla \phi_\lambda|^2 - \frac{\Delta a_\lambda}{a_\lambda}] + eV = 0 .$$

Setting $n_\lambda = a^2_\lambda$ and $\nabla \phi_\lambda = \frac{m^*_\lambda}{\hbar} u_\lambda$, we obtain formally a zero temperature Euler equation with a corrected potential term for the amplitude and phase of each wave function of the mixed state.

(a) $\partial_t n_\lambda + \nabla \cdot (n_\lambda u_\lambda) = 0$

(b) $\partial_t (m^*_\lambda n_\lambda u_\lambda) + \nabla \cdot (m^*_\lambda n_\lambda u_\lambda u_\lambda^T) + n_\lambda \nabla [eV - \frac{\hbar^2}{2m^*_\lambda} \sqrt{\frac{n_\lambda}{m^*_\lambda}}] = 0, \quad V = V^C + V^B .$

In the derivation of (8)(b) we have made use of the fact that each individual $u_\lambda$ is a gradient field. The corrected potential $V^Q = V^C + V^B - \frac{\hbar^2}{2m^*_\lambda} \sqrt{\frac{n_\lambda}{m^*_\lambda}}$ is called the Bohm potential [8]. Alternatively, the quantum correction in the potential can be included as a pressure term by writing the momentum equation (8)(b) as

$$\partial_t (m^*_\lambda n_\lambda u_\lambda) + \nabla \cdot [n_\lambda (m^*_\lambda u_\lambda u_\lambda^T) - \frac{\hbar^2}{4m^*_\lambda} \nabla^2 \ln(n_\lambda))] + cn_\lambda \nabla V = 0, \quad V = V^C + V^B .$$

So, to reproduce the Schrödinger equation (1) one solves the system (8) for each energy $\lambda$ and computes the total densities according to the statistics used in (3). Each of the equations (9) represents the movement of a single electron, and is therefore given at zero temperature. The temperature of the ensemble has to be computed from the variance among the different $u_\lambda$. Although (8) has the appearance of a fluid equation it is completely equivalent to the Schrödinger equation, and therefore is not really a macroscopic or fluid approximation. Fluid approximations at finite temperature are obtained from (8) in the following, more or less ad
hoc, way [15], [16], [17] [18]. Starting from a single nonlinear Schrödinger equation, where
the potential is written as $V = V^C + V^B + V^{NLS}(|\psi|^2)$ the same procedure as outlined above
can be repeated, leading to

$$\partial_t (m*u) + \nabla \cdot [n(m*uuT - \frac{\hbar^2}{4m*} \nabla^2 \ln(n)) + en\nabla(V^C + V^B + V^{NLS}(n)) = 0 \ .$$

The nonlinear potential term $\nabla V^{NLS}$ can be absorbed into a stress tensor by introducing
the function $P(n)$ as $P_0(n) = -e \int n \frac{d}{dn} V^{NLS}(n) \ dn$, giving

$$\partial_t (m*u) + \nabla \cdot [n(m*uuT - \frac{\hbar^2}{4m*} \nabla^2 \ln(n)) - P_0(n)I] + en\nabla(V^C + V^B) = 0 \ . \quad (10)$$

This is the approach taken in [15], [16] and [17] and basically assumes that, somewhat in the
spirit of density functional theory, the properties of the electron ensemble can be replaced
by an effective single electron picture, using the right potential. So there is a one to one
relation between each scalar form the stress tensor $P = P_0 I$ and the nonlinear potential. So
c.f. a constant temperature $P_0 = -nKT$ would correspond to a logarithmic nonlinearity in
the Schrödinger equation $eV^{NLS} = KT \ln(n)$.

### 3.2 Thermodynamic Approximations

In this paper we take a different approach, namely to derive effective potential as well as fluid
approximations via a thermodynamic argument. We start from the Wigner - Boltzmann -
Poisson equation (7) and transform it first to center of mass coordinates. We define the
mean velocity $u(x,t)$ of the ensemble by

$$\int_{\mathbb{R}^3} pf(x,p,t) \ dp = m*un(x,t), \quad n(x,t) = \int_{\mathbb{R}^3} f(x,p,t) \ dp \ ,$$

and choose $m*u$ as the origin of the coordinate system in momentum space, i.e. we choose
Lagrangian instead of Eulerian coordinates. Defining $f^L(x,p,t) = f(x,p + m*u(x,t),t)$ we
obtain the Wigner - Boltzmann equation

$$\partial_t f^L + \frac{1}{m*} \nabla_x \cdot [(p+m*u)f^L] - \nabla_p \cdot [f^L [m*\partial_t u + ((p+m*u) \cdot \nabla_x)u]] - \theta[V] f^L = Q^L(f^L) \ , \quad (11)$$

where we also shift the collision operator $Q$, i.e. $Q^L(f^L)(x,p,t) = Q(f)(x,p + m*u,t)$ holds.
Note that, by definition, the first order moment of $f^L$ vanishes identically, i.e. $\int p f^L \ dp = 0$
always holds. The basic idea is to choose approximations which are exact if the momentum
centered density $f^L$ is given by the thermal equilibrium density. In the classical case this
would correspond to

$$f^L(x,p,t) = \exp[-\frac{\beta|p|^2}{2m*} - e\beta V(x,t)] \quad (12)$$
with $\beta = (KT)^{-1}$ the equilibrium temperature. In quantum mechanics thermal equilibrium is characterized by the density matrix $\rho$ corresponding to the Wigner function $f^L$ being given by $F(\beta H)$ with $H = -\frac{\hbar^2}{2m^*} \Delta + eV$ and $F$ the statistical distribution function used, i.e. the Boltzmann or the Fermi - Dirac statistic. Using the definition (5) of the Wigner function this means

$$f^{eq}(x, p) = (2\pi)^{-3} \int_{\mathbb{R}^3} \rho^{eq}(x + \frac{\hbar}{2} \eta, x - \frac{\hbar}{2} \eta) \exp(i \eta \cdot p) \, d\eta,$$

$$\rho^{eq}(x, y) := \sum_{\lambda} F(\beta \lambda) \psi^{eq}_\lambda(x) \psi^{eq}_\lambda(y),$$

$$(-\frac{\hbar^2}{2m^*} \Delta + V) \psi^{eq}_\lambda = \lambda \psi^{eq}_\lambda.$$  

Of course $f^{eq}$ as well as the eigenfunctions $\psi^{eq}_\lambda$ will depend on the the time $t$ through the potential $V$, since we only assume a local in time equilibrium for a given time dependent potential. The Function $F(z)$ is either given by $F(z) = e^{\phi - z}$ in the case of Boltzmann statistics or by $F(z) = (1 + ce^{z - \phi})^{-1}$ in the case of Fermi - Dirac statistics.

**Effective Potentials:**

The effective potential approximation is now given by replacing the pseudo differential operator $\theta[V]$ in (11) by a classical counterpart

$$\theta[V]f \to \nabla_p \cdot [f \nabla_x V^Q(x, p, t)]$$

in such a way that (14) is exact for $f^L = f^{eq}$ given by (13). Because of the equilibrium property this implies that

$$f^{eq}(x, p) = \exp[-\frac{\beta |p|^2}{2m^*} - \beta eV^Q(x, p)]$$

holds.

**Quantum Hydrodynamics:**

To consider fluid approximations to the momentum centered Wigner - Boltzmann equation (11) we take, following [11],[12], the first three moments of (11), corresponding to mass momentum and energy, and obtain

$$\partial_t n + \nabla_x \cdot (nu) = 0$$

$$\partial_t (nm_* u) + \nabla_x \cdot (m_* n uu^T) - \nabla_x \cdot P + n \nabla_x V = \langle p Q^0 \rangle$$

$$\partial_t (3nKT) + \nabla_x \cdot (3nKT u + 2q) - 2Tr(P \nabla_x^T u) = \langle |p|^2 Q^0 \rangle,$$

where we define temperature $T$, stress tensor $P$ and heat flux $q$ in the usual way by

$$P = -\frac{1}{m_*} \int_{\mathbb{R}^3} pp^T f^L \, dp,$$

$$3nKT = -Tr(P) = \frac{1}{m_*} \int_{\mathbb{R}^3} |p|^2 f^L \, dp.$$
Here $K$ denotes the Boltzmann constant. The moment system (15) has to be closed by expressing the $P$ and $q$ in terms of the primary variables $n, u, T$ (charge, velocity, temperature). Models for the moments of the collision operator $Q$ have to be assumed and a variety is readily available in the literature [3], [28], the simplest being the relaxation time models

$$\langle pQ^0 \rangle = -\frac{n m_* u}{\tau_p}, \quad \langle |p|^2 Q^0 \rangle = \frac{3nK(T_0 - T)}{\tau_w}$$

with relaxation times $\tau_p$ and $\tau_w$. In the context of thermodynamic approximations this is done by setting again $f^L = f^{eq}$ and deriving the appropriate relations from (13). Note that the moment system (15) is identical to the classical moment system, leading to the compressible Euler equations, since the first three moments of the pseudo differential operator $\theta[V]f$ coincide with the moments of its classical counterpart $\nabla_p \cdot (f\nabla_x V)$. The quantum corrections therefore will appear only in the closure relations. In the classical case, with $f^L$ given by (12), one obtains

$$P = -nKT Id, \quad q = 0,$$

where $Id$ denotes the identity matrix, and the quantum corrections will enter only through the difference between (12) and (13). For the purpose of actual numerical discretization it is beneficial to write (15)(c) in conservative form, expressing the balance of the total energy instead of the thermal energy $3nKT$. Multiplying (15)(a) with $-m_*|u|^2$ and (15)(b) with $2u^T$ and adding to (15)(c) gives

$$\partial_t (m_* n|u|^2 + 3nT) + \nabla_x \cdot [m_* n|u|^2 u + 3nTu + 2q] - 2\nabla_x \cdot (Pu) + 2nu^T \nabla_x V = \langle (2u^T p + \frac{|p|^2}{m_*})Q^0 \rangle = \frac{1}{m_*} \langle |p + m_*u|^2 Q^0 \rangle$$

### 4 Thermodynamic Equilibrium

This section is concerned with the derivation of the equilibrium density $f^{eq}$ in (13). Although (13) completely defines the thermodynamic equilibrium in terms of the Hamiltonian, and therefore in terms of the potential $V$, we will need an explicit expression for $f^{eq}$ in order to obtain a formula for the effective quantum potential $V^Q$ in (14) and the closure moments $P$ and $q$ in (15). Of course, it is not possible to find an exact expression for $f^{eq}$, and we will need some kind of approximate formula. These approximations will be the subject of the next two sections. In this section we will give a form of the equilibrium density which is appropriate for the derivation of such approximations, i.e. some alternative way to solve the eigenvalue problem for the wave functions $\psi^{eq}_\lambda$. For the case of Boltzmann statistics this alternative way is given by the Bloch equation. The Bloch equation is a formal way to
replace the computation of the exponential of a density matrix by the solution of a parabolic differential equation. We define the equilibrium density matrix \( \rho^{eq}(x, y, \beta) \) as the exponential of the Hamiltonian by

\[
\rho^{eq}(x, y, \beta) = \sum_{\lambda} \exp[\beta(\phi - \lambda)] \psi^{eq}_\lambda(x)^* \psi^{eq}_\lambda(y) ,
\]

where the \( \psi^{eq}_\lambda \) are the normalized eigenfunctions of the Hamiltonian \( H = -\frac{\hbar^2}{2m^*} \Delta + eV \).

Evaluating \( \rho^{eq} \) at \( \beta = 0 \) gives, because of the mutual orthogonality of the eigenfunctions, \( \rho^{eq}(x, y, 0) = \delta(x - y) \). Differentiating \( \rho^{eq} \) with respect to the inverse temperature \( \beta \) gives

\[
\partial_\beta \rho^{eq}(x, y, \beta) = \sum_{\lambda} \exp[\beta(\phi - \lambda)](\phi - \lambda) \psi^{eq}_\lambda(x)^* \psi^{eq}_\lambda(y) .
\]

Applying the Hamiltonian \( H \) to the density matrix \( \rho^{eq} \) gives

\[
H \rho^{eq} = \sum_{\lambda} \exp[\beta(\phi - \lambda)] \lambda \psi^{eq}_\lambda(x)^* \psi^{eq}_\lambda(y) = \rho^{eq}H ,
\]

Because of the eigenfunction property of the \( \psi^{eq}_\lambda \) and the fact that \( \rho^{eq} \) commutes with the Hamiltonian. Thus the equilibrium density matrix \( \rho^{eq} \) satisfies the initial value problem

\[
\partial_\beta \rho^{eq}(x, y, \beta) = -\frac{1}{2}(H \rho^{eq} + \rho^{eq}H) + \phi \rho^{eq}, \quad \rho^{eq}(x, y, 0) = \delta(x - y) ,
\]

where \( H \rho^{eq} \) denotes the usual matrix product, or, evaluating the Hamiltonian,

\[
\partial_\beta \rho^{eq}(x, y, \beta) = \frac{\hbar^2}{4m^*}((\Delta_x + \Delta_y) \rho^{eq} - \frac{e}{2}[V(x) + V(y)] \rho^{eq} + \phi \rho^{eq}), \quad \rho^{eq}(x, y, 0) = \delta(x - y) . \quad (17)
\]

(17) is called the Bloch equation for the equilibrium density matrix.

**Remark:** We have symmetrized the Bloch equation (17) in order to make it obvious that the equilibrium density matrix \( \rho^{eq} \) will always be real and self adjoint.

So we have replaced the solution of the eigenvalue problem for the Hamiltonian \( H \) by solving the parabolic initial value problem (17) in twice as many spatial dimensions. This does not make the problem computationally easier, but more amenable to asymptotic analysis. Physically the Bloch equation has the interpretation that we start from a totally uncorrelated state at infinite temperature (\( \beta = 0 \)) and smooth out to a state at finite temperature (\( \beta > 0 \)).

Computing the Wigner transform \( f^{eq} \) of \( \rho^{eq} \), according to (13), gives the initial value problem

\[
\partial_\beta f^{eq}(x, p, \beta) = \frac{\hbar^2}{8m^*} \Delta_x f^{eq} - \frac{|p|^2}{2m^*} f^{eq} - e\omega[V] f^{eq} + \phi f^{eq}, \quad f^{eq}(x, p, 0) = \hbar^{-3} , \quad (18)
\]

with the pseudo differential operator \( \omega \) given by

\[
\omega[V] = \frac{1}{2}[V(x + \frac{\hbar}{2i} \nabla_p) + V(x - \frac{\hbar}{2i} \nabla_p)]
\]

or

\[
\omega[V]f(x, p) = \frac{1}{2}(2\pi)^{-3} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} [V(x + \frac{\hbar}{2} \eta) + V(x - \frac{\hbar}{2} \eta)]f(x, q) \exp[i\eta \cdot (p - q)] dq d\eta
\]

11
5 Approximations to Thermal Equilibrium

We now turn to the derivation of approximate formulas for the thermodynamic equilibrium state given by the solution of the Bloch equation (18). The advantage of the Bloch equation is that asymptotic formulas for the equilibrium state can be found more easily from (18) than from the solution of the eigenvalue problem. We will use two different types of asymptotic formulas. The first is derived from the semiclassical regime (i.e. $\hbar \to 0$). This leads essentially again to the Bohm potential. The disadvantage of the semiclassical approximation is twofold: First it assumes that the Planck constant is small compared to the involved spatial and time scales. This is in general not the case for small quantum devices such as c.f. resonant tunneling diodes. Second, the resulting approximations involve higher derivatives of the potential, i.e. they hold only for sufficiently smooth potentials and are not valid in the vicinity of discontinuous potential barriers. We therefore present another approximation, which assumes neither smooth potentials nor the semiclassical regime but instead is based on the assumption that the unbounded operators in the Bloch equation will dominate the bounded operators.

5.1 Semiclassical Approximations

Assuming a sufficiently smooth potential $V$ the pseudo differential operator $\omega$ in (18) can be expanded in powers of $\hbar$, giving

$$\omega[V] f = V f - \frac{\hbar^2}{8} Tr[(\nabla_x^2 V)(\nabla_p^2 f)] + O(\hbar^4).$$

For reasons of notational convenience we introduce the parameter $\alpha = \frac{\hbar^2}{8m}\$ and write the expanded equation (18) as

$$\partial_\beta f^eq(x, p, \beta) = \alpha \Delta_x f^eq - \frac{|p|^2}{2m}\ f^eq - eV f + e\alpha m\ Tr[(\nabla_x^2 V)(\nabla_p^2 f)] + \phi f^eq, \quad f^eq(x, p, 0) = h^{-3}$$

We expand the equilibrium Wigner function $f^eq$ as $f^eq = h^{-3}(f_0 + \alpha f_1 + \ldots)$ and obtain for the zero and first order term

(a) \quad $\partial_\beta f_0(x, p, \beta) = -\frac{|p|^2}{2m} f_0 - eV f_0 + \phi f_0, \quad f_0(x, p, 0) = 1$ \hspace{1cm} (19)

(b) \quad $\partial_\beta f_1(x, p, \beta) = \Delta_x f_0 - \frac{|p|^2}{2m} f_1 - eV f_1 + e\alpha m\ Tr[(\nabla_x^2 V)(\nabla_p^2 f_0)] + \phi f_1, \quad f_1(x, p, 0) = 0$

The zero order term can be computed immediately from (19)(a) as

$$f_0(x, p, \beta) = \exp[-\frac{\beta |p|^2}{2m} + \beta(\phi - eV)].$$ \hspace{1cm} (20)
In order to compute $f_1$ form (19)(b) via variation of constants we set $f_1 = g f_0$ and obtain

$$\partial_\beta g(x, p, \beta) = \frac{\Delta_x f_0}{f_0} + \frac{e m_*}{f_0} \nabla_x^2 V : \nabla_p^2 f_0, \quad g(x, p, 0) = 0$$

as the initial value problem for $g$. Using the formula (20) for $f_0$ we compute the terms on the right hand side as

$$\partial_\beta g(x, p, \beta) = -\beta e \Delta_x V + \beta^2 e^2 |\nabla_x V|^2 + e m_* Tr \left[ (\nabla_x^2 V) \left( -\frac{\beta}{m_*} I + \frac{\beta^2}{m_*^2} p p^T \right) \right], \quad g(x, p, 0) = 0$$

or

$$\partial_\beta g(x, p, \beta) = -2\beta e \Delta_x V + \beta^2 e^2 |\nabla_x V|^2 + \frac{e \beta^2}{m_*} Tr \left[ (\nabla_x^2 V) p p^T \right], \quad g(x, p, 0) = 0$$

and after integration

$$g(x, p, \beta) = -\beta^2 e \Delta_x V + \frac{\beta^3 e^2}{3} |\nabla_x V|^2 + \frac{e \beta^3}{3 m_*} Tr \left[ (\nabla_x^2 V) p p^T \right] \cdot$$

This gives the approximate formula for the thermal equilibrium Wigner function $f^{eq}$ as

$$\hbar^3 f^{eq}(x, p, \beta) =$$

$$\exp \left[ -\frac{\beta |p|^2}{2 m_*} + \beta (\phi - eV) \right] \left\{ 1 - \frac{\beta^2 e \hbar^2}{8 m_*} \Delta_x V + \frac{\beta^3 e^2 \hbar^2}{24 m_*} |\nabla_x V|^2 + \frac{e \beta^3 \hbar^2}{24 m_*^2} Tr \left[ (\nabla_x^2 V) p p^T \right] \right\} + O(\hbar^4) \cdot$$

Clearly, (21) is an order $\hbar^2$ correction to the classical thermal equilibrium given by the Maxwellian. The total number of electrons in the equilibrium state is given by the Quasi Fermi potential $\phi$.

### 5.2 Born Approximations

As mentioned, the disadvantage of the approximate expression (21) for the thermal equilibrium Wigner function is that it involves higher order derivatives of the potential $V$ which, in the case of discontinuous barrier potentials, will only exist in a distributional sense. We therefore give a different type of asymptotic solution. The basic assumption is that the unbounded operator (the Laplacian) in (18) dominates the potential term. So we formally set $V = \epsilon V_\epsilon$ and expand the solution of (18) in powers of the parameter $\epsilon$. It should be pointed out that $\epsilon$ is a purely formal parameter and that the procedure below corresponds to the Born approximation of $e^{-\beta H}$ [13], [14]. So we expand the solution of

$$\partial_\beta f^{eq}(x, p, \beta) = \alpha \Delta_x f^{eq} - \frac{|p|^2}{2 m_*} f^{eq} - e \epsilon \omega[V_\epsilon] f^{eq} + \phi f^{eq}, \quad f^{eq}(x, p, 0) = \hbar^{-3} \quad (22)$$
in powers of the artificial parameter \( \varepsilon \). \( \alpha \) in (22) denotes again the quantity \( \frac{\hbar^2}{8m_*} \), but now is not assumed to be small. Similarly to the previous case, we again set \( f^{eq} = \hbar^{-3}(f_0 + \varepsilon f_1) \), and obtain

\[
\text{(a) } \partial_\beta f_0(x, p, \beta) = \alpha \Delta_x f_0 - \frac{|p|^2}{2m_*} f_0 + \phi f_0, \quad f_0(x, p, 0) = 1 \tag{23}
\]

\[
\text{(b) } \partial_\beta f_1(x, p, \beta) = \alpha \Delta_x f_1 - \frac{|p|^2}{2m_*} f_1 - e \omega[V_\varepsilon] f_0 + \phi f_1, \quad f_1(x, p, 0) = 0
\]
as the equations determining \( f_0 \) and \( f_1 \). Again, (23)(a) can be solved immediately, giving

\[
f_0 = \exp[\beta \phi - \beta \frac{|p|^2}{2m_*}]
\]

Solving (23)(b) turns out to be a bit more technically involved since we have to use the Greens function for the heat operator. This is done best by Fourier transforming (23)(b) in space, and writing \( f_1 \) as a pseudo differential operator acting (in the spatial direction) on the potential \( V_\varepsilon \). We define

\[
g(\xi, p, \beta) = (2\pi)^{-3} \int_{\mathbb{R}^3} f_1(x, p, \beta) \exp(-i\xi \cdot x) \ dx, \quad \hat{V}(\xi) = (2\pi)^{-3} \int_{\mathbb{R}^3} V_\varepsilon(x) \exp(-i\xi \cdot x) \ dx
\]

and obtain

\[
\partial_\beta g(\xi, p, \beta) = -\alpha |\xi|^2 g - \frac{|p|^2}{2m_*} g - e R(\xi, p, \beta) \hat{V} + \phi g, \quad g(\xi, p, 0) = 0 \tag{25}
\]
as the initial value problem defining the Fourier transform \( g(\xi, p, \beta) \), where the term \( R(\xi, p, \beta) \hat{V} \) is the Fourier transform in the spatial direction of \( \omega[V_\varepsilon] f_0 \).

\[
R(\xi, p, \beta) \hat{V}(\xi) = (2\pi)^{-3} \int_{\mathbb{R}^3} \omega[V_\varepsilon] f_0(x, p, \beta) e^{-i\xi \cdot x} \ dx
\]

holds. A direct calculation, using the definition of the pseudo differential operator \( \omega[V_\varepsilon] \) yields

\[
R(\xi, p, \beta) = \frac{1}{2} \exp[\beta \phi] (2\pi)^{-3/2} \sum_{\sigma = \pm 1} \int_{\mathbb{R}^3} \exp[i \sqrt{\frac{\beta}{m_*} \eta \cdot (p + \sigma \frac{\hbar}{2} \xi)} - \frac{|\eta|^2}{2}] \ d\eta
\]

\[
R(\xi, p, \beta) = \frac{1}{2} \exp[\beta \phi] \sum_{\sigma = \pm 1} \exp[-\frac{\beta}{2m_*} |p + \sigma \frac{\hbar}{2} \xi|^2]
\]

(26)

Inserting (26) in (25) and solving the ordinary differential equation gives

\[
g(\xi, p, \beta) = -e \beta \hat{V}(\xi) \exp[-\alpha |\xi|^2 - \frac{\beta |p|^2}{2m_*} + \beta \phi] \int_0^1 \cosh[\frac{\gamma \beta \hbar}{2m_*} \cdot \xi] \ d\gamma
\]

Reversing the Fourier transforms, we obtain the first order term \( f_1 \) as

\[
f_1(x, p, \beta) =
\]
\[-e \beta (2\pi)^{-3} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \int_0^1 V_\varepsilon(y) \exp[-\alpha \beta |\xi|^2 - \frac{\beta |p|^2}{2m_*} + \beta \phi] \cosh \left( \frac{\gamma \hbar}{2m_*} p \cdot \xi \right) \exp[i \xi \cdot (x-y)] \, d\gamma dy d\xi \]

which we write, formally, as a pseudo differential operator $S$ in the spatial direction, acting on the potential $V_\varepsilon$:

\[
f_1(x, p, \beta) = -\exp[\beta \phi] S(-i \nabla_x, p, \beta) V_\varepsilon
\]

with the symbol $S$ given by

\[
S(\xi, p, \beta) = \frac{e \beta}{2} \sum_{\sigma=\pm} \int_0^1 \exp[\alpha \beta (\gamma^2 - 1)|\xi|^2 - \frac{\beta}{2m_*} |p + \sigma \gamma \hbar \xi|^2] \, d\gamma, \quad \alpha = \frac{\hbar^2}{8m_*}
\]

So, combining (24) and (27) and setting $V = \varepsilon V_\varepsilon$ again, the Born approximation approximation formula to the thermal equilibrium state is given by

\[
h^3 f_{eq}(x, p, \beta) \approx \exp(\beta \phi)[\exp(-\frac{\beta |p|^2}{2m_*}) - S(-i \nabla_x, p, \beta)V]
\]

with the symbol of the operator $S$ given by (28). Note that $S$ is actually a smoothing operator. It can be shown that for large $\xi$ the symbol $S$ behaves like $O\left(\frac{1}{|\xi|^2}\right)$. Therefore the term $SV$ will actually have two more classical derivatives than the original potential $V$.

6 Effective Potentials and Particle Discretizations

The goal of the effective potential approach is to incorporate quantum corrections into particle based simulators by replacing the classical forces moving the electron during the free flight by modified quantum forces. So, in the semiclassical Boltzmann equation

\[
\partial_t f + \nabla_x \left( \frac{1}{m_*} pf \right) - \nabla_p (ef \nabla_x V) = Q(f)
\]

the field term $-e \nabla_x V(x, t)$ is replaced by a modified quantum field term $-e \nabla_x V^Q(x, p, t)$. Given an existing particle based simulator, this is, at least in theory, easily achieved by computing the trajectories of electrons via

\[
\frac{d}{dt} x = \frac{1}{m_*} p, \quad \frac{d}{dt} p = -e \nabla_x V^Q(x, p),
\]

where the time dependence of the potential is frozen during the free flight phase. This approach is computationally much easier than trying to solve the Wigner equation (6) directly by a particle based method (see [6], [7] [29] and reference therein). There, however, are two important points which should be mentioned here

1. Representing the density function as a simultaneous $\delta-$ function in space and momentum violates the Heisenberg principle.
2. Just modifying the forces will invariably result in a strictly positive phase space density $f$ while the Wigner function can become locally nonpositive.

So, this approach is not an exact representation of quantum mechanics. It can, however, represent certain quantum effects, such as the nonlocal interaction of electrons with barrier potentials and even tunneling. On the other hand, it allows for the immediate incorporation of all the relevant semiclassical scattering mechanisms usually included in Monte Carlo simulations. The use of more quantum mechanically correct collision operators is still very much an open research topic [2], [4], [10]. Using the Bohm potential from Section 3.1 the effective quantum potential $V^Q$ is given by

$$eV^Q = eV - \frac{\hbar^2}{2m_\ast} \frac{\Delta \sqrt{n}}{\sqrt{n}}. \tag{30}$$

The same quantum corrected potential can be obtained from the semiclassical approximation to the thermal equilibrium state in Section 5.1. The evaluation of (30) poses a significant challenge in the context of particle based discretizations. The number density $n$ is given by a superposition of $\delta-$functions in physical space of which the third derivative has to be computed to evaluate the force. This can be either done by computing $n$ on a mesh using the usual cloud in cell approach and then using a difference discretization, or by computing the third derivatives directly in a meshless approach [32], [33]. The first approach requires a considerable amount of smoothing.

Alternatively, one can use the Born approximation to the thermodynamic equilibrium state in Section 5.2 by requiring that replacing term $\theta[V]f$ by $\nabla_{p}(f \nabla_x V^Q)$ in the Wigner equation (6) is exact in thermal equilibrium, i.e. by writing

$$\hbar^3 f^{eq} = \exp[\beta \phi - \frac{\beta |p|^2}{2m_\ast} - \beta eV^Q(x, p, \beta)] ,$$

with $f^{eq}$ given by (29). This gives the effective quantum potential $V^Q$ as

$$V^Q(x, p, \beta) = \frac{1}{\beta e} \exp(\frac{\beta |p|^2}{2m_\ast})S(-i\nabla_x, p, \beta)V(x) , \tag{31}$$

with the symbol of the operator $S$ given, according to (28), by

$$S(\xi, p, \beta) = \frac{e\beta}{2} \sum_{\sigma=\pm 1} \int_0^1 \exp[\alpha \beta (\gamma^2 - 1)|\xi|^2 - \frac{\beta}{2m_\ast} |p + \sigma \gamma \hbar \frac{p \cdot \xi}{2m_\ast}|] d\gamma, \quad \alpha = \frac{\hbar^2}{8m_\ast}$$

The symbol $S$ can be simplified to give

$$S(\xi, p, \beta) = e\beta \exp(-\frac{\beta |p|^2}{2m_\ast} - \alpha \beta |\xi|^2) \int_0^1 \cosh(\frac{\gamma \beta \hbar p \cdot \xi}{2m_\ast}) d\gamma. \tag{32}$$
Combining (31) and (32) gives

\[ V^Q(x, p, \beta) = \exp(\alpha \beta |\nabla_x|^2) \frac{2m_*}{i\beta \hbar p \cdot \nabla_x} \sinh(\frac{i\beta \hbar p \cdot \nabla_x}{2m_*}) V(x) \]

or

(a) \[ V^Q(x, p, \beta) = (2\pi)^{-3} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \Gamma(\xi, p, \beta) V(y) \exp[i\xi \cdot (x - y)] \, dy \, d\xi \] (33)

(b) \[ \Gamma(\xi, p, \beta) = \exp(-\frac{\beta \hbar^2 |\xi|^2}{8m_*}) \frac{2m_*}{\beta \hbar p \cdot \xi} \sinh(\frac{\beta \hbar p \cdot \xi}{2m_*}) \]

The quantum potential (33) is a more sophisticated version of the smoothed potential, based on wave packet analysis, presented in [27]. We note that, in general, the evaluation of smoothed potentials in the context of particle based methods is a quite expensive enterprise, since the smoothed potentials are computed as convolutions of the Coulomb and barrier potential \( V^C, V^B \) with a Gaussian type kernel, and have to be evaluated for every electron. This is therefore done by tabulating the potential and subsequent interpolation. Since the smoothed potential in (33) is given in terms of Fourier transforms, the use of fast Fourier transform algorithms is essential for this task. Alternatively, in certain device structures, one can use the effective potential \( V^Q \) only for the barrier part and take the action of the Coulomb potential to be classical, since in many applications the quantum action of the Coulomb potential is negligible [30].

We conclude this section by giving the smoothed effective quantum potential (33) for the case of a single potential step. Multiple steps and barriers can be easily combined from the single step case. We assume the potential step to occur along a plane in three dimensional space through a point \( r \) with unit normal vector \( \mathbf{n} \). So the potential gradient is given by \( \nabla_x V = V_0 \delta((x - r) \cdot \mathbf{n}) \mathbf{n} \) with \( V_0 \) the height of the step. Inserting this potential gradient into (33), we obtain

\[ \nabla_x V^Q(x + r, p, \beta) = \mathbf{n} (2\pi)^{-3} V_0 \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \Gamma(\xi, p, \beta) \delta(y \cdot \mathbf{n}) \exp[i\xi \cdot (x - y)] \, dy \, d\xi \]

\[ \nabla_x V^Q(x + r, p, \beta) = \mathbf{n} (2\pi)^{-3} V_0 \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \Gamma(\xi_0 \mathbf{n} + \xi, p, \beta) \delta(y_0) \exp[i\xi_0 (\mathbf{n} \cdot x - y_0) + i\xi \cdot (x - y)] \, dy_0 \, d\xi_0 \, d\xi \]

where we have decomposed the variables \( \xi \) and \( y \) into their components parallel \((\xi_0 \mathbf{n}, y_0 \mathbf{n})\) and orthogonal \((\xi, y)\) to \( \mathbf{n} \). The integral with respect to \( y_0 \) yields a \( \delta \)-function in \( \xi_0 \), giving

\[ \nabla_x V^Q(x + r, p, \beta) = \mathbf{n} (2\pi)^{-1} V_0 \int \Gamma(\xi_0 \mathbf{n}, p, \beta) \exp[i\xi_0 \mathbf{n} \cdot x] \, d\xi_0 \] (34)

with, according to (33)(b), \( \Gamma(\xi_0 \mathbf{n}, p, \beta) \) given by

\[ \Gamma(\xi_0 \mathbf{n}, p, \beta) = \exp(-\frac{\beta \hbar^2 \xi_0^2}{8m_*}) \frac{2m_*}{\beta \hbar \xi_0 \mathbf{n} \cdot \xi_0} \sinh(\frac{\beta \hbar \xi_0 \mathbf{n} \cdot \xi_0}{2m_*}), \quad p_0 := p \cdot \mathbf{n}. \]
Inserting this into (34) gives
\[
\nabla_x V^Q(x, r, p, \beta) = n(2\pi)^{-1}V_0 \int_0^1 \int_\mathbb{R} \exp\left(-\frac{\beta h^2 \xi_0^2}{8m^*}\right) \cosh\left(\frac{\gamma \beta p_0 \xi_0}{2m^*}\right) \exp[i \xi_0 \cdot x] \, d\xi_0 \, d\gamma = \\
n(2\pi)^{-1}V_0 \left[ \sum_{\sigma=\pm 1} \int_0^1 \int_\mathbb{R} \exp\left[-\frac{\beta h^2 \xi_0^2}{8m^*} + i \xi_0 (n \cdot x - \frac{i \sigma \gamma \beta p_0}{2m^*})\right] \, d\xi_0 \, d\gamma \right]
\]

and finally,
\[
\nabla_x V^Q(x, p, \beta) = n(2\pi)^{-1/2}V_0 \sqrt{\frac{m^*}{\beta h^2}} \sum_{\sigma=\pm 1} \int_0^1 \exp\left[-\frac{2m^*}{\beta h^2} (n \cdot (x - r) - \frac{i \sigma \gamma \beta p_0}{2m^*})^2\right] \, d\gamma . \quad (35)
\]

### 7 Quantum Hydrodynamics

Given an approximate solution to the thermal equilibrium Wigner function \(f^{eq}\) one can now proceed to find equations of state for the stress tensor \(P\) and the heat flux \(q\) in (15) in Section 3 by computing \(P\) and \(q\) from the equilibrium density \(f^{eq}\). This is done by using a functional expansion of the equilibrium Wigner function. First, we note that the solution \(f^{eq}\) of the Bloch equation (18) will always be an even function of the momentum \(p\). Therefore the odd moments of \(f^{eq}\), including the heat flux, vanish. As in the classical case, a nonzero heat flux has to be derived by using the collision operator in a Chapman - Enskog type procedure. The stress tensor \(P\) is derived by a functional expansion of the equilibrium Wigner function \(f^{eq}\) in Section 5, using either the semiclassical or the Born approximation.

#### 7.1 The semiclassical closure

In the semiclassical approximation of the equilibrium density is given in terms of an expansion of the parameter \(\alpha = \frac{\hbar^2}{8m^*}\). Consequently, we express the stress tensor \(P\) as a function of the density \(n\) of the asymptotic form
\[
P = P_0(n, \beta) + \alpha P_1(n, \beta) + O(\alpha^2) . \quad (36)
\]

Since the charge density \(n\) is itself expanded in powers of \(\alpha\), i.e. computing the zero order moment of (21) with respect to \(p\) we obtain \(n = n_0 + \alpha n_1 + O(\alpha^2)\), we have, up to terms of order \(\alpha^2\)
\[
P = P_0(n_0, \beta) + \alpha \partial_n P_0(n_0, \beta) + \alpha P_1(n_0, \beta) + O(\alpha^2) .
\]

From (21) we have
\[
\hbar^3 f^{eq} = \quad (37)
\]

18
\[
\exp\left[-\frac{\beta|p|^2}{2m_*} + \beta(\phi - eV)\right] \{1 - \beta^2 e\alpha \Delta_x V + \frac{\beta^3 e^2 \alpha}{3} |\nabla_x V|^2 + \frac{e\beta^3 \alpha}{3m_*} Tr[(\nabla_x^2 V)p p^T]\} + O(\alpha^2)
\]

Computing the integral of (37) with respect to momentum one obtains

\[
(a) \quad n = n_0 \left[1 + \frac{\alpha \beta^3 e^2}{3} |\nabla_x V|^2 - \frac{2\alpha e \beta^2}{3} \Delta_x V + O(\alpha^2)\right],
\]

(38)

\[
(b) \quad n_0 := \hbar^{-3} \int_{\mathbb{R}^3} \exp\left[-\frac{\beta|p|^2}{2m_*} + \beta(\phi - eV)\right] dp
\]

and therefore the first order term \(n_1\) is given by

\[
n_1 = n_0 \left[\frac{\beta^3 e^2}{3} |\nabla_x V|^2 - \frac{2\alpha e \beta^2}{3} \Delta_x V\right].
\]

Computing the integral in (38), we see that \(n_0\) equals a constant times \(\exp(-e\beta V)\). Therefore we can replace derivatives of \(e\beta V\) in the expression for \(n_1\) by derivatives of \(-\ln n_0\) and obtain

\[
n_1 = n_0 \left[\frac{\beta^3 e^2}{3} |\nabla_x \ln n_0|^2 + \frac{2\beta}{3} \Delta_x \ln n_0\right].
\]

Computing the stress tensor according to (16) from (37) we obtain

\[
P_{jk} = -\frac{1}{m_*} \int_{\mathbb{R}^3} p_j p_k f_{eq} dp = -n_0 \left[\delta_{jk} \left(\frac{1}{\beta} - \frac{2}{3} \alpha e \Delta_x V + \frac{\alpha \beta^2 e^2}{3} |\nabla_x V|^2 + \frac{2\alpha e \beta}{3} \partial_{jk} V\right)\right]
\]

up to terms of order \(\alpha^2\). This gives for \(P_0\)

\[
P_0(n_0, \beta) = -\frac{n_0}{\beta} I, \quad \partial_n P_0(n_0, \beta) = -\frac{1}{\beta} I,
\]

where again \(I\) denotes the identity matrix. Inserting the above into (36) yields

\[
-\frac{n_0}{\beta} I - \alpha n_0 \left[\frac{1}{3} |\nabla_x \ln n_0|^2 + \frac{2}{3} \Delta_x \ln n_0\right] I + \alpha P_1(n_0, \beta) =
\]

\[
-2n_0 \left[\frac{1}{\beta} + \frac{2}{3} \alpha \Delta_x \ln n_0 + \frac{\alpha}{3} |\nabla_x \ln n_0|^2\right] - \frac{2\alpha}{3} \nabla^2 \ln n_0,
\]

which yields for the functional form of \(P_1\)

\[
P_1(n, \beta) = \frac{2n}{3} \nabla^2 \ln n,
\]

or, altogether

\[
P(n, \beta) = -nK T I + \frac{n\hbar^2}{12m_*} \nabla^2 \ln n, \quad K T := \frac{1}{\beta}
\]

(39)
7.2 Smoothed Potential QHD based on the Born Approximation

The same procedure, namely using the asymptotic expression for the thermal equilibrium Wigner function $f^{eq}$ to derive a functional expansion of the stress tensor $P$ can be applied to the asymptotic form (29) of $f^{eq}$, derived from the Born approximation. However, since the zero order term of $f^{eq}$ in (29) does not depend on the potential, the potential cannot in zero’th order be expressed in terms of the charge density $n$. Therefore, the functional form of the stress tensor has to depend explicitly on the potential $V$ in the Born approximation. Thus we make the ansatz

$$P = P_0(n, \beta) + \varepsilon P_1(n, \beta, V) + O(\varepsilon^2)$$

where again, in the same spirit as in Section 5.2, we write formally $V = \varepsilon V$. Expanding the charge density $n$ in powers of $\varepsilon$ again gives, as before

$$P = P_0(n_0, \beta) + \varepsilon \partial_n P_0(n_0, \beta) n_1 + \varepsilon P_1(n_0, V) + O(\varepsilon^2), \quad n = n_0 + \varepsilon n_1 + O(\varepsilon^2) \quad (40)$$

From (29) we obtain

$$n_0 = \hbar^{-3} \exp(\beta \phi) \int_{\mathbb{R}^3} \exp\left(-\frac{\beta |p|^2}{2m_*}\right) dp, \quad n_1 = -\hbar^{-3} \exp(\beta \phi) \int_{\mathbb{R}^3} S(-i\nabla_x, p, \beta) V_\varepsilon(x) \, dp$$

Using the definition (28) of the operator $S$, we obtain

$$n_1 = -n_0 \Gamma^0(-i\nabla_x, \beta)V_\varepsilon, \quad \Gamma^0(\xi, \beta) = e^\beta \int_0^1 \exp[\alpha \beta (\gamma^2 - 1)|\xi|^2] \, d\gamma, \quad \alpha = \frac{\hbar^2}{8m_*}.$$

Computing now the stress tensor from (29) gives

$$P = -\frac{n_0}{\beta} I + \varepsilon n_0 \Gamma^2(-i\nabla_x, \beta)V_\varepsilon$$

with the symbol of $\Gamma^2$ given by

$$n_0 \Gamma^2_{jk}(\xi, \beta) = \frac{1}{m_*} \hbar^{-3} e^{\beta \phi} \int_{\mathbb{R}^3} S(\xi, p, \beta)p_j p_k \, dp$$

Shifting variables $p \rightarrow p - \frac{\sigma \hbar}{2} \xi$ in the definition (28) of $S$ and computing the integrals gives

$$n_0 \Gamma^2_{jk}(\xi, \beta) = \frac{e^\beta}{m_*} \hbar^{-3} e^{\beta \phi} \int_0^1 \int_{\mathbb{R}^3} \exp[\alpha \beta (\gamma^2 - 1)|\xi|^2] - \frac{\beta}{2m_*} |p|^2 (p_j p_k + \frac{\hbar^2 \gamma^2}{4} |\xi_j \xi_k|) \, dp \, d\gamma$$

or

$$\Gamma^2_{jk}(\xi, \beta) = \frac{e^\beta}{m_*} \int_0^1 \exp[\alpha \beta (\gamma^2 - 1)|\xi|^2] (\frac{m_*}{\beta} \delta_{jk} + \frac{\hbar^2 \gamma^2}{4} |\xi_j \xi_k|) \, d\gamma$$

$$\Gamma^2_{jk}(\xi, \beta) = \frac{1}{\beta} \Gamma^{00}(\xi, \beta) \delta_{jk} + \frac{e^\beta}{m_*} \int_0^1 \exp[\alpha \beta (\gamma^2 - 1)|\xi|^2] \frac{\hbar^2 \gamma^2}{4} \xi_j \xi_k \, d\gamma$$
This gives for the functional expression $P_0$

$$P_0(n_0, \beta) = -\frac{n_0}{\beta} I, \quad \partial_n P_0(n_0, \beta) = -\frac{1}{\beta} I$$

(41)

and (40) becomes

$$-\frac{n_0}{\beta} I + \varepsilon_n \frac{1}{\beta} \Gamma_0(\xi, \beta) V \varepsilon_I - \varepsilon_n \frac{e\beta}{m^*} \int_0^1 \exp[\alpha\beta(1 - \gamma^2)|\nabla_x|^2] \frac{\hbar^2 \gamma^2}{4} \, d\gamma \nabla_x^2 V \varepsilon =$$

$$-\frac{n_0}{\beta} I + \varepsilon_n \frac{1}{\beta} \Gamma_0(-i\nabla_x, \beta) V \varepsilon_I + \varepsilon P_1(n_0, \beta, V \varepsilon)$$

$$P_1(n_0, \beta, V \varepsilon) = -n_0 \frac{e\beta \hbar^2}{4m^*} \int_0^1 \gamma^2 \exp[\alpha\beta(1 - \gamma^2)|\nabla_x|^2] \, d\gamma \nabla_x^2 V \varepsilon$$

(42)

Thus, setting $\varepsilon V \varepsilon = V$ and combining (41) and (42), we obtain for the stress tensor the approximation

(a) $P(n, \beta, V) = -nKT I - n - \frac{e\hbar^2}{4m^*KT} \nabla^2 \tilde{V}$

(b) $\tilde{V}(x) = \int_0^1 \gamma^2 \exp[\frac{\hbar^2 \beta}{8m^*}(1 - \gamma^2)|\nabla_x|^2] \, d\gamma V(x) =$

$$(2\pi)^{-3} \int_0^1 \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \gamma^2 \exp[\frac{\hbar^2 \beta}{8m^*}(\gamma^2 - 1)|\xi|^2] V(y) \exp[i\xi \cdot (x - y)] \, dyd\xi d\gamma$$

Remark:
We should point out the similarities between the different hydrodynamic models arising from the Bohm potential, and thermodynamic closures produced by the semiclassical and the Born approximation. The form of the pressure tensor obtained from the semiclassical approximation of the thermal equilibrium state (39) is essentially the same as that obtained from the combination of the Bohm potential with the nonlinear Schrödinger equation (10) in Section 3.1, except for a factor 3 in the quantum correction to the stress tensor. Again the same form is obtained in the smoothed potential closure (43), except that the density $n$ is replaced by $\exp(-\beta e\tilde{V})$ in the quantum correction to the stress tensor. This represents a very significant difference. First, the semiclassical approximation relied on Taylor expanding the potential which, in the presence of barriers, is a discontinuous function, although the higher derivatives of the potential do not appear in the final result. The derivation of the smoothed potential formula (43) does not rely on differentiating the potential. On the other hand the type of the resulting equations is quite different. The presence of higher order derivatives of the density $n$ in (10) and (39) results in the quantum hydrodynamic equations becoming dispersive, while (43) is strictly hyperbolic.
8 Applications

8.1 Effective potentials in short channel MOSFETS

For quite some time, the dimensions of semiconductor devices have been scaled aggressively in order to meet the demands of reduced cost per function on a chip used in modern integrated circuits. There are some problems associated with device scaling, however. Critical dimensions, such as transistor gate length and oxide thickness, are reaching physical limitations. Considering the manufacturing issues, photolithography becomes difficult as the feature sizes approach the wavelength of ultraviolet light. In addition, it is difficult to control the oxide thickness when the oxide is made up of just a few monolayers. In addition to the processing issues, there are also some fundamental device issues. As the oxide thickness becomes very thin, the gate leakage current due to tunneling increases drastically. This significantly affects the power requirements of the chip and the oxide reliability. Short-channel effects (SCEs), such as drain-induced barrier lowering (DIBL) and the Early effect in bipolar junction transistors (BJTs), degrade the device performance. Hot carriers also degrade device reliability. The control of the density and location of the dopant atoms in the MOSFET channel and source/drain region to provide a high on-off-current ratio becomes a problem in the 50 nm node devices where there are no more than 100 impurity atoms in the device active region. Yet another issue that becomes prominent in nano-scale MOSFET devices is the quantum-mechanical nature of the charge description which, in turn, gives rise to inversion layer capacitance comparable to the oxide capacitance. This, in turn, degrades the total gate capacitance and, therefore, the device transconductance.

To capture the role of the quantum-mechanical size-quantization effects, we have applied the effective potential approach discussed in Section 6 in conjunction with a Monte Carlo particle-based simulation scheme. The Monte Carlo model, used in the transport portion of the simulator, is based on the usual Si band-structure for three-dimensional electrons in a set of non-parabolic D-valleys with energy-dependent effective masses. The six conduction band valleys are included through three pairs: valley pair 1 pointing in the (100) direction, valley pair 2 in the (010) direction, and valley pair 3 in the (001) direction. The explicit inclusion of the longitudinal and transverse masses is important and this is done in the program using the Herring-Vogt transformation [21]. Intravalley scattering is limited to acoustic phonons. For the intervalley scattering, we include both g- and f-phonon processes. It is important to note that, by group symmetry considerations, the zeroth-order low-energy f-and g-phonon processes are forbidden. Nevertheless, three zeroth-order f-phonons and three zeroth-order g-phonons with various energies are usually assumed [22]. We have taken into account this selection rule and have considered two high-energy f- and g-phonons and two low-energy f- and g-phonons. The high-energy phonon scattering processes are included via the usual zeroth-order interaction term, and the two low-energy phonons are treated via a first-order process [9]. The first-order process is not really important for low-energy electrons.
but gives a significant contribution for high-energy electrons. The low-energy phonons are important in achieving a smooth velocity saturation curve, especially at low temperatures. The phonon energies and coupling constants in our model are determined so that the experimental temperature-dependent mobility and velocity-field characteristics are consistently recovered [19]. At present, impact ionization is not included in the model as it does not affect the results discussed here.

In solving Poisson’s equation, the Monte Carlo simulation is used to obtain charge distribution in the device. We use the Incomplete Lower-Upper decomposition method for the solution of the 3D Poisson equation. After solving for the potential, the electric fields are calculated and used in the free-flight portion of the Monte Carlo transport kernel. The field due to the barrier potential, discussed in Section 6, is precomputed at the beginning of the simulation and added to the field arising from the classical Hartree potential. Regarding the charges obtained from the EMC simulation, they are usually distributed within the continuous mesh cell instead of on the discrete grid points. The particle mesh method (PM) is used to perform the switch between the continuum in a cell and discrete grid points at the corners of the cell. The procedures of PM coupling is outlined below:

1. Assign the charges from within the continuous mesh cell onto the discrete grid points.
2. Solve Poisson equation for potentials at those points.
3. Calculate electric fields at those points from the potential profiles and obtain the electric fields for particles within the cell. Add the barrier field to account for quantum-mechanical size quantization effects.

The charge assignment to each mesh-point depends on the particular scheme that is used. A proper scheme must ensure proper coupling between the charged particles and the Coulomb forces acting on the particles. Therefore, the charge assignment scheme must maintain zero self-forces and a good spatial accuracy of the forces. To achieve this, two major methods have been implemented in the present version of the code: the Nearest-Grid-Point (NGP) scheme and the Cloud-in-Cell (CIC) scheme. The cloud-in-cell scheme (CIC) produces a smoother force interpolation, but introduces self-forces on non-uniform meshes. These issues have been dealt with extensively by Hockney and Eastwood [20], and quite recently by Laux [23].

The device current is determined by using two different, but consistent methods. First, by keeping track of the charges entering and exiting each terminal, the net number of charges over a period of the simulation can be used to calculate the terminal current. The method is quite noisy, due to the discrete nature of the electrons. In the second method, the sum of the electron velocities in a portion of the device is used to calculate the current. For this purpose, the device is divided into several sections along the x-axis (along the semiconductor/oxide
interface). The number of electrons, and their corresponding velocities, are added up for each section after each free-flight time step. The total x-velocity in each section is then averaged over several time steps to determine the current for that section. The total device current is determined from the average of several sections, which gives a much smoother result than that based on counting the terminal charges. By breaking the device into sections, individual section currents can be compared to verify that the currents are uniform. In addition, sections near the source and drain regions may have a high y- and z-components in their velocity and should be excluded from the current calculations. Finally, by using several sections in the channel, the average energy and velocity of electrons along the channel is checked to ensure proper physical characteristics.

A snapshot of the electron distribution in the channel for a 25 nm gate-length MOSFET device with substrate doping density of 21019 cm$^{-3}$, oxide thickness of 1.5 nm and doping of the source and drain regions of 1019 cm$^{-3}$, are shown in Figure 1. We use $V_G=V_D = 1.2$ V in these simulations. There are several noteworthy features that can be deduced from these simulation results. First, the inclusion of the barrier potential gives rise to non-uniform channel. Second, the electrons are displaced from the interface proper, the displacement being the largest for the low-energy electrons. Finally, the average electron energy is increased by about 0.2 eV, thus accounting for the so-called band-gap widening effect. The current-voltage characteristics of this device structure are shown in Figure 2. We find more than 40% decrease in the current when quantum-mechanical size-quantization effects are included in the model. This is a consequence of the threshold voltage reduction due to the degradation of the total gate capacitance. Note that similar reduction in the drain current is obtained when surface-roughness and Coulomb scattering are included in the model.

## 8.2 Smooth QHD Simulation of the Resonant Tunneling Diode

We present simulations of a GaAs resonant tunneling diode with Al$_{0.3}$Ga$_{0.7}$As double barriers at 300 K. The barrier height is equal to 280 meV. The diode consists of $n^+$ source (at the left) and drain (at the right) regions with the doping density $N = 10^{18}$ cm$^{-3}$, and an $n$ channel with $N = 5 \times 10^{15}$ cm$^{-3}$. The channel is 20nm long, the barriers are 2.5 nm wide, and the quantum well between the barriers is 5 nm wide. Note that the device has 5 nm spacers between the barriers and the contacts.

In Fig. 3, the experimental signal of quantum resonance—negative differential resistance, a region of the current-voltage curve where the current decreases as the applied voltage is increased—is displayed. Simulations have demonstrated that it is difficult to achieve NDR at 300 K with the $O(\hbar^2)$ QHD model, while NDR is experimentally observed at 300 K and is predicted by the smooth QHD simulations.
Figure 1: Snapshot of electron distribution.
Figure 2: I-V Characteristics.

Figure 3: Current density in kiloamps/cm$^2$ vs. voltage using the smooth QHD model for the RTD at 300 K. The barrier height is 280 meV.

Figure 4: Smooth effective potential $U$ in eV for applied voltages between 0 and 400 millivolts for 280 meV double barriers at 300 K. $x$ is in 5 Angstrom.
The smooth QHD model involves a smoothing of the classical potential over both space and temperature—the extra degree of smoothing is just enough to insure that the gradient of the smooth effective potential in the momentum transport equation is continuous. The smooth QHD equations contain at worst a step function discontinuity. To see this, we define the 1D smooth effective potential in the momentum conservation equation as the most singular part of $V - F_{11}$:

$$U = V + \frac{\hbar^2}{4mT_0} \frac{d^2V}{dx^2}.$$  

(44)

As can be proved using Fourier transforms, the smooth effective potential is smoother by two degrees than the classical potential $V$; i.e., if $V$ has a discontinuity, then $U$ is once differentiable.

Fig. 4 illustrates the smooth effective potential $U$ for the resonant tunneling diode for applied voltages 0–400 millivolts. At zero applied voltage, the 280 meV barrier height is effectively reduced to 120 meV by the smoothing in the effective potential. The effective quantum well is parabolic near its minimum.

The resonant peak of the current-voltage curve occurs as the electrons tunneling through the first barrier come into resonance with the energy levels of the quantum well. The resonant peak in Fig. 3 occurs at 150 millivolts. As the voltage bias increases above 150 millivolts, the resonance effect rapidly decreases because the right barrier height is progressively reduced and electrons tunnel out of the well through the thin portion of the effective parabolic well.

9 Conclusions

Particle based simulations, using effective potentials, and quantum fluid models can be an effective tool to include quantum effects into the semiclassical transport picture. Smoothed potential approaches are better able to capture the nonlocal interaction of electrons with potential barriers than effective potential approaches based on gradient expansions alone, such as the Bohm potential. While the implementation of smooth quantum potentials is a non-trivial numerical problem, it is still considerably less expensive than solving full quantum kinetic equations in more than one spatial dimension. As a practical matter, the use of effective potential allows for the merger of quantum transport phenomena with semiclassical collision mechanisms via a moderate modification in existing Monte Carlo simulators.
References


