Transport models for semiconductors

Dr. Nicola Zamponi
Vienna University of Technology

Summer term 2015

Lecture notes from a lecture series at Vienna University of Technology

Contents

1 Introduction

1.1 Aim of the course

The main goal of this course is to derive several models (in practice, PDEs) describing the electron flow through a semiconductor device due to an applied voltage. The main transport phenomena, that have to be accounted for in the derivation process and which depend in general on the particular device considered, are:

- *diffusion*: namely, the tendency of a distribution of some physical quantiity to spread in space until it reaches an “equilbrium” configuration;
- *drift*: i.e. the tendency of the same distribution to translate in space due to some force field;
- *scattering*: interaction between two or more particles, or between particles and atoms;

- *quantum mechanical phenomena*: electrons are quantum mechanical objects and do not obey to the rules of classical mechanics, so in many cases the formalism of quantum mechanics will be unavoidable, althought in some situations it is possible to describe quantum systems with a so-called *semi-classic* formalism.

In order to choose the right model for a given device, it is necessary to pay attention to some key physical parameters, namely:

- *device size*: tipically the device length is of the order of magnitude of 100 nm;
- *number of free electrons*: usually of the order of magnitude of 10^4;
- *mean free time*: the average time interval between two consecutive collisions experienced by a particle;
- *ambient temperature*: usually around 300 K.

The transport models that will be derived in this course will be (systems of) nonlinear PDEs, which will not be solvable with analytical methods; the only possible way to solve such equations will be through numerical methods. For this reason, another variable that will have to be considered throughout the course is the computational effort required to find a numerical solution for the considered models. The ideal compass of the modeling activity will be to compromise between accuracy in the description of the relevant physical phenomena and simplicity of the models.

We warn the attentive reader that the arguments employed in the derivation of the model equations in this course will be usually purely formal. Even if we will write many times the well-known mathematical keywords “Theorems”, “Proposition”, “Lemma” and so on throughout these lecture notes, they will not have the same meaning that they have in Analysis courses; for example we will not specify, in general, the regularity assumptions on the involved functions, nor the precise notions of convergence when limits of sequences of functions are taken, nor the exact functional setting of the considered problems; as a matter of fact, rigorous proofs of many of the results stated in these lecture notes are nowhere to be found in literature. Such is, after all, the spirit of Mathematical Modeling.

1.2 Preliminary considerations

First of all: what is a semiconductor?
We give two definitions of a semiconductor: a “macroscopic” one and a “microscopic” one. Let us start with the macroscopic definition.

Definition 1. A semiconductor is a solid material with a conductivity between 10^{-6} and 10^5 S/m.

To understand the above definition, it is necessary to know what the conductivity of a material is. According to Ohm’s law, the magnitude I of the electric current flowing through a body subject to a voltage V is given by $I = V/R$, where R is the electric resistance of the body. Moreover, it holds $R = r\ell/A$, where A is the section of the body, ℓ is the length of the body, and r is an intrinsic property of the material composing the body which is called *resistivity.* The reciprocal $1/r$ of the resistivity is the *conductivity.*

Definition 1 actually states that a semiconductor is a material with a conductivity higher than the one of an insulator and lower than the one of a metal (this is, after all, the reason of the name "semiconductor"). The "microscopic" definition is based instead upon the concept of *band gap*, which is related to the intimate electronic properties of the material.

Definition 2. A semiconductor is a solid material with a band gap greater than zero and lower than 4 eV.

The notion of band gap will be explained to the reader in the next sections. However, we point out that a band gap equal to zero is typical of metals, while a high band gap yields an insulator.

So, now that it is more or less clear what a semiconductor is, the next question is: how to describe the electron flow through a semiconductor?
To have an intuitive idea of the answer, let us just consider a very simple situation: let us describe a single electron as a *classical* particle with mass m, position x, velocity v, subject to an electric force $F = -q\nabla V(x)$ where q is the charge of the electron and V is the electric potential. According to classical mechanics, the system obeys *Newton's law*: $m\ddot{x} = F$ or, equivalently, *Hamilton's equations*:

$$\dot{x} = v, \qquad \dot{v} = \frac{F}{m} = -\frac{q}{m}\nabla V(x). \tag{1}$$

Eq. (1) is an example of *microscopic* model, that is, a model that is focused on single particles. This approach would require us to solve eq. (1) for each electron. Since there are in average 10^4 free electrons in a semiconductor device, the numerical solution of such a system would be very expensive. Thus we search for alternative models. Our key idea is that we are interested in describing the particle *ensamble* rather than a single particle; for this reason, it appear meaningful to employ a *statistical* description of the system. We define the *distribution function* of the system as the function $f : \mathbb{R}^3 \times \mathbb{R}^3 \times [0, \infty)$, $f = f(x, p, t)$ depending on position $x \in \mathbb{R}^3$, momentum $p \in \mathbb{R}^3$, time $t > 0$ representing the particle density of the system in the phase space at time t; more precisely,

$$\forall \Omega \subset \mathbb{R}^3 \times \mathbb{R}^3, \qquad \iint_\Omega f(x,p,t)dxdp \; = \text{number of electrons with position } x \text{ and momentum } p \text{ satisfying } (x,p) \in \Omega \text{ at time } t.$$

In particular $\iint_{\mathbb{R}^3\times\mathbb{R}^3} f(x,p,t)dxdp$ is the total number of electrons of the ensamble.

Let us assume that the initial distribution $f_I = f(\cdot, \cdot, 0)$ of the system is known (it's quite a big assumption, actually). We want to write an evolution equation for f. The starting point is Liouville's Theorem: since the divergence of the right-hand side of (1) is zero (namely: $\nabla_x \cdot v + \nabla_v \cdot F(x)/m = 0$ due to the force F being positional), then the distribution function f of the system is constant along trajectories:

$$f(x(t), v(t), t) = f_I(x(0), v(0)) \quad t > 0, \qquad \forall (x(t), v(t)) \text{ solution of (1).} \tag{2}$$

Let us differentiate both sides of (2) with respect to time. We get:

$$0 = \frac{d}{dt} f(x(t), v(t), t) = \partial_t f + \dot{x} \cdot \nabla_x f + \dot{v} \cdot \nabla_v f, \qquad t > 0. \tag{3}$$

By applying (1) we conclude:

$$\partial_t f + v \cdot \nabla_x f - \frac{q}{m} \nabla V \cdot \nabla_v f = 0 \qquad x \in \mathbb{R}^3, \quad v \in \mathbb{R}^3, \quad t > 0. \tag{4}$$

Equation (4) is the *Liouville equation.* It provides a *mesoscopic* description of the system, i.e. a description involving a distribution in the phase space; such a model can be regarded as intermediate between microscopic and macroscopic models.

Equation (4) must be considered together with the initial condition $f(\cdot, \cdot, 0))f_I$. It is a scalar PDE for a function in the phase space, which means, a 6-dimensional space; this means that the numerical solution of such a problem is quite expensive. Moreover, in spite of the assumption we made, the initial datum f_I is known, this is usually not true in real situations. Finally, the distribution f contains much more informations than we need: in fact, we are not interested in knowing the distribution function of the system at each point of the phase space, but rather we are interested in only some macroscopic (i.e. *measurable*) quantities, mainly the *particle density* $n(x, t) = \int_{\mathbb{R}^3} f(x, v, t) dv$ and the *current density* $J(x, t) = \int_{\mathbb{R}^3} v f(x, v, t) dv$. So, we wish to write equations involving only these quantities (which are called *moments* of the distribution function). Let us start by integrating the Liouville equation (4) in $\mathbb{R}^3$ with respect to v. By keeping in mind the definitions of n and J and recalling that the potential V does not depend on v, we obtain:

$$\partial_t n + \nabla_x \cdot J - \frac{q}{m} \nabla V \cdot \int_{\mathbb{R}^3} \nabla_v f dv = 0. \tag{5}$$

If f tends to 0 quick enough as $|v| \to \infty$, the divergence theorem yields $\int_{\mathbb{R}^3} \nabla_v f dv = 0$. Thus (5) becomes:

$$\partial_t n + \nabla_x \cdot J = 0. \tag{6}$$

This is the *continuity equation.* It is a balance of mass; it states that, given an arbitrary domain $\Omega \subset \mathbb{R}^3$, the change per unit time of the particle number inside Ω equals the flow of incoming particles minus the flow of outgoing particles.[1]

---

[1] To see this, just integrate (6) in Ω and apply the divergence theorem (ν is the outward normal to $\partial\Omega$):

$$\partial_t \int_\Omega n dx = - \int_{\partial\Omega} J \cdot \nu d\sigma = J^{incoming} - J^{outgoing},$$

with $J^{incoming} = - \int_{\partial\Omega} \min\{J \cdot \nu, 0\} d\sigma$, $J^{outgoing} = \int_{\partial\Omega} \max\{J \cdot \nu, 0\} d\sigma$.

Now we face a problem, that we will encounter again many times in the course: the so-called *closure problem.* Equation (6) is not closed, namely it is not possible to write J as a function of n without making further assumptions. The reason lies in the definition of n, J in terms of f: since we have almost no information on f, we can't possibly hope to find an explicit relation involving only n and J unless we make some ad-hoc hypothesis. We will see in the rest of the course the exact nature of such assumptions; for now we just state that in some suitable situation we can write $J = -(\nabla n - n\nabla V)$. By plugging this expression inside (6) we get the well-known *drift-diffusion equation*:

$$\partial_t n = \operatorname{div}(\nabla n - n\nabla V). \tag{7}$$

The potential V can be considered as a given function (in this case (7) is linear), or be self-consistently defined by the Poisson equation:

$$\lambda_D^2 \Delta V = n, \tag{8}$$

where $\lambda_D > 0$ is a constant parameter called *Debye length.*

Eqs. (7), (8) are the simplest model for electron transport in semiconductor and, for this reason, also the most used model in semiconductor industry. This success comes in spite of the physical limitations of the model; in fact, the Poisson-drift-diffusion equations provide an accurate description of the system only if:

- the size of the device is at least 1000 nm;
- the system is close to thermal equilibrium, meaning that the current density should be small and the temperature constant;
- the applied voltage is small.

Of course, the drift-diffusion equation (7) is not the only model for electron transport that can be derived. Another model can be obtained by retaining the continuity equation (6) and finding an equation for J. With this goal in mind, let us multiply eq. (4) times v and integrate it in $\mathbb{R}^3$ with respect to v:[2]

$$\partial_t J + \nabla_x \cdot \int_{\mathbb{R}^3} v \otimes v f dv - \frac{q}{m} \int_{\mathbb{R}^3} v(\nabla_x V \cdot \nabla_v) f dv = 0. \tag{9}$$

Let us consider a generic component $i \in \{1, 2, 3\}$ of the third integral in (9). The divergence theorem allows to write:

$$\int_{\mathbb{R}^3} v_i (\nabla_x V \cdot \nabla_v) f dv = \sum_{j=1}^{3} \partial_{x_j} V \int_{\mathbb{R}^3} v_i \partial_{v_j} f dv = -\sum_{j=1}^{3} \partial_{x_j} V \int_{\mathbb{R}^3} \delta_{ij} f dv = -n \partial_{x_i} V. \tag{10}$$

---

[2] In the equation which follows, the symbol $\otimes$ denotes the *tensor product* between vectors: if $a \in \mathbb{R}^N$, $b \in \mathbb{R}^M$, then $a \otimes b \in \mathbb{R}^{N\times M}$ with components $(a \otimes b)_{ij} = a_i b_j$.

However, we can't compute the other integral, namely $\int_{\mathbb{R}^3} v \otimes v f dv$: we face again the closure problem, and we are forced to make some assumption in order to express such term as a function of n, J. We will see later on that in suitable circumstances it holds $\int_{\mathbb{R}^3} v \otimes v f dv = J \otimes J/n + nI$. By plugging this relation into (9) and exploiting (10) we conclude:

$$\partial_t J + \nabla \cdot \left(\frac{J \otimes J}{n} \right) + \nabla n + \frac{q}{m} n \nabla V = 0. \tag{11}$$

Eqs. (6), (11) are the *Euler (or hydrodynamic) equations* known in gas dynamics. The term $\nabla \cdot \left(\frac{J\otimes J}{n}\right)$ is the convection; the pressure of the system equals n; $\frac{q}{m} n \nabla V$ is a force term.

The drift-diffusion equation and the Euler equations are examples of *macroscopic* transport models, that is, models involving macroscopic, observable quantities. In the rest of the course we will explain precisely how the models we introduced here can be derived, and how the quantum mechanical nature of the electrons can be taken into account.

2 Basic Physical Background

2.1 The cristal structure of solids

It is not possible to describe electron transport in semiconductors without having at least some basic knowledge about the structure of a semiconductor. Thus, in this subsection we present briefly the cristal structure of solids. For more details about the subject see [6].

An ideal solid is, by definition, an infinite 3-dimensional array of atoms disposed in a lattice $L \subset \mathbb{R}^3$, which is called *Bravais lattice*:

$$L = \{ n_1 a_1 + n_2 a_2 + n_3 a_3 \ : \ n_1, n_2, n_3 \in \mathbb{Z} \}.$$

The given (linearly independent) vectors $a_1, a_2, a_3 \in \mathbb{R}^3$ are called *primitive vectors* of the lattice.

Another concept, that is rather useful in solid state physics, is the concept of *reciprocal lattice*, which is the lattice $L^* \subset \mathbb{R}^3$ defined as:

$$L^* = \{ n_1 a_1^* + n_2 a_2^* + n_3 a_3^* \ : \ n_1, n_2, n_3 \in \mathbb{Z} \}, \qquad a_i^* \cdot a_j = 2\pi \delta_{ij} \qquad (i,j = 1,2,3). \tag{12}$$

The primitive vectors a_1^*, a_2^*, a_3^* of the reciprocal lattice L^* can be easily computed from definition (12):

$$a_1^* = \frac{2\pi a_2 \times a_3}{a_1 \cdot (a_2 \times a_3)}, \qquad a_2^* = \frac{2\pi a_3 \times a_1}{a_1 \cdot (a_2 \times a_3)}, \qquad a_3^* = \frac{2\pi a_1 \times a_2}{a_1 \cdot (a_2 \times a_3)}. \tag{13}$$

A *primitive cell* of the Bravais lattice L is any connected set $D \subset \mathbb{R}^3$ having volume equal to $|a_1 \cdot (a_2 \times a_3)|$ and with the property $\cup_{x \in L} (D + x) = \mathbb{R}^3$. Notice that $|a_1 \cdot (a_2 \times a_3)|$

is the volume of the parallelepiped with edges a_1, a_2, a_3. The notion of primitive cell for the reciprocal lattice L^* is defined in a completely analogous way.

The primitive cell of L (or L^*) that is closest, among all the primitive cells, to the origin of the chosen reference frame is called *Wigner-Seitz cell.* The Wigner-Seitz cell of the reciprocal lattice is referred to as *first Brillouin zone.*

The vectors of L and L^* can be seen as conjugate variables, like e.g. time and frequency. In fact, let $x \in L$ and $k \in L^*$. Clearly $x = \sum_{i=1}^{3} n_i a_i$ and $k = \sum_{i=1}^{3} m_i a_i^*$ for some integer coefficients $n_1, n_2, n_3, m_1, m_2, m_3$. Thus:

$$e^{ik\cdot x} = \exp\left(i \sum_{i,j=1}^{3} n_i m_j a_i \cdot a_j^* \right) = \exp\left(i \sum_{i,j=1}^{3} n_i m_j \, 2\pi \delta_{ij} \right) = e^{2\pi N i},$$

with $N = \sum_{i=1}^{3} n_i m_i \in \mathbb{Z}$. So $e^{ik\cdot x} = 1$. We point out that x, being a vector in the physical space, has the dimension of a length, while the relation $e^{ik\cdot x} = 1$ clearly implies that k has the dimension of an inverse length; for this reason, the vectors in the reciprocal lattice are called *wavevectors.*

The reciprocal lattice is not merely a mathematical construct, but it has a nice physical interpretation, which is related to the phenomenon of X-ray diffraction in solids. A common way to explore the cristal structure of solids is the analysis of the reflection pattern of X-ray radiation scattered by some solid body. Such reflection patterns are quite different from the ones produced by fluids: in fact, in cristalline materials, intense peaks of scattered radiation (*Bragg peaks*) are observed in correspondence of sharply defined wavelengths and incident directions. These peaks can be accounted for by exploiting the cristal structure of solids: it is possible to show that *a Bragg peak is obtained if and only if the difference between the wavevector of the reflected radiation and the wavevector of the incident radiation is a vector of the reciprocal lattice.*

2.2 The postulates of Quantum Mechanics

Electrons are quantum mechanical objects; it is therefore meaningful to give a short overview of quantum mechanics before addressing the problem of electron transport in semiconductors. For more details about the subject see e.g. [51].

In the following we consider a relatively simple quantum mechanical system, that is, a single quantum particle in the usual 3-dimensional space; however, the considerations below extend naturally to more complex situations.

1 The state of the system is represented by a normalized vector ψ of the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^3, \mathbb{C})$. The vector ψ is called the *wavefunction* of the system. The quantity $|\psi|^2$ represent the *probability density* associated to the system; namely, for all $\Omega \subset \mathbb{R}^3$, the integral $\int_\Omega |\psi|^2 dx$ is the probability of finding the particle in Ω.

2 The observables of the system are represented by linear self-adjoint operators acting on the Hilbert space $\mathcal{H}$. The fundamental observables position and momentum are

given respectively by: $X\psi \equiv x\psi$, $P\psi = -i\hbar\nabla_x\psi$ for all $\psi \in \mathcal{H}$, where $\hbar \approx 10^{-34}\, J{\cdot}s$ is the Dirac constant. Any classical observable $\gamma = \gamma(x,p)$ has a quantum counterparts intuitively defined by $\gamma(X,P) = \gamma(x, -i\hbar\nabla_x)$.

3 If ψ is the state of the system, then the result of a measurement of a quantum observable is an element of the spectrum $\sigma(A)$ of the quantum operator A associated to the observable. For every $\lambda \in \sigma(A)$, the probability of obtaining λ as a result of the measurement equals $\|\mathbb{P}_{A,\lambda}\psi\|^2_{\mathcal{H}}$, where $\mathbb{P}_{A,\lambda} : \mathcal{H} \to \mathcal{H}$ is the projection operator onto the eigenspace associated to λ. *The state of the system changes due to the measurement;* after the measurement, the new state of the system is $\psi' = \mathbb{P}_{A,\lambda}\psi / \|\mathbb{P}_{A,\lambda}\psi\|_{\mathcal{H}}$.

4 The evolution of the system is described by the *Schrödinger equation*:

$$i\hbar\partial_t\psi = H\psi, \qquad x \in \mathbb{R}^3,\ t > 0,$$

where ∂_t is the partial derivative with respect to time and H is the Hamiltonian operator. If the particle has mass m and is subject to a force $F = -\nabla V(x)$, then $H = \frac{P^2}{2m} + V(X) = -\frac{\hbar^2}{2m}\Delta + V(x)$.

A few remarks about the postulates stated above.

While classical systems are deterministically described, the description of a quantum system is instead probabilistic. The outcome of a measurement of some quantum observable A is, rather then a number, a random variable with density [3] $\|\mathbb{P}_{A,\lambda}\psi\|^2_{\mathcal{H}} = (\psi, \mathbb{P}_{A,\lambda}\psi)$, with ψ being the state of the system.

The mathematical expectation of a measurement of A is $\mathbb{E}[A;\psi] = (\psi, A\psi)_{\mathcal{H}}$. This formula can be motivated, in the case of discrete spectrum $\sigma(A) = \{\lambda_k \ :\ k \in \mathbb{N}\}$, in the following way:

$$\mathbb{E}[A;\psi] = \sum_{k\in\mathbb{N}} \lambda_k(\psi, \mathbb{P}_{A,\lambda_k}\psi)_{\mathcal{H}} = \left(\psi, \sum_{k\in\mathbb{N}} \lambda_k \mathbb{P}_{A,\lambda_k}\psi\right)_{\mathcal{H}} = (\psi, A\psi)_{\mathcal{H}}.$$

Notice that, since self-adjoint operators have real spectrum, then Postulate 3 is coherent with the fact that the result of a measurement of any observable is always a real number.

While in classical mechanics it is possible (in principle) to measure a system without perturbing it, in quantum mechanics this is not possible. Every measurement of any observable A alters the state of the system in such a way that, if λ is the outcome of the measurement and V_λ is the eigenspace of A associated to λ, then the wavefunction ψ loses its components orthogonal to V_λ (*collapse of the wavefunction*). Thus it is not possible to know the state of a quantum system before a measurement, but only after it.

---

[3] Since $\mathbb{P}_{A,\lambda}$ is a projection operator then $\mathbb{P}^2_{A,\lambda} = \mathbb{P}_{A,\lambda}$. Being $\mathbb{P}_{A,\lambda}$ also self-adjoint then:

$$\|\mathbb{P}_{A,\lambda}\psi\|^2_{\mathcal{H}} = (\mathbb{P}_{A,\lambda}\psi, \mathbb{P}_{A,\lambda}\psi)_{\mathcal{H}} = (\psi, \mathbb{P}^2_{A,\lambda}\psi)_{\mathcal{H}} = (\psi, \mathbb{P}_{A,\lambda}\psi)_{\mathcal{H}}.$$

While in classical mechanics it is always possible (in theory) to measure two different observables simultaneously with arbitrarily high precision, this is not the case in quantum mechanics. Analytically speaking, this is a consequence of the noncommutativity of the composition of linear operators on Hilbert spaces: given two linear operators $A, B : \mathcal{H} \to \mathcal{H}$, it is not necessarily true that $AB = BA$. It is possible to show that, if A, B obey the so-called *canonical commutation rule*, namely $[A, B] \equiv AB - BA = i\hbar$, then the product between the uncertainties in the measurements of A and B cannot be lower than $\hbar/2$: $\triangle A \cdot \triangle B \geq \hbar/2$. Since X and P obey indeed to the canonical commutation rule, it follows that *it is not possible to determine with arbitrarily low uncertainty both the position and the momentum of a quantum particle (Heisemberg uncertainty principle).* On the other hand, when A, B are *compatible* observables, i.e. $AB = BA$ holds, then it is possible to measure both observables simultaneously with arbitrarily small uncertainty (in theory).

The expression $\gamma(X, P) = \gamma(x, -i\hbar\nabla_x)$ is ambiguous. For example, it is not clear how to define the quantum analogue of $\gamma(x, p) = x \cdot p$, because while $x \cdot p = p \cdot x$, the operators X, P do not commute, so $X \cdot P$, $P \cdot X$, $(X \cdot P + P \cdot X)/2$ are three possible distinct interpretations of $\gamma(X, P)$. Several rules have been defined (called *quantization rules*), which associate to every classical observable its quantum counterpart. A possible quantization rule is the *Weyl quantization*, according to which the quantum counterpart A_γ of a classical observable $\gamma = \gamma(x, p)$ is defined by:

$$(A_\gamma \psi)(x) = (2\pi\hbar)^{-3} \int_{\mathbb{R}^3 \times \mathbb{R}^3} \gamma\left(\frac{x+y}{2}, p\right) \psi(y) e^{i(x-y)\cdot p/\hbar} dy dp, \qquad \forall \psi \in \mathcal{H}, \quad \forall x \in \mathbb{R}^3.$$

It is possible to show that X, P given in Postulate 2 can be derived by means of the rule above. Moreover, the following (formal) limit holds: $\lim_{\hbar \to 0} A_\gamma \psi = \gamma\psi$ for all $\psi \in \mathcal{H}$. This means that, in the formal limit $\hbar \to 0$, the quantum mechanics reduces to the classical mechanics.

The solutions ψ of the time-dependent Schrödinger equation having the form $\psi(x, t) = e^{iEt/\hbar}\phi(x)$, where $E \in \mathbb{R}$ is an eigenvalue of H and ϕ is an eigenstate of the Hamiltonian corresponding to the eigenvalue E, are called *stationary states* of the system, and the corresponding eigenvalue-eigenvector problem $H\phi = E\phi$ for H is called *stationary Schrödinger equation.* There are two main reasons why such solutions are important. First, notice that, for every linear self-adjoint operator A on $\mathcal{H}$ and every stationary state $\psi(x, t) = e^{iEt/\hbar}\phi(x)$ of the system it holds:

$$(\psi(t), A\psi(t))_{\mathcal{H}} = \int_{\mathbb{R}^3} \overline{e^{iEt/\hbar}\phi} A e^{iEt/\hbar}\phi dx = \int_{\mathbb{R}^3} \overline{\phi} A\phi dx = (\phi, A\phi)_{\mathcal{H}}, \qquad t > 0.$$

Thus, if a quantum system is in a stationary state, the mathematical expectation of any observable associated to the system is constant in time (hence the name “stationary state”). Moreover, since multiplying the wavefunction of the system times any phase factor does not alter the modulus of the wavefunction itself, then also the probability density $|\psi|^2$ of the system is constant in time if the state of the system is stationary.

The second reason for considering stationary states is related to semigroup theory. In fact, the solution of the time dependent Schrödinger equation can be written as $\psi(t) = e^{-itH/\hbar}\psi(0)$, for $t \in \mathbb{R}$, where $e^{-itH/\hbar}$ is the group associated to H.[4] Let us assume, for the sake of simplicity, that H has discrete spectrum $\sigma(H) = (E_k)_{k\in\mathbb{N}}$ and each eigenvalue E_k has multiplicity 1, that is, the eigenspace associated to E_k is spanned by only one vector ϕ_k. Thus:

$$\psi(t) = e^{-itH/\hbar}\psi(0) = \sum_{k\in\mathbb{N}} e^{-itE_k/\hbar}\mathbb{P}_{H,E_k}\psi(0) = \sum_{k\in\mathbb{N}} e^{-itE_k/\hbar}(\phi_k, \psi(0))\phi_k.$$

This means that, if we can solve the stationary Schrödinger equation (i.e. we can find (E_k, ϕ_k) for $k \in \mathbb{N}$), then we can also solve the time-dependent Schrödinger equation.

Exercise. Prove that the Schrödinger equation preserves the normality of the wavefunction, that is: if $\int_{\mathbb{R}^3} |\psi(0)|^2 dx = 1$ then $\int_{\mathbb{R}^3} |\psi(t)|^2 dx = 1$ for all $t \in \mathbb{R}$.

We show now that the quantum formalism that we have just introduced allows us to derive an evolution equation for the probability density $|\psi|^2$ of a quantum particle described by the wavefunction ψ, that is formally identical to the continuity equation (6). Let us differentiate $|\psi|^2$ with respect to time:

$$\partial_t|\psi|^2 = \overline{\psi}\partial_t\psi + \psi\partial_t\overline{\psi} = -\frac{i}{\hbar}\left(\overline{\psi}H\psi - \psi H\overline{\psi}\right). \tag{14}$$

Since $H = -\frac{\hbar^2}{2m}\Delta + V$ for a single quantum particle with mass m subject to a force $F = -\nabla V$, it follows:

$$\partial_t|\psi|^2 = -\frac{1}{i\hbar}\frac{\hbar^2}{2m}\left(\overline{\psi}\Delta\psi - \psi\Delta\overline{\psi}\right) = -\frac{\hbar}{m}\frac{1}{2i}\nabla\cdot\left(\overline{\psi}\nabla\psi - \psi\nabla\overline{\psi}\right),$$

and so:

$$\partial_t|\psi|^2 + \nabla\cdot\left(\frac{\hbar}{m}\Im(\overline{\psi}\nabla\psi)\right) = 0. \tag{15}$$

Eq. (15) is formally identical to (6) with $n = |\psi|^2$ and $J = \frac{\hbar}{m}\Im(\overline{\psi}\nabla\psi)$. It is a balance equation for the probability density n of the system; the vector J can be regarded as a probability current density. Intuitively, eq. (15) represents a bridge between quantum mechanics and fluid models for electron transport.

Example 2.1 (free electron)**.** We consider a free electron in the one-dimensional space subject to no potential: $V \equiv 0$. We wish to determine the allowed energy values of the system. The stationary Schrödinger equation for the system looks like:

$$-\frac{\hbar^2}{2m}\psi'' = E\psi, \qquad x \in \mathbb{R}. \tag{16}$$

---

[4] Being H self-adjoint, then $iH/\hbar$ is skew-adjoint, and thus generates a unitary group $e^{-itH/\hbar}$ in $\mathcal{H} = L^2(\mathbb{R}^3, \mathbb{C})$ (Stone's theorem).

If $\omega \in \mathbb{C}$ satisfies

$$\omega^2 = -\frac{2mE}{\hbar^2} \tag{17}$$

then the solutions of (16) can be written as:

$$\psi(x) = C_1 e^{\omega x} + C_2 e^{-\omega x}, \qquad x \in \mathbb{R}. \tag{18}$$

We are interested only in the physically reasonable solutions of (16). Any such solution must clearly be bounded for $|x| \to \infty$; this means that $\Re\omega = 0$, so $\omega = ik$ for some $k \in \mathbb{R}$. From (17) it follows $E = \hbar^2 k^2/2m$; in particular E must be nonnegative. However, this appears to be the only limitations for the energy spectrum: the system energy can assume all values between 0 and ∞; the spectrum of H is continuous.

Example 2.2 (infinite square-well potential)**.** Now we consider an electron that is confined between two parallel walls, and we try to determine the allowed energy values of the system. We consider this problem as one-dimensional by neglecting what is happening in the directions parallel to the walls. We choose the reference frame in such a way that the x axis is orthogonal to the walls, and these walls are placed at points 0 and L, respectively. Clearly any stationary state of the system will be given by the restriction of (18) to the interval $(0, L)$:

$$\psi(x) = C_1 e^{\omega x} + C_2 e^{-\omega x}, \qquad 0 \le x \le L. \tag{19}$$

This time we cannot invoke any boundedness argument to determine ω, since the function ψ given by (19) is always bounded in $[0, L]$. We can, however, exploit the fact that the system in confined in the interval $[0, L]$: as a consequence $\psi(0) = \psi(L) = 0$. The constraint $\psi(0) = 0$ implies $C_2 = -C_1$; the relation $\psi(L) = 0$ leads to:

$$0 = C_1 e^{\omega L} + C_2 e^{-\omega L} = C_1(e^{\omega L} - e^{-\omega L}). \tag{20}$$

Clearly $C_1 \neq 0$ (otherwise also $C_2 = -C_1 = 0$ and so $\psi \equiv 0$, which is not acceptable). Thus (20) implies $e^{2\omega L} = 1$, which means $2\omega L = 2\pi i n$ for some $n \in \mathbb{Z}$, that is $\omega = n\pi i/L$. By plugging this result into (17) we get:

$$E = -\frac{\omega^2\hbar^2}{2m} = \frac{\pi^2\hbar^2}{2mL^2} n^2. \tag{21}$$

Since $n \in \mathbb{Z}$, (21) means that the energy E can assume only a discrete set of (positive) values: that is, the spectrum of H is discrete.

2.3 Electrons in a semiconductor cristal

A semiconductor material is is made up by nuclei and electrons. The latter ones fall into two cathegories according to their energy: the *core electrons*, which have negative energy and are bound to the nuclei, and the *valence electrons*, which have positive energy and are flow freely throughout the semiconductor. A nucleous and the core electrons orbiting around it form an *ion.* The electronic properties of a semiconductor are determined by the valence electrons and their relation with the ions.

The valence electrons ensemble can be described by a wavefunction $\psi : \mathbb{R}^{3M} \times \mathbb{R} \to \mathbb{C}$, $\psi = \psi(x,t)$, with $x = (x_1, \ldots, x_N)$, and $x_j \in \mathbb{R}^3$ representing the position of the $j-$th electron, for $j = 1, \ldots, N$. To write the Hamiltonian of the system one has to take into account three phenomenon: ion vibrations, electron-ion interaction, electron-electron scattering. To make things simpler, we neglect ion vibrations: we assume that the ions are fixed and in equilibrium. Thus the Hamiltonian of the system has the form:

$$H = -\frac{\hbar^2}{2m} \sum_{j=1}^{N} \Delta_{x_j} + H_{ei} + H_{ee}, \tag{22}$$

where H_{ei}, H_{ee} take into account the electron-ion interaction and the electron-electron scattering, respectively. Let us focus on these two terms. Clearly $H_{ei} = q \sum_{j=1}^{N} V_{ei}(x_j)$, where V_{ei} is the electrostatic potential generated by the static ion distribution:

$$V_{ei}(y) = \sum_{k=1}^{M} \frac{Q}{4\pi\varepsilon_0 |y - R_k|}, \qquad y \in \mathbb{R}^3,$$

where ε_0 is the electric permittivity, M is the number of ions, Q is the charge of an ion and $R_k \in \mathbb{R}^3$ is the position of the $k-$th ion, for $k = 1, \ldots, M$. It follows that:

$$H_{ei} = \sum_{j=1}^{N} \sum_{k=1}^{M} \frac{qQ}{4\pi\varepsilon_0 |x_j - R_k|}. \tag{23}$$

The contribution from the electron-electron scattering has the form:[5]

$$H_{ee} = \frac{1}{2} \sum_{\substack{j,k=1 \\ j \neq k}}^{N} H_{ee,j,k},$$

where $H_{ee,j,k}$ represents the contribution of the pair of electrons with labels j, k, which is given by:

$$H_{ee,j,k} = \frac{q^2}{4\pi\varepsilon_0 |x_j - x_k|}, \qquad j, k = 1, \ldots, N, \quad j \neq k. \tag{24}$$

Thus:

$$H_{ee} = \frac{1}{2} \sum_{\substack{j,k=1 \\ j \neq k}}^{N} \frac{q^2}{4\pi\varepsilon_0 |x_j - x_k|}. \tag{25}$$

---

[5] The factor 1/2 is the equation below takes into account the obvious fact that the couples (j,k) and (k,j) give the same contribution.

By putting (22)–(25) together we conclude:

$$H = -\frac{\hbar^2}{2m}\sum_{j=1}^{N}\Delta_{x_j} + \sum_{j=1}^{N}\sum_{k=1}^{M}\frac{qQ}{4\pi\varepsilon_0|x_j - R_k|} + \frac{1}{2}\sum_{\substack{j,k=1\\ j\neq k}}^{N}\frac{q^2}{4\pi\varepsilon_0|x_j - x_k|}. \tag{26}$$

Since the number N of electrons is rather big, the solution of the stationary Schrödinger equation for H (namely $H\psi = E\psi$) is numerically very expensive. So we look for a way to simplify the problem. The strategy that we will follow is:

1. we will replace the term H_{ee} describing the electron-electron interactions with a single-particle effective potential, reducing the $3M-$dimensional problem to a 3-dimensional one (Hartree-Fock approximation);

2. we will reduce the solution of the Schrödinger equation in the whole space $\mathbb{R}^3$ to the solution of the Schrödinger equation in the Wigner-Seitz cell of the Bravais lattice (Bloch decomposition).

2.3.1 The Hartree-Fock approximation.

Let us explain the first approximation. We preliminary observe that, if electron-electron interactions can be neglected, then the problem becomes much simpler, namely:

- the wavefunction ψ factorizes into a product of single-particle wavefunctions:

$$\psi(x) = \prod_{j=1}^{N}\psi_j(x_j), \quad \psi_j : \mathbb{R}^3 \to \mathbb{C} \quad (j = 1, \ldots, N); \tag{27}$$

- the Hamiltonian H equals the sum of single-particle Hamiltonians:

$$H = \sum_{j=1}^{N} H_j, \quad H_j = -\frac{\hbar^2}{2m}\Delta_{x_j} + \sum_{k=1}^{M}\frac{qQ}{4\pi\varepsilon_0|x_j - R_k|}, \quad (j = 1, \ldots, N); \tag{28}$$

- the Schrödinger equation decouples into a system of independent single-particle Schrödinger equations: $H_j\psi_j = E_j\psi_j$, $j = 1, \ldots, N$.

In fact, if ψ, H satisfy (27), (28):

$$H\psi = \sum_{j=1}^{N} H_j \prod_{k=1}^{N}\psi_k(x_k) = \sum_{j=1}^{N}\prod_{\substack{k=1\\ k\neq j}}^{N}\psi_k\, H_j\psi_j;$$

this means that, if $H_j\psi_j = E_j\psi_j$ for $j = 1, \ldots, N$, then $H\psi = E\psi$ with $E = \sum_{j=1}^{N} E_j$. Conversely, if $H\psi = E\psi$, let us write $E = \sum_{j=1}^{N} E_j$ for some $E_1, \ldots, E_N \in \mathbb{R}$. It follows:

$$0 = \frac{1}{\psi}(H\psi - E\psi) = \sum_{k=1}^{N}\frac{1}{\psi_j}(H_j\psi_j - E_j\psi_j),$$

and since ψ_j, $H_j\psi_j$ depend only on x_j we conclude that $H_j\psi_j = E_j\psi_j$ for $j = 1, \ldots, N$.

The idea of the Hartree-Fock method is that the product ansatz (27) makes sense even in presence of electron-electron interactions. In this case, however, (28) does not hold true; instead,

$$H = \sum_{j=1}^{N} H_j + \frac{1}{2} \sum_{\substack{j,k=1 \\ j\neq k}}^{N} H_{ee,j,k}, \tag{29}$$

with $H_1, \ldots, H_N$ and $H_{ee,j,k}$ given by (28), (24), respectively. To determine the functions $\psi_1, \ldots, \psi_N$ in (27) we assume that the system is in the *ground state*, which means, its energy is minimal. Thus $(\psi_1, \ldots, \psi_N)$ will be the solution of the constraint minimization problem:

$$\begin{aligned} &\text{minimize } (\psi, H\psi) \text{ under the constraints } \|\psi_j\|^2_{L^2} = 1 \quad (j = 1, \ldots, N), \\ &\text{with } \psi,\ H \text{ given by (27), (29), respectively.} \end{aligned} \tag{30}$$

We solve problem (30) with the Lagrange multipliers method. If $\psi(x) = \prod_{j=1}^{N} \psi_j(x_j)$ solves (30), then parameters $E_1, \ldots, E_N \in \mathbb{R}$ exist such that:

$$\delta(\psi, H\psi) - \sum_{j=1}^{N} E_j \delta\|\psi_j\|^2_{L^2} = 0. \tag{31}$$

In (31) the symbol δ denotes the *variation* of a functional $H \to \mathbb{R}$. Thanks to (29) the left-hand side of (31) can be rewritten as:

$$\begin{aligned} &\delta \sum_{j=1}^{N} (\psi_j, H_j\psi_j) + \delta \frac{1}{2} \sum_{\substack{j,k=1 \\ j\neq k}}^{N} \int_{\mathbb{R}^6} H_{ee,j,k} |\psi_j|^2 |\psi_k|^2 dx_j dx_k - 2 \sum_{j=1}^{N} E_j (\delta\psi_j, \psi_j) \\ &\quad = 2 \sum_{j=1}^{N} (\delta\psi_j, H_j\psi_j) + 2 \sum_{\substack{j,k=1 \\ j\neq k}}^{N} \int_{\mathbb{R}^6} \overline{\delta\psi}_j H_{ee,j,k} \psi_j |\psi_k|^2 dx_j dx_k - 2 \sum_{j=1}^{N} E_j (\delta\psi_j, \psi_j) \\ &\quad = 2 \sum_{j=1}^{N} (\delta\psi_j, (H_j + V_{\text{eff}} - E_j)\psi_j), \end{aligned} \tag{32}$$

where

$$V_{\text{eff}}(r) = \int_{\mathbb{R}^3} \frac{q^2 |\psi_k(r')|^2}{4\pi\varepsilon_0 |r - r'|} dr', \qquad r \in \mathbb{R}^3. \tag{33}$$

Since the variation $\delta\psi$ in (32) is arbitrary, (31) implies:

$$-\frac{\hbar^2}{2m} \Delta\psi_j(r) + \sum_{k=1}^{M} \frac{qQ\psi_j(r)}{4\pi\varepsilon_0 |r - R_k|} + V_{\text{eff}}(r)\psi_j(r) = E_j\psi_j(r), \qquad r \in \mathbb{R}^3, \tag{34}$$

for $j = 1, \dots, N$. Eq. (34) is the *Hartree equation.* It is a single-particle Schrödinger equation (that is, a PDE for a scalar function of $r \in \mathbb{R}^3$) which incorporates the many-body effects through the effective potential V_{eff}. Notice that Δ in (34) is the 3-dimensional Laplacian.

Unfortunately, the Hartree equation has a flaw: in the derivation of (34) we neglected the *fermionic nature* of the electrons. Quantum particles fall into two cathegories: *bosons* and *fermions.* To understand the difference between the two classes, consider two undistinguishable quantum particles, labeled with indexes 1 and 2. The wavefunction of the system will be some function $\psi = \psi(x_1, x_2)$. What happens to ψ is the particle labels are exchanged? It can be shown that either $\psi(x_2, x_1) = \psi(x_1, x_2)$ or $\psi(x_2, x_1) = -\psi(x_1, x_2)$. Any quantum particle will choose *always* one of the two options. The particles for which $\psi(x_2, x_1) = -\psi(x_1, x_2)$ are called *fermions*, while the particles for which $\psi(x_2, x_1) = \psi(x_1, x_2)$ are called *bosons.*

It happens that electrons are fermions. So the wavefunction of an electron ensemble must clearly be antysimmetric, that is, $\psi(x_1, \dots, x_N)$ must change sign for every odd permutation of the variables $x_1, \dots, x_N$. However, there is no reason why the wavefunction defined in (27) should satisfy this property. Thus we have to consider a symmetrized version of (27), namely, in place of a factorized wavefunction we choose a suitable linear combinations of products like the one in (27):

$$\psi(x) = \sum_{(j_1,\dots,j_N)\in S_N} \operatorname{sign}(j_1, \dots, j_N) \prod_{k=1}^{N} \psi_k(x_{j_k}), \tag{35}$$

where S_N is the set of permutations of $\{1, \dots, N\}$. It is easy to see that ψ defined in (35) equals the determinant of the matrix $(\psi_j(x_k))_{j,k}$, and for this reason an expression like (35) is called a *Slater determinant.* If we solve (30) with ψ given by (35) we obtain in the end the *Hartree-Fock equation*:

$$\begin{aligned} -\frac{\hbar^2}{2m}\Delta\psi_j(r) + \sum_{k=1}^{M} \frac{qQ\psi_j(r)}{4\pi\varepsilon_0|r - R_k|} + V_{\text{eff}}(r)\psi_j(r) \qquad\qquad (36)\\ - q^2 \sum_{k\neq j,\ \text{spin}\ \|} \int_{\mathbb{R}^3} \frac{\overline{\psi}_k(r')\psi_j(r')}{|r - r'|} dr' \psi_k(r) = E_j\psi_j(r), \qquad r \in \mathbb{R}^3, \end{aligned}$$

with V_{eff} still given by (33). The derivation of (36) can be found in [11]. The additional term in (36):

$$(H_{ex}\psi)_j \equiv -q^2 \sum_{k\neq j,\ \text{spin}\ \|} \int_{\mathbb{R}^3} \frac{\overline{\psi}_k(r')\psi_j(r')}{|r - r'|} dr' \psi_k(r), \tag{37}$$

contains a summation over electrons with the same spin[6] and arises from “exchange interaction”, i.e. the fact that the particles are indistinguishable under the exchange of particle label.

---

[6] The spin is an intrinsic property of some quantum particles, a form of intrinsic angular momentum which does not have a classical counterpart [51].

2.3.2 The Bloch decomposition.

So, with the Hartree-Fock method we have obtained a Schrödinger equation for the single-particle Hamiltonian

$$H = -\frac{\hbar^2}{2m}\Delta + qV_L(x), \qquad x \in \mathbb{R}^3, \tag{38}$$

where V_L is a suitable effective potential. Given the periodic crystal structure of a semiconductor, one expects that also V_L is periodic. So we hope that the Schrödinger equation in the whole space can be reduced to a Schrödinger equation in a primitive cell of the Bravais lattice. This hope is not in vain:

Theorem 2.1 (Bloch)**.** *Let V_L a potential satisfying $V_L(x+y) = V_L(x)$ for all $x \in \mathbb{R}^3$, $y \in L$ (Bravais lattice). Then the eigenfunctions of the Hamiltonian H given by* (38) *can be chosen to have the following form:*

$$\psi_{n,k}(x) = e^{ik\cdot x}u_{n,k}(x), \qquad x \in \mathbb{R}^3, \quad k \in B \text{ (the Brillouin zone)}, \quad n \in \mathbb{N}, \tag{39}$$

for some functions $u_{n,k}$ satisfying: $u_{n,k}(x+y) = u_{n,k}(x)$ for all $x \in \mathbb{R}^3$, $y \in L$.

Proof. Let T_a be the translation operator: $(T_a\psi)(x) = \psi(x+a)$ for $a \in L$, $x \in \mathbb{R}^3$, $\psi \in L^2(\mathbb{R}^3, \mathbb{C})$. We claim that the eigenvalues of T_a have the form $e^{i\theta}$ for some $\theta \in \mathbb{R}$. In fact, if $T_a\psi = \lambda\psi$ and $\psi \neq 0$ then:

$$|\lambda|^2\|\psi\|^2_{L^2} = \|\lambda\psi\|^2_{L^2} = \|T_a\psi\|^2_{L^2} = \int_{\mathbb{R}^3} |\psi(x+a)|^2dx = \int_{\mathbb{R}^3} |\psi(x)|^2dx = \|\psi\|^2_{L^2},$$

so $|\lambda| = 1$.
It is trivial to see that $T_aT_{a'} = T_{a'}T_a$ for all $a, a' \in L$. Moreover, we claim that $HT_a = T_aH$ for all $a \in L$. In fact, thanks to the periodicity of V_L and the fact $a \in L$,

$$\begin{aligned}(T_aH\psi)(x) &= -\frac{\hbar^2}{2m}\Delta\psi(x+a) + qV_L(x+a)\psi(x+a)\\ &= -\frac{\hbar^2}{2m}\Delta\psi(x+a) + qV_L(x)\psi(x+a) = (HT_a\psi)(x).\end{aligned}$$

Being H, T_a self-adjoint operators, this means that H and T_a have the same eigenvectors for all $a \in L$ [5]. Let ψ be such a common eigenvector of H and T_a for all $a \in L$. In particular, for $j = 1, 2, 3$ there exist $\theta_j \in \mathbb{R}$ such that

$$T_{a_j}\psi = e^{i\theta_j}\psi \qquad (j = 1, 2, 3). \tag{40}$$

Let now, for $K \in L^*$ arbitrary,

$$k = -\frac{1}{2\pi}\sum_{\ell=1}^{3}\theta_\ell a^*_\ell + K, \qquad u(x) = e^{ik\cdot x}\psi(x). \tag{41}$$

It is easy to see that $e^{i(k\cdot a_j+\theta_j)}=1$, for $j=1,2,3$. Thus it follows:

$$u(x+a_j)=e^{ik\cdot x}e^{ik\cdot a_j}\psi(x+a_j)=e^{ik\cdot x}e^{ik\cdot a_j}(T_{a_j}\psi)(x)=e^{ik\cdot x}e^{i(k\cdot a_j+\theta_j)}\psi(x)=u(x).$$

So $u(x+y)=u(x)$ for all $x\in\mathbb{R}^3$, $y\in L$. By choosing $K=\sum_{j=1}^3 K_j a_j^*$ with K_j a minimizer of the function $s\in\mathbb{R}\mapsto\left|s-\frac{\theta_j}{2\pi}\right|\in\mathbb{R}$, we conclude that $k\in B$. $\square$

Let us make some remarks about Bloch's theorem. First, we point out that the index $n\in\mathbb{N}$ appears in the theorem because there are many solutions to the Schrödinger equation. In fact, if ψ satisfies (39) then the eigenvalue problem $H\psi=E\psi$ becomes:

$$\frac{\hbar^2}{2m}(-i\nabla+k)^2u_{n,k}(x)+qV_L(x)u_{n,k}(x)=E_n(k)u_{n,k}(x),\qquad x\in\mathbb{R}^3,$$

with the periodicity condition

$$u_{n,k}(x+y)=u_{n,k}(x)\qquad x\in\mathbb{R}^3,\quad y\in L.$$

The above-stated problem can be regarded as an eigenvalue problem for a self-adjoint operator in a bounded domain D (a primitive cell of the lattice L). So we expect to have an infinite (countable) family of solutions with *discretely* spaced eigenvalues: $n\in\mathbb{N}$.

Relation $k\mapsto E_n(k)$ is called *dispersion relation*; it can be proved to be a continuous map and L^*-periodic ($E_n(k+K)=E_n(k)$ for $k\in\mathbb{R}^3$, $K\in L^*$). The image of $\overline{B}$ through E_n, namely $E_n(\overline{B})=\{E_n(k)\,:\,k\in\overline{B}\}$, is the $n-$th *energy band.* If the set $\mathbb{R}\setminus\cup_{n\in\mathbb{N}}E_n(\overline{B})$ is nonempty, then its connected components are called *energy gaps*. Finally, if there is only one energy gap, then the nearest energy bands below and above the energy gap are called *valence band* and *conduction band*, respectively.

Bloch's theorem introduces a wavevector k, to which a momentum $p=\hbar k$ can be associated. However, p is not the (mathematical expectation of the) electron momentum. In fact, for $\psi_{n,k}$ of the form (39) it holds:

$$-i\hbar\nabla\psi_{n,k}=-i\hbar\nabla(e^{ik\cdot x}u_{n,k}(x))=\hbar k\psi_{n,k}-i\hbar e^{ik\cdot x}\nabla u_{n,k}(x)\neq\hbar k\psi_{n,k},$$

so ψ is not a momentum eigenstate. Neverthless, the quantity $p=\hbar k$, called *crystal momentum*, plays an important role in the problem of the motion of electrons in periodic potentials, as we will see in the rest of the course.

We conclude this section by pointing out that now the reader is finally able to understand the definition of semiconductor anticipated in the Introduction:

Definition. A semiconductor is a solid material with a band gap greater than zero and lower than 4 eV.

2.4 The Semi-Classical Picture

We will present some equations which look like Newton's laws of classical mechanics, but incorporate also quantum effects.

2.4.1 Semi-Classical equations of motion.

Let us begin by considering a Bloch electron in a semiconductor subject to an external potential V. The particles will be described by a wavefunction $\psi_{n,k}(x)$ satisfying the family of eigenvalue-eigenvector problems depending on $k \in B$:

$$-\frac{\hbar^2}{2m}\Delta\psi_{n,k} + q(V_L + V)\psi_{n,k} = E_n(k)\psi_{n,k} \qquad \text{in } D, \tag{42}$$

and admitting the decomposition:

$$\psi_{n,k}(x) = e^{ik\cdot x}u_{n,k}(x) \qquad x \in D. \tag{43}$$

Here D is a primitive cell of the lattice, V_L is an effective potential with the periodicity of the Bravais lattice, k is the electron's (pseudo)wavevector and $n \in \mathbb{N}$ is the band index. We assume, for the sake of simplicity, that the electron stays in the same band n during the evolution of the system. For this reason we can drop the index n from $E_n(k)$ and $\psi_{n,k}$.

If the electron were a classical particle, then one could associate to it a position $x = x(t)$ and a velocity $v = v(t)$ in a straightforward way. However, even if the electron is a quantum particle, we can associate to it a "semi-classical trajectory" $x(t)$ defined by:

$$\dot{x} = \frac{\langle P\rangle}{m}, \tag{44}$$

where $\langle P\rangle$ is the mathematical expectation of the momentum operator P:

$$\langle P\rangle \equiv \int_D \overline{\psi_k(x)}(-i\hbar\nabla_x)\psi_k(x)dx.$$

The wavevector k is also to be considered a function of time: $k = k(t)$. The quantity $\langle P\rangle$ is a function of k, thus (44) is a relation between $x(t)$ and $k(t)$.

The mathematical expectation of P admits a simple form:

Lemma 2.1. *It holds:*

$$\langle P\rangle = \frac{m}{\hbar}\nabla_k E(k).$$

Proof. Since (43) holds, it follows $\nabla_k\psi_k = ix\psi_k + e^{ik\cdot x}\nabla_k u_k$. Thus differentiating (42) with respect to k yields:

$$\begin{aligned} -\frac{\hbar^2}{2m}\Delta_x(ix\psi_k + e^{ik\cdot x}\nabla_k u_k) + q(V_L + V)(ix\psi_k + e^{ik\cdot x}\nabla_k u_k) \\ = \psi_{n,k}\nabla_k E(k) + E(k)(ix\psi_k + e^{ik\cdot x}\nabla_k u_k). \end{aligned} \tag{45}$$

Clearly $\Delta_x(ix\psi_k) = 2i\nabla_x\psi_k + ix\Delta_x\psi_k$, so from (45) it follows:

$$ix\left(-\frac{\hbar^2}{2m}\Delta_x\psi_k + q(V_L + V)\psi_k - E(k)\psi_k\right) - \frac{i\hbar^2}{m}\nabla_x\psi_k \tag{46}$$

$$+\left(-\frac{\hbar^2}{2m}\Delta_x + q(V_L+V) - E(k)\right)(e^{ik\cdot x}\nabla_k u_k) = \psi_{n,k}\nabla_k E(k).$$

The first term on the left-hand side of (46) vanishes thanks to (42), so:

$$-\frac{i\hbar^2}{m}\nabla_x\psi_k + (H - E(k))\,(e^{ik\cdot x}\nabla_k u_k) = \psi_{n,k}\nabla_k E(k), \tag{47}$$

where $H = -\frac{\hbar^2}{2m}\Delta_x + q(V_L+V)$ is the Hamiltonian. Let us multiply (47) times $\overline{\psi_k}$ and integrate in D:

$$-\frac{i\hbar^2}{m}\int_D \overline{\psi_k}\nabla_x\psi_k + \int_D \overline{\psi_k}\,(H - E(k))\,(e^{ik\cdot x}\nabla_k u_k)dx = \nabla_k E(k). \tag{48}$$

The self-adjointness of H and (42) imply:

$$\int_D \overline{\psi_k}\,(H - E(k))\,(e^{ik\cdot x}\nabla_k u_k)dx = \int_D \overline{(H\psi_k - E(k)\psi_k)}\; e^{ik\cdot x}\nabla_k u_k dx = 0.$$

So (48) becomes:

$$-\frac{i\hbar^2}{m}\int_D \overline{\psi_k}\nabla_x\psi_k = \nabla_k E(k),$$

which implies the statement. □

By putting together (44) and Lemma 2.1 we obtain:

$$\dot{x} = \frac{1}{\hbar}\nabla_k E(k) = \nabla_p \tilde{E}(p), \tag{49}$$

where $\tilde{E}(p) \equiv E(k) = E(p/\hbar)$ is the energy written as a function of the pseudo-momentum $p = \hbar k$. Eq. (49) strikingly resembles one of the Hamilton equations of classical mechanics, namely $\dot{x} = \nabla_p\mathscr{H}$, linking the time derivative of the position with the (classical) Hamiltonian $\mathscr{H}$ of the system. At this point it is natural to ask whether there is a semiclassical analogue of the other Hamilton equation, that is $\dot{p} = -\nabla_x\mathscr{H}$. As a matter of fact, it can be shown that:

$$\hbar\dot{k} = -q\nabla_x V. \tag{50}$$

The proof of (50) is quite difficult; see e.g. [6, p. 220], [40, p. 39]. Anyway, we can derive a weaker relation in a simple way by assuming that the total energy of the system $E_{tot} = E(k) + qV(x)$ is constant along the trajectories of the system:

$$0 = \frac{dE_{tot}}{dt} = \nabla_k E(k)\cdot\dot{k} + q\nabla_x V(x)\cdot\dot{x} = \left(\dot{k} + \frac{q}{\hbar}\nabla_x V(x)\right)\cdot\nabla_k E(k).$$

So $\hbar\dot{k} + q\nabla_x V(x)$ is orthogonal to $\nabla_k E(k)$. This is coherent with (50), although it is by no means of proof of it.

Notice that (50) does not contain the effective potential V_L: the evolution of the pseudo-momentum p depends only on the applied potential.

2.4.2 Effective mass and parabolic band approximation.

Eq. (49) shows that the semiclassical velocity $\dot{x}$ is proportional to $\langle P \rangle$: $\langle P \rangle = m\dot{x}$. What is the relation between the pseudo-momentum $\hbar k$ and $\dot{x}$? Is there a scalar, or a 3×3 matrix m^* such that $\hbar k = m^* \dot{x}$? Let us assume that such a matrix exists. Then from (49) it follows:

$$\hbar \dot{k} = m^* \ddot{x} = \frac{m^*}{\hbar} \frac{\partial^2 E(k)}{\partial k^2} \dot{k},$$

which implies, since $\dot{k}$ is arbitrary:

$$(m^*)^{-1} = \frac{1}{\hbar^2} \frac{\partial^2 E(k)}{\partial k^2}. \tag{51}$$

Eq. (51) gives the definition of the *effective mass* m^*. Notice that m^* is, in general, a 3×3 symmetric matrix. If k_0 is a point of local minimum for $E(k)$ (that is, a minimum of the conduction band), then $m^*(k_0)$ is positive definite. So the spectral theorem allows us to find a suitable orthogonal coordinates system in which $m^*(k_0)$ is diagonal:

$$m^*(k_0) = \begin{pmatrix} m_1^* & 0 & 0 \\ 0 & m_2^* & 0 \\ 0 & 0 & m_3^* \end{pmatrix}.$$

So, thanks to Taylor's formula, the conduction band can be approximated, in a neighborhood of the minimum, by:[7]

$$E(k) \approx E_c + \frac{\hbar^2}{2} \left(\frac{k_1^2}{m_1} + \frac{k_2^2}{m_2} + \frac{k_3^2}{m_3} \right),$$

where E_c is the minimum value of the conduction band.
A semiconductor is called *isotropic* if $m_1 = m_2 = m_3 = m_e^*$. In such a case:

$$E(k) \approx E_c + \frac{\hbar^2 |k|^2}{2m_e^*}. \tag{52}$$

Eq. (52) is called *parabolic band approximation.* It means that, in an isotropic semiconductor, the energy of an electron near a minimum of the conduction band equals the energy of a free quantum particle in a vacuum having mass equal to the effective mass m^*.

What happens if we evaluate (51) in a maximum of the valence band? Then clearly m^* is negative definite. Physicists dislike things like negative masses, so they found a way to avoid this concept. The idea is: if we change the sign of both m^* and $\dot{x}$ in the relation $\hbar k = m^* \dot{x}$, then such relation remains valid and the newly defined effective mass is positive definite. However, a change in sign of the velocity $\dot{x}$ means that the particles move, under the influence of an electric field, in the opposite direction compared to the electrons;

---

[7] We can assume, without loss of generality, that $k_0 = 0$.

namely, they behave like "positively charged electrons". These particles are called *holes*. Physically speaking, a hole is a vacant orbital in the valence band, which behaves like a particle with positive charge. An electron moving in a direction can be equivalently viewed as a hole moving in the opposite direction. Simmetrically to what happens near the bottom of the conduction band, near the top of the valence band the band energy can be approximated by:

$$E(k) \approx E_v - \frac{\hbar^2 |k|^2}{2m_h^*}, \tag{53}$$

where E_v is the energy at the valence band maximum, and m_h^* is the hole effective mass near the maximum of the valence band.

2.5 Semiconductor Statistics

Given a system of N electrons in a semiconductor of finite size in thermal equilibrium (i.e. no current flow), how many electrons are in a given energy band? More in general, how are electrons distributed with respect to their energy?

Answering these questions means to analyze the *occupation probability* $f^N(E)$ for the system, which can be heuristically defined as follows. Given an arbitrary (allowed) energy value E,

$$f^N(E)dE = \text{number of electrons with energy in the range } [E, E + dE].$$

Recall that the allowed energy values in a semiconductor have the form $E = E_n(k)$, where $n \in \mathbb{N}$ is a band index and $k \in B$ (the Brillouin zone) is a wavevector. So it seems natural to state that the number of electrons in the $n-$th energy band equals the sum of $f^N(E_n(k))$ over all possible wavevectors k. To have a more handy expression for this quantity it is possible to take the continuous limit, so the sum becomes an integral. The result is:

$$N_n \equiv \text{number of electrons in the } n\text{−th energy band} = \frac{2\mathcal{V}}{(2\pi)^3} \int_B f^N(E_n(k))dk, \tag{54}$$

where $\mathcal{V}$ denotes the volume of the semiconductor. The term $(2\pi)^3$ is related to the volume of the Brillouin zone, while the factor 2 takes into account the two possible states of the electron spin. For details see [33, Section 1.6].

So we have answered the first question. The second one can be restated as: how does $f^N(E)$ look like? To derive an expression for the occupation probability[8], one has to keep in mind that:

- the electrons are undistinguishable;
- each quantum state can be occupied by at most two electrons with opposite spin (Pauli exclusion principle).

---

[8] The following derivation can be found in [6, p. 40]. For details about statistical mechanics see [35].

Thus to build an $N-$electron state one must fill N different one-electron levels.[9] In particular, the probability f_i^N of finding an electron in the level i, when the $N-$electron system is in thermal equilibrium, is simply the sum of the independent probabilities of finding an electron in any of the $N-$electron states in which the level i is occupied:

$$f_i^N = \sum_{\substack{\alpha\,:\,\alpha\,\text{is an } N-\text{electron state,}\\ \text{level } i \text{ occupied in state } \alpha}} P_N(E_\alpha^N), \tag{55}$$

where E_α^N is the energy of the $N-$electron state α. Here $P_N(E)$ is the probability of finding an electron with energy E when the $N-$electron system is in thermal equilibrium (at temperature T), and is given by:

$$P_N(E) = e^{-(E-F_N)/k_BT}, \tag{56}$$

where k_B is the Boltzmann constant and F_N is the Helmotz free energy for the system. We can evaluate f_i^N in three steps.

1 Pauli exclusion principle implies that the probability of finding an electron in level i equals 1 minus the probability of finding no electrons at level i. So:

$$f_i^N = 1 - \sum_{\substack{\alpha\,:\,\alpha\,\text{is an } N-\text{electron state,}\\ \text{level } i \text{ unoccupied in state } \alpha}} P_N(E_\alpha^N). \tag{57}$$

2 By taking any $(N+1)-$electron state in which the electron level i is occupied, we can build an $N-$electron state in which level i is unoccupied, simply by removing the electron from the $i-$th level and leaving all the other levels unaltered. Conversely, any $N-$electron state in which level i is unoccupied can be obtained in this manner from any $(N+1)-$electron state in which the electron level i is occupied. Clearly the energy of an $(N+1)-$electron state in which the electron level i is occupied and the corresponding $N-$electron state in which level i is unoccupied differ only by an amount $\mathcal{E}_i$ equals to the energy of the only one-electron level whose occupation is different in the two states. So:

$$\begin{aligned}&\{\text{Energies of } N-\text{electron states in which level } i \text{ is unoccupied}\}\\ &= \{\text{Energies of } (N+1)-\text{electron states in which level } i \text{ is occupied}\} - \mathcal{E}_i.\end{aligned}$$

By taking into account this correspondence we can rewrite (57) as:

$$f_i^N = 1 - \sum_{\substack{\alpha\,:\,\alpha\,\text{is an } (N+1)-\text{electron state,}\\ \text{level } i \text{ occupied in state } \alpha}} P_N(E_\alpha^{N+1} - \mathcal{E}_i). \tag{58}$$

---

[9] Here each level is specified by the electron's wavevector k and the projection of its spin along some axis.

By exploiting (56) we obtain:

$$P_N(E_\alpha^{N+1} - \mathcal{E}_i) = e^{(\mathcal{E}_i - q\mu)/k_B T} P_{N+1}(E_\alpha^{N+1}), \qquad q\mu = F_{N+1} - F_N.$$

The quantity μ is the above equation is the *chemical potential* at temperature T, and q is the electron charge. By substituting the above relation into (58) we get:

$$f_i^N = 1 - e^{(\mathcal{E}_i - q\mu)/k_B T} f_i^{N+1}. \tag{59}$$

3 We are interested in systems with a large number of electrons. Being f_i^N a sum of many one-electron contributions, it seems reasonable to approximate f_i^{N+1} with f_i^N if N is big enough. Thus, by replacing f_i^{N+1} with f_i^N in (59) we conclude:

$$f_i^N = 1 - e^{(\mathcal{E}_i - q\mu)/k_B T} f_i^N,$$

which can be explicitly solved for f_i^N, yielding:

$$f_i^N = \frac{1}{1 + e^{(\mathcal{E}_i - q\mu)/k_B T}}.$$

Since f_i^N (probability of finding an electron in level i when the $N-$electron state is in thermal equilibrium) depends only on the energy $\mathcal{E}_i$ of level i, we can write $f_i^N = f^N(\mathcal{E}_i)$. So we found an expression for the occupation probability:

$$f^N(E) = \frac{1}{1 + e^{(E - q\mu)/k_B T}}. \tag{60}$$

Eq. (60) is called the *Fermi-Dirac* distribution function. The dependence of f^N from the total number of electrons N is hidden inside the chemical potential μ, which must be in such a way that the following constraint holds:

$$N = \sum_{n\in\mathbb{N}} N_n = \sum_{n\in\mathbb{N}} \frac{2\mathcal{V}}{(2\pi)^3} \int_B f^N(E_n(k))dk = \sum_{n\in\mathbb{N}} \frac{2\mathcal{V}}{(2\pi)^3} \int_B \frac{dk}{1 + e^{(E_n(k) - q\mu)/k_B T}}. \tag{61}$$

Let us play a bit with the Fermi-Dirac distribution function. What happens when $T \to 0^+$? It is immediate to see that the following pointwise convergence relation holds:

$$f(E) \to \chi_{(-\infty,q\mu)}(E) = \begin{cases} 1 & \text{for } E < q\mu \\ 1/2 & \text{for } E = q\mu \\ 0 & \text{for } E > q\mu \end{cases}.$$

So, at tero temperature, all states having an energy smaller than $q\mu$ are filled, while all other states are empty. This behaviour can be understood with the help of the Pauli exclusion principle, according to which a quantum state cannot be occupied by more than one electron: thus an electron ensemble at zero temperature (that is, with minimal total energy) can be built *by filling each quantum state with one electron, starting with the states with lowest energy.* The energy of the state filled by the last particle equals $E_F \equiv q\mu$, which is called *Fermi energy.* However, a zero temperature system is a fairly ideal object; at positive temperature (that is, in real physical situations), electrons have a nonzero probability of jumping to higher energy levels due to thermal excitation, although this probability tends exponentially to zero as $E \to \infty$.

It is a basic fact of analysis that $(1 + e^x)^{-1} \to e^{-x}$ as $x \to \infty$. So, under the assumption $E - E_F \gg k_B T$, the Fermi-Dirac distribution can be approximated by the *Maxwell-Boltzmann distribution*:

$$f^{MB}(E) = e^{-(E-E_F)/k_B T}. \tag{62}$$

Semiconductors whose electron distribution can be effectively approximated by f^{MB} are called *nondegenerate*, otherwise are termed *degenerate* (and in this latter case the Fermi-Dirac distribution f has to be employed).

2.5.1 Electrons and holes densities.

Eq. (54) gives us the number of electrons in a given energy band. The ratio between this number and the volume of the semiconductor is called *electron density* in the energy band:

$$n_j = \frac{N_j}{\mathcal{V}} = \frac{2}{(2\pi)^d} \int_B f(E_j(k))dk, \tag{63}$$

with f given by (60). The above integral can be reformulated in the energy space by employing the Dirac delta distribution. [10] In fact, since:

$$f(E_j(k)) = \int_{\mathbb{R}} f(E)\delta(E - E_j(k))dE,$$

it follows:

$$n_j = \frac{2}{(2\pi)^d} \int_B \int_{\mathbb{R}} f(E)\delta(E - E_j(k))dEdk = \int_{\mathbb{R}} DOS_j(E)f(E)dE, \tag{64}$$

where:

$$DOS_j(E) = \frac{2}{(2\pi)^d} \int_B \delta(E - E_j(k))dk \tag{65}$$

---

[10] The Dirac delta is the linear functional δ on $C(\mathbb{R})$ defined by $\langle \delta, g \rangle = g(0)$ for all $g \in C(\mathbb{R})$. This definition is often rewritten in terms of the *symbolic* integral:

$$\int_{\mathbb{R}} \delta(x)g(x)dx = g(0) \qquad \forall g \in C(\mathbb{R}).$$

We stress the fact that the above integral is not a true integral, since δ is not a function, but a distribution. See e.g. [5, 38] for details.

is the so-called *density of states* of the $j-$th band.

In a similar way we can compute the density of holes p_j in the $j-$th band. Since holes are nothing but empty states, then their distribution function is simply given by $1-f(E)$; thus, in analogy with (64) we deduce:

$$p_j = \int_{\mathbb{R}} DOS_j(E)(1-f(E))dE.$$

In particular, the electron density n in the conduction band and the hole density p in the valence band can be written as:

$$n = \int_B DOS_c(E)f(E)dE, \qquad p = \int_B DOS_v(E)(1-f(E))dE,$$

with DOS_c, DOS_v densities of states in the conduction and valence band, respectively.

For nondegenerate semiconductors, under the parabolic band approximation, n, p can be computed explicitly.

Lemma 2.2. *Let $d=3$. Assume that the energy in the conduction and valence bands is given by $E_c + \frac{\hbar^2|k|^2}{2m_e^*}$, $E_v - \frac{\hbar^2|k|^2}{2m_h^*}$ respectively (parabolic band approximation), and that $E_c - q\mu, q\mu - E_v \gg k_BT$ (Maxwell-Boltzmann approximation).*[11] *Then:*

$$n = N_c^* e^{(q\mu - E_c)/k_BT}, \qquad p = N_v^* e^{(E_v - q\mu)/k_BT}, \tag{66}$$

where the effective densities of states N_c^*, N_v^* in the conduction and valence band (respectively) are given by:

$$N_c^* = 2\left(\frac{m_e^* k_B T}{2\pi\hbar^2}\right)^{3/2}, \qquad N_v^* = 2\left(\frac{m_h^* k_B T}{2\pi\hbar^2}\right)^{3/2}. \tag{67}$$

Proof. We prove the equation for n; the relation for p can be shown in a similar way.

By using spherical coordinates, eq. (65) can be restated as: [12]

$$DOS_c(E) = \frac{2}{(2\pi)^3}\int_{\mathbb{R}^3} \delta\left(E - E_c - \frac{\hbar^2|k|^2}{2m_e^*}\right)dk = \frac{1}{\pi^2}\int_0^\infty \delta\left(E - E_c - \frac{\hbar^2\rho^2}{2m_e^*}\right)\rho^2 d\rho;$$

---

[11] Recall that the distribution function of the holes is $1-f(E)$. Since:

$$1 - \frac{1}{1+e^x} = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}} \approx e^x \qquad \text{if } x \to -\infty,$$

the Maxwell-Boltzmann approximation for the hole distribution rewrites as $q\mu - E \gg k_BT$.

[12] Notice that $f(E) \to 0$ exponentially as $E \to \infty$, and that, according to the parabolic band approximation, the band energy diverges quadratically as $|k| \to \infty$. For this reason, we can replace the Brillouin zone B in the integral involving the density of states with the whole space $\mathbb{R}^3$, since the additional contribution to the electron density brought by this approximation will be negligible.

with the further change of variable $\rho = \frac{\sqrt{2m_e^*}}{\hbar}\sqrt{z}$ the above integral becomes:

$$
\begin{aligned}
DOS_c(E) &= \frac{1}{2\pi^2}\left(\frac{\sqrt{2m_e^*}}{\hbar}\right)^3 \int_0^\infty \delta\left(E - E_c - z\right)\sqrt{z}dz \\
&= \frac{1}{2\pi^2}\left(\frac{\sqrt{2m_e^*}}{\hbar}\right)^3 \int_{\mathbb{R}} \delta\left(E - E_c - z\right)\chi_{(0,\infty)}(z)\sqrt{z}dz \\
&= \frac{1}{2\pi^2}\left(\frac{\sqrt{2m_e^*}}{\hbar}\right)^3 \chi_{(0,\infty)}(E - E_c)\sqrt{E - E_c}.
\end{aligned}
$$

It follows:

$$
\begin{aligned}
n &= \int_{\mathbb{R}} DOS_c(E) f^{MB}(E) dE \\
&= \frac{1}{2\pi^2}\left(\frac{\sqrt{2m_e^*}}{\hbar}\right)^3 \int_{\mathbb{R}} \chi_{(0,\infty)}(E - E_c)\sqrt{E - E_c} e^{-(E-q\mu)/k_BT} dE \\
&= \frac{1}{2\pi^2}\left(\frac{\sqrt{2m_e^*}}{\hbar}\right)^3 \int_{E_c}^\infty \sqrt{E - E_c} e^{-(E-q\mu)/k_BT} dE.
\end{aligned}
$$

With the change of variables $x = (E - E_c)/k_BT$ we get:

$$
n = \frac{1}{2\pi^2}\left(\frac{\sqrt{2m_e^*}}{\hbar}\right)^3 (k_BT)^{3/2} e^{-(E_c-q\mu)/k_BT} \int_0^\infty \sqrt{x} e^{-x} dx.
$$

It is straightforward to see that $\int_0^\infty \sqrt{x}e^{-x}dx = \sqrt{\pi}/2$. We conclude:

$$
n = \frac{1}{4\pi^{3/2}}\left(\frac{\sqrt{2m_e^*}}{\hbar}\right)^3 (k_BT)^{3/2} e^{-(E_c-q\mu)/k_BT},
$$

which yields the statement for n. ☐

We point out that the factor 2 in front of N_c^*, N_v^* takes into account the 2 possible states of the spin.

2.5.2 Intrinsic semiconductors and doping.

A semiconductor with no impurities is called an *intrinsic* semiconductor. Clearly, in this case electrons in the conduction band can only come from valence band levels leaving a vacancy behind them; that is, there is a one-to-one correspondence between electrons in the conduction band and holes in the valence band. In particular, the electron density in the conduction band n equals the hole density in the valence band p. Their common value n_i is called *intrinsic density*. Under the assumption of Lemma 2.2 it holds:

$$
n_i = \sqrt{np} = \sqrt{N_c^* N_v^*} e^{(E_v - E_c)/k_BT} = \sqrt{N_c^* N_v^*} e^{-E_g/k_BT},
$$

where $E_g = E_c - E_v$ is the energy gap. From the above relation and Lemma 2.2 we can deduce the Fermi energy of an intrinsic semiconductor:

$$q\mu = E_c + k_B T \log \frac{n}{N_c^*} = E_c + k_B T \log \frac{n_i}{N_c^*} = \frac{E_c + E_v}{2} + \frac{3}{4} k_B T \log \frac{m_h^*}{m_e^*}.$$

This means that at zero temperature the Fermi energy lies exactly in the middle of the energy gap, while for positive temperatures the correction to this value is of the order of $k_B T$. Moreover, since in most semiconductors at room temperature $E_g \gg k_B T$, it follows:

$$E - q\mu \geq E_c - q\mu = \frac{E_g}{2} - \frac{3}{4} k_B T \log \frac{m_h^*}{m_e^*} \gg k_B T,$$

which confirms the validity of the Maxwell-Bolzmann approximation.

Intrinsic semiconductors have indeed nice properties; however, it happens usually in applications that the intrinsic density is too small to result in a significant conductivity for nonzero temperature. To counter this inconvenient, the so-called process of *doping* is employed, which consists in replacing some atoms in the semiconductor crystal with atoms providing free electrons in the conduction band or free holes in the valence band, thus increasing the semiconductor conductivity. Such impurities are called *donors* if they supply additional electrons to the conduction band, while are called *acceptors* if they supply additional holes to the valence band. A semiconductor doped with donors is called an *n-type* semiconductor, while a semiconductor doped with acceptors is called a *p-type* semiconductor. If $N_d(x)$, $N_a(x)$ denote the densities of donors and acceptors (respectively), then $C(x) = N_a(x) - N_d(x)$ is called *doping profile*. The charge density ρ is given by:

$$\rho = q(n + N_d - p - N_a) = q(n - p - C).$$

Thus the Poisson equation, relating the electrostatic potential V generated by the charge distribution in the semiconductor and the charge density ρ, reads as:

$$-\varepsilon_s \Delta V = \rho = q(n - p - C),$$

where ε_s is the semiconductor permittivity.

3 Kinetic transport equations

We have seen in Subs. 1.2 that the *classical* motion of an ensamble of M electrons in a vacuum can be described by Newton's equations:

$$\dot{x} = v, \qquad \dot{v} = F/m, \qquad t > 0,$$

where the position vector $x \in \mathbb{R}^{3M}$, the velocity vector $v \in \mathbb{R}^{3M}$ and the force vector $F \in \mathbb{R}^{3M}$ decompose as $x = (x_1, \ldots, x_M)$, $v = (v_1, \ldots, v_M)$, $F = (F_1, \ldots, F_M)$, with $x_i \in \mathbb{R}^3$, $v_i \in \mathbb{R}^3$ position and velocity of the $i-$th particle, respectively, and $F_i \in \mathbb{R}^3$

electric force acting on the $i-$th particle, for $i = 1, \dots, M$. The electric force F_i, in case of conservative electric field, takes the form $F_i = -q\nabla_{x_i} V$, with V the electric potential and q the electron charge. The distribution function of the system $f = f(x, v, t)$ satisfies the Liouville equation in the $6M-$dimensional position-velocity phase space:

$$\partial_t f + v \cdot \nabla_x f + \frac{F}{m} \cdot \nabla_v f = 0, \qquad (x, v) \in \mathbb{R}^{6M}, \quad t > 0. \tag{68}$$

Alas, electrons in a semiconductor cannot be described classically, so (68) does not hold in the situations we are interested in. So the attentive Reader might wonder why we are losing time in writing equations that we do not plan to use; the answer is, that an equation similar to (68) actually holds for an electron ensemble in a semiconductor, and that such model can be obtained from a set of motion equations that is analogue to Newton's laws of classical mechanics. This is the topic of the next subsection.

3.1 Semi-classical Liouville equation.

In Subs. 2.4 we have seen that the motion of an electron ensemble in a semiconductor can be *semi-classically* described by the following equations:

$$\hbar \dot{x}_i = \nabla_{k_i} E_n(k_i), \qquad \hbar \dot{k}_i = F_i \qquad i = 1, \dots, M, \tag{69}$$

where E_n is the energy of the $n-$th band and $k_i \in B$ is the pseudo-wavevector, which belongs to the Brillouin zone B. Let $k = (k_1, \dots, k_M) \in \mathbb{R}^{3M}$. We consider now f as a function of (x, k, t) instead of (x, v, t). For the sake of simplicity, we assume that the electrons remain in the same energy band during the evolution of the system, so that we can drop the index n from E_n. To derive an equation for $f(x, k, t)$ we proceed like in Subs. 1.2. The starting point is the fact that f is constant along trajectories:

$$f(x(t), k(t), t) = f(x(0), k(0), 0) \qquad t > 0.$$

By differentiating both side of the above equality with respect to time we deduce:

$$0 = \frac{d}{dt} f(x(t), k(t), t) = \partial_t f + \dot{x} \cdot \nabla_x f + \dot{k} \cdot \nabla_k f,$$

which, by exploiting (69), can be rewritten as:

$$\partial_t f + \frac{1}{\hbar} \nabla_{k_i} E_n(k_i) \cdot \nabla_x f + \frac{1}{\hbar} F \cdot \nabla_k f = 0, \qquad x \in \mathbb{R}^{3M}, \ k \in B^M, \ t > 0. \tag{70}$$

Eq. (70) is called *semiclassical Liouville equation.* It must be complemented with initial conditions as well as boundary conditions, since B is bounded; usually periodic boundary conditions are chosen, mainly for simplicity reasons.

We point out that, in the parabolic band approximation the semi-classical Liouville equation reduces to the classical one. In fact, since $E(k) = \hbar^2 |k|^2 / 2m^*$, then $\hbar^{-1} \nabla_k E(k) = \hbar k / m^* = v$.

3.2 Semi-classical Vlasov equation.

Just like its classical cousin, the main disadvantage of the semi-classical Liouville equation is that it is an equation in a $6M-$dimensional space, where M is the number of electrons which, in physical situations, is very high. So the computational effort required to solve (70) with numerical methods is often prohibitive. But, luckily, we can circumvent this problem by deriving a single-particle equation, involving a distribution in the usual 6-dimensional phase space, that is equivalent to (70) under certain assumptions. The strategy that we will follow is:

(a) assume that the electric force has a suitable structure;

(b) integrate the Liouville equation in sub-phase spaces, thus obtaining a hierarchy of M equations;

(c) take the limit $M \to \infty$ (infinite number of particles);

(d) reduce the solution of the (infinite) hierarchy for the many particles system to the solution of a single-particle (nonlinear) equation, in which the many-particles effects are accounted for by a suitable effective field.

The Reader can confront this procedure with the derivaton of the Hartree-Fock equation in Subs. 2.3.

Let us first define, for $a \in \mathbb{N}$, $a < M$, the *sub-ensemble densities*[13] at time $t > 0$:

$$f^{(a)}(x_1 \dots x_a, k_1 \dots k_a, t) = \frac{1}{(4\pi^3)^{M-a}} \int_{(\mathbb{R}^3 \times B)^{M-a}} f \; dx_{a+1} \dots dx_M dk_{a+1} \dots dk_M,$$

and the initial sub-ensemble densities:

$$f_I^{(a)}(x_1 \dots x_a, k_1 \dots k_a) = \frac{1}{(4\pi^3)^{M-a}} \int_{(\mathbb{R}^3 \times B)^{M-a}} f_I \; dx_{a+1} \dots dx_M dk_{a+1} \dots dk_M.$$

We make now the following hypothesis:

1. **(Force field structure)** The force F_i acting on the particle i is the sum of a contribution from an external field E_{ext} and a contribution from an interaction field E_{int}:

$$F_i(x,t) = qE_{ext}(x_i,t) + \frac{q}{4\pi^3} \sum_{\substack{j=1 \\ j \neq i}}^{M} E_{int}(x_i, x_j),$$

where the interaction field E_{int} is anti-symmetric:

$$E_{int}(x_i, x_j) = -E_{int}(x_j, x_i) \qquad i, j = 1, \dots, M.$$

---

[13] The Reader can notice the similarity with the definition of the "marginal distributions" of Probability Theory.

This assumption means that the force F does not depend on k; in particular, no magnetic field is considered in this model. The anti-symmetry of E_{int} is simply the consequence of the action-reaction principle: the force exerted by the particle j on the particle i equals *minus* the force exerted by the particle i on the particle j.

2. **(Long range interactions)** The interaction field E_{int} is of order $1/M$ as $M \to \infty$, and the limit $E_0 = \lim_{M\to\infty} ME_{int}$ exists.

3. **(Undistinguishable particles)** The initial density is independent of the numbering of the particles:

$$f_I(x_1 \dots x_M, k_1 \dots k_M) = f_I(x_{\pi(1)} \dots x_{\pi(M)}, k_{\pi(1)} \dots k_{\pi(M)})$$

for any permutation π of the indices $1, \dots, M$. This, along with Assumption 1 on F, implies that the density is independent of the numbering of the particles *at any time instant* $t > 0$:

$$f(x_1 \dots x_M, k_1 \dots k_M, t) = f(x_{\pi(1)} \dots x_{\pi(M)}, k_{\pi(1)} \dots k_{\pi(M)}, t)$$

for any permutation π of the indices $1, \dots, M$. The meaning of this hypothesis is that each particle is identical to the other ones.

4. **(Initial chaos assumption)** For any $a < M$, the sub-ensemble initial density $f_I^{(a)}$ factorizes as:

$$f_I^{(a)}(x_1 \dots x_a, k_1 \dots k_a) = \prod_{i=1}^{a} f_I^*(x_i, k_i),$$

for some given function f_I^*. This means that at initial time the motion of the particles of a subensemble is decoupled, so that each particle moves independently from the other ones. See [16, Chap. 2.3] for more details.

Now it is time to integrate the Liouville equation in the sub-phase spaces:

$$\{(x_{a+1} \dots x_M, k_{a+1} \dots k_M) \in \mathbb{R}^{6(M-a)}\}, \qquad a = 1, \dots, M-1.$$

Thanks to Assumptions 1, 3 and to the fact that the interaction field $E_{int}(x_i, x_j)$ has the same form independently of the particular couple of particles (i, j) considered, we obtain after straightforward computations:

$$\partial_t f^{(a)} + \sum_{j=1}^{a} v(k_j) \cdot \nabla_{x_j} f^{(a)} + \frac{q}{\hbar} \sum_{j=1}^{a} E_{ext}(x_j, t) \cdot \nabla_{k_j} f^{(a)} + \frac{q}{\hbar} \sum_{j,\ell=1}^{a} E_{int}(x_j, x_\ell) \cdot \nabla_{k_j} f^{(a)}$$

$$+ \frac{q}{4\pi^3\hbar} \sum_{j=1}^{a} \operatorname{div}_{k_j} \int_{\mathbb{R}^3 \times B} (M-a) E_{int}(x_j, \bar{x}) f^{(a+1)}(x_1 \dots x_a, \bar{x}, k_1 \dots k_a, \bar{k}, t)\, d\bar{x} d\bar{k} = 0, \quad (71)$$

where $v(k) = \hbar^{-1} \nabla_k E(k)$ is the semiclassical velocity. Eq. (71) is called *the BBGKY hierarchy*, named after Bogoliubov, Born, Green, Kirkwood, Yvon.

Now we take the limit $M \to \infty$ in (71). We assume that $f^{(a)}$ converge in some sense to some limit as $M \to \infty$, and we denote such limit, with a slight abuse of notation, again with $f^{(a)}$. Thanks to Assumption 2 eq. (71) becomes:

$$\partial_t f^{(a)} + \sum_{j=1}^{a} v(k_j) \cdot \nabla_{x_j} f^{(a)} + \frac{q}{\hbar} \sum_{j=1}^{a} E_{ext}(x_j, t) \cdot \nabla_{k_j} f^{(a)}$$
$$+ \frac{q}{4\pi^3 \hbar} \sum_{j=1}^{a} \operatorname{div}_{k_j} \int_{\mathbb{R}^3 \times B} E_0(x_j, \bar{x}) f^{(a+1)}(x_1 \dots x_a, \bar{x}, k_1 \dots k_a, \bar{k}, t) d\bar{x} d\bar{k} = 0. \qquad (72)$$

Eq. (72) must be considered together with the initial condition:

$$f^{(a)}(x_1 \dots x_a, k_1 \dots k_a, 0) = f_I^{(a)}(x_1 \dots x_a, k_1 \dots k_a). \qquad (73)$$

We prove now the following result, stating that the solution of the many-particles hierarchy (72), (73) can be reduced to the solution of a one-particle equation:

Theorem 3.1 (Semi-classical Vlasov equation)**.** *Let the assumption 1–4 hold, and let f^* solve the* semiclassical Vlasov equation*:*

$$\partial_t f^* + v(k) \cdot \nabla_x f^* + \frac{q}{\hbar} E_{\mathit{eff}} \cdot \nabla_k f^* = 0, \qquad x \in \mathbb{R}^3,\ k \in B,\ t > 0, \qquad (74)$$

provided with the initial and boundary conditions:

$$\begin{aligned} &f^*(x,k,0) = f_I^*(x,k) \qquad x \in \mathbb{R}^3,\ k \in B, && (75)\\ &f^*(x,k,t) = f^*(x,-k,t) \qquad x \in \mathbb{R}^3,\ k \in \partial B,\ t > 0, && \end{aligned}$$

where f_I^* is as in Assumption 4, the effective field E_{eff} is given by:

$$E_{\mathit{eff}}(x,t) = E_{ext}(x,t) + \int_{\mathbb{R}^3} n(\bar{x},t) E_0(x,\bar{x}) d\bar{x}, \qquad (76)$$

and n is the electron density:

$$n(x,t) = \int_B f^*(x,k,t) \frac{dk}{4\pi^3}.$$

Then the family of functions:

$$f^{(a)} = \prod_{i=1}^{a} f^*(x_i, k_i, t), \qquad a \in \mathbb{N}, \qquad (77)$$

is a solution of (72)*,* (73)*.*

Proof. Let us define

$$Q_i = \prod_{\substack{j=1\\ j\neq i}}^{a} f^*(x_j, k_j, t) \qquad i = 1, \dots, a.$$

Let us evaluate (74) for $(x,k) = (x_i, k_i)$, then multiply it times Q_i and sum for $i = 1, \dots, a$. It holds:

$$\sum_{i=1}^{a} Q_i \partial_t f^*(x_i, k_i, t) = \sum_{i=1}^{a} \prod_{\substack{j=1\\ j\neq i}}^{a} f^*(x_j, k_j, t) \partial_t f^*(x_i, k_i, t) = \partial_t \prod_{i=1}^{a} f^*(x_i, k_i, t) = f^{(a)}.$$

Analogously,

$$\sum_{i=1}^{a} Q_i v(k_i) \cdot \nabla_{x_i} f^*(x_i, k_i, t) = \sum_{i=1}^{a} v(k_i) \cdot \nabla_{k_i} f^{(a)},$$

$$\sum_{i=1}^{a} Q_i E_{ext}(x_i, t) \cdot \nabla_{k_i} f^*(x_i, k_i, t) = \sum_{i=1}^{a} E_{ext}(x_i, t) \cdot \nabla_{k_i} f^{(a)}.$$

Moreover,

$$\begin{aligned}
&\sum_{i=1}^{a} Q_i \left(\int_{\mathbb{R}^3 \times B} f^*(\bar{x}, \bar{k}, t) E_0(\bar{x}, \bar{k}) d\bar{x} d\bar{k} \right) \cdot \nabla_{k_i} f^*(x_i, k_i, t) \\
&\quad = \sum_{i=1}^{a} \operatorname{div}_{k_i} \int_{\mathbb{R}^3 \times B} \prod_{j=1}^{a} f^*(x_j, k_j, t) f^*(x_i, k_i, t) E_0(x_i, \bar{x}) d\bar{x} d\bar{k} \\
&\quad = \sum_{i=1}^{a} \operatorname{div}_{k_i} \int_{\mathbb{R}^3 \times B} f^{(a+1)}(x_1 \dots x_a, \bar{x}, k_1 \dots k_a, \bar{k}, t) d\bar{x} d\bar{k}.
\end{aligned}$$

By putting the above equations together we conclude that (77) solves (72). Assumption 4 ensures that (73) holds as well. □

So the evolution of an electron ensemble in a semiconductor can be described semiclassically by means of a one-particle equation, which incorporates the many-particle effects through an effective field. The drawback of this approach is that the Vlasov equation, contrarily to the Liouville equation, is nonlinear, since the effective field depends on the distribution f through the electron density n. Eq. (74) is thus a scalar equation with a nonlocal quadratic nonlinearity. It is valid under the assumption that the interactions are long-range; in particular, we have not considered any scattering mechanism in the derivation of the equations. Collisions will be taken into consideration in the next subsection.

3.3 Semi-Classical Boltzmann equation

In deriving the semiclassical Vlasov equation (74) we have neglected collisional processes. But collisions do exist; so we must take them into account eventually. This is conventionally done by adding a term on the right-hand side of (74), which represents the rate of change of the density f due to collisions:

$$\partial_t f + v(k) \cdot \nabla_x f + \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k f = Q(f), \qquad x \in \mathbb{R}^3,\ k \in B,\ t > 0, \tag{78}$$

where the effective field E_{eff} is given by (76). We will now derive an expression for Q. We assume that:

- the scattering of particles is instantaneous;
- scattering only affects the pseudo-wavevector k.

It is convenient to consider the quantity $P(x, k' \to k, t)$, defined as the rate at which a particle with position x and wavevector k' changes its wavevector to k due to a collision, at time t. Thus $Q(f)$, the rate of change of f due to collisions, evaluated at (x, k, t), will be equal to the sum (integral), with respect to all possible wavevectors $k' \in B$, of $P(x, k' \to k, t)$ (gain term) minus $P(x, k \to k', t)$ (loss term):

$$(Q(f))(x, k, t) = \int_B (P(x, k' \to k, t) - P(x, k \to k', t))\, dk'. \tag{79}$$

It is intuitively clear that $P(x, k' \to k, t)$ should be proportional to $f(x, k', t)$. It is perhaps less evident that $P(x, k' \to k, t)$ should also be proportional to $1 - f(x, k, t)$; however, this is a consequence of the Pauli exclusion principle: since every quantum state can only be occupied by at most one electron, it follows that $P(x, k' \to k, t)$ must vanish if there are no available states with wavevector k, that is, if $f(x, k, t) = 1$. So:

$$P(x, k' \to k, t) = s(x, k' \to k) f(x, k', t)(1 - f(x, k, t)), \tag{80}$$

where $s(x, k' \to k)$ is called *scattering rate*. Let us define the shorthands: $f = f(x, k, t)$, $f' = f(x, k', t)$. From (79), (80) we obtain:

$$(Q(f))(x, k, t) = \int_B (s(x, k' \to k) f'(1 - f) - s(x, k \to k') f(1 - f'))\, dk'. \tag{81}$$

Equation (78), with Q given by (81), is called *semiclassical Boltzmann equation*. It is a nonlinear equation with quadratic nonlinearities in both the potential and the collisional term; moreover, these terms are also nonlocal, that is, any compactly-supported perturbation of f alters the value of V and Q in any point of the phase space. To be precise, V is nonlocal in the $x-$space (being given by the convolution of $n - C$ with $1/|x|$, the fundamental harmonic of $\mathbb{R}^3$), while Q is nonlocal in the $k-$space (since it is an integral with respect to the wavevector).

In the case of Coulomb fields:

$$E_0(x, y) = \frac{q}{4\pi\varepsilon_s} \frac{x - y}{|x - y|^3} = -\nabla_x \left(\frac{q}{4\pi\varepsilon_s} \frac{1}{|x - y|} \right), \tag{82}$$

$$E_{ext}(x,t) = -\frac{q}{4\pi\varepsilon_s}\int_{\mathbb{R}^3} C(y)\frac{x-y}{|x-y|^3}dy = -\nabla_x\left(-\frac{q}{4\pi\varepsilon_s}\int_{\mathbb{R}^3}\frac{C(y)}{|x-y|}dy\right), \tag{83}$$

it is easy to see that curl $E_{\text{eff}} = 0$, $\varepsilon_s \text{div}\, E_{\text{eff}} = q(n - C)$; as a consequence, $E_{\text{eff}} = -\nabla V$ for some potential V satisfying the Poisson equation:

$$-\varepsilon_s \Delta V = q(n - C). \tag{84}$$

Eqs. (78), (84), coupled with the definition (81) of Q and the relation $E_{\text{eff}} = -\nabla V$, are called the *Boltzmann-Poisson system.*

Let us summarize the features of the semiclassical Boltzmann equation.

- The semiconductor Boltzmann equation is a single-particle model of a many-particle system; interactions between particles are incorporated through an effective field.
- Quantum mechanical phenomena are accounted for through the semiclassical approximation: electrons are assumed to be described by the semiclassical equations of motion.
- Collisions are assumed to be binary, instantaneous, and local in space.
- The Boltzmann equation provides a statistical description of the electron ensemble, which is valid only if the number of particles in the system is large enough.

3.3.1 More about scattering.

The structure of the scattering rate $s(x, k' \to k, t)$ can be made more explicit by considering more specific collision processes. As a matter of fact, the most important scattering events in semiconductor crystals are:

- carrier-carrier (electron-hole, electron-electron, hole-hole) scattering;
- ionized impurity scattering;
- electron-phonon scattering.

Among these three scattering mechanism, the most important is the third one. Thus we consider here only the electron-phonon scattering, and we neglect (for the sake of simplicity) the other types of scattering.

What is a phonon? At positive temperatures, the ions in a semiconductor crystal vibrate around equilibrium positions. These vibrations are quantized; the quantum of lattice vibration is called a *phonon.* Actually, a phonon is not a physical particle, but a *virtual* particle that has been invented by physicists to describe lattice vibrations and their interactions with the charge carriers. Ions can vibrate in both the physical space and the momentum space: the phonons that arise from such vibrations are called *acoustic* and *optic*, respectively. Intuitively, acoustic phonons can be thought as sound waves, while optic phonons can be regarded as electromagnetic waves (they are indeed able to interact with light). Phonons are bosons (that is, the wavefunction of an ensamble of indistinguishable phonons is symmetric with respect to the exchange of the particles labels), and thus the phonon occupation number $\mathcal{N}(\hbar\omega)$ (the number of phonons with energy $\hbar\omega$ at temperature T, ω being the frequency) is computed from Bose-Einstein statistics:

$$\mathcal{N}(\hbar\omega) = \frac{1}{e^{\hbar\omega/k_B T} - 1}. \tag{85}$$

Let us suppose that an electron in the Bloch state k' with energy $E(k')$ experiences a collision with a phonon with energy $\hbar\omega$. Due to this collision, the electron can change its state from k' to k if the (absolute value of the) energy variation $E(k) - E(k')$ equals $\hbar\omega$, namely:

$$E(k) - E(k') = \pm\hbar\omega,$$

where the plus and minus signs hold in case of photon absorption or emission, respectively. Since the scattering rate s is nonzero only if the above relation is satisfied, we can rewrite s in the following way:

$$s(x, k' \to k) = \sigma_{emi}(x, k, k')\delta(E(k') - E(K) - \hbar\omega) + \sigma_{abs}(x, k, k')\delta(E(k') - E(K) + \hbar\omega),$$

for some suitable functions $\sigma_{emi}(x, k, k')$, $\sigma_{abs}(x, k, k')$. Actually, it is possible to show (with quantum mechanical arguments, see [54] for details) that s takes the form:

$$s(x, k' \to k) = \sigma_0(x, k, k')\Big((1 + \hat{N})\delta(E' - E + \hbar\omega) + \hat{N}\delta(E' - E - \hbar\omega)\Big), \tag{86}$$

where $E \equiv E(k)$, $E' \equiv E(k')$, and $\sigma_0(x, k, k')$ is symmetric in k, k'. The first delta represents the contribution of scattering events in which a phonon is absorbed, while the second delta accounts for collisions characterized by phonon emissions. In the case of acoustic phonons at room temperature, $\hbar\omega$ can be neglected with respect to the electron energy. Under this approximation s becomes:

$$s_{ac}(x, k' \to k) = \sigma_{ac}(x, k, k')\delta(E' - E), \tag{87}$$

with $\sigma_{ac} = (2\hat{N} + 1)\sigma_0$. The symmetry of the delta distribution ("$\delta(x) = \delta(-x)$") implies:

$$\int_B (s_{ac}(x, k' \to k)f'f - s_{ac}(x, k \to k')ff')\, dk' = \int_B \sigma_{ac}\delta(E' - E)(f'f - ff')dk' = 0,$$

so the collision operator becomes linear:

$$(Q_{ac}(f))(x, k, t) = \int_B \sigma_{ac}\delta(E' - E)(f' - f)dk'. \tag{88}$$

3.3.2 Approximations of the collision operator and their properties.

We consider here two approximations of the collision operator: the low-density and elastic approximations.

The low-density approximation is based upon these two assumptions:

- the quadratic (w.r.t. f) terms inside Q are small and can be neglected;
- the so-called *principle of detailed balance* [6] holds:

$$s(x,k'\to k)f'_{eq}(1-f_{eq})-s(x,k\to k')f_{eq}(1-f'_{eq})=0.$$

In the above equation f_{eq} is the Fermi-Dirac distribution:

$$f_{eq}(k)=\frac{1}{1+e^{(E(k)-E_F)/k_BT}}.$$

The principle of detailed balance states that the scattering probability at local thermal equilibrium vanishes. An (easy to verify) consequence of this assumption is:

$$\frac{s(x,k'\to k)}{M(k)}=\frac{s(x,k\to k')}{M(k')},$$

which means that the so-called *cross section* $\sigma(x,k,k')=s(x,k'\to k)/M(k)$ is symmetric in k, k'. From this fact and our assumptions it follows that $Q(f)$ can be approximated by the *low-density collision operator*:

$$(Q_0(f))(x,k,t)=\int_B \sigma(x,k,k')(Mf'-M'f)dk'. \tag{89}$$

The following proposition states several properties of Q_0.

Proposition 3.1 (Low-density collision operator)**.** *For all functions $f\geq 0$ and $\chi:\mathbb{R}\to\mathbb{R}$, with χ nondecreasing, it holds:*

$$\int_B Q_0(f)dk=0, \tag{90}$$

$$\int_B Q_0(f)\chi\left(\frac{f}{M}\right)dk\leq 0, \tag{91}$$

$$Q_0(f)=0 \quad\Leftrightarrow\quad \exists g=g(x):\ f(x,k)=g(x)M(k). \tag{92}$$

Proof. Let $\chi:\mathbb{R}\to\mathbb{R}$ a (smooth) function. Let us consider:

$$\int_B Q_0(f)\chi(f/M)dk=\int_B \sigma MM'\left(\frac{f'}{M'}-\frac{f}{M}\right)\chi\left(\frac{f}{M}\right)dk'dk.$$

Since σ is symmetric in k, k', by exchanging k and k' in the integral on the right-hand side of the above equation we get:

$$\int_B Q_0(f)\chi(f/M)dk = -\int_B \sigma M M' \left(\frac{f'}{M'} - \frac{f}{M}\right) \chi\left(\frac{f'}{M'}\right) dk'dk.$$

By summing up the previous two equations we obtain:

$$2\int_B Q_0(f)\chi(f/M)dk = -\int_B \sigma M M' \left(\frac{f'}{M'} - \frac{f}{M}\right)\left(\chi\left(\frac{f'}{M'}\right) - \chi\left(\frac{f}{M}\right)\right) dk'dk. \tag{93}$$

If we take $\chi \equiv 1$ in (93) we obtain (90). If χ is an arbitrary nondecreasing function, then the right-hand side of (93) is clearly nonpositive, so (91) follows. Finally, if $Q_0(f) = 0$, then from (93) with $\chi(x) = x$ we obtain:

$$\int_B \sigma M M' \left(\frac{f'}{M'} - \frac{f}{M}\right)^2 dk'dk = 0,$$

which proves the implication $\Rightarrow$ in (92). The implication $\Leftarrow$ follows immediately from the definition (89) of Q_0. ☐

Another approximation of the collision operator Q is constituted by the *elastic collision operator*:

$$(Q_{el}(f))(x,k,t) = \int_B \sigma(x,k,k')\delta(E - E')(f' - f)dk'. \tag{94}$$

We assume that the cross section σ is positive and symmetric in k, k'. This operator has the following properties:

$$\int_B Q_{el}(f)dk = \int_B Q_{el}(f)E(k)dk = 0, \tag{95}$$

$$\int_B Q_{el}(f)f dk \leq 0, \tag{96}$$

$$Q_{el}(f) = 0 \quad \Leftrightarrow \quad \exists F = F(x,\mathcal{E},t) : \ f(x,k,t) = F(x,E(k),t). \tag{97}$$

The proof of (95)–(97) is similar to the proof of (90)–(92) and thus we omit it; instead, we wish to point out the physical meaning of (90), (95). By integrating the semiclassical Boltzmann equation (78) in $\mathbb{R}^3 \times B$ with respect to x, k, we obtain:

$$\partial_t n = \int_{\mathbb{R}^3 \times B} Q(f)dxdk.$$

If $Q = Q_0$, then (90) implies that $\partial_t n = 0$. Thus (90) expresses the conservation of total mass.

Let us now assume that $E_{ext} = 0$ (no external forces) and let us define the energy density ne as $ne = \int fE(k)dk/4\pi^3$. By multiplying (78) times $E(k)$ and then integrating it in $\mathbb{R}^3 \times B$ with respect to x, k, we obtain:

$$\partial_t(ne) = \int_{\mathbb{R}^3 \times B} Q(f)E(k)dxdk.$$

If $Q = Q_{el}$, then (95) implies that $\partial_t(ne) = 0$ and, since (90) is contained into (95), $\partial_t n = 0$ as well. Thus (95) expresses the conservation of total mass *and* total energy.

4 Drift-Diffusion Equations

It is time that we put what we studied in the previous section into practice, namely, that we derive some macroscopic transport model for semiconductors. This section is devoted to the derivation of drift-diffusion equations, which, among all semiconductor fluid models, are the simplest ones, and were derived from the first time by Van Roosbroeck in 1950 [55].

Everything begins with the semiclassical Boltzmann equation (78) coupled with the Poisson equation (84) (the effective field E_{eff} being given by $E_{\text{eff}} = -\nabla V$) and provided with an initial condition:

$$\begin{aligned}
\partial_t f + v(k)\cdot\nabla_x f - \frac{q}{\hbar}\nabla_x V \cdot \nabla_k f &= Q(f) && x \in \mathbb{R}^3,\ k \in B,\ t > 0, && (98)\\
-\varepsilon_s \Delta V &= q(n - C) && x \in \mathbb{R}^3,\ t > 0, && (99)\\
n &= \int_B f \frac{dk}{4\pi^3} && x \in \mathbb{R}^3,\ t > 0, &&\\
f(x,k,0) &= f_I(x,k) && x \in \mathbb{R}^3,\ k \in B. &&
\end{aligned}$$

In (98), $v(k) = \hbar^{-1}\nabla_k E(k)$ is the average velocity in the considered energy band, $E = E(k)$ is the energy, $Q(f)$ is the collision operator, $n = n(x,t)$ is the electron density, $C = C(x)$ is the doping profile, B is the Brillouin zone. The drift-diffusion model is derived from the kinetic equation (99) under the assumption that the timescale of collisions is much shorter than the timescale of transport, which is much shorter than the reference timescale (this statement will be clear in the following subsection).

4.1 Scaling of the Boltzmann-Poisson system

In mathematical physics, it is important to know what is big and what is small. For this reason, we are going to scale eqs. (98), (99). Let us define the following reference quantities (in the following, $T > 0$ is the lattice temperature):

- the domain diameter L;
- the mean free path λ, which is the average distance traveled by a particle between two subsequent collisions;

- the reference length $\lambda_0 = \sqrt{L\lambda}$;
- the reference velocity $v_0 = \sqrt{k_B T/m^*}$;
- the reference potential $U = k_B T/q$;
- the reference time $\tau_0 = L/v_0$, which is the average time required by a particle to cross the domain;
- the mean free time $\tau = \lambda/v_0$, which is the average time interval between two subsequent collisions experienced by a particle;
- the reference wave vector $k_0 = m^* v_0/\hbar$;
- the reference electron density $n_0 = k_0^3$.

Notice that $m^* v_0^2/2 = qU/2 = k_B T/2$, that is, the kinetic, electric and thermal energy (respectively) are of the same order of magnitude. This implies, in particular, that the electric fields are small.

We scale the variables as follows:

$$\begin{gathered}t = t_0 t_s, \quad x = x_0 x_s, \quad k = k_0 k_s, \quad v(k) = v_0 v_s(k_s), \\ V = UV_s, \quad n = n_0 n_s, \quad C = n_0 C_s, \quad Q(f) = \tau^{-1} Q_s(f).\end{gathered}$$

where the quantities with index s are a-dimensional. By plugging the above relations inside (98), (99) and playing with the physical constants we obtain:

$$\alpha^2 \partial_{t_s} f + \alpha \left(v_s(k_s) \cdot \nabla_{x_s} f - \nabla_{x_s} V \cdot \nabla_{k_s} f \right) = Q_s(f), \tag{100}$$

$$-\lambda_D^2 \Delta V_s = n_s - C_s, \tag{101}$$

where:

$$\alpha = \frac{\lambda}{\lambda_0} = \frac{\text{mean free path}}{\text{reference length}}, \qquad \lambda_D = \sqrt{\frac{\varepsilon_s U}{q \lambda_0^2 k_0^3}} \qquad \text{(scaled Debye length).} \tag{102}$$

We make the following assumptions [43, 45]:

1 **(Parabolic band approximation)** The band energy $E(k)$ takes the form: $E(k) = E_c + \hbar^2 |k|^2 / 2m^*$, where E_c is the conduction band minimum. In particular, the average velocity v is given by $v = \hbar k/m^*$.

2 **(Low density collision operator)** The (unscaled) collision operator is given by:

$$Q(f) = \int_B \sigma(x, k, k')(Mf' - M'f) dk',$$

where the cross-section $\sigma(x, k, k')$ is positive and symmetric in k and k', $M(k) = e^{-E(k)/k_B T}$ is the (unscaled) Maxwellian, $f = f(x, k, t)$, $f' = f(x, k', t)$. So $Q(f)$ is linear in f; moreover, the presence of the Maxwellian in place of the Fermi-Dirac distribution implies that we are considering nondegenerate materials.

3 **(Frequent collisions)** The parameter α (defined in (102)) is small. This means that collisions occur frequently in the material.

Under the above-written assumptions, the scaled collision operator Q becomes:

$$Q(f) = \int_B s(x,k,k')(Mf' - M'f)dk', \quad s(x,k,k') = \frac{2}{(2\pi)^{3/2}} e^{-E_c/k_B T} \sigma(x,k,k'), \qquad (103)$$

and the adimensional Maxwellian is $M = 2^{-1}(2\pi)^{3/2} e^{-|k|^2/2}$ (notice that $\int_{\mathbb{R}^3} M dk/4\pi^3 = 1$). Finally, we point out that the parabolic band approximation allows us to replace the Brillouin zone B in the collision integral (103) with the whole space $\mathbb{R}^3$, since the Maxwellian M tends to zero very quickly as $|k| \to \infty$, and thus the contribution of $\mathbb{R}^3 \backslash B$ to the integral will be fairly small (we assume here that f is bounded at infinity, or shows at most a polynomial growth).

Remark. Assumption 3 in the above list clarifies the statement at the beginning of this section: in deriving the drift-diffusion model, we consider a situation in which the timescale of collisions is much shorter than the timescale of transport, which is much shorter than the reference timescale. In fact, in the scaled Boltzmann equation (100), due to the smallness of α, the collision term dominates the transport term, which on turn dominates the time derivative of w.

4.2 Properties of the low-density collision operator

In this subsection we study the kernel $N(Q)$ and the range $R(Q)$ of the collision operator (103), defined (in the usual way) as:

$$N(Q) = \{f \ : \ Q(f) = 0\}, \quad R(Q) = \{g \ : \ \text{there exists an } f \text{ such that } Q(f) = g\}.$$

Let us define the so-called *collision frequency*:

$$S(x,k) = \int_{\mathbb{R}^3} s(x,k,k')M(k')dk',$$

and the Hilbert space:

$$X = \left\{ f \in L^2(\mathbb{R}^3) \ : \ \int_{\mathbb{R}^3} f^2 S M^{-1} dk < \infty \right\},$$

provided with the scalar product:

$$(f_1, f_2)_X = \int_{\mathbb{R}^3} f_1 f_2 S M^{-1} dk \qquad f_1, f_2 \in X.$$

Thanks to Prop. 3.1, we know that $N(Q) = \text{Span}\{M\}$. What about $R(Q)$? The following result answers this question:

Lemma 4.1. *For a given g, the equation $Q(f) = g$ has a solution $f \in X$ if and only if $\int_{\mathbb{R}^3} gdk = 0$. Moreover, there is only one solution f satisfying $\int_{\mathbb{R}^3} fSdk = 0$.*

Proof. We symmetrize Q in order to deal with an operator acting on L^2 functions. Let $f_s = \sqrt{S/M}f$, $Q_s(f_s) = (SM)^{-1/2}Q(f)$. Straightforward computations lead to:

$$Q_s(f_s) = \int_{\mathbb{R}^3} s(x,k,k') \left(\frac{MM'}{SS'}\right)^{1/2} f'_s dk' - f_s.$$

Since $s(x,k,k') = s(x,k',k)$, then $Q_s : L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^3)$ is symmetric. Moreover, it is possible to prove that $R(Q_s)$ is closed; as a consequence, a known result [12, Thr. 2.19] implies that $R(Q_s) = N(Q_s^*)^\perp = N(Q_s)^\perp$. It follows that $Q(f) = g$ has a solution if and only if $g \in N(Q_s)^\perp$, that is:

$$\int_{\mathbb{R}^3} g_s h dk = 0 \qquad \forall f \in N(Q_s).$$

Since $N(Q) = \text{Span}\{M\}$, then $N(Q_s) = \text{Span}\{\sqrt{SM}\}$, so the above relation becomes:

$$0 = \int_{\mathbb{R}^3} g_s \sqrt{SM} dk = \int_{\mathbb{R}^3} gdk.$$

To prove the uniqueness part, we exploit the following fact: a constant $c > 0$ exists such that

$$-\int_{\mathbb{R}^3} Q_s(f_s) f_s dk \geq c\|f_s\|^2_{L^2} \qquad \forall f \in N(Q_s)^\perp. \tag{104}$$

Property (104) is called *hypocoercivity* (since it is a coercivity property holding not on the whole Hilbert space X, but only on a smaller space). For the proof see [45] or [18, Lemma 10]. This property implies that Q_S in one-to-one on:

$$\begin{aligned} N(Q_s)^\perp &= \left\{ f_s \ : \ \int_{\mathbb{R}^3} f_s w dk = 0 \quad \forall w \in N(Q_s) = \text{Span}\{\sqrt{SM}\} \right\} \\ &= \left\{ f_s \ : \ \int_{\mathbb{R}^3} f_s \sqrt{SM} dk = 0 \right\}, \end{aligned}$$

which means, thanks to the definition of f_s, that Q is one-to-one on $\left\{ f \ : \ \int_{\mathbb{R}^3} fSdk = 0 \right\}$. This finishes the proof. $\square$

A consequence of the previous lemma is that the equations

$$Q(h_i) = k_i M(k), \qquad i = 1, 2, 3, \tag{105}$$

have solutions, since $\int_{\mathbb{R}^3} k_i M(k) dk = 0$ (M is even). We state (without proving it) the following technical result:

Lemma 4.2. *Assume that, for all orthogonal matrices* $A \in \mathbb{R}^{3\times 3}$*, it holds:*

$$s(x, Ak, Ak') = s(x, k, k') \qquad x, k, k' \in \mathbb{R}^3.$$

Then there exists a scalar function $\mu_0 = \mu_0(x) \geq 0$ *such that the solutions* h_i *of* (105) *satisfy:*

$$\int_{\mathbb{R}^3} k_i h_j \frac{dk}{4\pi^3} = -\mu_0 \delta_{ij} \qquad (i, j = 1, 2, 3).$$

4.3 Derivation of the drift-diffusion equations

We are now ready to derive the drift-diffusion model. We denote with (f_α, V_α) a solution of the scaled Boltzmann-Poisson system:

$$\alpha^2 \partial_t f_\alpha + \alpha \left(k \cdot \nabla_x f_\alpha - \nabla_x V_\alpha \cdot \nabla_k f_\alpha \right) = Q_\alpha(f_\alpha), \tag{106}$$

$$-\lambda_D^2 \Delta V_\alpha = n_\alpha - C, \tag{107}$$

$$n_\alpha = \int_{\mathbb{R}^3} f_\alpha \frac{dk}{4\pi^3}.$$

We assume that (f, V) exists such that $(f_\alpha, V_\alpha) \to (f, V)$ as $\alpha \to 0$ in some suitable (distributional) sense, along with their derivatives. The derivation of the model is made up by three steps.

Step 1. Let us take the formal limit $\alpha \to 0$ in (106). it follows that $Q(f) = 0$; since $N(Q) = \mathrm{Span}\{M\}$, this implies that $f(x, k, t) = n(x, t) M(k)$ with $n = \int_{\mathbb{R}^3} f \frac{dk}{4\pi^3}$.

Step 2. The result obtained in the previous step[14] allows us to write $f_\alpha = nM + \alpha g_\alpha$ for some g_α that is bounded as $\alpha \to 0$. This is the so-called *Chapman-Enskog expansion.* We assume that g_α converges to some g as $\alpha \to 0$. Let us insert this expansion into (106). Thanks to the fact that $Q(nM) = 0$, after a division by α we deduce:

$$\alpha \partial_t f_\alpha + k \cdot \nabla_x (nM) - \nabla_x V_\alpha \cdot \nabla_k (nM) + \alpha \left(k \cdot \nabla_x g_\alpha - \nabla_x V_\alpha \cdot \nabla_k g_\alpha \right) = Q(g_\alpha).$$

Since $\nabla_k M = -kM$, taking the limit $\alpha \to 0$ in the above equation yields:

$$Q(g) = (\nabla_x n + n \nabla_x V) \cdot kM. \tag{108}$$

Step 3. Let us integrate (106) with respect to k and divide by α^2:

$$\partial_t \int_{\mathbb{R}^3} f_\alpha \frac{dk}{4\pi^3} + \alpha^{-1} \int_{\mathbb{R}^3} \left(k \cdot \nabla_x f_\alpha - \nabla_x V_\alpha \cdot \nabla_k f_\alpha \right) \frac{dk}{4\pi^3} = \alpha^{-2} \int_{\mathbb{R}^3} Q_\alpha(f_\alpha) \frac{dk}{4\pi^3}. \tag{109}$$

One of the properties of Q_α is mass conservation: $\int_{\mathbb{R}^3} Q_\alpha(f) dk = 0$ for all f. Moreover the divergence theorem implies that $\int_{\mathbb{R}^3} \nabla_k f_\alpha dk = 0$, while from the expansion $f_\alpha = nM + \alpha g_\alpha$

---

[14] We also assume that f_α depends smoothly on α.

and the fact that M is an even function of k it follows that $\int_{\mathbb{R}^3} kMdk = 0$. Thus (109) becomes:

$$\partial_t \int_{\mathbb{R}^3} f_\alpha \frac{dk}{4\pi^3} + \operatorname{div}_x \int_{\mathbb{R}^3} kg_\alpha \frac{dk}{4\pi^3} = 0. \tag{110}$$

Since $\int_{\mathbb{R}^3} f_\alpha dk/4\pi^3 = n + \alpha \int_{\mathbb{R}^3} g_\alpha dk/4\pi^3$, by taking the limit $\alpha \to 0$ in (110) we obtain:

$$\partial_t n + \operatorname{div}_x \int_{\mathbb{R}^3} kg \frac{dk}{4\pi^3} = 0. \tag{111}$$

From (108) we deduce that $g = (\nabla_x n + n\nabla_x V) \cdot h$ with $h = (h_1, h_2, h_3)$ and $Q(h_i) = k_i M$ for $i = 1, 2, 3$. From Lemma 4.2 it follows that:

$$\int_{\mathbb{R}^3} k_i g \frac{dk}{4\pi^3} = \sum_{j=1}^{3} \left(\int_{\mathbb{R}^3} k_i h_j dk \right) \left(\partial_{x_j} n + n\partial_{x_j} V \right) = -\mu_0 \left(\partial_{x_i} n + n\partial_{x_i} V \right). \tag{112}$$

By putting together (111), (112) we conclude:

$$\partial_t n + \operatorname{div}_x J_n = 0, \qquad J_n = -\mu_0(\nabla_x n + n\nabla_x V). \tag{113}$$

By taking the limit $\alpha \to 0$ in (107) we get:

$$-\lambda_D^2 \Delta V = n - C. \tag{114}$$

Eqs. (113), (114) constitute the scaled *drift-diffusion model*. Going back to the physical variables, eqs. (113), (114) become:

$$\partial_t n + q^{-1}\operatorname{div}_x J_n = 0, \qquad J_n = -q\mu_n(U\nabla_x n + n\nabla_x V), \tag{115}$$

$$-\varepsilon_s \Delta V = q(n - C), \tag{116}$$

where $\mu_n = (q\tau/m^*)\mu_0$.

Recall that the derivation of the drift-diffusion model relies on the following physical assumptions:

- the mean free path λ_0 is much smaller than the reference length λ;
- the electric potential is of the order of $|U| = 0.026$ V.

In spite of the above limitations, the drift-diffusion model is also used for higher applied voltages. In practice, it gives reasonable results as long as the characteristic length is bigger than $\lambda \approx 10^{-7}$ m.

4.4 The bipolar model

We have derived a drift-diffusion model for electrons. But also holes contribute to the total current density. Since holes are identical to electrons, except for the fact that they have positive charge, then (115) can be rewritten for the hole density p as:

$$\partial_t p - q^{-1}\mathrm{div}_x J_p = 0, \qquad J_p = q\mu_p(U\nabla_x p - p\nabla_x V), \tag{117}$$

where $\mu_p \geq 0$ is the hole mobility. Usually in applications it holds $\mu_p < \mu_n$ (for example, in silicon: $\mu_n = 1400$ cm^2/Vs, $\mu_p = 450$ cm^2/Vs). Notice that the sign of the drift term $p\nabla_x V$ in (117) is opposite to the sign of the corresponding term $-n\nabla_x V$ in (115); this reflects the fact that electrons and holes move in opposite directions due to the electric field.

In order to describe the process of generation-recombination of electrons and holes, we employ the so-called *Shockley-Read-Hall recombination-generation term* (see [11, Chap. 10] or [29, Chap. 10] for details):

$$R(n,p) = \frac{np - n_i^2}{\tau_p(n+n_d) - \tau_n(p+p_d)}. \tag{118}$$

Here τ_n, τ_p are the carrier life times,

$$n_d = N_c e^{(E_t - E_c)/k_B T}, \qquad p_d = N_v e^{(E_v - E_t)/k_B T}, \tag{119}$$

$$n_i = \sqrt{n_d p_d} = 2\left(\frac{\sqrt{m_e^* m_v^*}k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-(E_c - E_v)/2k_B T}. \tag{120}$$

The quantities N_c, N_v are the carrier effective densities of states, n_i is the intrinsic density, and E_t is called *trap energy level*: it is an energy level in the forbidden band region, caused by crystal impurities, which facilitates the formation of electron-hole pairs, since the jump from the valence to the conduction band can be split into two parts.

The following system of PDEs:

$$\partial_t n + q^{-1}\mathrm{div}_x J_n = -R(n,p), \qquad J_n = -q\mu_n(U\nabla_x n + n\nabla_x V), \tag{121}$$

$$\partial_t p - q^{-1}\mathrm{div}_x J_p = -R(n,p), \qquad J_p = q\mu_p(U\nabla_x p - p\nabla_x V), \tag{122}$$

$$-\varepsilon_s \Delta V = q(n - p - C), \tag{123}$$

with $R(n,p)$ given by (118)–(120), constitute the *bipolar model*. Notice that the Poisson equation changes as well, since the total carrier density $n - p$ must be taken into account in place of the electron density.

4.4.1 Thermal equilibrium state

A thermal equilibrium state is a time-independent solution $(n^{\mathrm{eq}}, p^{\mathrm{eq}}, V^{\mathrm{eq}})$ of (121)–(123) with zero current flow:

$$\partial_t n^{\mathrm{eq}} = \partial_t p^{\mathrm{eq}} = 0, \quad J_n^{\mathrm{eq}} = J_p^{\mathrm{eq}} = 0 \quad \text{in } \Omega.$$

This implies that $R(n^{\mathrm{eq}}, p^{\mathrm{eq}}) = 0$ and so $n^{\mathrm{eq}} p^{\mathrm{eq}} = n_i^2$. Moreover:

$$\begin{aligned} 0 &= U\nabla n^{\mathrm{eq}} + n^{\mathrm{eq}}\nabla V = n^{\mathrm{eq}}\nabla(U\log n^{\mathrm{eq}} + V^{\mathrm{eq}}), \\ 0 &= U\nabla p^{\mathrm{eq}} - p^{\mathrm{eq}}\nabla V^{\mathrm{eq}} = p^{\mathrm{eq}}\nabla(U\log p^{\mathrm{eq}} - V^{\mathrm{eq}}). \end{aligned}$$

Since it is physically reasonable to assume that n^{eq}, p^{eq} are positive, it follows that $U\log n^{\mathrm{eq}} + V^{\mathrm{eq}}$, $U\log p^{\mathrm{eq}} - V^{\mathrm{eq}}$ are constant. So:

$$n^{\mathrm{eq}} = e^{(\alpha - V^{\mathrm{eq}})/U}, \qquad p^{\mathrm{eq}} = e^{(\beta + V^{\mathrm{eq}})/U},$$

for some constants α, β. Since V^{eq} is defined up to an additive constant, we can replace V^{eq} with $V^{\mathrm{eq}} + (\alpha - \beta)/2$, obtaining:

$$n^{\mathrm{eq}} = e^{(\gamma - V^{\mathrm{eq}})/U}, \qquad p^{\mathrm{eq}} = e^{(\gamma + V^{\mathrm{eq}})/U},$$

with $\gamma = (\alpha + \beta)/2$. Because of $n^{\mathrm{eq}} p^{\mathrm{eq}} = n_i^2$, it follows $e^{\gamma} = n_i$. So:

$$n^{\mathrm{eq}} = n_i e^{-V^{\mathrm{eq}}/U}, \quad p^{\mathrm{eq}} = n_i e^{V^{\mathrm{eq}}/U}. \tag{124}$$

By plugging (124) into (123) we obtain an elliptic PDE for V^{eq}:

$$-\varepsilon_s \Delta V^{\mathrm{eq}} = q(n_i e^{-V^{\mathrm{eq}}/U} - n_i e^{V^{\mathrm{eq}}/U} - C). \tag{125}$$

4.4.2 Boundary conditions for the bipolar model

When considering (121)–(123) in a bounded domain $\Omega \subset \mathbb{R}^3$ it is necessary to impose suitable boundary conditions. The usual procedure consists in splitting the boundary of Ω into two parts: the Dirichlet boundary Γ_D (which models the contacts) and the Neumann boundary Γ_N (which describes the insulating boundary segments). The carrier densities and the potential are prescribed on Γ_D:

$$n = n_D, \quad p = p_D, \quad V = V_D \quad \text{on } \Gamma_D, \tag{126}$$

while the normal components of the current densities and the electric field vanish on Γ_N:

$$J_n \cdot \nu = J_p \cdot \nu = \nabla V \cdot \nu = 0 \quad \text{on } \Gamma_N.$$

Thanks to (121), (122), this means:

$$\nabla n \cdot \nu = \nabla p \cdot \nu = \nabla V \cdot \nu = 0 \quad \text{on } \Gamma_N. \tag{127}$$

In order to determine n_D, p_D, V_D we employ the notion of thermal equilibrium. We assume that:

- the total space charge vanishes on Γ_D: $n_D - p_D - C = 0$;
- the densities are in thermal equilibrium on Γ_D;

- the boundary potential is the superposition of the *built-in potential* V_{bi} and the applied voltage V_{app}:

$$V_D = V_{bi} + V_{app}. \tag{128}$$

From these assumptions and (124) it follows that

$$n_D = n_i e^{-V_{bi}/U}, \quad p_D = n_i e^{V_{bi}/U}.$$

From the relations $n_D p_D = n_i^2$ and $n_D - p_D - C = 0$ we deduce:

$$n_D = \frac{1}{2}\left(C + \sqrt{C^2 + 4n_i^2}\right), \quad p_D = \frac{1}{2}\left(-C + \sqrt{C^2 + 4n_i^2}\right), \tag{129}$$

and so:

$$V_{bi} = U \log\left(\frac{C}{2n_i} + \sqrt{\frac{C^2}{4n_i^2} + 1}\right). \tag{130}$$

Thus the bipolar model (121)–(123) is to be solved with the boundary conditions (126)–(130). In an analogue way, eq. (125) must be coupled with the boundary conditions:

$$V^{\mathrm{eq}} = V_{bi} \quad \text{on } \Gamma_D, \qquad \nabla V^{\mathrm{eq}} \cdot \nu = 0 \quad \text{on } \Gamma_N,$$

with V_{bi} given by (130).

5 Hydrodynamic equations

The starting point is, like in the previous section, the semiconductor Boltzmann-Poisson system:

$$\partial_t f + v(k) \cdot \nabla_x f - \frac{q}{\hbar} \nabla_x V \cdot \nabla_k f = Q(f), \quad x \in \mathbb{R}^3,\ k \in \mathbb{R}^3,\ t > 0, \tag{131}$$

$$-\varepsilon_s \Delta V = q(\langle f \rangle - C), \quad x \in \mathbb{R}^3,\ t > 0, \tag{132}$$

where $v(k) = \hbar k/m^*$ is the group velocity, m^* is the electron effective mass, k is the pseudo-wavevector, $V = V(x,t)$ is the electric potential, and we have defined the shorthand $\langle f \rangle \equiv \int_{\mathbb{R}^3} f dk/4\pi^3$ for all functions f. The attentive Reader should have noticed that we have made the parabolic band approximation: $E(k) = \hbar^2 |v|^2/m^*$, so that $v(k) = \nabla_k E(k)/\hbar = \hbar k/m^*$ and the Brillouin zone can be replaced by the whole space $\mathbb{R}^3$.

The hydrodynamic model is derived from the kinetic equation (132) under the assumption that the timescale of collisions is much shorter than the timescale of transport, which is of the same order as the reference timescale. We will clarify this statement in a few moments.

We assume that the collision operator Q can be written as the sum of two operators Q_0, Q_1, representing two different types of collisions and characterized by different mean free times: $Q(f) = Q_0(f) + Q_1(f)$.

We wish to scale eqs. (131), (132). To this purpose, let us define the following reference values:

- the reference length λ, which is the mean free path corresponding to Q_1;
- the mean free path λ_0 corresponding to Q_0;
- the reference wavevector $k_0 = \sqrt{m^* k_B T_L/\hbar^2}$; notice that the corresponding value for the energy is $\hbar^2 k_0^2/m^* = k_B T_L$, so the reference energy equals the thermal energy $k_B T_L$;
- the reference potential $U = k_B T_L/q$;
- the reference times $\tau = \lambda/\sqrt{k_B T_L/m^*}$, $\tau_0 = \lambda_0/\sqrt{k_B T_L/m^*}$.

We then define the a-dimensional variables:

$$x = \lambda x_s, \quad t = \tau t_s, \quad k = k_0 k_s,$$

and the a-dimensional functions:

$$V = U V_s, \quad Q_0(f) = \frac{1}{\tau_0} Q_{0,s}(f), \quad Q_1(f) = \frac{1}{\tau} Q_{1,s}(f).$$

By defining

$$\alpha = \frac{\lambda_0}{\lambda} = \frac{\text{mean free path for } Q_0}{\text{mean free path for } Q_1},$$

inserting the above definitions into (131), (132) and carrying out some straightforward computations we conclude:[15]

$$\partial_t f_\alpha + k \cdot \nabla_x f_\alpha - \nabla_x V_\alpha \cdot \nabla_k f_\alpha = \frac{1}{\alpha} Q_0(f_\alpha) + Q_1(f_\alpha), \quad x \in \mathbb{R}^3,\ k \in \mathbb{R}^3,\ t > 0, \tag{133}$$

$$-\lambda_D \Delta V_\alpha = \langle f_\alpha \rangle - C, \quad x \in \mathbb{R}^3, \tag{134}$$

We assume that Q_0 and Q_1 have the following properties:

- Q_0 conserves mass, momentum, and energy:

  $$\int_{\mathbb{R}^3} Q_0(f) dk = \int_{\mathbb{R}^3} k_i Q(f) dk = \int_{\mathbb{R}^3} |k|^2 Q_0(f) dk = 0, \quad i = 1, 2, 3,$$

  for all functions f.

---

[15] We remove the no more necessary subscript s for better readability and insert a subscript α in f and V to remember the Reader their dependence on α.

- The kernel of Q_0 is spanned by *local* Maxwellians:

$$N(Q_0) = \{f \ : \ f = M[n,u,T] \ \text{ for some } n,u,T\},$$
$$M[n,u,T](k) \equiv \frac{1}{2}\left(\frac{2\pi}{T}\right)^{3/2} n e^{-|u-k|^2/2T}.$$

- Q_1 is given by the low-density operator with a cross-section σ depending only on x (see Subs. 3.3.2):

$$Q_1(f) = \int_{\mathbb{R}^3} \sigma(x)\left(\mathcal{M}f' - \mathcal{M}'f\right) dk' = \sigma(x)\left(\mathcal{M}\int_{\mathbb{R}^3} f'dk' - f\right),$$

  where $\mathcal{M}$ is the *global* Maxwellian: $\mathcal{M}(k) = (1/2)(2\pi)^{3/2}e^{-|k|^2/2}$. We point out that, as stated in Prop. 3.1, Q_1 preserves mass: $\int_{\mathbb{R}^3} Q_1(f)dk = 0$ for all functions f.

We also assume that $\alpha = \lambda_0/\lambda$ is small, that is, the collisions described by Q_0 are much more frequent than the collisions described by Q_1.

Remark. The assumption on the smallness of α clarifies the statement at the beginning of this section: in deriving the hydrodynamic model, we consider a situation in which the timescale of collisions is much shorter than the timescale of transport, which is comparable to the reference timescale. In fact, in the scaled Boltzmann equation (100), due to the smallness of α, the collision term dominates both the transport term and the time derivative of w, which are of the same order.

Our goal is to derive a macroscopic model for the moments $\langle f\rangle$ (electron density), $\langle vf\rangle$ (current density), $\langle(|k|^2/2)f\rangle$ (energy density). The procedure to derive this model is made up by three steps.

1. Let us multiply (133) times $1, k, |k|^2/2$ and integrate in $\mathbb{R}^3$ with respect to k. We obtain a set of moment equations:

$$\partial_t \langle f_\alpha\rangle + \operatorname{div}_x \langle k f_\alpha\rangle = 0, \tag{135}$$
$$\partial_t \langle k f_\alpha\rangle + \operatorname{div}_x \langle k\otimes k f_\alpha\rangle + \langle f_\alpha\rangle \nabla_x V = \langle k Q_1(f_\alpha)\rangle, \tag{136}$$
$$\partial_t \left\langle (|k|^2/2) f\right\rangle + \operatorname{div}_x \left\langle (|k|^2/2) k f_\alpha\right\rangle + \langle k f_\alpha\rangle\cdot\nabla_x V = \left\langle (|k|^2/2) Q_1(f_\alpha)\right\rangle. \tag{137}$$

   Eqs. (135)–(137) are not closed, namely, the integrals $\langle k\otimes k f_\alpha\rangle$, $\langle(|k|^2/2)kf_\alpha\rangle$ cannot be written in terms of $\langle f_\alpha\rangle$, $\langle v f_\alpha\rangle$, $\langle(|k|^2/2)f_\alpha\rangle$. This is the so-called *closure problem.* We will solve this problem by taking the formal limit $\alpha\to 0$ in the scaled Boltzmann equation (133) and in the moment equations (135)–(137).

2. Let us take the limit $\alpha\to 0$ in (133). We assume that $f \equiv \lim_{\alpha\to 0} f_\alpha$ exists. It follows that $Q_0(f) = 0$; thanks to the assumptions on Q_0, this imples that f is a Maxwellian: $f = M[n,u,T]$ for some n, u, T.

   A straightforward computation yields:

$$\langle M[n,u,T]\rangle = n,$$

$$\begin{aligned}\langle kM[n,u,T]\rangle &= nu,\\ \langle k\otimes kM[n,u,T]\rangle &= nu\otimes u + nTI,\\ \left\langle k(|k|^2/2)M[n,u,T]\right\rangle &= \frac{1}{2}nu(|u|^2+5T),\\ \langle kQ_1(M[n,u,T])\rangle &= -\sigma nu,\\ \left\langle \frac{|k|^2}{2}Q_1(M[n,u,T])\right\rangle &= \sigma n\left(\frac{3}{2}-\frac{|u|^2}{2}-\frac{3}{2}T\right).\end{aligned}$$

3. Let us take the limit $\alpha \to 0$ in (135)–(137) and exploit the results obtained in the previous step:

$$\begin{aligned}\partial_t n + \operatorname{div}_x(nu) &= 0,\\ \partial_t(nu) + \operatorname{div}_x(nu\otimes u) + \nabla(nT) + n\nabla_x V &= -\sigma nu,\\ \partial_t\left(\frac{n|u|^2}{2}+\frac{3}{2}nT\right) + \operatorname{div}_x\left(\frac{nu}{2}(|u|^2+5T)\right) + nu\cdot\nabla_x V &= \sigma n\left(\frac{3}{2}-\frac{|u|^2}{2}-\frac{3}{2}T\right).\end{aligned}$$

By defining $J = nu$ (current density), $e = |u|^2/2 + (3/2)T$ (total energy per particle), and $\tau_0(x) = 1/\sigma(x)$ (relaxation time), and taking the limit $\alpha \to 0$ in the Poisson equation (134) we finally obtain the (scaled) *hydrodynamic model*:

$$\partial_t n + \operatorname{div}_x J = 0, \tag{138}$$

$$\partial_t J + \operatorname{div}_x\left(\frac{J\otimes J}{n}\right) + \nabla(nT) + n\nabla_x V = -\frac{J}{\tau_0}, \tag{139}$$

$$\partial_t(ne) + \operatorname{div}_x(J(e+T)) + J\cdot\nabla_x V = -\frac{n}{\tau_0}\left(e-\frac{3}{2}\right), \tag{140}$$

$$-\lambda_D \Delta V = n - C. \tag{141}$$

Going back to the physical variables, the above system takes the form:

$$\partial_t n + \frac{1}{q}\operatorname{div}_x J = 0, \tag{142}$$

$$\partial_t J + \frac{1}{q}\operatorname{div}_x\left(\frac{J\otimes J}{n}\right) + \frac{qk_B}{m^*}\nabla(nT) + \frac{q^2}{m^*}n\nabla_x V = -\frac{J}{\tau_{rel}}, \tag{143}$$

$$\partial_t(ne) + \frac{1}{q}\operatorname{div}_x(J(e+k_BT)) + J\cdot\nabla_x V = -\frac{n}{\tau_{rel}}\left(e-\frac{3}{2}k_BT_L\right), \tag{144}$$

$$-\varepsilon_s\Delta V = q(n-C), \tag{145}$$

with $\tau_{rel} = \tau\tau_0$.
The hydrodynamic model describes the evolution of the electron density n, the electron current density J and the electron energy density ne in a timescale that is shorter than in the drift-diffusion model, under the assumption that some scattering phenomena (described by the operator Q_0) happen much more frequently than other scattering phenomena (described by Q_1); the mean free time for this second class of collisions is assumed to be of the order of the considered timescale.

5.1 Relaxation-time limits

We are going to show that by performing the so-called *relaxation-time limits* in the hydrodynamic model it is possible to derive a new model (the energy-transport equations) and to recover the drift-diffusion equations. Our starting point is a generalization of eqs. (142)–(144) (we consider V as a given function, since it does not affect the computations that follow):

$$\partial_t n + \frac{1}{q}\text{div}\, J = 0, \tag{146}$$

$$\partial_t J + \frac{1}{q}\text{div}\left(\frac{J \otimes J}{n}\right) + \frac{qk_B}{m^*}\nabla(nT) + \frac{q^2}{m^*}n\nabla_x V = -\frac{J}{\tau_p}, \tag{147}$$

$$\partial_t(ne) + \frac{1}{q}\text{div}\,(J(e + k_B T)) + J \cdot \nabla_x V - \text{div}\,(\kappa n T \nabla T) = -\frac{n}{\tau_e}\left(e - \frac{3}{2}k_B T_L\right). \tag{148}$$

As the attentive Reader has surely noticed, there are two differences between (142)–(144) and (146)–(148):

1. we have introduced two different relaxation times: a momentum relaxation time τ_p, and an energy relaxation time τ_e;
2. we have added the term $-\text{div}\,(\kappa n T \nabla T)$ on the left-hand side of (148) which describes heat conduction ($\kappa = \kappa(x) > 0$ is given).

The point in choosing two different relaxation mechanisms is that we wish to consider two different timescales for energy and momentum relaxation. The next step in our argument is a (diffusive) scaling of (146)–(148). As it is always the case in this kind of procedure, we introduce some reference values:

- the reference length λ (the device diameter);
- the reference particle density C_m;
- the reference potential $U = k_B T_L/q$;
- the reference current density $J_0 = qC_M\lambda/\tau$;
- the reference time $\tau = m^*\lambda/\tau_p k_B T_L$.

Notice that the relation:

$$(k_B T_L)^2 = m^*\left(\frac{\lambda}{\tau}\right)^2 \cdot m^*\left(\frac{\lambda}{\tau_p}\right)^2$$

means that the thermal energy $k_B T_L$ is of the same order of the geometric average of the kinetic energies required to cross the device in times τ and τ_p.

We define the a-dimensional quantities:

$$x = \lambda x_s, \quad t = \tau t_s, \quad n = C_m n_s, \quad J = J_0 J_s, \quad V = U V_s, \quad T = T_L T_s.$$

We also define the a-dimensional parameters $\alpha = \tau_p/\tau$, $\beta = \tau_e/\tau$.

By plugging the above relations into (142)–(144) and carrying out the straightforward computations we obtain:

$$\partial_t n + \operatorname{div} J = 0, \tag{149}$$

$$\alpha \partial_t J + \alpha \operatorname{div} \left(\frac{J \otimes J}{n} \right) + \nabla(nT) + n \nabla_x V = -J, \tag{150}$$

$$\partial_t(ne) + \operatorname{div} \left(J(e+T) \right) + J \cdot \nabla_x V - \operatorname{div} (\kappa_0 n T \nabla T) = -\frac{n}{\beta} \left(e - \frac{3}{2} \right), \tag{151}$$

$$e = \alpha \frac{|J|^2}{2n^2} + \frac{3}{2} T. \tag{152}$$

Let us now take the (formal) limit $\alpha \to 0$ in the above equations (with $\beta > 0$ fixed). From (152) we obtain $e = (3/2)T$; from this relation and (149)–(151) we deduce:

$$\partial_t n + \operatorname{div} J = 0, \quad J = -\nabla(nT) - n \nabla_x V, \tag{153}$$

$$\partial_t \left(\frac{3}{2} nT \right) + \operatorname{div} \left(\frac{5}{2} JT - \kappa_0 n T \nabla T \right) + J \cdot \nabla_x V = -\frac{3}{2} \frac{n}{\beta} (T - 1). \tag{154}$$

Eqs. (153), (154) constitute the so-called *energy-transport model.* It describes the evolution of the electron density n and the electron energy density $\frac{3}{2} nT$ under the assumption that the momentum relaxation time is much smaller than the energy relaxation time.

What happens if we take the limit $\beta \to 0$ in the energy-transport model? Well, eq. (154) becomes the simple relation $T = 1$, which can be plugged into (153), thus obtaining:

$$\partial_t n + \operatorname{div} J = 0, \quad J = -\nabla n - n \nabla V,$$

which is the already known the drift-diffusion model. So the drift-diffusion model can be seen as the limit of the hydrodynamic equations (149)–(151) under the assumption that the relaxation times for both energy and momentum are small compared to the considered timescale, which is coherent with the fact that the drift-diffusion model describes the evolution of the system in a longer timescale than the hydrodynamic equations.

6 Microscopic quantum models

Until now, we have used a semiclassical formalism to describe the motion of an ensemble of electrons in a semiconductor; in this section we present instead some tools and concepts that are proper to statistical quantum mechanics.

6.1 Mixed states, density matrices, and density operators

We have seen in subsection 2.2 that a quantum system is described by a wavefunction, satisfying the Schrödinger equation. However, it happens often that the state of the system is not known, and only partial informations about it are available to the observer. Namely, this latter might know that the possible states of the system are given by *a sequence* of (normalized) wavefunctions $(\psi_j)_{j\in\mathbb{N}} \subset L^2(\mathbb{R}^3)$,[16] each of them satisfying the Schrödinger equation $i\hbar\partial_t\psi_j = H\psi_j$, and that the system can be found in the state j with some probability $\lambda_j \geq 0$, with $\sum_{j=1}^{\infty}\lambda_j = 1$. In such a case, the system is said to be in a *mixed state*, while quantum systems that are described by only one wavefunction are said to be in a *pure state*. In order to describe a quantum system in a mixed state, it is convenient to define the so-called *density matrix* of the system:

$$\rho(x,y) = \sum_{j\in\mathbb{N}} \lambda_j \psi_j(x)\overline{\psi_j(y)}, \tag{155}$$

and the *density operator* $\hat{\rho}$ associated to it:

$$(\hat{\rho}\phi)(x) = \int_{\mathbb{R}^3} \rho(x,y)\phi(y)dy, \qquad \phi \in L^2(\mathbb{R}^3). \tag{156}$$

The density operator is positive, self-adjoint, compact, and has trace equal to 1: $\mathrm{Tr}(\hat{\rho}) = 1$.[17] Moreover, it holds in general that $\mathrm{Tr}(\hat{\rho}^2) \leq 1$, with the equality holding if and only if $\rho(x,y) = \psi(x)\overline{\psi(y)}$ for some $\psi \in L^2(\mathbb{R}^3)$, that is, if and only if the system is in a pure state. In this case, ψ is the wavefuction of the system. We point out that, given an arbitrary positive, self-adjoint and compact operator $\hat{\rho}$, thanks to the spectral theorem, a complete orthonormal system $(\psi_j)_{j\in\mathbb{N}}$ of L^2 exists, which is made up by eigenvectors for $\hat{\rho}$. Thus $\hat{\rho}$ can be written as $\hat{\rho} = \sum_{j\in\mathbb{N}}\lambda_j\psi_j(\psi_j,\cdot)$; in particular, the integral representation (156) holds, with the kernel ρ given by (155).

In a mixed state, the particle density n and the current density J are defined in a natural way by:

$$n(x,t) = 2\sum_{j=1}^{\infty}\lambda_j|\psi_j(x,t)|^2, \qquad J = 2\sum_{j=1}^{\infty}\lambda_j\frac{q\hbar}{m}\Im\left(\overline{\psi_j(x)}\nabla_x\psi_j(x)\right),$$

where the factor 2 in front of the expression accounts for the two spin degrees of freedom. The same quantities can be expressed in terms of the density matrix in the following way:

$$n(x,t) = 2\rho(x,x,t), \qquad J(x,t) = \frac{iq\hbar}{m}\left(\nabla_r - \nabla_s\right)\rho(r,s,t)|_{r=s=x}. \tag{157}$$

---

[16] We consider here a single particle quantum system.

[17] The trace of an operator A on some Hilbert space H is defined as $\mathrm{Tr}(A) = \sum_{j\in\mathbb{N}}(u_j, Au_j)$, where $(u_j)_{j\in\mathbb{N}}$ is any complete orthonormal set for H.

It is also possible to show[18] that the following equation, called the *Liouville-Von Neumann equation*, holds:

$$i\hbar\partial_t\rho(x,y,t) = (H_x - H_y)\rho(x,y,t), \qquad x,y \in \mathbb{R}^3,\ t>0, \tag{158}$$

with H_x, H_y being the Hamiltonian operator acting on the variables x, y, respectively. Eq. (158) admits an equivalent formulation in terms of the density operator:

$$i\hbar\partial_t\hat{\rho} = [H,\hat{\rho}] \equiv H\hat{\rho} - \hat{\rho}H.$$

6.2 Many particle Wigner equation

We are about to present an equation that can be seen as a quantum counterpart of the classical Liouville equation. Our starting point is the Liouville-Von Neumann equation for an ensemble of M electrons with mass m in a vacuum, formulated in terms of the system density matrix ρ:

$$i\hbar\partial_t\rho(r,s,t) = (H_r - H_s)\rho(r,s,t), \qquad r,s \in \mathbb{R}^{3M},\ t>0. \tag{159}$$

We define the *Wigner function* associated to the system as:

$$w(x,p,t) = \int_{\mathbb{R}^{3M}} \rho\left(x+\frac{y}{2}, x-\frac{y}{2}\right) e^{-iy\cdot p/\hbar}dy.$$

The mapping $\rho \mapsto w$ is called *Wigner-Weyl transform.* The Wigner function satisfies the following *many particle Wigner equation*:

$$\partial_t w + \frac{p}{m}\cdot\nabla_x w - q\Theta[V]w = 0, \qquad x,p \in \mathbb{R}^{3M},\ t>0, \tag{160}$$

where $\Theta[V]$ is the pseudo-differential operator given by:

$$(\Theta[V]w)(x,p,t) = \int_{\mathbb{R}^{3M}} \frac{i}{\hbar}\left(V\left(x+\frac{y}{2},t\right) - V\left(x-\frac{y}{2},t\right)\right) w(x,p',t)e^{iy\cdot(p-p')/\hbar}\frac{dp'dy}{(2\pi\hbar)^{3M}}.$$

The drift term $(p/m)\cdot\nabla_x w$ is the quantum analogue of the classical transport term of the Liouville equation. The nonlocal term $q\Theta[V]w$ models the contribution of the electric potential. The nonlocality means that the electrons are influenced by a potential barrier even *before* they actually reach it.

The operator $\Theta[V]$ looks ugly, but it has at least some interesting propertiers, namely:

- if $V(x,t)$ is quadratic is x, then $\Theta[V] = \nabla_x V\cdot\nabla_p$.
- for general potentials $V(x,t)$, in the semiclassical regime ($\hbar\to 0$) it holds: $\Theta[V] = \nabla_x V\cdot\nabla_p + O(\hbar^2)$.

---

[18] The proof is a straightforward computation which exploits the fact that each wavefunction ψ_j satisfies a Schrödinger equation.

The Wigner function w can be seen as the quantum counterpart of the classical Boltzmann distribution. In fact, the particle density n and the current density J can be written as moments of w:

$$n(x,t) = \frac{2}{(2\pi\hbar)^{3M}} \int_{\mathbb{R}^{3M}} w(x,p,t)dp, \qquad J(x,t) = \frac{2}{(2\pi\hbar)^{3M}} \frac{q}{m} \int_{\mathbb{R}^{3M}} w(x,p,t)pdp.$$

In general, it can be shown that the mathematical expectation of a quantum observable $A_\gamma = \mathrm{Op}(\gamma)$ (Op being the Weyl quantization) equals $\int_{\mathbb{R}^{3M}} \gamma w dp/(2\pi\hbar)^{3M}$. Thus the Wigner distribution plays in statistical quantum mechanics a role analogue to the one played by the Boltzmann distribution in the statistical classical mechanics. However, there is at least an important difference between the two objects: while the Boltzmann distribution is always nonnegative, this is not true for the Wigner distribution: this latter can vanish in sets of positive Lebesgue measure. But, in any case, the Wigner equation preserves the positivity of the particle density n if $n(\cdot, 0) > 0$.[19]

6.3 Quantum Vlasov equation

The many-particle Wigner equation has the same drawback as the classical Liouville equation: it has to be solved in a very high-dimensional space, so the numerical solution of the problem is almost unfeasible. However, we will now show that a single-particle equation contains all the informations about the evolution of the many-particle system.

Let us consider an ensemble of M electrons with mass m moving in a vacuum under the action of a potential $V(x,t)$ ($x = (x_1, \ldots, x_M)$, with $x_i \in \mathbb{R}^3$ being the position of the $i-$th electron), and let ρ the system density matrix. We assume that:

1. the potential V is the sum of single-particle external potentials and two particle interaction potentials:

$$V(x,t) = \sum_{j=1}^{M} V_{ext}(x_i,t) + \sum_{\substack{j,k=1 \\ j\neq k}}^{M} V_{int}(x_j,x_k);$$

2. the limit $V_0 \equiv \lim_{M\to\infty} MV_{int}$ exists and is finite;

3. the electrons are initially indistinguishable, that is

$$\rho_I(r_1,\ldots,r_M,s_1,\ldots,s_M) = \rho(r_1,\ldots,r_M,s_1,\ldots,s_M,0)$$

is independent of the particle labels;

---

[19] As a matter of fact, the Wigner distribution cannot be *too* negative: it is possible to show that so-called *Husimi regularization* of w, namely the convolutin between w and a gaussian with standard deviation of order $\hbar$, is always nonnegative. This can be interpreted in light of the indetermination principle: w becomes positive if averaged on a set of radius $\sim \hbar$.

4. the initial sub-ensemble density matrices

$$\rho_I^{(a)}(r_1,\dots,r_a,s_1,\dots,s_a) = \int_{\mathbb{R}^{3(M-a)}} \rho_I|_{(r_{a+1}\dots r_M)=(s_{a+1}\dots s_M)=(u_{a+1}\dots u_M)} du_{a+1}\dots du_M$$

   can be factorized for all $a = 1,\dots,M-1$ is the following way:

$$\rho_I^{(a)}(r_1,\dots,r_a,s_1,\dots,s_a) = \prod_{j=1}^{a} R_I(r_j,s_j),$$

   for some given function R_I.

As a consequence, both $V(x_1,\dots,x_M,t)$ and $\rho(r_1,\dots,r_M,s_1,\dots,s_M,t)$ are invariant with respect to the exchange of the particle labels, that is, the electrons are indistinguishable at all time.

Under these assumptions, the dynamics of the many-particle system can be described by a single particle equation. By proceeding like in the classical case, we can prove the following:

Theorem 6.1. *Let W be a solution of the* quantum Vlasov equation*:*

$$\partial_t W + \frac{p}{m}\cdot\nabla_x W + q\Theta[V_{\mathit{eff}}]W = 0, \qquad x,p\in\mathbb{R}^3,\ t>0,$$

where the effective potential V_{eff} is defined by:

$$V_{\mathit{eff}}(x,t) = V_{ext}(x,t) + \int_{\mathbb{R}^3} n(x',t)V_0(x,x')dx', \quad n(x,t) = \frac{2}{(2\pi\hbar)^3}\int_{\mathbb{R}^3} W(x,p,t)dp.$$

Let R the density matrix associated to W via the Wigner-Weyl transform, and let, for all $a>0$, $\rho^{(a)}$ be defined by the so-called Hartree ansatz:

$$\rho^{(a)}(r_1,\dots,r_a,s_1,\dots,s_a) = \prod_{j=1}^{a} R(r_j,s_j).$$

Then $\rho^{(a)}$ is a solution of the limit BBGKY Liouville-von Neumann hierarchy, obtained from the many-particle Wigner equation by evaluation in $(r_{a+1},\dots,r_M) = (s_{a+1},\dots,s_M) = (u_{a+1},\dots,u_M)$, integration in $\mathbb{R}^{3(M-a)}$ w.r.t. $(u_{a+1},\dots,u_M)$, and limit $M\to\infty$.

So, we have reduced the many-particles problem to a single-particle equation, but there is a price to pay, namely: while the many-particle Wigner equation was linear, the quantum Vlasov equation is not; it is a nonlinear and nonlocal equation. Moreover, in contrast to the classical Vlasov equation, the quantum Vlasov equation does not preserve the nonnegativity of the solution; it does preserve, however, the nonnegativity of the particle density, if the Wigner function is nonnegative at initial time. Finally, we point out that, since $\Theta[V] = \nabla_x V\cdot\nabla_p + O(\hbar^2)$, the quantum Vlasov equation becomes the classical Vlasov equation in the (formal) limit $\hbar\to 0$.

6.4 Wigner-Boltzmann equation

Until now, we have considered only ballistic motion of quantum particles; but, if the characteristic device length is large compared to the mean free path of the electrons, collisions do happen and must be taken into account. Scattering phenomena can be accounted for by adding a suitable collision operator on the right-hand side of a kinetic equation; an example in this sense is given by the *Wigner-Boltzmann equation*:

$$\partial_t w + \frac{p}{m} \cdot \nabla_x w - q\Theta[V_{\text{eff}}]w = Q(w), \tag{161}$$

which models an open quantum system, that is, a quantum system which interacts with some environment, e.g. a phonon thermal bath, through the collision operator $Q(w)$. V_{eff} is the effective potential defined in Thr. 6.1.

The Reader may wander how does the collision operator $Q(w)$ look like; some possible choices of Q are listed below.

- **Wigner-BGK operator.** The simplest choice of Q is given by the *Bhatnagar-Gross-Krook (BGK)* or relaxation-time operator:

  $$Q(w) = \frac{1}{\tau}\left(\frac{n}{n^{eq}} w^{eq} - w\right),$$

  where w^{eq} is the thermal equilibrium distribution[20], n is the particle density, and n^{eq} is the equilibrium density:

  $$n = \frac{2}{(2\pi\hbar)^3} \int_{\mathbb{R}^3} w dp, \qquad n^{eq} = \frac{2}{(2\pi\hbar)^3} \int_{\mathbb{R}^3} w^{eq} dp.$$

  This collision operator simply describes the relaxation of the system towards the thermal equilibrium. An easy way to see this is by considering (161) in the uniform case (that is, w is $x-$independent) with vanishing effective field:

  $$\partial_t w = \frac{1}{\tau}\left(\frac{n}{n^{eq}} w^{eq} - w\right).$$

  The solution to the above-written ODE is $w(t) = e^{-t/\tau}w(0) + (n/n^{eq})w^{eq}(1 - e^{-t/\tau})$, which indeed converges to $(n/n^{eq})w^{eq}$ as $t \to \infty$.

- **Caldera-Leggett operator.** It was derived by Caldera and Leggett [13] and reads as:

  $$Q(w) = D_{pp}\Delta_p w + 2\gamma \text{div}_p(pw).$$

  The parameter $\gamma > 0$ describes friction phenomena, while $D_{pp} > 0$ is a self-diffusion coefficient. The Wigner-Boltzmann equation with this collision operator is also known as *quantum Brownian motion* or *quantum Langevin equation* and it is relevant in the study of interaction between light and matter [17].

---

[20] The way w^{eq} is defined will be explained in the subsequent subsections.

- **Wigner-Fokker-Planck operator.** The quantum Langevin equation does not preserve the positivity of the density operator along its time evolution, which is unphysical; to address this issue, the Caldera-Leggett operator was modified into the *quantum Fokker-Planck operator*:

  $$Q(w) = D_{pp}\Delta_p w + 2D_{pq}\mathrm{div}_x(\nabla_p w) + D_{qq}\Delta_x w + 2\gamma \mathrm{div}_p(pw).$$

  The nonnegative diffusion coefficients D_{pp}, D_{pq}, D_{qq} satisfy the so-called *Lindblad* condition:

  $$D_{pp}D_{qq} - D_{pq} \geq \frac{\gamma^2}{4},$$

  which guarantees that the density matrix remains positive along its evolution. The Wigner equation (161) with the Wigner-Fokker-Planck operator is called *Wigner-Fokker-Planck* operator and models an electron ensemble interacting dissipatively with an ideal thermal bath, given by the semiconductor lattice, whose ions are represented as harmonic oscillators.

7 Quantum macroscopic equations

In this section we will present some macroscopic models for quantum systems. We will see that such models can be obtained by means of a suitable (hydrodynamic or diffusive) scaling of the Wigner equations, similarly to the classical case, but can also be derived from the Schrödinger equation with simple algebraic manipulations.

7.1 The Madelung equations

Let us consider a single electron with mass m and charge q in a vacuum subject to an electric potential V. The wavefunction ψ of the system obeys the Schrödinger equation:

$$i\hbar\partial_t\psi = -\frac{\hbar^2}{2m}\Delta\psi + qV\psi, \quad t>0, \qquad \psi(\cdot,0) = \psi_I \quad \text{in } \mathbb{R}^3.$$

We assume that $\psi(x,t) \neq 0$ for all $x \in \mathbb{R}^3$, $t \geq 0$. We can thus write the wavefunction ψ and the initial wavefunction ψ_I in WKB form:

$$\psi = \sqrt{n}e^{iS/\hbar}, \qquad \psi_I = \sqrt{n_I}e^{iS_I/\hbar},$$

where the phases S, S_I are well defined (up to integer multiples of $2\pi\hbar$), given the positivity of n and n_I. Let us insert the above-written ansatz in the Schrödinger equation:

$$\begin{aligned} i\hbar e^{iS/\hbar}\left(\frac{\partial_t n}{2\sqrt{n}} + \sqrt{n}\frac{i}{\hbar}\partial_t S\right) &= -\frac{\hbar^2}{2m}\mathrm{div}\left(e^{iS/\hbar}\left(\frac{\nabla n}{2\sqrt{n}} + \sqrt{n}\frac{i}{\hbar}\nabla S\right)\right) + qV\sqrt{n}e^{iS/\hbar} \\ &= -e^{iS/\hbar}\left(\frac{\hbar^2}{2m}\mathrm{div}\left(\frac{\nabla n}{2\sqrt{n}} + \sqrt{n}\frac{i}{\hbar}\nabla S\right) + \frac{i\hbar}{2m}\nabla S\cdot\left(\frac{\nabla n}{2\sqrt{n}} + \sqrt{n}\frac{i}{\hbar}\nabla S\right) - qV\sqrt{n}\right), \end{aligned}$$

which means:

$$i\hbar\frac{\partial_t n}{2\sqrt{n}} - \sqrt{n}\partial_t S + \frac{\hbar^2}{2m}\Delta\sqrt{n} + \frac{i\hbar}{m}\nabla\sqrt{n}\cdot\nabla S + \frac{i\hbar}{2m}\sqrt{n}\Delta S - \frac{\sqrt{n}}{2m}|\nabla S|^2 - qV\sqrt{n} = 0.$$

Separating real and imaginary part in the above equation yields:

$$\begin{aligned} \frac{\partial_t n}{2\sqrt{n}} + \frac{1}{m}\nabla\sqrt{n}\cdot\nabla S + \frac{1}{2m}\sqrt{n}\Delta S = 0, \\ -\sqrt{n}\partial_t S + \frac{\hbar^2}{2m}\Delta\sqrt{n} - \frac{\sqrt{n}}{2m}|\nabla S|^2 - qV\sqrt{n} = 0, \end{aligned}$$

which is equivalent to:

$$\partial_t n = -\frac{1}{m}\nabla n\cdot\nabla S - \frac{1}{m}n\Delta S = -\frac{1}{m}\mathrm{div}\,(n\nabla S), \tag{162}$$

$$\partial_t S = \frac{\hbar^2}{2m}\frac{\Delta\sqrt{n}}{\sqrt{n}} - \frac{1}{2m}|\nabla S|^2 - qV. \tag{163}$$

Since the current J is given by:

$$J = \frac{\hbar}{m}\Im(\overline{\psi}\nabla\psi) = \frac{1}{m}n\nabla S,$$

eq. (162) can be rewritten as:

$$\partial_t n + \mathrm{div}\, J = 0. \tag{164}$$

Moreover, let us take the time derivative of J:

$$\partial_t J = \frac{\partial_t n}{m}\nabla S + \frac{n}{m}\nabla\partial_t S = -\frac{J}{n}\mathrm{div}\, J + \frac{n}{m}\nabla\left(\frac{\hbar^2}{2m}\frac{\Delta\sqrt{n}}{\sqrt{n}} - \frac{m}{2n^2}|J|^2 - qV\right),$$

which means:

$$\begin{aligned} \partial_t J + \frac{q}{m}n\nabla V - n\frac{\hbar^2}{2m^2}\nabla\left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) &= -\frac{J}{n}\mathrm{div}\, J - \frac{n}{2}\nabla\left(\frac{|J|^2}{n^2}\right) \\ &= -\frac{J}{n}\mathrm{div}\, J - J\cdot\nabla\left(\frac{J}{n}\right) = -\mathrm{div}\left(\frac{J\otimes J}{n}\right). \end{aligned} \tag{165}$$

Eqs. (164), (165) are the so-called *Madelung equations*:

$$\partial_t n + \mathrm{div}\, J = 0, \quad \partial_t J + \mathrm{div}\left(\frac{J\otimes J}{n}\right) + \frac{q}{m}n\nabla V - n\frac{\hbar^2}{2m^2}\nabla\left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) = 0. \tag{166}$$

The Madelung equations are an alternative formulation of the Schrödinger equation, and thus describe the evolution of a quantum system in a pure state. They are the quantum analogue of the classical pressureless Euler equations of gas dynamics. There latter can be formally recovered from (166) by letting $\hbar \to 0$ (semiclassical limit). The quantity:

$$V^B = -\frac{\hbar^2}{2m}\frac{\Delta\sqrt{n}}{\sqrt{n}},$$

which appears in the quantum contribution in (166), is called *Bohm potential.* Notice that, since:

$$-n\frac{\hbar^2}{2m^2}\nabla\left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) = -\frac{\hbar^2}{4m^2}\text{div}\,\left(n\nabla\otimes\nabla\log n\right),$$

the quantum term in (166) can be represented as the divergence of the nondiagonal quantum pressure tensor

$$P^Q = -\frac{\hbar^2}{4m}n\nabla\otimes\nabla\log n.$$

We also point out that, since $J/n = \nabla S$, the initial velocity must be irrotational for consistency with the derivation of the model.

A question that arises from the previous derivation is: is it possible to reconstruct the wavefunction ψ in terms of the moments n, J? This is a particular case of a more general problem, the so-called *Pauli problem*: is it possible to reconstruct a pure quantum state by knowing a finite set of measurements of the state? We don't mean to answer this question here, but we point out that, for the case that we have considered, the answer is negative: in fact, in order to derive (166), we assumed that $n > 0$ (or, equivalently, that $\psi \neq 0$) at all times, in order to be able to define the phase S in an unique way. We did not consider vacuum, namely, we ruled out the possibility that $\psi(\cdot, t) = 0$ in a positive measure set, for some $t \geq 0$. Since vacuum does exist, determining the wavefunction ψ in terms of n, J is in general impossible.

7.2 A mixed-state quantum hydrodynamic model

We have seen in the previous subsection that the Madelung equations describe pure states. We will now derive a hydrodynamic model for a statistical mixture of quantum particles (a mixed state). Let this state be represented by a sequence of wavefunctions $(\psi_j)_{j\in\mathbb{N}} \subset L^2(\mathbb{R}^3)$ and occupation probabilites $(\lambda_j)_{j\in\mathbb{N}} \subset [0,1]$, with $\sum_{j=1}^{\infty}\lambda_j = 1$, where each wavefunction satisfies the Schrödinger equation:

$$i\hbar\partial_t\psi_j = -\frac{\hbar^2}{2m}\Delta\psi_j + qV\psi_j, \quad t > 0, \qquad \psi_j(\cdot,0) = \psi_j^0 \quad \text{in } \mathbb{R}^3. \tag{167}$$

Let the single-state particle density n_j and the single-state current density J_j be defined as:

$$n_j = |\psi_j|^2, \qquad J_j = \frac{\hbar}{m}\Im\left(\overline{\psi_j}\nabla\psi_j\right),$$

while the total particle density n and the total current density J are given by:

$$n = \sum_{j=1}^{\infty}\lambda_j n_j, \qquad J = \sum_{j=1}^{\infty}\lambda_j J_j.$$

We wish to find hydrodynamic equations for n, J. The starting point is the Madelung system (166), which holds true with n, J replaced by n_j, J_j:

$$\partial_t n_j + \operatorname{div} J_j = 0, \quad \partial_t J_j + \operatorname{div}\left(\frac{J_j \otimes J_j}{n_j}\right) + \frac{q}{m} n_j \nabla V - n_j \frac{\hbar^2}{2m^2} \nabla \left(\frac{\Delta\sqrt{n_j}}{\sqrt{n_j}}\right) = 0,$$

for $j \in \mathbb{N}$. In we multiply both equations times λ_j and sum in j we obtain:

$$\partial_t n + \operatorname{div} J = 0, \tag{168}$$

$$\partial_t J + \operatorname{div}\left(\sum_{j=1}^{\infty} \lambda_j \frac{J_j \otimes J_j}{n_j}\right) + \frac{q}{m} n_j \nabla V - \frac{\hbar^2}{2m^2} \sum_{j=1}^{\infty} \lambda_j n_j \nabla \left(\frac{\Delta\sqrt{n_j}}{\sqrt{n_j}}\right) = 0. \tag{169}$$

It is convenient to define the following quantities, having the physical dimension of velocities:

$$u_j^{cu} = \frac{J_j}{n_j}, \quad u^{cu} = \frac{J}{u}, \quad u_j^{os} = \frac{\hbar}{2m} \nabla \log n_j, \quad u^{os} = \frac{\hbar}{2m} \nabla \log n.$$

We call u_j^{cu}, u^{cu} “current velocities” and u_j^{os}, u^{os} “osmotic velocities”. It holds:

$$\begin{aligned}
&\sum_{j=1}^{\infty} \lambda_j \frac{J_j \otimes J_j}{n_j} = \sum_{j=1}^{\infty} \lambda_j n_j u_j^{cu} \otimes u_j^{cu} \\
&= \sum_{j=1}^{\infty} \lambda_j n_j \left((u_j^{cu} - u^{cu}) \otimes (u_j^{cu} - u^{cu}) + u^{cu} \otimes u_j^{cu} + u_j^{cu} \otimes u^{cu} - u^{cu} \otimes u^{cu} \right) \\
&= n \sum_{j=1}^{\infty} \lambda_j \frac{n_j}{n} (u_j^{cu} - u^{cu}) \otimes (u_j^{cu} - u^{cu}) + \frac{J \otimes J}{n}.
\end{aligned}$$

Moreover:

$$\begin{aligned}
&\frac{\hbar^2}{2m^2} \sum_{j=1}^{\infty} \lambda_j n_j \nabla \left(\frac{\Delta\sqrt{n_j}}{\sqrt{n_j}}\right) = \frac{\hbar^2}{4m^2} \sum_{j=1}^{\infty} \lambda_j \operatorname{div}\left((\nabla \otimes \nabla) n_j - \frac{\nabla n_j \otimes \nabla n_j}{n_j} \right) \\
&= \frac{\hbar^2}{4m^2} \operatorname{div}\left((\nabla \otimes \nabla) n\right) \\
&\quad - \sum_{j=1}^{\infty} \lambda_j \operatorname{div}\left(n_j (u_j^{os} - u^{os}) \otimes (u_j^{os} - u^{os}) + u^{os} \otimes u_j^{os} + u_j^{os} \otimes u^{os} - u^{os} \otimes u^{os} \right) \\
&= \frac{\hbar^2}{4m^2} \operatorname{div}\left((\nabla \otimes \nabla) n - \frac{\nabla n \otimes \nabla n}{n} \right) - \operatorname{div}\left(\sum_{j=1}^{\infty} \lambda_j n_j (u_j^{os} - u^{os}) \otimes (u_j^{os} - u^{os}) \right) \\
&= \frac{\hbar^2}{2m^2} n \nabla \left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) - \operatorname{div}\left(n \sum_{j=1}^{\infty} \lambda_j \frac{n_j}{n} (u_j^{os} - u^{os}) \otimes (u_j^{os} - u^{os}) \right)
\end{aligned}$$

Let us define the tensors:

$$\theta^{cu} = \sum_{j=1}^{\infty} \lambda_j \frac{n_j}{n} (u_j^{cu} - u^{cu}) \otimes (u_j^{cu} - u^{cu}),$$
$$\theta^{os} = \sum_{j=1}^{\infty} \lambda_j \frac{n_j}{n} (u_j^{os} - u^{os}) \otimes (u_j^{os} - u^{os}),$$
$$\theta = \theta^{cu} + \theta^{os}.$$

We refer to θ^{cu}, θ^{os} as "current temperature" and "osmotic temperature", respectively. The quantity θ represents the total temperature. Thus we can rewrite (168), (169) as:

$$\partial_t n + \operatorname{div} J = 0, \quad \partial_t J + \operatorname{div} \left(\frac{J \otimes J}{n} + n\theta \right) + \frac{q}{m} n \nabla V - n \frac{\hbar^2}{2m^2} \nabla \left(\frac{\Delta \sqrt{n}}{\sqrt{n}} \right) = 0. \quad (170)$$

Eq. (170) is a system of hydrodynamic equations describing the evolution of a quantum system in a mixed state. Such a system is not closed, since the temperature tensor θ cannot be expressed as a function of n, J without further assumptions. So we face the closure problem once more. A possible way of closing the system is by assuming that $\theta = T\mathrm{Id}$, for some positive constant temperature T; the model obtained with this hypothesis is called *isothermal quantum hydrodynamic model.* Another possiblity is to assume that $\theta = T(n)\mathrm{Id}$ with $T(n) = T_0 n^\alpha$ for some $T_0 > 0$ and $\alpha > 0$; the corresponding model is referred to as *isentropic quantum hydrodynamic model.*

7.3 Quantum thermal equilibrium

In subs. 6.4 we have mentioned the quantum thermal equilibrium Wigner distribution w^{eq}; it is time to define this object in a (more or less) precise way. The quantum thermal equilibrium distribution, also called *quantum Maxwellian* for its formal resemblance with the Maxwellian of classical statistical mechanics, is the Wigner distribution that maximizes the quantum entropy functional [21, 22]. Recall the definition of the Wigner-Weyl transform $\mathcal{W} : L^2(\mathbb{R}^3 \times \mathbb{R}^3) \to L^2(\mathbb{R}^3 \times \mathbb{R}^3)$:

$$(\mathcal{W}\rho)(x,p) = \int_{\mathbb{R}^3} \rho\left(x + \frac{y}{2}, x - \frac{y}{2}\right) e^{-iy\cdot p/\hbar} dy.$$

Remember also that a density matrix ρ can be identified with the density operator $\hat{\rho}$ defined as:

$$(\hat{\rho}\phi)(x) = \int_{\mathbb{R}^3} \rho(x,y)\phi(y)dy \qquad \forall \phi \in L^2(\mathbb{R}^3).$$

Conversely, any density operator $\hat{\rho}$ (that is, any linear, self-adjoint, compact operator on L^2 such that $\mathrm{Tr}(\hat{\rho}) = 1$ and $\mathrm{Tr}(\hat{\rho}^2) < \infty$) can be identified with its kernel ρ, which is a density matrix. So we can lift $\mathcal{W}$ to an operator (which will be denoted again by $\mathcal{W}$) such that:

$$(\mathcal{W}\hat{\rho})(x,p) = \int_{\mathbb{R}^3} \rho\left(x + \frac{y}{2}, x - \frac{y}{2}\right) e^{-iy\cdot p/\hbar} dy,$$

where $\hat{\rho}$ is the density operator associated to the density matrix ρ. This definition allows us to associate to any function $\phi : \mathbb{R} \to \mathbb{R}$ a corresponding nonlocal quantum operator Φ defined in the following way:

$$\Phi(w) = \mathcal{W}(\phi(\mathcal{W}^{-1}(w))) \qquad \text{for all Wigner functions } w.$$

The meaning of the above-written definition is: given some Wigner function $w \in L^2(\mathbb{R}^3 \times \mathbb{R}^3)$, in order to build $\Phi(w)$ we first apply the inverse Wigner transform $\mathcal{W}^{-1}$ to w and obtain a density operator $\hat{\rho} = \mathcal{W}^{-1}w$, then we compute $\phi(\hat{\rho})$ with the help of spectral theory,[21] and finally we apply $\mathcal{W}$ to this density operator to obtain a Wigner function, which is the element $\Phi(w)$ that we wanted to compute. By choosing $\phi = \exp$ and $\phi = \log$ we obtain the so called *quantum exponential* and *quantum logarithm*:

$$\operatorname{Exp}(w) = \mathcal{W}(\exp(\mathcal{W}^{-1}(w))), \quad \operatorname{Log}(w) = \mathcal{W}(\log(\mathcal{W}^{-1}(w))) \quad \text{for all Wigner functions } w.$$

It is possible to prove that Log is the inverse mapping of Exp, and that the Fréchet derivatives of $\operatorname{Exp}(w)$, $\operatorname{Log}(w)$ are $\operatorname{Exp}(w)$ and w^{-1}, respectively. Moreover, it is also possible to show that

$$\operatorname{Exp}(w) = \exp(w) + O(\hbar^2), \qquad \operatorname{Log}(w) = \log(w) + O(\hbar^2),$$

in the (formal) semiclassical limit $\hbar \to 0$. The $O(\hbar^2)$ correction in the semiclassical expansion of Exp has been explicitly computed; see [33, Lemma 12.4] for details.

Let w be a solution of the Wigner-Boltzmann equation (161), and let us define:

$$\langle g \rangle = \frac{2}{(2\pi\hbar)^3} \int_{\mathbb{R}^3} g(p)dp \qquad \text{for all functions } g = g(p).$$

The von Neumann entropy of the state described by w is:

$$S(w) = -\frac{2}{(2\pi\hbar)^3} \int_{\mathbb{R}^3 \times \mathbb{R}^3} w(x,p,\cdot) \left((\operatorname{Log} w)(x,p,t) - 1 + \frac{h(x,p)}{k_B T} \right) dxdp,$$

where $h = \mathcal{W}H = \frac{|p|^2}{2m} + qV$ is the classical symbol of the Hamiltonian operator H and $T > 0$ is the temperature of a thermal bath in equilibrium with the system. Notice that, while the classical entropy is an integral in the momentum (or velocity) space, the quantum entropy is an integral in the whole phase space; this is due to the nonlocal nature of quantum mechanics.

---

[21] If $A : H \to H$ is a compact self-adjoint operator on a Hilbert space H, and $f : \mathbb{R} \to \mathbb{R}$ is some real function, we can give a meaning to $f(A)$ in a natural way. Being A compact and self-adjoint, we can write $A = \sum_{k \in \mathbb{N}} \lambda_k \Pi_k$, where λ_k is the k−th (nonzero) eigenvalue of A, and Π_k is the projection operator in the eigenspace associated to λ_k. We define $f(A)$ simply as:

$$f(A) = \sum_{k \in \mathbb{N}} f(\lambda_k)\Pi_k.$$

Let $(k_j(p))_{j=1,\dots,N}$ weight functions, and let $m_j = \langle k_j w(x,p,\cdot)\rangle$ for $j = 0,\dots,N$. The quantities m_j are moments of the Wigner function w. The *quantum thermal equilibrium distribution* or *quantum Maxwellian* M related to the moments $m_0,\dots,m_N$ is defined as the solution of the (formal) constrained maximization problem:

$$S(M) = \max\{S(f) \ : \ \langle f k_j\rangle = m_j \ \ (j = 0,\dots,N)\}. \tag{171}$$

Lemma 7.1. *The solution of* (171)*, if it exists, takes the form:*

$$M = Exp\left(-\frac{h}{k_B T} + \sum_{j=0}^{N} \lambda_j k_j\right), \tag{172}$$

where $\lambda_0(x,t), \cdot, \lambda_N(x,t)$ *are Lagrange multipliers.*

Proof. Let us define the Lagrangian functional

$$F(w,\lambda_0,\dots,\lambda_N) = S(w) + \sum_{j=0}^{N} \int_{\mathbb{R}^3} \lambda_j(m_j - \langle k_j w\rangle)dx.$$

If M is a solution to (171), then there exist $\lambda_0^*,\dots,\lambda_N^*$ such that $(M,\lambda_0^*,\dots,\lambda_N^*)$ is a critical point for F:

$$\frac{dF}{dw}(M,\lambda_0^*,\dots,\lambda_N^*) = 0, \qquad \frac{dF}{d\lambda_j}(M,\lambda_0^*,\dots,\lambda_N^*) = 0 \ \ (j = 0,\dots,N). \tag{173}$$

The Fréchet derivative of S is given by:

$$\left\langle \frac{d}{dw}S(w),\varphi\right\rangle = -\frac{2}{(2\pi\hbar)^3}\int_{\mathbb{R}^3\times\mathbb{R}^3}\left(\operatorname{Log} w + \frac{h}{k_B T}\right)\varphi dx dp, \quad \forall \varphi \in L^2(\mathbb{R}^3\times\mathbb{R}^3).$$

So (173) becomes:

$$\int_{\mathbb{R}^3\times\mathbb{R}^3}\left(\operatorname{Log} w + \frac{h}{k_B T} - \sum_{s=0}^{N}\lambda_s k_s\right)\varphi dx dp = \int_{\mathbb{R}^3}(\langle k_j M\rangle - m_j)\xi dx = 0,$$

for all $\varphi \in L^2(\mathbb{R}^3\times\mathbb{R}^3)$, $\xi \in L^2(\mathbb{R}^3)$, $j = 0,\dots,N$. It follows:

$$\operatorname{Log} w + \frac{h}{k_B T} - \sum_{s=0}^{N}\lambda_s k_s = 0, \qquad \langle k_j M\rangle = m_j \ \ (j = 0,\dots,N),$$

which, thanks to the fact that Exp is the inverse function of Log, implies the statement. □

The standard choice of the moments $m_0,\dots,m_N$ is:

$$m_0 = \langle w\rangle, \qquad m_j = \langle p_j w\rangle \ \ (j = 1,2,3), \qquad m_4 = \left\langle \frac{|p|^2}{2m} w\right\rangle.$$

In this case the term $h/k_B T$ appearing in (172) can be absorbed into the sum $\sum_{j=0}^{N}\lambda_j k_j$ with a suitable re-definition of the Lagrange multipliers $\lambda_0,\dots,\lambda_N$, and M takes the form:

$$M = \operatorname{Exp}\left(\lambda_0 + \sum_{j=1}^{3}\lambda_j p_j + \lambda_4\frac{|p|^2}{2m}\right). \tag{174}$$

7.4 Quantum hydrodynamic models and quantum Maxwellians

Let us consider the Wigner-Boltzmann-Poisson system:

$$\partial_t w + \frac{p}{m} \cdot \nabla_x w - q\Theta[V]w = Q(w), \tag{175}$$

$$-\varepsilon_s \Delta V = q(n - C), \qquad n = \frac{2}{(2\pi\hbar)^3} \int_{\mathbb{R}^3} w dp. \tag{176}$$

We proceed like in Section 5. We apply the same hydrodynamic scaling, and we assume that the collision operator Q can be written as the sum of two operators Q_0, Q_1, representing two different types of collisions and characterized by different mean free times: $Q(f) = Q_0(f) + Q_1(f)$. To be precise, we assume that the collisions modeled by Q_0 happen much more frequently than the collision described by Q_1. The resulting scaled Wigner-Boltzmann-Poisson system reads as:

$$\alpha(\partial_t w + p \cdot \nabla_x w - \Theta[V]w) = Q_0(w) + \alpha Q_1(w), \tag{177}$$

$$-\lambda_D^2 \Delta V = n - C, \qquad n = \frac{2}{(2\pi\varepsilon)^2} \int_{\mathbb{R}^3} w dp, \tag{178}$$

where ε is the scaled Planck constant. We assume that Q_0 and Q_1 have the following properties:

- Q_0 conserves mass, momentum, and energy:

  $$\int_{\mathbb{R}^3} Q_0(f) dp = \int_{\mathbb{R}^3} p_i Q(f) dp = \int_{\mathbb{R}^3} |p|^2 Q_0(f) dp = 0, \quad i = 1, 2, 3,$$

  for all functions f.

- The kernel of Q_0 is spanned by quantum Maxwellians:

  $$N(Q_0) = \{f \ : \ f = M[A, v, T] \ \text{ for some } A(x), v(x), T(x)\},$$
  $$M[A, v, T](p) \equiv \text{Exp}\left(A - \frac{|p - v|^2}{2T}\right).$$

- Q_1 preserves mass: $\int_{\mathbb{R}^3} Q_1(f) dp = 0$ for all functions f.

We point out that the functions A, v, T can be written in terms of the moments $n = \langle M \rangle$, $J = \langle pM \rangle$, $ne = \langle (|p|^2/2) M \rangle$ of M through inversion of the relations:

$$n = \langle M[A, v, T] \rangle, \quad J = \langle pM[A, v, T] \rangle, \quad ne = \left\langle (|p|^2/2) M[A, v, T] \right\rangle.$$

To derive the hydrodynamic model we are interested in, we need to know how to compute the moments of the nonlocal term $\Theta[V]w$.

Lemma 7.2. *For all Wigner functions $w = w(x,p,t)$ it holds:*

$$\langle \Theta[V]w\rangle = 0, \quad \langle p\Theta[V]w\rangle = -n\nabla_x V, \quad \left\langle \frac{|p|^2}{2}\Theta[V]w\right\rangle = -\langle pw\rangle \cdot \nabla_x V,$$

$$\left\langle \frac{|p|^2}{2}p\Theta[V]w\right\rangle = -\left(\langle p\otimes p\ w\rangle + \left\langle \frac{|p|^2}{2}w\right\rangle I\right)\nabla_x V + \frac{\varepsilon^2}{8}\langle w\rangle \nabla_x\Delta_x V.$$

Hints of the proof. The scaled pseudodifferential operator $\Theta[V]$ is given by:

$$(\Theta[V]w)(x,p) = \int_{\mathbb{R}^3\times\mathbb{R}^3} \frac{i}{\varepsilon}\left(V\left(x+\frac{y}{2}\right) - V\left(x-\frac{y}{2}\right)\right) w(x,p')e^{iy\cdot(p-p')/\varepsilon}\frac{dp'dy}{(2\pi\varepsilon)^3}.$$

For any function $\mu = \mu(p)$ it holds:

$$\begin{aligned}&\int_{\mathbb{R}^3} \mu\Theta[V]wdp \\ &= \int_{\mathbb{R}^3\times\mathbb{R}^3} \frac{i}{\varepsilon}\left(V\left(x+\frac{y}{2}\right) - V\left(x-\frac{y}{2}\right)\right) w(x,p')\left(\int_{\mathbb{R}^3}\mu(p)e^{iy\cdot(p-p')/\varepsilon}\frac{dp}{(2\pi\varepsilon)^3}\right)dp'dy. \quad (179)\end{aligned}$$

It alos holds (formally):

$$\int_{\mathbb{R}^3} e^{iy\cdot(p-p')/\varepsilon}\frac{dp}{(2\pi\varepsilon)^3} = \delta(y), \quad (180)$$

$$\int_{\mathbb{R}^3} pe^{iy\cdot(p-p')/\varepsilon}\frac{dp}{(2\pi\varepsilon)^3} = -i\varepsilon e^{-iy\cdot p'/\varepsilon}\nabla\delta(y), \quad (181)$$

$$\int_{\mathbb{R}^3} |p|^2e^{iy\cdot(p-p')/\varepsilon}\frac{dp}{(2\pi\varepsilon)^3} = -\varepsilon^2 e^{-iy\cdot p'/\varepsilon}\Delta\delta(y). \quad (182)$$

The statement follows by replacing $\mu(p) = 1, p, |p|^2/2$ inside (179) and exploiting (180)–(182). □

We are now ready to derive a *closed* system of quantum hydrodynamic equations for *mixed* states. Let (w_α, V_α) be the solution of (177)–(178). Let $w = \lim_{\alpha\to 0} w_\alpha$, $V = \lim_{\alpha\to 0} V_\alpha$ (we assume these limits exist). By taking the limit $\alpha \to 0$ in (175) we get $Q_0(w) = 0$, which, due to the assumptions on Q_0, yields $w = M[A,v,T]$ for some functions $A(x,t), v(x,t), T(x,t)$, which depend on the moments $n = \langle w\rangle$, $J = nu = \langle pw\rangle$, $ne = \langle(|p|^2/2)w\rangle$ of w through the relations:

$$\begin{pmatrix} n \\ nu \\ ne\end{pmatrix} = \frac{2}{(2\pi\varepsilon)^3}\int_{\mathbb{R}^3} \operatorname{Exp}\left(A - \frac{|p-v|^2}{2T}\right)\begin{pmatrix} 1 \\ p \\ \frac{|p|^2}{2}\end{pmatrix} dp.$$

Multiplying (175) times $1, p, |p|^2/2$, integrating in p yields and taking the limit $\alpha \to 0$ yields:

$$\partial_t n + \operatorname{div}(nu) = 0, \quad (183)$$

$$\partial_t(nu) + \operatorname{div}\, \langle p \otimes pM \rangle + n\nabla V = \langle pQ_1(M) \rangle \,, \tag{184}$$

$$\partial_t(ne) + \operatorname{div} \left\langle \frac{|p|^2}{2} pM \right\rangle + nu \cdot \nabla V = \left\langle \frac{|p|^2}{2} Q_1(M) \right\rangle . \tag{185}$$

By defining the *quantum stress tensor* P and the *quantum heat flux* q:

$$P = \langle (p-u) \otimes (p-u) M \rangle \,, \quad q = \left\langle \frac{1}{2}(p-u)|p-u|^2 M \right\rangle , \tag{186}$$

we can rewrite the following integrals:

$$\langle p \otimes p\ M \rangle = P + nu \otimes u, \quad \left\langle \frac{|p|^2}{2} pM \right\rangle = (P + neI)u + q,$$

and thus (183)–(185) takes the form:

$$\partial_t n + \operatorname{div}\,(nu) = 0, \tag{187}$$

$$\partial_t(nu) + \operatorname{div}\, (nu \otimes u + P) + n\nabla V = \langle pQ_1(M) \rangle \,, \tag{188}$$

$$\partial_t(ne) + \operatorname{div}\, ((P + neI)u + q) + nu \cdot \nabla V = \left\langle \frac{|p|^2}{2} Q_1(M) \right\rangle . \tag{189}$$

Eqs. (187)–(189) are a set of nonlinear and nonlocal quantum hydrodynamic equations, describing the evolution of particle density n, momentum density nu, energy density ne for a quantum mixture (that is, a quantum system in a mixed state). This model is quite involved, and rather difficult to treat numerically. It can be simplified by performing the semiclassical expansion of the quantum Maxwellian, which allows to obtain a set of local equations. We will put this idea into practice in the next sections.

7.5 Local quantum hydrodynamic equations

The hydrodynamic equations (186)–(189) are nonlocal, because the relations between the Lagrange multipliers A, v, T involve the quantum exponential, which is indeed a nonlocal operator. In order to make the model simpler (namely, local) we assume that the scaled Plack constant ε is small and expand the quantum Maxwellian in powers of ε (*semiclassical expansion*). This procedure is showed in the Appendix A; the curious Reader can also see [34, Lemma 3.4] for further details. The quantum Maxwellian can be approximated, up to $O(\varepsilon^4)$, by a standard gaussian distribution in the velocity space times a polynomial factor; this makes it possible to compute explicitly the quantum pressure P and the quantum heat flux q. However, the expressions for P, q with this method are, albeit local, still quite involved; they can be much simplified by assuming:

- the temperature T is slowly varying: $\nabla \log T = O(\varepsilon^2)$;
- the vorticity (that is, the skew-symmetric part of the velocity gradient) is small: $\partial_x u - (\partial_x u)^\top = O(\varepsilon^2)$.

In the end, we are left with the following approximated expressions for P, q:

$$P = nTI - \frac{\varepsilon^2}{12} n (\nabla \otimes \nabla) \log n + O(\varepsilon^4), \qquad q = -\frac{\varepsilon^2}{8} n \Delta u + O(\varepsilon^4), \tag{190}$$

where the temperature T depends on the moments n, nu, ne through the relation:

$$ne = \frac{3}{2} nT + \frac{1}{2} n |u|^2 - \frac{\varepsilon^2}{24} n \Delta \log n.$$

Eqs. (187)–(189) with P, p given by (190) are a set of local quantum hydrodynamic equations. They constitute a local-in-space approximation of (186)–(189), which is meaningful provided that the scaled Planck constant ε, the (relative) temperature gradient $\nabla T/T$ and vorticity $\partial_x u - (\partial_x u)^\top$ are small enough.

7.6 Quantum drift-diffusion equations

A quantum drift-diffusion model can be derived from the Wigner-Boltzmann equation:

$$\partial_t w + \frac{p}{m} \cdot \nabla_x w - q\Theta[V] w = Q(w), \tag{191}$$

under the assumption that the collision operator Q is given by the BGK operator:

$$Q(w) = \frac{\tilde{M}[n] - w}{\tau},$$

where $\tilde{M}[n]$ is the quantum Maxwellian having particle density equal to n, which is defined as the particle density associated to the Wigner function w:

$$\tilde{M}[n] = \operatorname{Exp}\left(A - \frac{|p|^2}{2m}\right), \quad \frac{2}{(2\pi\hbar)^3} \int_{\mathbb{R}^3} \tilde{M}[n] dp = n \equiv \frac{2}{(2\pi\hbar)^3} \int_{\mathbb{R}^3} w dp.$$

The function A in the above equation is, as usual, a Lagrange multiplier. We proceed like in the derivation of the classical drift-diffusion model. The starting point in the derivation is a diffusive scaling of (191):

$$\alpha \partial_t w_\alpha + p \cdot \nabla_x w_\alpha - \Theta[V] w_\alpha = \frac{M[n] - w_\alpha}{\alpha}, \tag{192}$$

where $\alpha > 0$ is the scaled mean free path, the scaled Maxwellian $M[n]$ reads as:

$$M[n] = \operatorname{Exp}\left(A - \frac{|p|^2}{2}\right), \quad \frac{2}{(2\pi\varepsilon)^3} \int_{\mathbb{R}^3} M[n] dp = n \equiv \frac{2}{(2\pi\varepsilon)^3} \int_{\mathbb{R}^3} w dp,$$

and ε is the scaled Plack constant.

We assume that α is small. We are thus authorized to the limit $\alpha \to 0$ in (192), which yields $w_\alpha \to M[n]$. Next, we try the following Chapman-Enskog ansatz: $w_\alpha = M[n] + \alpha g_\alpha$, for some g_α that is bounded as $\alpha \to 0$. By inserting this ansatz into (192) we obtain an expression for g_α:

$$g_\alpha = -\alpha \partial_t w_\alpha - p \cdot \nabla_x w_\alpha + \Theta[V] w_\alpha,$$

which, due to the relation $\lim_{\alpha \to 0} w_\alpha = M[n]$, implies

$$\lim_{\alpha \to 0} g_\alpha = -p \cdot \nabla_x M[n] + \Theta[V] M[n]. \tag{193}$$

Recalling that $\int_{\mathbb{R}^3} \Theta[V] w dp = 0$ for all Wigner functions w, let us integrate (192) w.r.t. p:

$$\partial_t n + \frac{1}{\alpha} \operatorname{div}_x \langle p w_\alpha \rangle = 0, \tag{194}$$

where $\langle f \rangle \equiv 2(2\pi\varepsilon)^{-3} \int_{\mathbb{R}^3} f dp$ for all $f = f(x, p, t)$. However, $w_\alpha = M[n] + \alpha g_\alpha$, and $\langle p M[n] \rangle = 0$ because $M[n]$ is even w.r.t. p;[22] so by taking the limit $\alpha \to 0$ in (194) and exploiting (193) we obtain:

$$\partial_t n + \operatorname{div} J_n = 0, \quad J_n = - \langle p \otimes p \ M[n] \rangle + \langle p \Theta[V] M[n] \rangle .$$

By applying Lemma 7.2 we conclude:

$$\partial_t n + \operatorname{div} J_n = 0, \quad J_n = - \langle p \otimes p \ M[n] \rangle - n \nabla V. \tag{195}$$

This is the *quantum drift-diffusion model* that we were looking for. Notice that, just like the quantum hydrodynamic equations (186)–(189), eq. (195) is a nonlocal model, which makes it difficult to treat it analytically and numerically.

It is possible to obtain a local quantum drift-diffusion model by expanding semiclassically the quantum Maxwellian in powers of ε, exactly as we did in order to obtain the local hydrodynamic equations (187)–(190) (the curious Reader can see [33, Sec. 12] for details about the derivation):

$$\partial_t n + \operatorname{div} J_n = 0, \quad J_n = -\nabla n - n \nabla V + \frac{\varepsilon^2}{6} n \nabla \left(\frac{\Delta \sqrt{n}}{\sqrt{n}} \right). \tag{196}$$

We point out that the Bohm potential $V_B = -\frac{\varepsilon^2}{6} \frac{\Delta \sqrt{n}}{\sqrt{n}}$ is present also in (196), just as it appeared in the Madelung equations (166) (actually, the two quantities are the same up to a numerical factor): in both models, it is proportional to the square of the scaled Planck constant and it carries the informations about the quantum nature of the system.

---

[22] It can be shown that the quantum exponential operator preserves parity: if $f = f(x,p)$ is even w.r.t. p, then also $\operatorname{Exp}(f)$ has the same property.

A How to derive a semiclassical expansion of the quantum Maxwellian

Consider the scaled quantum Maxwellian:

$$M = \mathrm{Exp}\,(-h) = \mathcal{W}e^{-H}, \qquad H \equiv \mathcal{W}^{-1}h. \tag{197}$$

In (197) $h = h(x,p)$ equals the (classical symbol of the) system Hamiltonian plus some additional terms containing some Lagrange multipliers, which depend on some moments $m_0, \dots, m_N$ of M through relations of the form:

$$\int_{\mathbb{R}^3} k_j(p) M dp = m_j \quad (j = 0, \dots, N), \tag{198}$$

where $k_0(p), \dots, k_N(p)$ are suitable weight functions. A standard example in this sense is given by (174). In this Appendix an idea is presented, which allows to derive a semiclassical expansion (i.e. an expansion in powers of the scaled Planck constant ε) of the quantum Maxwellian M. We point out that this procedure works for general functions $h = h(x,p)$, not just quadratic in p. For further details, see [59].

The starting point is to relate M with the semigroup $S(\beta)$ generated by $-H$:

$$M = f(\beta)|_{\beta=1}, \quad f(\beta) \equiv \mathcal{W}S(\beta), \qquad S(\beta) \equiv e^{-\beta H} \quad \forall \beta > 0.$$

The next step consists in taking the derivative of S w.r.t. β. It holds:

$$\partial_\beta S(\beta) = -HS(\beta), \qquad \beta > 0.$$

Let us apply $\mathcal{W}$ to both sides of the above equation and rewrite it in terms of h and f:

$$\partial_\beta f(\beta) = -\mathcal{W}((\mathcal{W}^{-1}h)(\mathcal{W}^{-1}f(\beta))). \tag{199}$$

Let us introduce the so-called *Moyal product*:

$$f\#g = \mathcal{W}((\mathcal{W}^{-1}f)(\mathcal{W}^{-1}g)), \qquad \forall f = f(x,p),\ g = g(x,p).$$

With this notation (199) can be rewritten as:

$$\partial_\beta f(\beta) + h\#f(\beta) = 0, \qquad \beta > 0. \tag{200}$$

The Moyal product admits a semiclassical expansion:

$$\# = \sum_{k=0}^{\infty} \varepsilon^k \#^k,$$

where the binary operators $\#^k$ are known for all $k \geq 0$. The first three terms in the above expansion are given by:

$$f\#^0 g = fg, \qquad f\#^1 g = \frac{i}{2}\left(\nabla_x f \cdot \nabla_p g - \nabla_p f \cdot \nabla_x g\right),$$

$$f\#^2 g = -\frac{1}{8}\sum_{j,k=1}^{3}\left(\partial^2_{x_j x_k} f\ \partial^2_{p_j p_k} g - 2\partial^2_{x_j p_k} f\ \partial^2_{p_j x_k} g + \partial^2_{p_j p_k} f\ \partial^2_{x_j x_k} g\right),$$

for all functions f, g. This information allows us to obtain the second order semiclassical expansion of $f(\beta)$ (and thus also of $M = f(1)$). In fact, by writing $f(\beta) = f^0(\beta) + \varepsilon f^1(\beta) + \varepsilon^2 f^2(\beta) + O(\varepsilon^3)$ and plugging this ansatz into (200) we get:

$$\begin{aligned}
\partial_\beta f^0(\beta) + h\#^0 f^0(\beta) &= 0,\\
\partial_\beta f^1(\beta) + h\#^0 f^1(\beta) + h\#^1 f^0(\beta) &= 0,\\
\partial_\beta f^2(\beta) + h\#^0 f^2(\beta) + h\#^1 f^1(\beta) + h\#^2 f^0(\beta) &= 0,
\end{aligned}$$

for all $\beta > 0$. The above equations must be considered together with initial conditions. Since $f(0) = 1$, it follows that $f^0(0) = 1$, while $f^1(0) = f^2(0) = 0$. Recalling that $\#^0$ is simply the usual product, the problem that must be solved is:

$$\partial_\beta f^0(\beta) + h f^0(\beta) = 0 \quad (\beta > 0), \qquad f^0(0) = 1, \tag{201}$$
$$\partial_\beta f^1(\beta) + h f^1(\beta) = -h\#^1 f^0(\beta) \quad (\beta > 0), \qquad f^1(0) = 0, \tag{202}$$
$$\partial_\beta f^2(\beta) + h f^2(\beta) = -h\#^1 f^1(\beta) - h\#^2 f^0(\beta) \quad (\beta > 0), \qquad f^2(0) = 0. \tag{203}$$

The above system is actually a hierarchy of equations, each one of them depending only on the terms appearing in the equations of lower order. The solution of (201) is clearly $f^0(\beta) = \exp(-\beta h)$, which means that the dominant term in the semiclassical expansion of the quantum Maxwellian M is nothing else but the classical Maxwellian. In view of this fact we find that $h\#^1 f^0(\beta) \equiv 0$, so the right-hand side of the ODE in (202) vanishes. Since $f^1(0) = 0$, this implies that $f^1(\beta) \equiv 0$. We are left with (203), which rewrites as:

$$\partial_\beta f^2(\beta) + h f^2(\beta) = -h\#^2\left(e^{-\beta h}\right) \quad (\beta > 0), \qquad f^2(0) = 0. \tag{204}$$

As the Reader surely knows, the solution to (204) is:

$$f^2(\beta) = -\int_0^\beta e^{-(\beta - s)h}\left(h\#^2\left(e^{-sh}\right)\right) ds, \qquad \beta > 0. \tag{205}$$

It can be shown that the third order correction of $f(\beta)$ is zero, just like $f^1(\beta)$.

Summarizing up, we have obtained the following second-order semiclassical expansion of the quantum Maxwellian M:

$$M = e^{-h}\left\{1 - \varepsilon^2\int_0^1 e^{sh}\left(h\#^2\left(e^{-sh}\right)\right) ds\right\} + O(\varepsilon^4).$$

By plugging the above expression inside (198) it is possible to obtain explicit forms for the Lagrange multipliers appearing inside h, provided that suitable expressions for h and $k_0(p), \dots, k_N(p)$ are available.

References

[1] R. Adler, A. Smith, and R. Longini. Introduction to Semiconductor Physics. John Wiley & Sons, New York, 1964.

[2] L. Arkeryd. On the Boltzmann equation. Arch. Rat. Mech. Anal. 45 (1971), 1-34.

[3] A. Arnold. The relaxation-time von Neumann-Poisson equation. In: O. Mahrenholtz and R. Mennicken (eds.), Proceedings of ICIAM 95, Hamburg. Z. Angew. Math. Mech. 76 Supp. 2 (1996), 293-296.

[4] A. Arnold. Mathematical properties of quantum evolution equations. In: G. Allaire, A. Arnold, P. Degond, and T. Hou (eds.), Quantum Transport Modelling, Analysis and Asymptotics, Lecture Notes Math. 1946, 45-110. Springer, Berlin, 2008.

[5] D. Arnold. Functional Analysis. Lecture Notes, Penn State University, USA, 1997.

[6] N. Ashcroft and N. Mermin. Solid State Physics. Sanners College, Philadelphia, 1976.

[7] P. Bechouche, N. Mauser, and F. Poupaud. Semiclassical limit for the Schrödinger-Poisson equation in a crystal. Commun. Pure Appl. Math. 54 (2001), 851-890.

[8] N. Ben Abdallah and P. Degond. On a hierarchy of macroscopic models for semiconductors. J. Math. Phys. 37 (1996), 3308-3333.

[9] F. Bloch. Über die Quantenmechanik der Elektronen in Kristallgittern. Z. Phys. 52 (1928), 555-600.

[10] L. Boltzmann. Weitere Studien über das W?armegleichgewicht unter Gasmolekülen. Sitzungsberichte Akad. Wiss. Wien 66 (1872), 275-370. Translation: Further studies on the thermal equilibrium of gas molecules. In: S. Brush (ed.), Kinetic Theory, Vol. 2, 88-174. Pergamon Press, Oxford, 1966.

[11] K. Brennan. The Physics of Semiconductors. Cambridge University Press, Cambridge, 1999.

[12] H. Brezis. Functional analysis, Sobolev spaces and partial differential equations. Springer Science & Business Media, 2010.

[13] A. Caldeira and A. Leggett. Path integral approach to quantum Brownian motion. Physica A 121 (1983), 587-616.

[14] T. Carleman. Sur la théorie de l'equation intégro-différentielle de Boltzmann. Acta Mathematica 60 (1933), 91-146.

[15] F. Castella, L. Erdös, F. Frommlet, and P. Markowich. Fokker-Planck equations as scaling limits of reversible quantum systems. J. Stat. Phys. 100 (2000), 543-601.

[16] C. Cercignani, R. Illner, and M. Pulvirenti. The Mathematical Theory of Dilute Gases. Springer, New York, 1994.

[17] C. Cohern-Tannoudji, J. Dupont-Roc, and G. Grynberg. Processus dinteraction entre photons et atomes. Savoirs actuels, Intereditions/Editions du CNRS, 1988.

[18] E. S. Daus, A. Jngel, C. Mouhot and N. Zamponi. Hypocoercivity for a linearized multi-species Boltzmann system. arXiv preprint (2015), arXiv:1504.05416.

[19] R. Dautray and J.-L. Lions. Mathematical Analysis and Numerical Methods for Science and Technology. Springer, Berlin, 1985.

[20] J. Davies. The Physics of Low-Dimensional Semiconductors. Cambridge University Press, Cambridge, 1998.

[21] P. Degond and C. Ringhofer. A note on quantum moment hydrodynamics and the entropy principle. C. R. Acad. Sci. Paris, S?er. I 335 (2002), 967-972.

[22] P. Degond and C. Ringhofer. Quantum moment hydrodynamics and the entropy principle. J. Stat. Phys. 112 (2003), 587-628.

[23] L. Dìosi. On high-temperature Markovian equation for quantum Brownian motion. Europhys. Letters 22 (1993), 1-3.

[24] R. DiPerna and P.-L. Lions. On the Cauchy problem for Boltzmann equations: global existence and weak stability. Ann. Math. 130 (1989), 321-366.

[25] H. Federer. Geometric Measure Theory. Springer, New York, 1969.

[26] I. Gasser and P. Markowich. Quantum hydrodynamics, Wigner transforms and the classical limit. Asympt. Anal. 14 (1997), 97-116.

[27] I. Gasser, P. Markowich, D. Schmidt, and A. Unterreiter. Macroscopic theory of charged quantum fluids. In: Mathematical Problems in Semiconductor Physics, Pitman Res. Notes Math. Ser. 340, 42-75. Longman, Harlow, 1995.

[28] F. Golse and F. Poupaud. Limite fluide des èquations de Boltzmann des semiconducteurs pour une statistique de Fermi-Dirac. Asympt. Anal. 6 (1992), 135-160.

[29] M. Grundmann. The Physics of Semiconductors. Springer, Berlin, 2006.

[30] R. Hudson. When is the Wigner quasiprobability nonnegative? Reports Math. Phys. 6 (1974), 249-252.

[31] K. Jackson and W. Schröter (eds.). Handbook of Semiconductor Technology, Vol. 1. Wiley-VCH, Weinheim, 2000.

[32] A. Jüngel. A steady-state quantum Euler-Poisson system for semiconductors. Commun. Math. Phys. 194 (1998), 463-479.

[33] A. Jüngel. Transport Equations for Semiconductors. Lecture Notes in Physics, Vol. 773. Springer, Berlin, 2009.

[34] A. Jüngel, D. Matthes, and J.-P. Milišič. Derivation of new quantum hydrodynamic equations using entropy minimization. SIAM J. Appl. Math. 67 (2006), 46-68.

[35] A. I. Khinchin. Mathematical foundations of statistical mechanics. Courier Corporation, 1949.

[36] C. Kittel. Introduction to Solid State Physcis. 7th edition, John Wiley & Sons, New York, 1996.

[37] N. Kluksdahl, A. Kriman, D. Ferry, and C. Ringhofer. Self-consistent study of the resonant tunneling diode. Phys. Rev. B 39 (1989), 7720-7735.

[38] E. Lieb and M. Loss. Analysis. American Mathematical Society 14, 1996.

[39] P.-L. Lions and T. Paul. Sur les mesures de Wigner. Rev. Mat. Iberoamer. 9 (1993), 553-618.

[40] M. Lundstrom. Fundamentals of Carrier Transport. 2nd edition, Cambridge University Press, Cambridge, 2000.

[41] E. Madelung. Quantentheorie in hydrodynamischer Form. Z. Physik 40 (1927), 322-326.

[42] P. Markowich, N. Mauser, and F. Poupaud. A Wigner-function approach to (semi) classical limits: electrons in a periodic potential. J. Math. Phys. 35 (1994), 1066-1094.

[43] P. Markowich, C. Ringhofer, and C. Schmeiser. Semiconductor Equations. Springer, Vienna, 1990.

[44] F. Poupaud. On a system of nonlinear Boltzmann equations of semiconductors physics. SIAM J. Appl. Math. 50 (1990), 1593-1606.

[45] F. Poupaud. Diffusion approximation of the linear semiconductor Boltzmann equation: analysis of boundary layers. Asympt. Anal. 4 (1991), 293-317.

[46] M. Reed and B. Simon. Methods of Modern Mathematical Physics IV: Analysis of Operators. Academic Press, New York, 1978.

[47] C. Ringhofer. Computational methods for semiclassical and quantum transport in semiconductor devices. Acta Numerica (1997), 485-521.

[48] U. Rössler. Solid State Theory. An Introduction. Springer, Berlin, 2004.

[49] F. Scheck. Theoretische Physik 2. Springer, Berlin, 2006.

[50] K. Seeger. Semiconductor Physics. An Introduction. Springer, Berlin, 2004.

[51] R. Shankar. Principles of Quantum Mechanics. Vol. 233, Plenum Press, New York, 1994.

[52] M. Taylor. Pseudodifferential Operators. Princeton University Press, Princeton, 1981.

[53] S. Weigert. How to determine a quantum state by measurements: the Pauli problem for a particle with arbitrary potential. Phys. Rev. A 53 (1996), 2078-2083.

[54] W. Wenckebach. Essentials of Semiconductor Physics. John Wiley & Sons, Chichester, 1999.

[55] W. van Roosbroeck. Theory of flow of electron and holes in germanium and other semiconductors. Bell Syst. Techn. J. 29 (1950), 560-607.

[56] E. Wigner. On the quantum correction for the thermodynamic equilibrium. Phys. Rev. 40 (1932), 749-759.

[57] C. Wilcox. Theory of Bloch waves. J. d'Analyse Math. 33 (1978), 146-167.

[58] R. Wyatt. Quantum Dynamics with Trajectories. Springer, New York, 2005.

[59] N. Zamponi. Some fluid-dynamic models for quantum electron transport in graphene via entropy minimization. Kinet. Relat. Models 5.1 (2012), 203-221.

[60] E. Zeidler. Nonlinear Functional Analysis and Its Applications, Vol. II. Springer, New York, 1990.