Transport Equations for Semiconductors

Prof. Dr. Ansgar Jüngel
Institut für Mathematik
Johannes Gutenberg-Universität Mainz

Winter 2004

Preliminary Lectures Notes (Feb. 2005)

Contents

1 Introduction 3

2 Basic Semiconductor Physics 8
2.1 Semiconductor crystals 8
2.2 The semi-classical picture 21
2.3 The $k \cdot p$ method 28
2.4 Semiconductor statistics 33

3 Classical Kinetic Models 46
3.1 The Liouville equation 46
3.2 The Vlasov equation 52
3.3 The Boltzmann equation 58
3.4 The bipolar Boltzmann model 68

4 Classical Fluid Models 72
4.1 Derivation of the drift-diffusion equations 72
4.2 Derivation of the hydrodynamic equations 84
4.3 Derivation of the spherical harmonics expansion equations 91
4.4 Derivation of the energy-transport equations 103
4.5 Relaxation-time limits 114

5 Quantum Kinetic Models 119
5.1 The Schrödinger equation 119

5.2 The quantum Liouville equation 121
5.3 The quantum Vlasov and Boltzmann equation 125
5.4 Open quantum systems 131

6 Quantum Fluid Models 136

6.1 Zero-temperature quantum hydrodynamic equations 136
6.2 Quantum hydrodynamics and the Schrödinger equation 138
6.3 Quantum hydrodynamics and the Wigner equation 141
6.4 Derivation of the quantum drift-diffusion equations 150

1 Introduction

The modern computer and telecommunication industry relies heavily on the use and development of semiconductor devices. The first semiconductor device (a germanium transistor) has been built in 1947 by Bardeen, Brattain and Shockley, who have been awarded the Nobel prize in 1956. In the following decades, a lot of different devices for special applications have been invented; for instance, semiconductor lasers, solar cells, light-emitting diodes (LED), metal-oxide semiconductor (MOS) transistors, quantum dots, to name only a few.

A very important fact of the success of the semiconductor technology is that the device length is much smaller than that of previous electronic devices (like tube transistors). The first transistor of Bardeen, Brattain and Shockley had a characteristic length (the emitter-collector length) of 20 μm, compared to the size of a few centimeters of a tube transistor. The first Intel processor 4004, built in 1971, consisted of 2250 transistors, each of them with a characteristic length of 10 μm. This length has been reduced to 90 nm for the transistors in the Pentium 4 processor (put on the market in June 2004). Modern quantum-based devices (like tunneling diodes) have structures of only a few nanometer length. Clearly, on such scales, the physical phenomena have to be described by equations from quantum mechanics.

Usually, a semiconductor device can be considered as a device which needs an *input* (an electronic signal or light) and produces an *output* (light or an electronic signal). The device is connected to the electric circuit by contacts at which a voltage (potential difference) is applied. We are mainly interested in devices which produce an electronic signal, for instance the current of electrons through the device, generated by the applied bias. In this situation, the input parameter is the applied voltage and the output parameter is the electron current through one contact. The relation between these two physical quantities is called *current-voltage characterestic.* It is a curve in the two-dimensional current-voltage space and does neither need to be a monotone mapping nor a function (but a relation).

The main objective of these lecture notes is to derive mathematical models which describe the electron flow through a semiconductor device due to the application of a voltage. Depending on the device structure, the main transport phenomena may be very different, caused by diffusion, drift, scattering, or quantum mechanical effects. Moreover, usually a large number of electrons is flowing through a device such that a particle-like description using fluid-dynamical or kinetic equations seems to be appropriate. On the other hand, electrons in a semiconductor crystal are quantum mechanical objects such that a wave-like description using the Schrödinger equation is necessary. For this reason, we have to devise different mathematical models which are able to describe the important physical phenomena for a particular situation or for a particular device. Moreover, since in some cases we are not interested in all the available physical information, we need simpler models which help to reduce the computation cost in the numerical simulations.

This leads to a *hierarchy* of semiconductor models. Roughly speaking, we distinguish three classes of semiconductor models: quantum models, kinetic models, and fluid-dynamical (macroscopic) models. In order to give some flavor of these models, we explain these three view points, quantum, kinetic, and fluid-dynamical, in the following in a simplified framework.

The fluiddynamical view. Consider an ensemble of electrons in a spatial domain $\Omega \subset \mathbb{R}^3$ under the influence of an electric field. We wish to find equations which describe the evolution of the number $N_\Omega(t)$ of electrons and the electric field. For this, we first introduce the *electron density* $n(x,t)$ at time t by

$$N_\Omega(t) = \int_\Omega n(x,t)\,dx.$$

Another important physical quantity is the *electron current density* $J(x,t)$. We assume that it is given as the sum of the drift current, $q\mu_k nE$, and the diffusion current, $qD_n\nabla n$,

$$J = qD_n\nabla n + q\mu_n nE, \tag{1.1}$$

where q is the elementary charge, μ_n the mobility and D_n the diffusivity of the electrons. Physically, it is reasonable to assume that the temporal change of the electron number is equal to the (normal component of the) current flow through the semiconductor boundary

$$\frac{dN_\omega}{dt} = \frac{1}{q}\int_{\partial\omega} J\cdot\nu\,ds,$$

where $\omega \subset \Omega$ and ν is the exterior unit normal to $\partial\omega$. The divergence theorem implies

$$\int_\omega \frac{\partial n}{\partial t}\,dx = \frac{dN_\omega}{dt} = \frac{1}{q}\int_\omega \mathrm{div} J\,dx \qquad \text{for all } \omega \subset \Omega.$$

Therefore,

$$\frac{\partial n}{\partial t} - \frac{1}{q}\mathrm{div} J = 0 \qquad \text{in } \Omega. \tag{1.2}$$

This equation expresses the conservation of mass.

It remains to find an equation for the electrostatic potential. We start with the Maxwell equations

$$\mathrm{curl}\ E = 0, \qquad \mathrm{div} D = \varrho \qquad \text{in } \mathbb{R}^3, \tag{1.3}$$

valid for vanishing magnetic fields. Here, D is the displacement vector and ϱ the total space charge. The first equation provides the existence of a potential V such that $E = -\nabla V$.

The function V is called the *electrostatic potential.* We assume that the displacement vector D and the electric field E are related by

$$D = \varepsilon_s E,$$

where ε_s is the semiconductor permittivity. Generally, ε_s is a matrix and depends on the spatial variable. In an isotropic homogeneous semiconductor, ε_s is a scalar. The total space charge ϱ is given by

$$\varrho = -qn + qC(x),$$

where $C(x)$ is the concentration of the fixed charged background ions in the semiconductor crystal and will be subject of later investigation (see Section 2.4). Thus, the second equation in (1.3) gives

$$\varepsilon_s \Delta V = \mathrm{div}(-\varepsilon_s \nabla V) = -\mathrm{div} D = q(n - C(x)), \tag{1.4}$$

if ε_s is a constant scalar (the semiconductor permittivity). This equation is called the *Poisson equation.*

We have motivated that the electron density $n(x,t)$ and the electrostatic potential $V(x,t)$ are solutions of (1.1), (1.2), and (1.4). This set of equations,

$$\frac{\partial n}{\partial t} - \mathrm{div}(D_n \nabla n - \mu_n n \nabla V) = 0, \qquad \varepsilon_s \Delta V = q(n - C(x)), \qquad x \in \mathbb{R}^3,\ t > 0,$$

is called the *drift-diffusion model* and is one of the most important semiconductor models.

The kinetic view. Consider a single electron in the semiconductor and interpret it as a *particle* at position $x(t)$ with velocity $v(t)$. These functions are a solution of Newton's equations

$$\frac{dx}{dt} = v, \qquad m\frac{dv}{dt} = F, \qquad t > 0, \tag{1.5}$$
$$x(0) = x_0, \qquad v(0) = v_0,$$

where m is the electron mass and F a force given by $F = -qE$. The *state* of the electron is given by a distribution function $f(x,v,t)$, i.e. a probability density in the (x,v)-phase space. More precisely,

$$\int_B f(x,v,t)\,dx\,dv$$

is the probability at time t to find the electron in the phase space set B. It is reasonable to assume that $f(x,v,t)$ does not change along the trajectory $(x(t), v(t))$ of the electron,

$$f(x(t), v(t), t) = f(x_0, v_0, 0) \qquad \text{for all } t > 0.$$

This implies that

$$0 = \frac{d}{dt} f(x(t), v(t), t) = \frac{\partial f}{\partial t} + \nabla_x f \cdot \frac{dx}{dt} + \nabla_v f \cdot \frac{dv}{dt},$$

and, by (1.5),

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f - \frac{q}{m} E \cdot \nabla_v f = 0 \qquad \text{in } \mathbb{R}^3. \tag{1.6}$$

This equation is called the *Liouville equation.*

Macroscopic quantities, like the electron density $n(x,t)$ and the electron current density, are defined in terms of the distribution function by

$$\begin{aligned} n(x,t) &= \int_{\mathbb{R}^3} f(x,v,t)\,dv, \\ J(x,t) &= -q \int_{\mathbb{R}^3} f(x,v,t) v\,dv. \end{aligned}$$

The macroscopic particle velocity $u(x,t)$ is then given by

$$u(x,t) = \frac{\int f(x,v,t) v\,dv}{\int f(x,v,t)\,dv} = -\frac{J(x,t)}{qn(x,t)}. \tag{1.7}$$

We notice that equation (1.2) can be easily derived from the Liouville equation. Indeed, integrating (1.6) over $v \in \mathbb{R}^3$ yields, by the divergence theorem,

$$\begin{aligned} 0 &= \int_{\mathbb{R}^3} \left(\frac{\partial f}{\partial t} + \mathrm{div}_x(vf) - \frac{q}{m} \mathrm{div}_v(Ef) \right) dv \\ &= \frac{\partial}{\partial t} \int_{\mathbb{R}^3} f\,dv + \mathrm{div}_x \int_{\mathbb{R}^3} vf\,dv \\ &= \frac{\partial n}{\partial t} - \frac{1}{q} \mathrm{div}_x J. \end{aligned}$$

The quantum view. We consider now a single electron interpreted as a *wave.* Then it is described by the complex-valued wave function $\psi(x,t)$ which is a solution of the *Schrödinger equation*

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi - qV(x)\psi, \quad t > 0, \qquad \psi(x,0) = \psi_0(x), \quad x \in \mathbb{R}^3. \tag{1.8}$$

Here, i is the complex unit with $i^2 = -1$, $\hbar = h/2\pi$ the reduced Planck constant, and $V(x)$ the (time-independent) electrostatic potential. The measurable, macroscopic quantities $n(x,t)$ and $J(x,t)$ are defined by

$$n(x,t) = |\psi(x,t)|^2, \qquad J(x,t) = -\frac{\hbar q}{m} \mathrm{Im}\,(\overline{\psi} \nabla \psi), \tag{1.9}$$

where $\mathrm{Im}\,(z)$ denotes the imaginary part of a complex number z and $\overline{z}$ its conjugate number.

Usually, the Schrödinger equation (1.8) is solved by the ansatz

$$\psi(x,t) = w(t)u(x).$$

Inserting this expression into (1.8) and dividing by ψ gives

$$i\hbar\frac{w_t}{w} = -\frac{\hbar^2}{2m}\frac{\Delta u}{u} - qV(x).$$

The left-hand side only depends on t, the right-hand side only on x. Therefore, both sides must be constant. We call this constant E (since physically, it has the unit of an energy). The solution of

$$i\hbar\frac{dw}{dt} = Ew$$

is given by $w(t) = \exp(-iEt/\hbar)$ (neglecting the integration constant which is put into $u(x)$). The function $u(x)$ solves

$$-\frac{\hbar^2}{2m}\Delta u - qV(x)u = Eu \qquad \text{in } \mathbb{R}^3. \tag{1.10}$$

This is the stationary Schrödinger equation for a single electron with energy E. Mathematically, (1.10) is an eigenvalue problem and solutions to (1.8) are eigenfunctions with eigenvalues E. Then, the solution to (1.8) is given by

$$\psi(x,t) = e^{-iEt/\hbar}u(x),$$

which oscillates in time with a frequency $\omega = E/\hbar$.

The above models are all stated or motivated in a very simplified situation. For a more precise description, the following questions have to be answered:

- How can the above models be derived (and not only motivated) from *basic principles*?
- How can the motion of *many electrons* be modeled?
- What is the influence of the *semiconductor crystal* on the motion of the electrons?
- How can *collisions* of the electrons with the crystal atoms or with other particles can be taken into account?
- Do there exist *relations* between the above models?

These questions will be answered (at least partially) in the following chapters.

2 Basic Semiconductor Physics

In this chapter we present a short summary of the physics and main properties of semiconductors. Only those subjects relevant to the subsequent chapters are included here. We refer to [1, 3, 40, 61] for introductory textbooks of solid-state and semiconductor physics and to [11, 37, 45, 50, 65] for more advanced expositions.

2.1 Semiconductor crystals

What is a semiconductor? Historically, the term "semiconductor" has been used to denote solid materials whose conductivity is much larger than that for insulators but much smaller than that for metals, measured at room temperature. A modern and more precise definition is that a semiconductor is a solid with an *energy gap* larger than zero and smaller than about 4eV (electron volt). Metals have no energy gap, whereas it is usually larger than 4eV in insulators. In order to understand the notion "energy gap", we have to introduce to the crystal structure of solids.

A solid is made of an infinite three-dimensional array of atoms arranged according to a lattice

$$L = \{n_1\vec{a}_1 + n_2\vec{a}_2 + n_3\vec{a}_3 : n_1, n_2, n_3 \in \mathbb{Z}\} \subset \mathbb{R}^3,$$

where $\vec{a}_1, \vec{a}_2, \vec{a}_3$ are the basis vectors of L, called *primitive vectors* of the lattice (see Figure 2.1). The set L is also called a *Bravais lattice*. The lattice atoms generate a periodic electrostatic potential V_L,

$$V_L(x+y) = V_L(x) \qquad \text{for all } x \in \mathbb{R}^3,\ y \in L,$$

which is the superposition of the Coulomb potentials

$$V_j(x) = -\frac{1}{4\pi}\,\frac{q^2}{|x-x_j|}$$

of the crystal atoms located at $x = x_j$ (see Figure 2.2). The state of an electron moving in this periodic potential is described by an eigenfunction $\psi(x)$ of the stationary Schrödinger equation (1.10):

$$-\frac{\hbar^2}{2m}\Delta\psi - qV_L(x)\psi = E\psi \qquad \text{in } \mathbb{R}^3, \tag{2.1}$$

where $\psi : \mathbb{R}^3 \to \mathbb{C}$ is the (stationary) wave function, and as in Chapter 1, $\hbar = h/2\pi$ is the reduced Planck constant, m the electron mass (at rest), q the elementary charge, and E the energy (or the eigenvalue corresponding to ψ). We illustrate this equation and its solutions by two examples.

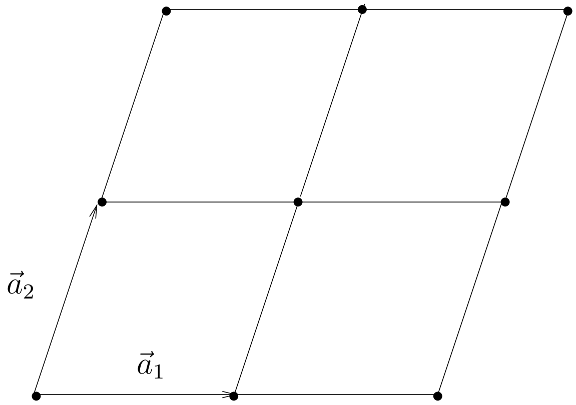

Figure 2.1: Illustration of a two-dimensional lattice L.

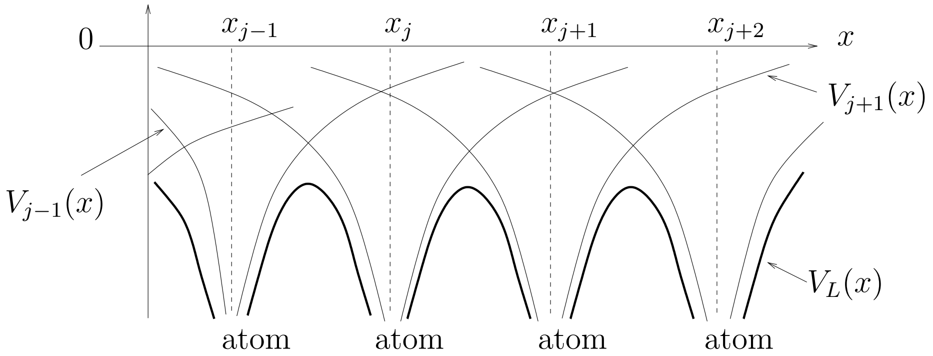

Figure 2.2: Potentials $V_j(x)$ of a single atom at $x = x_j$ and net potential $V_L(x)$ of a one-dimensional crystal lattice.

Example 2.1 (Motion of a free electron)
Consider a free electron moving in a one-dimensional vacuum, i.e. $V_L(x) = 0$ for all $x \in \mathbb{R}$. Then the eigenfunctions of

$$-\frac{\hbar^2}{2m}\psi'' = E\psi \qquad \text{in } \mathbb{R} \tag{2.2}$$

are given by

$$\psi_k(x) = Ae^{ik\cdot x} + Be^{-ik\cdot x}, \qquad x \in \mathbb{R},$$

where $k^2 = 2mE/\hbar^2$, and the eigenvalues are

$$E = E(k) = \frac{\hbar^2 k^2}{2m}, \qquad k \in \mathbb{R}.$$

To be precise, k is complex but the purely imaginary part $i\gamma$ ($\gamma \in \mathbb{R}$) leads to solutions of the type $\exp(\pm ikx) = \exp(\mp\gamma x)$ for $x \in \mathbb{R}$. The integral

$$\int_{\mathbb{R}} |\psi(x)|^2 \, dx$$

is the particle mass (or particle number) and should be finite. However,

$$\int_{\mathbb{R}} |\exp(\mp\gamma x)|^2 \, dx$$

is never finite. Therefore, only $k \in \mathbb{R}$ gives physically reasonable solutions.

We have found that the eigenvalue problem (2.2) has infinitely many solutions parametrized by $k \in \mathbb{R}$ and corresponding to different energies $E(k)$. The functions $\exp(\pm ikx)$ are called *plane waves*. Thus, the eigenstates of a free particle are plane waves. □

Example 2.2 (Infinite square-well potential)
The infinite square-well potential is a one-dimensional structure of length L with a vanishing potential inside the well and with an infinite potential at its boundaries. As the potential is confining an electron to the inner region, we have to solve the Schrödinger equation (2.1) in the interval $(0, L)$ with boundary conditions

$$\psi(0) = \psi(L) = 0$$

and potential $V_L(x) = 0$ for $x \in (0, L)$ (see Figure 2.3). The eigenfunctions to (2.1) are as in Example 2.1

$$\psi(x) = Ae^{ik\cdot x} + Be^{-ik\cdot x}$$

with $k^2 = 2mE/\hbar^2$ and $k \in \mathbb{R}$. The constants A and B are determined by the boundary conditions:

$$0 = \psi(0) = A + B, \qquad 0 = \psi(L) = Ae^{ikL} + Be^{-ikL}.$$

Inserting the first equation in the second one gives

$$0 = A(e^{ikL} - e^{-ikL}) = \frac{A}{2i} \sin kL.$$

Disregarding the trivial solution $A = 0$ (and hence $\psi \equiv 0$), the parameter k should be such that $\sin kL = 0$, i.e., kL is a multiple of π. Thus

$$k = \frac{n\pi}{L}, \qquad n \in \mathbb{N}_0.$$

The eigenfunctions become

$$\psi_k(x) = \widetilde{A} \sin kx, \qquad x \in [0, L],$$

where $\widetilde{A} = A/2i$, with eigenvalues

$$E(k) = \frac{\hbar^2 k^2}{2m}.$$

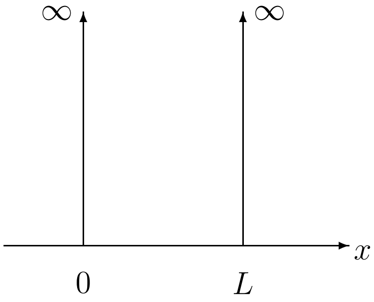

Figure 2.3: Infinite square-well potential.

The integration constant $\widetilde{A} \in \mathbb{R}$ can be determined by assuming that

$$\int_0^L |\psi_k(x)|^2 \, dx = 1$$

holds. A simple computation shows that $\widetilde{A} = \sqrt{2L}$. The system only allows *discrete* energy states. In particular, the parameter k can only take discrete values. ☐

Since the lattice potential V_L is periodic, one might hope that the whole-space Schrödinger problem (2.1) can be reduced to an eigenvalue problem on a cell of the lattice. Bloch's theorem states that this is indeed possible. Before we can formulare the result, we need some definitions.

Definition 2.3 [49, p. 12f.]

(1) The reciprocal lattice *(or dual lattice) L^* of L is defined by*

$$L^* = \{n_1\vec{a}_1^* + n_2\vec{a}_2^* + n_3\vec{a}_3^* : n_1, n_2, n_3 \in \mathbb{Z}\},$$

where the primitive vectors $\vec{a}_1^*, \vec{a}_2^*, \vec{a}_3^* \in \mathbb{R}^3$ *are uniquely given by the relation*

$$\vec{a}_m \cdot \vec{a}_n^* = 2\pi\delta_{mn}, \qquad m, n = 1, 2, 3. \tag{2.3}$$

(2) The connected set $D \subset \mathbb{R}^3$ is called primitive cell *of L (or L^*) if the volume of D equals the volume of the parallelepiped spanned by the basis vectors of L (or L^*),*

$$\text{vol } D = \vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3) \qquad (\textit{or} \text{ vol } D = \vec{a}_1^* \cdot (\vec{a}_2^* \times \vec{a}_3^*)),$$

and if the whole space $\mathbb{R}^3$ is covered by the union of translates of D by the primitive vectors (see Figure 2.4).

(3) The (first) Brillouin zone $B \subset \mathbb{R}^3$ *is the primitive cell of the reciprocal lattice L^* which consists of all points being closer to the origin than to any other point of L^* (see Figure 2.4):*

$$B = \{k \in \mathbb{R}^3 : |k| \leq \min_\ell |k + \ell|, \ \ell \in L^*, \ \ell \neq 0\}.$$

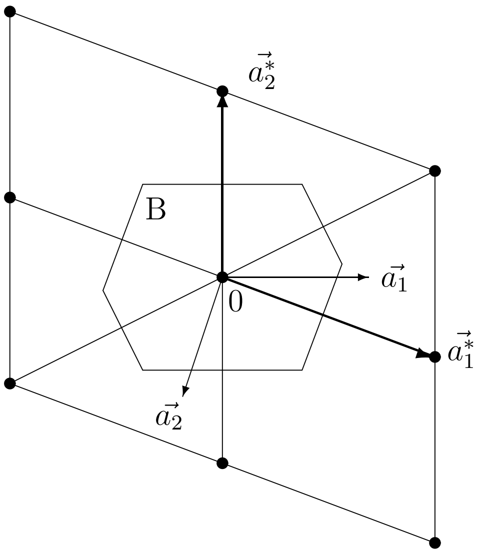

Figure 2.4: The primitive vectors of a two-dimensional lattice L and its reciprocal lattice L^* and the Brillouin zone B.

We give some explanations of the above definition. What is the meaning of the reciprocal lattice? The reciprocal lattice vectors and the direct lattice vectors can be seen as conjugate variables, like time and frequency are conjugate variables in signal analysis. In fact, let $x \in L$ and $k \in L^*$ be given such that

$$x = \sum_{m=1}^{3} \alpha_m \vec{a}_m \qquad \text{and} \qquad k = \sum_{n=1}^{3} \beta_n \vec{a}_n^*,$$

where $\alpha_m, \beta_n \in \mathbb{Z}$. Then, by (2.3),

$$e^{ik\cdot x} = \exp\left(i \sum_{m,n=1}^{3} \alpha_m \beta_n \cdot 2\pi\delta_{mn}\right) = \exp\left(2\pi i \sum_{m=1}^{3} \alpha_m \beta_n\right) = 1. \tag{2.4}$$

As x has the dimension of length, k has the dimension of inverse length and therefore, k is called a wave vector. (More precisely, k is called a *pseudo-wave vector*; see below). Physically, the reciprocal lattice appears in X-ray diffraction experiments on crystals. It can be shown that the peaks of intensity of the reflected X-ray are obtained when the change in momentum $\triangle k$ of the X-ray wave is an element of the reciprocal lattice [11, p. 404]. This allows to determine the structure of the crystal lattice.

How can the Brillouin zone of the reciprocal latice be constructed? Mathematically, the primitive vectors $\vec{a}_n^*$ of the Brillouin zone are given as the inverse of the matrix A whose columns are the vectors $\vec{a}_m$. More precisely, let $\vec{a}_n = (a_{1n}, a_{2n}, a_{3n})^\top, \vec{a}_m^* = (a_{1m}^*, a_{2m}^*, a_{3m}^*)^\top \in \mathbb{R}^3$ and $A^* = (\vec{a}_1^*, \vec{a}_2^*, \vec{a}_3^*)$, $A = (\vec{a}_1, \vec{a}_2, \vec{a}_3) \in \mathbb{R}^{3\times 3}$. Then (2.3) implies that

$$(A^\top A^*)_{mn} = \sum_{j=1}^{3} a_{jm} a^*_{jn} = \vec{a}_m \cdot \vec{a}_n = 2\pi\delta_{mn}$$

and thus $A^\top A^* = 2\pi I$, where I is the unit matrix of $\mathbb{R}^{3\times 3}$. Hence

$$A^* = 2\pi(A^\top)^{-1} = 2\pi(A^{-1})^\top. \tag{2.5}$$

Graphically, the Brillouin zone can be constructed as follows. Draw arrows from a lattice point to its nearest neighbors and cut them in half. Then the planes through these mid points are perpendicular to the arrows from the surface of the (bounded) Brillouin zone. In two space dimensions, the Brillouin zone is a hexagon or a square (see Figure 2.4). In three space dimensions, the zone is a polyhedron (a "capped" octahedron).

Lemma 2.4 *The volumes of a primitive cell D and its Brillouin zone B are related to*

$$\text{vol } B = \frac{(2\pi)^3}{\text{vol } D}.$$

Proof: With the above notations,

$$\begin{aligned}\text{vol } D &= \vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3) = \det(\vec{a}_1, \vec{a}_2, \vec{a}_3) = \det A,\\ \text{vol } B &= \det(\vec{a}^*_1, \vec{a}^*_2, \vec{a}^*_3) = \det A^*\end{aligned}$$

and hence, by (2.5),

$$\text{vol } B = \det[2\pi(A^{-1})^\top] = (2\pi)^3 \det(A^{-1}) = \frac{(2\pi)^3}{\det A} = \frac{(2\pi)^3}{\text{vol } D}. \qquad \square$$

Now we can formulate the Bloch theorem.

Theorem 2.5 (Bloch)
Let V_L be a periodic potential, i.e., $V_L(x+y) = V_L(x)$ for all $x \in \mathbb{R}^3$ and $y \in L$ (the Bravais lattice). Then the eigenfunctions of

$$-\frac{\hbar^2}{2m}\Delta\psi - qV_L(x)\psi = E\psi \qquad \text{in } \mathbb{R}^3 \tag{2.6}$$

can be written as

$$\psi(x) = e^{ik\cdot x} u(x) \tag{2.7}$$

for some $k \in B$ (the Brillouin zone) and some function $u(x)$ satisfying $u(x) = u(x+y)$ for all $x \in \mathbb{R}^3$, $y \in L$.

This theorem asserts that any eigenfunction of the Schrödinger equation is the product of a plane wave $e^{ik\cdot x}$ and a function having the periodicity of the lattice L. Before we can give a formal proof of this result, we need some preparations. In the following we do neither specify the underlying (Hilbert) space nor the domain of definition of the considered operators in order to simplify the presentation.

Definition 2.6 *Let T be an operator, defined on some Hilbert space X with scalar product $(\cdot,\cdot)$.*

(1) The adjoint operator T^* *is formally defined by*

$$(Tx, y) = (x, T^*y) \qquad \textit{for all } x, y \in X.$$

(2) The operator T is called normal *if $T^*T = TT^*$.*

Example 2.7 (1) Define the translation operator T_a for some $a \in \mathbb{R}^3$ by

$$(T_a\psi)(x) = \psi(x + a), \qquad x \in \mathbb{R}^3, \tag{2.8}$$

for $\psi \in L^2(\mathbb{R}^3; \mathbb{C})$, where the scalar product is given by

$$(\psi, \chi) = \int_{\mathbb{R}^3} \overline{\psi(x)}\chi(x)\,dx.$$

We claim that T_a is normal. From

$$(\psi, T_a\chi) = \int_{\mathbb{R}^3} \overline{\psi(x)}\chi(x+a)\,dx = \int_{\mathbb{R}^3} \overline{\psi(y-a)}\chi(y)\,dy = (T_{-a}\psi, \chi)$$

follows that $T_a^* = T_{-a}$. Furthermore,

$$T_{-a}T_a\psi = \psi = T_aT_{-a}\psi,$$

and thus, $(T_a)^{-1} = T_{-a} = T_a^*$. This implies that $T_a^*T_a = I = T_aT_a^*$, where I is the identity operator. Hence T_a is normal. (In fact, we have even shown that T_a is *unitary*, i.e. $T_a^* = (T_a)^{-1}$.)

Next we claim that the eigenvalues of T_a are given by $\lambda = e^{i\theta}$ for $\theta \in \mathbb{R}$. To see this, let $T_a\psi = \lambda\psi$ and set $\|\psi\|^2 = (\psi, \psi)$. Then

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

and thus, $|\lambda| = 1$ or $\lambda = e^{i\theta}$ with $\theta \in \mathbb{R}$.

(2) We state without proof that the *Hamilton operator* H, defined by

$$H\psi = -\frac{\hbar^2}{2m}\Delta\psi - qV_L(x)\psi \tag{2.9}$$

for appropriate ψ, is also normal. (In fact, H is *self-adjoint*, i.e. $H^* = H$.) $\square$

Theorem 2.8 (Spectral theorem for normal operators)
Let T be a normal operator on the Hilbert space X. Then there exists a orthonormal basis of eigenvectors of T.

Theorem 2.9 *Let S and T be two normal operators on the Hilbert space X such that $ST = TS$. Then there exists an (orthonormal) basis of X whose elements are eigenvectors for both S and T.*

In Example 2.7 and Theorems 2.8 and 2.9 we have simplified the situation. To be precise, the Hamilton operator is defined for all ψ in a dense subset of the underlying Hilbert space and the operators in Theorems 2.8 and 2.9 have to be compact (i.e., if the sequence (x_n) is bounded then (Tx_n) and (Sx_n) contain convergent subsequences).

We are now able to turn to the (formal) proof of Bloch's theorem.

Proof of Theorem 2.5: By Example 2.7, the translation operator (2.8) and the Hamilton operator (2.9) are normal. Moreover, since for all $a \in L$,

$$\begin{aligned}(T_a H\psi)(x) &= T_a\left(-\frac{\hbar^2}{2m}\Delta\psi - qV_L\psi\right)(x) = -\frac{\hbar^2}{2m}\Delta\psi(x+a) - qV_L(x+a)\psi(x+a) \\ &= -\frac{\hbar^2}{2m}\Delta\psi(x+a) - qV_L(x)\psi(x+a) = H(\psi(x+a)) \\ &= (HT_a\psi)(x),\end{aligned}$$

by Theorem 2.9, there exists a basis of eigenfunctions for both T_a and H. By Example 2.7 (1), for given primitive vector $\vec{a}_j \in L$, there exists $\theta_j \in \mathbb{R}$ such that

$$T_{-\vec{a}_j}\psi = e^{i\theta_j}\psi, \tag{2.10}$$

and ψ is also an eigenfunction of H. We set

$$k = -\frac{1}{2\pi}\sum_{j=1}^{3}\theta_j \vec{a}_j^*,$$

where $\vec{a}_j^* \in L^*$. Then (2.4) implies that

$$k \cdot \vec{a}_j = -\theta_j. \tag{2.11}$$

We define $u(x) = e^{-ik\cdot x}\psi(x)$, $x \in \mathbb{R}^3$. We have to show that $u(x+y) = u(x)$ for all $x \in \mathbb{R}^3$ and $y \in L$. Since every $y \in L$ is a linear combination of the $\vec{a}_j$, it is sufficient to prove the periodicity for $y = \vec{a}_j$. We obtain, using (2.10) and (2.11),

$$\begin{aligned}u(x) &= e^{-ik\cdot x}\psi(x) = e^{-ik\cdot x}(T_{-\vec{a}_j}\psi)(x+\vec{a}_j) = e^{-ik\cdot x}e^{i\theta_j}\psi(x+\vec{a}_j) \\ &= e^{-ik\cdot x}e^{i\theta_j}e^{ik(x+\vec{a}_j)}u(x+\vec{a}_j) = e^{i(\theta_j + k\cdot\vec{a}_j)}u(x+\vec{a}_j) = u(x+\vec{a}_j).\end{aligned}$$

It remains to show that k can be restricted to the Brillouin zone. Let $k \in \mathbb{R}^3$ and decompose $k = k_B + \ell$, where $k_B \in B$ and $\ell \in L^*$ is a point in the reciprocal lattice closest to k (see Figure 2.5). Then

$$\psi(x) = e^{ik\cdot x}u(x) = e^{ik_B\cdot x}\phi(x), \tag{2.12}$$

where $\phi(x) = e^{i\ell \cdot x}u(x)$ satisfies, in view of (2.4),

$$\phi(x+y) = e^{i\ell\cdot x}\, e^{i\ell\cdot y}u(x+y) = e^{i\ell\cdot x}u(x) = \phi(x)$$

for all $x \in \mathbb{R}^3$ and $y \in L$. Now, the representation (2.12) is of the form as stated in the theorem. $\square$

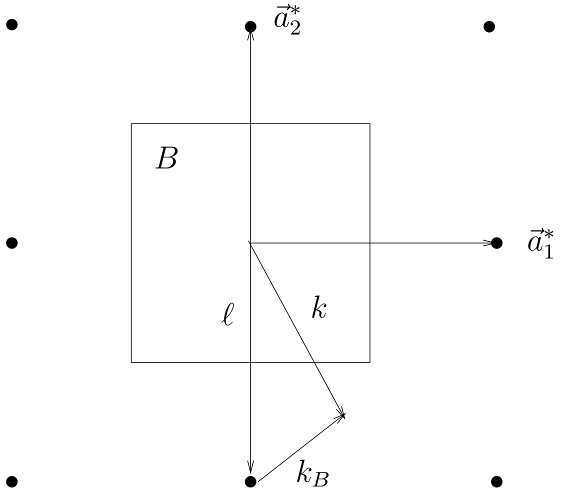

Figure 2.5: Illustration of $k = k_B + \ell$ with $k_B \in B$ and $\ell \in L^*$ for the proof of Theorem 2.5.

Inserting the decomposition (2.7) into the Schrödinger equation (2.6) shows that $u(x)$ satisfies the eigenvalue problem

$$-\frac{\hbar^2}{2m}(\Delta u + 2ik\cdot\nabla u) + \left(\frac{\hbar^2}{2m}|k|^2 - qV_L(x)\right)u = Eu \qquad \text{in } D, \tag{2.13}$$

with periodicity conditions

$$u(x+y) = u(x), \qquad x \in \mathbb{R}^3,\ y \in L, \tag{2.14}$$

since

$$\begin{aligned} Eu = e^{-ik\cdot x}E\psi &= e^{-ik\cdot x}\left[-\frac{\hbar^2}{2m}\Delta\left(e^{ik\cdot x}u\right) - qV_L(x)e^{ik\cdot x}u\right] \\ &= -\frac{\hbar^2}{2m}\left(\Delta u + 2ik\cdot\nabla u - |k|^2u\right) - qV_L(x)u. \end{aligned}$$

For each $k \in B$, there exists a sequence of eigenfunctions $u = u_{n,k}$ and eigenvalues $E = E_n(k)$ of (2.13)–(2.14). (This follows from the self-adjointness of the operator defined by the left-hand side of (2.13); see [56] for details). In particular, the functions $(u_{n,k})_{n\in\mathbb{N}}$ form an orthonormal basis of the underlying Hilbert space (see Theorem 2.8). Then we can introduce the so-called *Bloch functions*

$$\psi_{n,k}(x) = e^{ik\cdot x}u_{n,k}(x).$$

They satisfy the Schrödinger equation in the primitive cell D of L,

$$-\frac{\hbar^2}{2m}\Delta\psi_{n,k} - qV_L(x)\psi_{n,k} = E_n(k)\psi_{n,k} \qquad \text{in } D \tag{2.15}$$

with pseudo-periodic boundary conditions

$$\psi_{n,k}(x+y) = e^{ik\cdot y}\psi_{n,k}(x), \qquad x, x+y \in \partial D. \tag{2.16}$$

In some sense, the functions $\psi_{n,k}$ are plane waves which are modulated by a periodic function $u_{k,n}$ taking into account the influence of the atom lattice. This also explains why k is termed *pseudo-wave vector*. In fact, k appears in distorted plane waves and is thus not a real wave vector.

The function $k \mapsto E_n(k)$ is called *dispersion relation* or the n-th *enery band*. It shows how the energy of the n-th band depends on the (pseudo-)wave vector k. The union of ranges of E_n over $n \in \mathbb{N}$ is not necessarily the whole real line $\mathbb{R}$, i.e., there may exist energies E^* for which there is no $n \in \mathbb{N}$ and no $k \in B$ such that $E_n(k) = E^*$. The connected components of the set of energies with this non-existence property are called *energy gaps*. We illustrate this property by the following example.

Example 2.10 (Kronig-Penney model)
The Kronig-Penney model is a simple model representing a one-dimensional single-crystal lattice (see also [50, Sec. 3.1.2] or [11, Sec. 8.2]). The potential of the lattice atoms is modeled by the function

$$V_L(x) = \begin{cases} -V_0 & \text{if } -b < x \le 0 \\ 0 & \text{if } 0 < x \le a, \end{cases}$$

and V_L is extended to $\mathbb{R}$ with period $a+b$:

$$V_L(x) = V_L(x+a+b) \qquad \text{for } x \in \mathbb{R},$$

where $a, b > 0$, $V_0 < 0$ (see Figure 2.6). The lattice atoms are supposed to be at the positions $a/2 + n(a+b)$, $n \in \mathbb{Z}$.

In order to solve the Schrödinger equation (2.6) we make the Bloch decomposition

$$\psi(x) = e^{ikx}u(x),$$

where $u(x)$ is a $(a+b)$-periodic solution of (see (2.13))

$$-u'' - 2iku' + k^2u = \frac{2m}{\hbar^2}(E + qV_L)u \qquad \text{in } \mathbb{R}. \tag{2.17}$$

We proceed as in [50, Sec. 3.1.2]. First we solve (2.17) in the interval $(0,a)$. Then $V_L(x) = 0$, $x \in (0,a)$, and the ansatz $u(x) = e^{i\gamma x}$ ($x \in \mathbb{R}$) leads to

$$(\gamma^2 + 2k\gamma + k^2 - \alpha^2)e^{i\gamma x} = 0,$$

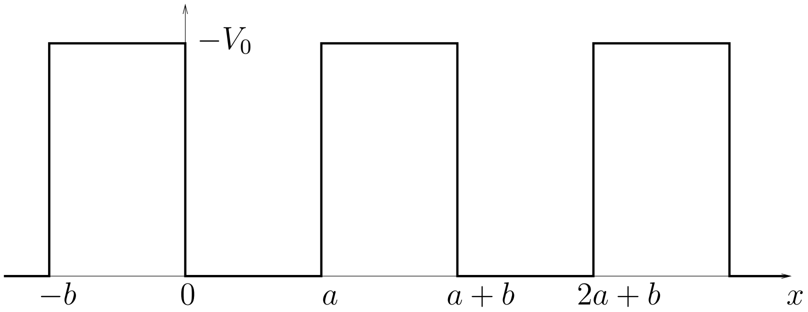

Figure 2.6: The periodic square-well potential $V_L(x)$ of the Kronig-Penney model.

where

$$\alpha = \sqrt{\frac{2mE}{\hbar^2}}.$$

The solutions of the above quadratic equation are given by $\gamma_{1/2} = -k \pm \alpha$ and therefore,

$$u_1(x) = Ae^{i(\alpha-k)x} + Be^{-i(\alpha+k)x}, \qquad x \in (0,a).$$

In the interval $(-b,0)$ we make again the ansatz $u(x) = e^{i\gamma x}$ yielding

$$(\gamma^2 + 2k\gamma + k^2 - \beta^2)e^{i\gamma x} = 0,$$

where

$$\beta = \sqrt{\frac{2m(E-qV_0)}{\hbar^2}}.$$

The solutions are $\gamma_{1/2} = -k \pm \beta$, and thus

$$u_2(x) = Ce^{i(\beta-k)x} + De^{-i(\beta+k)x}, \qquad x \in (-b,0).$$

We notice that β is purely imaginary if $E < qV_0$, i.e., the electrons are bound within the crystal, and β is real if $E > qV_0$.

The constants A, B, C, D are determined from the interface conditions. We assume that u is continuously differentiable and periodic on $\mathbb{R}$:

$$\begin{array}{rclcrcl} u_1(0) &=& u_2(0), & \qquad & u_1'(0) &=& u_2'(0), \\ u_1(a) &=& u_2(-b), & & u_1'(a) &=& u_2'(-b). \end{array}$$

We obtain the following four equations:

$$\begin{aligned} A + B - C - D &= 0, \\ (\alpha-k)A - (\alpha+k)B - (\beta-k)C + (\beta+k)D &= 0, \\ Ae^{i(\alpha-k)a} - Be^{-i(\alpha+k)a} - Ce^{-i(\beta-k)b} - De^{i(\beta+k)b} &= 0, \\ (\alpha-k)Ae^{i(\alpha-k)a} - (\alpha+k)Be^{-i(\alpha+k)a} - (\beta-k)Ce^{-i(\beta-k)b} + (\beta+k)De^{i(\beta+k)b} &= 0 \end{aligned}$$

for the unknowns A, B, C, D. This is a homogeneous linear system which has non-trivial solutions only if the determinant of the coefficient matrix vanishes. A very lenghty calculation shows that this condition is equivalent to

$$-\frac{\alpha^2+\beta^2}{2\alpha\beta}\sin(\alpha a)\sin(\beta b)+\cos(\alpha a)\cos(\beta b)=\cos(k(a+b)). \tag{2.18}$$

This equation relates the wave vector k to the energy E through the parameters α and β.

There are values of E for which there does not exist any k satisfying (2.18). In order to see this we assume that $E < qV_0$ such that β is purely imaginary and set $\beta = i\gamma$. Since $\sin(ix) = i\sinh(x)$ and $\cos(ix) = \cosh(x)$, (2.18) becomes

$$\frac{\gamma^2-\alpha^2}{2\alpha\gamma}\sin(\alpha a)\sinh(\gamma b)+\cos(\alpha a)\cosh(\gamma b)=\cos(k(a+b)). \tag{2.19}$$

Using

$$\lim_{\alpha\to 0}\frac{\sin(\alpha a)}{\alpha}=a,$$

we obtain for $E = 0$ (which implies $\alpha = 0$):

$$\frac{\gamma a}{2}\sinh(\gamma b)+\cosh(\gamma b)=\cos(k(a+b)).$$

Since $\sinh(\gamma b) > 0$ and $\cosh(\gamma b) > 1$, the left-hand side is strictly larger than one and thus, this equation cannot have a solution. By continuity, there is no solution in a neighborhood of $E = 0$.

We can compute the intervals for which no solution exists more easily by simplifying (2.19). For this, we let the potential barrier width $b \to 0$ and the barrier height $V_0 \to \infty$ such that the product $bV_0 \to \delta \in \mathbb{R}$. Then

$$\gamma b=\sqrt{\frac{2m(qV_0-E)b^2}{\hbar^2}}\to 0, \qquad \cosh(\gamma b)\to 1$$

and

$$\frac{(\gamma^2-\alpha^2)b}{2\alpha}\,\frac{\sinh(\gamma b)}{\gamma b}=\frac{m(qV_0-2E)b}{\hbar^2}\,\frac{\sinh(\gamma b)}{\gamma b}\to\frac{mq\delta}{\hbar^2}.$$

Thus, (2.19) becomes in the limit

$$f(\alpha a):=Q\frac{\sin(\alpha a)}{\alpha a}+\cos(\alpha a)=\cos(ka), \tag{2.20}$$

where $Q = mq\delta a/\hbar^2$. Figure 2.7 shows the function $f(\alpha a)$. In regions, where $|f(\alpha a)| \le 1$, there exist one or two solutions k of (2.20); in regions with $|f(\alpha a)| > 1$, no solution k exists. Every connected subset of $[0,\infty)\backslash R(E)$, where $R(E) = \{E_0 \ge 0 : \exists\, k \in \mathbb{R} : E(k) = E_0\}$, is an energy gap. $\square$

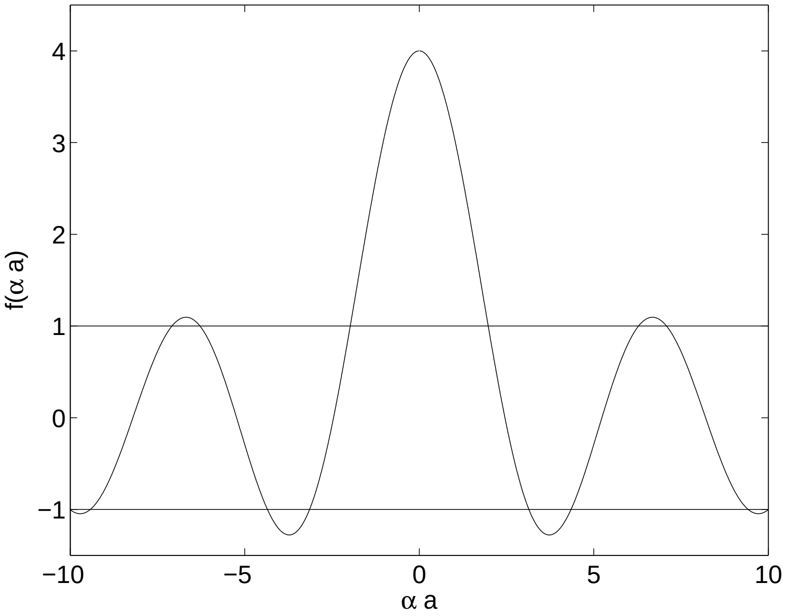

Figure 2.7: The function $f(\alpha a)$ of (2.20) depending on the energy αa.

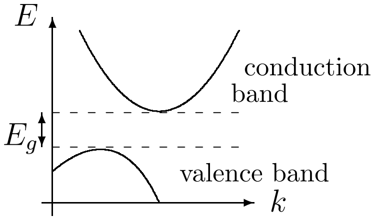

Figure 2.8: Schematic band structure with energy gap E_g.

The energy gap separates two energy bands. The nearest energy band below the energy gap (if it is unique) is called *valence band*, the nearest energy band above the energy gap is termed *conduction band* (see Figure 2.8).

Now we are able to state the definition of a semiconductor: it is solid with an energy gap whose value is positive and smaller than about 4eV. In Table 2.1 the values of the energy gaps for some commun semiconductor materials are collected.

Notice that the band structure of real crystals in three space dimensions is much more complicated than the one-dimensional situation of the Kronig-Penny model. Indeed, electrons traveling in different directions encounter different potential patterns, generated by the lattice atoms, and therefore the $E(k)$ diagram is strictly speaking a function of the three-dimensional wave vector k. In physics textbooks, usually a projection of the full $E(k)$ diagram is shown. For instance, Figure 2.9 shows the band structures of silicon and gallium arsenide. In place of the positive and negative k axes of the one-dimensional case,

Material	Symbol	Energy gap in eV
Silicon	Si	1.12
Germanium	Ge	0.67
Gallium arsenide	GaAl	1.42
Aluminium gallium arsenide	$\mathrm{Al}_{0.3}\mathrm{Ga}_{0.7}\mathrm{As}$	1.80
Gallium phosphide	GaP	2.20

Table 2.1: Energy gaps of selected semiconductors (from [3, Table 28.1] and [45, Fig. 1.14]).

two different crystal directions are shown, namely the $k = (0, 0, 1)^\top$ direction along the $+k$ axis (also called the Δ line) and the $k = (1, 1, 1)^\top$ direction along the $-k$ axis (also called the Λ line). The point $k = (0, 0, 0)$ is termed Γ point. The points at the boundary of the Brillouin zone in the Λ and Δ directions are called L and X points, respectively (see Figure 2.10).

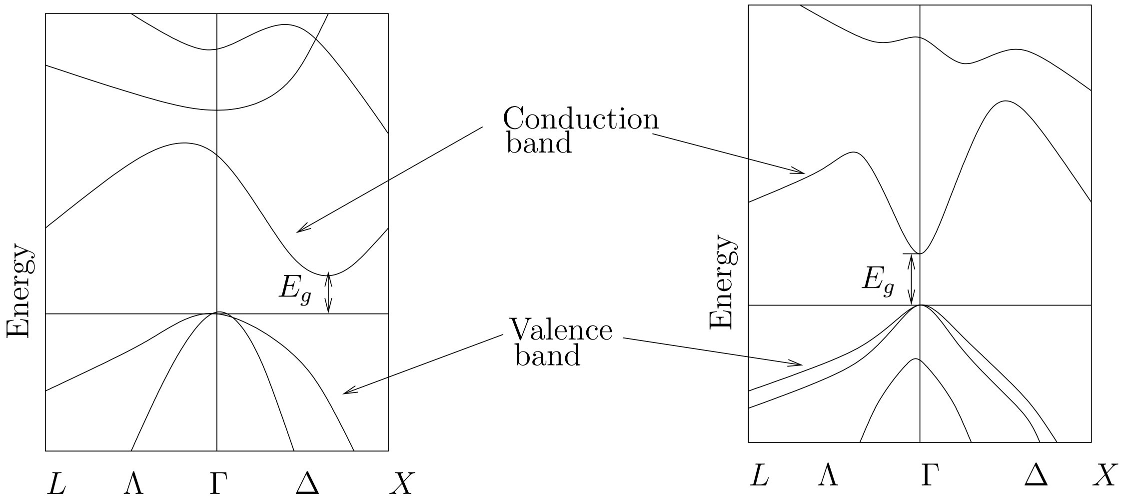

Figure 2.9: Schematic band structure of silicon (left) and gallium arsenide (right) (see [50, Fig. 3.23] or [65, Fig. 37 and 3.8]).

2.2 The semi-classical picture

The transport of electrons in semiconductors is based on the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m}\Delta\psi - q(V(x) + V_L(x))\psi,$$

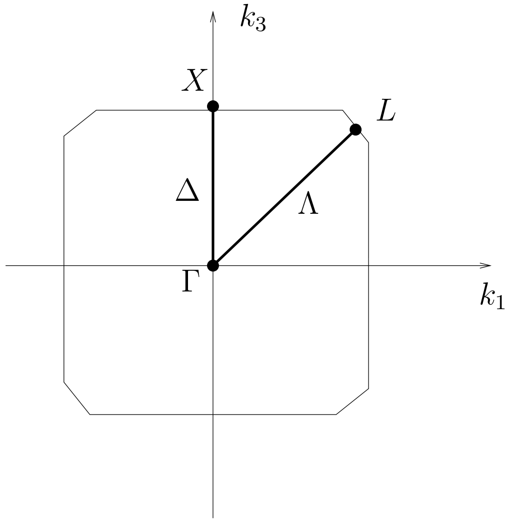

Figure 2.10: Projection of a simplified Brillouin zone on the (k_1, k_3) plane and illustration of the Γ, L, and X points and the Δ and Λ directories.

where $V_L(x)$ is the (periodic) potential due to the atoms of the crystal lattice and $V(x)$ is a (non-periodc) potential that is built-in or applied to the semiconductor. The solution of this equation is extremely difficult such that approximate models need to be used. One possibility is the semi-classical treatment which describes the carrier dynamics in the potential $V(x)$ by Newton's laws without explicitly treating the crystal potential $V_L(x)$. The influcence of $V_L(x)$ is indirectly taken into account by the use of the energy band structure in the description of the velocity and the mass of the carrier ensemble. Therefore, we will first motivate two formulas for the *mean electron velocity* and the *effective mass* as functions of the energy bands:

- The mean velocity (or group velocity of the wave packet) in the n-th band is given by

$$v_n(k) = \frac{1}{\hbar}\nabla_k E_n(k). \tag{2.21}$$

- The effective mass tensor m^* is defined by

$$(m^*)^{-1} = \frac{1}{\hbar}\frac{d^2 E_n}{dk^2}. \tag{2.22}$$

First, we motivate (2.21) by following [11, Sec. 8.1]. We omit the index n in the following and define the group velocity by

$$v(k) = \left(\int_D |\psi_k|^2\,dx\right)^{-1} \int_D \widehat{v}_k |\psi_k|^2\,dx, \tag{2.23}$$

where D is a primitive cell of the lattice and $\widehat{v}_k$ is the particle velocity,

$$\widehat{v}_k = -\frac{J_k}{qn_k} = \frac{\hbar}{m}\,\frac{\mathrm{Im}\,(\overline{\psi}_k\nabla\psi_k)}{|\psi_k|^2}. \tag{2.24}$$

The first equality follows from (1.7) and the second one from (1.9). The wave functions $\psi_k(x)$ are solutions of (2.15),

$$-\frac{\hbar^2}{2m}\Delta\psi_k - qV_L(x)\psi_k = E(k)\psi_k \qquad \text{in } D, \tag{2.25}$$

and can be decomposed by the Bloch theorem (see (2.7)),

$$\psi_k(x) = e^{ik\cdot x}u_k(x). \tag{2.26}$$

Differentiating (2.25) with respect to k and using (2.26) gives

$$\begin{aligned}
(\nabla_k E)\psi_k &= -\frac{\hbar^2}{2m}\Delta_x(\nabla_k\psi) - (qV_L+E)\nabla_k\psi_k \\
&= -\frac{\hbar^2}{2m}\Delta_x\left(e^{ik\cdot x}\nabla_k u_k + ix\psi_k\right) - (qV_L+E)\left(e^{ik\cdot x}\nabla_k u_k + ix\psi_k\right) \\
&= -\frac{\hbar^2}{2m}\left(\Delta_x(e^{ik\cdot x}\nabla_k u_k) + 2i\nabla_x\psi_k + ix\Delta_x\psi_k\right) - (qV_L+E)\left(e^{ik\cdot x}\nabla_k u_k + ix\psi_k\right) \\
&= \left(-\frac{\hbar^2}{2m}\Delta_x - (qV_L+E)\right)\left(e^{ik\cdot x}\nabla_k u_k\right) - \frac{i\hbar^2}{m}\nabla_x\psi_k \\
&\qquad + ix\left(-\frac{\hbar^2}{2m}\Delta_x - (qV_L+E)\right)\psi_k.
\end{aligned}$$

Observing that the last term vanishes in view of (2.25), multiplication of the resulting equation with $\overline{\psi}_k$ and integration over D yields

$$\begin{aligned}
&\nabla_k E\int_D |\psi_k|^2\,dx + \frac{i\hbar^2}{m}\int_D (\nabla_x\psi_k)\overline{\psi}_k\,dx \\
&\quad = \int_D \overline{\psi}_k\left(-\frac{\hbar^2}{2m}\Delta_x - (qV_L+E)\right)\left(e^{ik\cdot x}\nabla_k u_k\right)\,dx \\
&\quad = \int_D e^{ik\cdot x}\nabla_k u_k\left(-\frac{\hbar^2}{2m}\Delta_x - (qV_L+E)\right)\overline{\psi}_k\,dx \\
&\quad = 0,
\end{aligned}$$

where we have used integration by parts and again (2.25). The boundary integral in the integration-by-parts formula vanishes since u_k is periodic on ∂D. Thus

$$\nabla_k E = -\frac{i\hbar^2}{m}\,\frac{\int_D \overline{\psi}_k\nabla_x\psi_k\,dx}{\int_D |\psi_k|^2\,dx}.$$

Taking the real part of both sides of this equation gives

$$\nabla_k E = \frac{\hbar^2}{m} \frac{\operatorname{Im} \int_D \overline{\psi}_k \nabla_x \psi_k \, dx}{\int_D |\psi_k|^2 \, dx} = \hbar v(k),$$

employing (2.23) and (2.24). This shows (2.21).

The expression (2.21) has some consequences. The change of energy with respect to time equals the product of a force F and the velocity v_n:

$$\partial_t E_n(k) = F v_n(k) = \hbar^{-1} F \nabla_k E_n(k).$$

Since, by the chain rule, $\partial_t E_n(k) = \nabla_k E_n(k) \partial_t k$, we conclude that

$$F = \partial_t(\hbar k). \tag{2.27}$$

Newton's law states that the force equals the time derivative of the momentum p. This motivates the definition of the *crystal momentum*

$$p = \hbar k. \tag{2.28}$$

It should not be confused with the momentum operator $p_Q = -i\hbar\nabla_x$ of quantum mechanics.

Another consequence of (2.21) is formula (2.22). Indeed, differentiating (2.21) leads to

$$\partial_t v_n = \frac{1}{\hbar} \frac{d^2 E_n}{dk^2} \partial_t k = \frac{1}{\hbar^2} \frac{d^2 E_n}{dk^2} F,$$

using (2.27). The momentum p equals the product of velocity and (effective) mass, $p = m^* v_n$. Then, by Newton's law $F = \partial_t p = m^* \partial_t v_n$, and we infer that

$$(m^*)^{-1} = \frac{1}{\hbar^2} \frac{d^2 E_n}{dk^2}.$$

This equation is considered as a definition of the *effective mass* m^*. The right-hand side of this definition is the Hessian matrix, so the symbol $(m^*)^{-1}$ is also a (3×3) matrix.

The effective mass has the advantage that under some conditions, the behavior of the electrons in a crystal can be described as a free electron gas, for which $E(k) = \hbar^2 |k|^2 / 2m$ holds (see Example 2.1). In order to see this, we evaluate the Hessian of E_n near a local minimum (of the conduction band), i.e. $\nabla_k E_n(k_0) = 0$. Then $d^2 E_n(k_0)/dk^2$ is a symmetric, positive matrix which can be diagonalized and the diagonal matrix has positive entries. We assume that the coordinates are chosen such that $d^2 E_n(k_0)/dk^2$ is already diagonal,

$$\frac{1}{\hbar^2} \frac{d^2 E_n}{dk^2}(k_0) = \begin{pmatrix} 1/m_1^* & 0 & 0 \\ 0 & 1/m_2^* & 0 \\ 0 & 0 & 1/m_3^* \end{pmatrix}.$$

Assume that the energy values are shifted in such a way that $E_n(k_0) = 0$. (This is possible by fixing a reference point for the energy.) Let us assume that already $E_n(0) = 0$, otherwise define $\widetilde{E}_n(k) = E_n(k + k_0)$. If the function $k \mapsto E_n(k)$ is smooth, Taylor's formula then implies

$$\begin{aligned} E_n(k) &= E_n(0) + \nabla_k E_n(0) \cdot k + \frac{1}{2} k^\top \left(\frac{d^2 E_n}{dk^2}(0) \right) k + O(|k|^3) \\ &= \frac{\hbar^2}{2} \left(\frac{k_1^2}{m_1^*} + \frac{k_2^2}{m_2^*} + \frac{k_3^2}{m_3^*} \right) + O(|k|^3), \end{aligned}$$

where $k = (k_1, k_2, k_3)^\top$. If the effective masses are equal in all directions, i.e. $m^* = m_1^* = m_2^* = m_3^*$, we can write, neglecting higher-order terms

$$E_n(k) = \frac{\hbar^2}{2m^*} |k|^2. \tag{2.29}$$

This relation is valid for wave vectors k sufficiently close to a local band minimum (of the conduction band). The scalar m^* is called here the *isotropic effective mass.* Comparing this expression with the dispersion relation of a free electron gas,

$$E(k) = \frac{\hbar^2}{2m} |k|^2,$$

we infer that the energy of an electron near a band minimum equals the energy of a free electron in a vacuum where the (rest) electron mass m is replaced by the effective mass m^*.

The expression (2.29) is referred to as the *parabolic band approximation* and usually, the range of wave vectors is extended to the whole space, $k \in \mathbb{R}^3$. This simple model is appropriate for low applied fields for which the carriers are close to the conduction band minimum. For high applied fields, however, the higher-order terms in the above Taylor expansion cannot be ignored. In order to account for non-parabolic effects, often the *non-parabolic band approximation* in the sense of Kane is used (see [61, Sec. 2.1] or [50, (1.40)]):

$$E_n(1 + \alpha E_n) = \frac{\hbar^2}{2m^*} |k|^2, \tag{2.30}$$

where m^* is determined from (2.22) at the conduction band minimum at $k = 0$,

$$\alpha = \frac{1}{E_g} \left(1 - \frac{m^*}{m} \right)^2,$$

and E_g is the band gap. In Table 2.2 some values for α are shown. Formula (2.30) can be obtained from approximate solutions to (2.13) derived by the so-called $k \cdot p$ theory (see below).

Material	Si	Ge	GaAs
α in $(\mathrm{eV})^{-1}$	0.5	0.65	0.64

Table 2.2: Values of the non-parabolicity parameter α for some semiconductors (from [45, Table 1.1].

When we consider the effective mass definition (2.22) near a maximum (of the valence band), we find that the Hessian of E_n is negative definite. This would lead to a negative effective mass. However, in the derivation of the mean velocity and consequently of the effective mass, we have used in (2.24) that the charge of the electron is negative. Employing a positive charge leads again to a positive effective mass. The corresponding particles are called *holes* (or defect electrons). Physically, a hole is a vacant orbital in an otherwise filled valence band. Thus, the current flow in a semiconductor crystal comes from two sources: the flow of electrons in the conduction band and the flow of holes in the valence band. It is a convention to consider the motion of the valence band vacancies rather than the electrons moving from one vacant orbital to the next (see Figure 2.11).

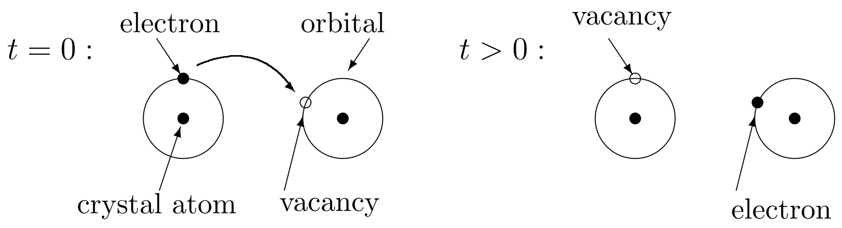

Figure 2.11: Motion of a valence band electron to a neighboring vacant orbital or, equivalently, of a hole in the inverse direction.

Close to the bottom $k = 0$ of the conduction band in an isotropic semiconductor, we obtain

$$E_n(k) = E_c + \frac{\hbar^2}{2m_e^*}|k|^2, \tag{2.31}$$

whereas near the top $k = 0$ of the valence band we have

$$E_n(k) = E_v - \frac{\hbar^2}{2m_h^*}|k|^2, \tag{2.32}$$

where E_c is the energy at the conduction band minimum, E_v the energy at the valence band maximum, m_e^* the effective electron mass, and m_h^* the effective hole mass. Clearly, the energy gap E_g is given by $E_g = E_c - E_v$ (see Figure 2.12). Some values for the effective masses of commun semiconductors can be found in Table 2.3.

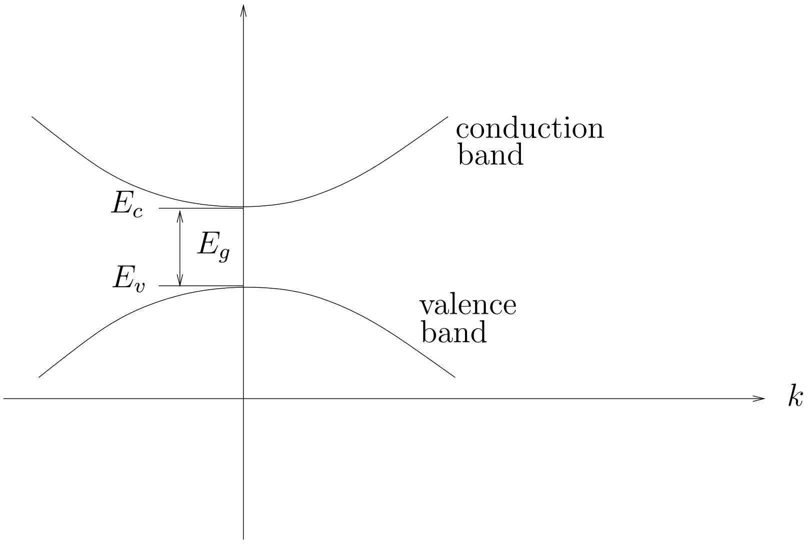

Figure 2.12: Schematic conduction and valence bands near the extrema at $k = 0$.

Material	m_e^*/m_e	m_h^*/m_h
Si	0.98	0.16
Ge	1.64	0.04
GaAs	0.067	0.082

Table 2.3: Relative effective electron and hole masses for some semiconductor material (from [32, Sec. 2.3.1]). The electron and hole masses at rest are denoted by m_e and m_h, respectively.

Now we come back to the *semi-classical picture*. In this picture, the motion of an electron in the n-th band is approximately described by a point particle moving with velocity $v_n(k)$. Denoting by $(x(t), v_n(k,t))$ the trajectory of an electron in the position-velocity phase space, we can write, by Newton's law,

$$\partial_t x = v_n = \frac{1}{\hbar}\nabla_k E_n, \qquad \partial_t(\hbar k) = F,$$

where F represents a driving force, for instance, $F = -q\nabla_x V_L$. Notice that band transitions are excluded since the band index n is fixed in the equations. In this picture, the wave packet of an electron is treated as a particle. However, by Heisenberg's uncertainty principle, the position and the momentum cannot have both sharp values. It is assumed that the uncertainty in the momentum is so small that the energy of the electron is sharply defined and that the uncertainty of the position is small compared to the distance over which the potential varies significantly.

If a *non-periodic (external) potential* is superimposed to the lattice potential V_L, the situation is much more complicated. Indeed, the Schrödinger equation (2.6) cannot be decomposed into the decoupled Schrödinger equation (2.15) and the energy bands are now coupled. However, it is usually assumed that the non-periodic potential is so weak that the coupling of the bands can be neglected, and then the above analysis remains approximately valid. In particular, we can use the semi-classical equations

$$\partial_t x = v_n = \frac{1}{\hbar}\nabla_k E_n, \qquad \partial_t(\hbar k) = -q\nabla_x V(x,t). \tag{2.33}$$

This semi-classical treatment will be used in the following chapters.

2.3 The $k \cdot p$ method

In the previous section we have seen that the mean velocity and the effective mass of the electrons in a semiconductor can be computed in the semi-classical picture from the energy band structure; see formulas (2.21) and (2.22). How can the band structure be computed? In this section we describe the $k \cdot p$ method which allows to derive an approximation of the energy $E_n(k)$ close to the bottom of the conduction band or close to the top of the valance band. The main assumption of this method is that the energy at $k = 0$ is known. Then $E_n(k)$ close to the Γ-point $k = 0$ can be computed using time-independent perturbation theory. We proceed in the following as in [11, Sec. 8.7] and [65, Sec. 4.1].

The starting point is the Schrödinger equation (2.13) for the functions $u_{n,k}$ of the Bloch function $\psi_{n,k} = e^{ik\cdot x}u_{n,k}$ (see Section 2.1), here written in the form

$$(H_0 + \varepsilon H_1)u_{n,k} = E_n(k)u_{n,k}, \tag{2.34}$$

where

$$H_0 = -\frac{\hbar^2}{2m}\Delta - qV_L(x)$$

is the single-electron Hamiltonian and

$$\varepsilon H_1 = -\frac{i\hbar^2}{m}k\cdot\nabla + \frac{\hbar^2}{2m}|k|^2$$

is considered to be a perturbation of H_0. Defining the quantum momentum operator $p_Q = -i\hbar\nabla$, we can formally write

$$\varepsilon H_1 = \frac{\hbar}{m}k\cdot p_Q + \frac{\hbar^2}{2m}|k|^2,$$

which explains the name of the $k \cdot p$ method.

We assume that $\varepsilon = |k|$ is small compared to one. Notice that for $k = 0$, the operator $H_0 + \varepsilon H_1$ reduces to the single-electron Hamiltonian. Furthermore, we suppose that the solutions of the eigenvalue problem

$$H_0 u_n^{(0)} = E_n^{(0)} u_n^{(0)} \qquad \text{in } D, \tag{2.35}$$

where D is the primitive cell, together with periodic boundary conditions are known. Since the operator H_0 is real, also the eigenfunctions $u_n^{(0)}$ are real. We will show the following result.

Theorem 2.11 *Let the solutions* $(u_n^{(0)})$ *to (2.35) form a non-degenerate orthonormal basis of* $L^2(D, \mathbb{C})$ *(i.e., all eigenspaces are one-dimensional). Then, up to second order in* ε*,*

$$E_n(k) = E_n^{(0)} + \frac{\hbar^2}{2} k^\top (m^*)^{-1} k \qquad (k \to 0), \tag{2.36}$$

where the matrix $(m^*)^{-1}$ *consists of the elements* $1/m^*_{j\ell}$ *with*

$$\frac{m}{m^*_{j\ell}} = \delta_{j\ell} - \frac{2\hbar^2}{m} \sum_{q \neq n} \frac{P_{qnj} P_{nq\ell}}{E_q^{(0)} - E_n^{(0)}} \tag{2.37}$$

and

$$P_{qnj} = \left(u_q^{(0)}, \frac{\partial u_n^{(0)}}{\partial x_j} \right) = \int_D u_q^{(0)} \frac{\partial u_n^{(0)}}{\partial x_j}\, dx.$$

Notice that the one-dimensionality of the eigenspaces implies that $E_q^{(0)} \neq E_n^{(0)}$ for all $q \neq n$ and so, (2.37) is defined. The symbol $(\cdot, \cdot)$ denotes the scalar product on $L^2(D; \mathbb{C})$.

Proof: We apply a perturbation method to (2.34) (see [11, Sec. 4.1] or [65, app. C.1.1]). For this, we develop

$$u_{n,k} = u_n^{(0)} + \varepsilon u_n^{(1)} + \varepsilon^2 u_n^{(2)} + \cdots, \qquad E_n(k) = E_n^{(0)} + \varepsilon E_n^{(1)} + \varepsilon^2 E_n^{(2)} + \cdots.$$

Inserting these expressions into (2.34) and equating terms with the same order of ε leads to

$$\varepsilon^0: \quad H_0 u_n^{(0)} = E_n^{(0)} u_n^{(0)}, \tag{2.38}$$

$$\varepsilon^1: \quad H_0 u_n^{(1)} + H_1 u_n^{(0)} = E_n^{(0)} u_n^{(1)} + E_n^{(1)} u_n^{(0)}, \tag{2.39}$$

$$\varepsilon^2: \quad H_0 u_n^{(2)} + H_1 u_n^{(1)} = E_n^{(0)} u_n^{(2)} + E_n^{(1)} u_n^{(1)} + E_n^{(2)} u_n^{(0)}. \tag{2.40}$$

The zeroth-order equation (2.38) clearly is the same as (2.35). In order to derive the first-order correction, we multiply (2.39) by $u_q^{(0)}$ and integrate over D. Then, observing that

$$(u_q^{(0)}, u_n^{(0)}) = \delta_{qn},$$

we obtain

$$(u_q^{(0)}, H_0 u_n^{(1)}) + (u_q^{(0)}, H_1 u_n^{(0)}) = E_n^{(0)}(u_q^{(0)}, u_n^{(1)}) + E_n^{(1)}\delta_{qn}.$$

Integrating by parts twice (or employing the self-adjointness of H_0) it follows

$$(u_q^{(0)}, H_0 u_n^{(1)}) = (H_0 u_q^{(0)}, u_n^{(1)}) = E_q^{(0)}(u_q^{(0)}, u_n^{(1)}),$$

and therefore,

$$(E_q^{(0)} - E_n^{(0)})(u_q^{(0)}, u_n^{(1)}) + (u_q^{(0)}, H_1 u_n^{(0)}) = E_n^{(1)}\delta_{qn}.$$

For $q = n$ this gives an expression for $E_n^{(1)}$ only depending on $(u_n^{(0)})_n$:

$$E_n^{(1)} = (u_n^{(0)}, H_1 u_n^{(0)}). \tag{2.41}$$

For $q \neq n$ we have

$$(u_q^{(0)}, u_n^{(1)}) = \frac{(u_q^{(0)}, H_1 u_n^{(0)})}{E_q^{(0)} - E_n^{(0)}}. \tag{2.42}$$

This is possible since the non-degeneracy assumption implies $E_q^{(0)} \neq E_n^{(0)}$ for all $q \neq n$. The sequence $(u_n^{(0)})$ is an orthonormal basis, so we can develop $u_n^{(1)}$ in this basis:

$$u_n^{(1)} = \sum_q (u_q^{(0)}, u_n^{(1)}) u_q^{(0)}. \tag{2.43}$$

In this sum we need an expression for the term $q = n$, $(u_n^{(0)}, u_n^{(1)})$. In fact, this term is not determinable from the above calculation and can be chosen freely. We make the choice $(u_n^{(0)}, u_n^{(1)}) = 0$. In view of (2.42), (2.43) becomes

$$u_n^{(1)} = \sum_{q \neq n} \frac{(u_q^{(0)}, H_1 u_n^{(0)})}{E_q^{(0)} - E_n^{(0)}} u_q^{(0)}.$$

Thus, up to first order, the eigenfunctions are given by

$$u_n^{(0)} + \varepsilon u_n^{(1)} = u_n^{(0)} + \sum_{q \neq n} \frac{(u_q^{(0)}, \varepsilon H_1 u_n^{(0)})}{E_q^{(0)} - E_n^{(0)}} u_q^{(0)}$$

and the eigenvalues are

$$E_n^{(0)} + \varepsilon E_n^{(1)} = E_n^{(0)} + (u_n^{(0)}, \varepsilon H_1 u_n^{(0)}).$$

Notice that these corrections only depend on the unperturbed eigenfunctions $u_n^{(0)}$ which are assumed to be known.

In order to derive the second-order correction, we multiply (2.40) by $u_q^{(0)}$ and integrate over D:

$$(u_q^{(0)}, H_0 u_n^{(2)}) + (u_q^{(0)}, H_1 u_n^{(1)}) = E_n^{(0)}(u_q^{(0)}, u_n^{(2)}) + E_n^{(1)}(u_q^{(0)}, u_n^{(1)}) + E_n^{(2)}\delta_{qn}.$$

As above, the first term on the left-hand side equals

$$(u_q^{(0)}, H_0 u_n^{(2)}) = E_q^{(0)}(u_q^{(0)}, u_n^{(2)}),$$

such that

$$(E_q^{(0)} - E_n^{(0)})(u_q^{(0)}, u_n^{(2)}) + (u_q^{(0)}, H_1 u_n^{(1)}) = E_n^{(1)}(u_q^{(0)}, u_n^{(1)}) + E_n^{(2)}\delta_{qn}.$$

Using (2.41) and (2.43), the case $q = n$ yields

$$\begin{aligned} E_n^{(2)} &= (u_n^{(0)}, H_1 u_n^{(1)}) - E_n^{(1)} \cdot (u_n^{(0)}, u_n^{(1)}) \\ &= \sum_q (u_q^{(0)}, u_n^{(1)}) \cdot (u_n^{(0)}, H_1 u_q^{(0)}) - (u_n^{(0)}, H_1 u_n^{(0)}) \cdot (u_n^{(0)}, u_n^{(1)}) \\ &= \sum_{q \neq n} (u_n^{(0)}, H_1 u_q^{(0)}) \cdot (u_q^{(0)}, u_n^{(1)}) \\ &= \sum_{q \neq n} (u_n^{(0)}, H_1 u_q^{(0)}) \frac{(u_q^{(0)}, H_1 u_n^{(0)})}{E_q^{(0)} - E_n^{(0)}}. \end{aligned}$$

In the last equation we employed (2.42). Thus, the second-order correction to the eigenvalues is

$$E_n^{(0)} + \varepsilon E_n^{(1)} + \varepsilon^2 E_n^{(2)} = E_n^{(0)} + (u_n^{(0)}, \varepsilon H_1 u_n^{(0)}) + \sum_{q \neq n} \frac{(u_n^{(0)}, \varepsilon H_1 u_q^{(0)})(u_q^{(0)}, \varepsilon H_1 u_n^{(0)})}{E_q^{(0)} - E_n^{(0)}}.$$

It remains to compute the scalar products. We write

$$\begin{aligned} (u_n^{(0)}, \varepsilon H_1 u_q^{(0)}) &= -\frac{i\hbar^2}{m} k \cdot (u_n^{(0)}, \nabla u_q^{(0)}) + \frac{\hbar^2}{2m}|k|^2 (u_n^{(0)}, u_q^{(0)}) \\ &= -\frac{i\hbar^2}{m} k \cdot P_{nq} + \frac{\hbar^2}{2m}|k|^2 \delta_{nq}, \end{aligned}$$

where

$$P_{nq} = \int_D \overline{u}_n^{(0)} \nabla u_q^{(0)}\, dx = \int_D u_n^{(0)} \nabla u_q^{(0)}\, dx,$$

since $u_n^{(0)}$ is real. The periodicity of $u_n^{(0)}$ on D gives

$$P_{nn} = \frac{1}{2} \int_D \operatorname{div}[(u_n^{(0)})^2]\, dx = 0,$$

and therefore,

$$(u_n^{(0)}, \varepsilon H_1 u_q^{(0)}) = \begin{cases} \dfrac{\hbar^2}{2m}|k|^2 & : n = q \\ -\dfrac{i\hbar^2}{m} k \cdot P_{nq} & : n \neq q. \end{cases}$$

This shows that $E_n(k)$ is, up to second order in ε,

$$\begin{aligned} E_n(k) &= E_n^{(0)} + \varepsilon E_n^{(1)} + \varepsilon^2 E_n^{(2)} \\ &= E_n^{(0)} + \frac{\hbar^2}{2m}|k|^2 - \frac{\hbar^4}{m^2} \sum_{q \neq n} \frac{(k \cdot P_{nq})(k \cdot P_{qn})}{E_q^{(0)} - E_n^{(0)}} \\ &= E_n^{(0)} + \frac{\hbar^2}{2m}|k|^2 - \frac{\hbar^4}{m^2} \sum_{q \neq n} \sum_{j,\ell} k_j k_\ell \frac{P_{nqj} P_{qn\ell}}{E_q^{(0)} - E_n^{(0)}} \\ &= E_n^{(0)} + \frac{\hbar^2}{2} \sum_{j,\ell} \frac{k_j k_\ell}{m^*_{j\ell}} \end{aligned} \tag{2.44}$$

which proves the theorem. □

Equation (2.44) shows that the first-order correction of the energy yields simply the free-electron mass. The second-order correction is needed to obtain an effective mass which is different from the free-electron mass. This is the reason why we computed the corrections up to second-order.

Theorem 2.11 can be applied to the bottom of the conduction band of most semiconductors since the eigenstates are non-degenerate (the energy bands do not cross; see Figure 2.9). However, the top of the valence band in all semiconductors is *degenerate* (the valence bands cross; see Figure 2.9) and hence, the above result does not hold. Mathematically, we have in such a situation several eigenfunctions with the same eigenvalue, for instance for $n \neq q$,

$$H_0 u_n^{(0)} = E_n^{(0)} u_n^{(0)}, \qquad H_0 u_q^{(0)} = E_q^{(0)} u_q^{(0)} \quad \text{but} \quad E_n^{(0)} = E_q^{(0)}.$$

Then, the expression (2.37) is not defined. It is still possible to derive a formula similar to (2.36) in the degenerate case by applying degenerate perturbation theory. The idea is to find a linear combination

$$\widetilde{u}_n^{(0)} = \sum_{\alpha=1}^{A} c_\alpha u_{n,\alpha}^{(0)}$$

of the eigenfunctions $u_{n,\alpha}^{(0)}$ with the same energy $E_n^{(0)}$ such that the nominator in the first-order correction

$$u_n^{(1)} = \sum_{q \neq n} \frac{(u_q^{(0)}, H_1 u_n^{(0)})}{E_q^{(0)} - E_n^{(0)}} u_q^{(0)}$$

vanishes if $E_q^{(0)} = E_n^{(0)}$. The problem is to find coefficients c_α such that $(\widetilde{u}_q^{(0)}, H_1 \widetilde{u}_n^{(0)}) = 0$ [11, Sec. 4.2]. It can be shown that the energies $E_n(k)$ are, up to second order, the eigenvalues of the matrix $H_{nk} \in \mathbb{R}^{A \times A}$ with elements [65, Sec. 4.1.4]

$$(H_{nk})_{\mu\nu} = E_n^{(0)} \delta_{\mu\nu} - \frac{i\hbar^2}{m} \sum_j k \widetilde{P}_{nnj}^{\mu\nu} + \frac{\hbar^2}{2} \sum_{j,\ell} \frac{k_j k_\ell}{(m^*)_{j\ell}^{\mu\nu}},$$

$\widetilde{P}^{j\ell}_{nq\mu}$ is defined similarly as in Theorem 2.11 and

$$\frac{m}{(m^*)^{\mu\nu}_{j\ell}} = \delta_{j\ell}\delta_{\mu\nu} - \frac{2\hbar^2}{m}\sum_{q\neq n}\sum_{\alpha=1}^{A}\frac{\widetilde{P}^{\mu\alpha}_{nqj}\widetilde{P}^{\alpha\nu}_{qn\ell}}{E^{(0)}_q - E^{(0)}_n}. \tag{2.45}$$

As a final remark we notice that analogous results as above can be derived for holes in the valence band. In this case, the energy $E_n(k)$ can be approximately written as

$$E_n(k) = E^c_n(0) - \frac{\hbar^2}{2}k^\top (m^*_h)^{-1}k,$$

where $E^c_n(0)$ is the top energy of the valence band and m^*_h is the effective mass tensor for the holes, similarly defined as above. In this case, $\widetilde{P}^{\mu\nu}_{nnj} = 0$ for all j, such that the linear term in k in (2.45) vanishes.

2.4 Semiconductor statistics

In this section we will answer the question how many electrons and holes are in a semiconductor which is in thermal equilibrium (i.e. no current flow)? With the mean number of electrons in a quantum state of energy E, $f(E)$, the answer is

$$N = 2\sum_n \sum_{k\in B} f(E_n(k)), \tag{2.46}$$

where the factor 2 takes into account the two possible states of the spin of an electron and B is the Brillouin zone. This leads to two questions:

- How can the sum over many k can be computed practically?
- How does the function f depend on the energy?

To answer the first question, we make the following observation. Consider a chain of $N+1$ atoms with distance d. Then the length of the chain is $L = Nd$. The Bloch function $\psi_k(x) = e^{ik\cdot x}u_k(x)$ (see Section 2.1) satisfies at the chain boundaries

$$\psi_k(0) = \psi_k(L)$$

and $u_k(x)$ is periodic, in particular, $u_k(0) = u_k(L)$. Then

$$u_k(0) = \psi_k(0) = \psi_k(L) = e^{ikL}u_k(L) = e^{ikL}u_k(0)$$

which implies that

$$kL = 2\pi j, \qquad j = 0, \dots, N-1.$$

The index j runs from 0 to $N-1$ since the one-dimensional Brillouin zone equals here $k \in B = [0, 2\pi/d)$ (see Lemma 2.4). Thus, the wave vector k can take one of the *discrete* values

$$k_j = \frac{2\pi j}{L} = \frac{2\pi j}{Nd} \qquad j = 0, \ldots, N-1.$$

Typically, $L = 1\mu\text{m} = 10^{-6}\text{m}$ and $d = 10^{-10}\text{m}$, so $N = 10^4$. Therefore, one can consider k to be continuous. The sum

$$\sum_{j=0}^{N-1} g(k_j)$$

for some function g then transforms to the integral

$$\sum_{j=0}^{N-1} g(k_j) = \sum_{j=0}^{N-1} g\left(\frac{2\pi j}{Nd}\right) \approx \int_0^N g\left(\frac{2\pi j}{Nd}\right) dj = \frac{L}{2\pi} \int_0^{2\pi/d} g(k)\, dk.$$

In d space dimensions, the factor $L/2\pi$ becomes $(L/2\pi)^d$. In the continuum limit $N \to \infty$ and $d \to 0$ such that Nd is finite we can extend the integration to $\mathbb{R}^d$ and write

$$\frac{\text{vol}(\Omega)}{(2\pi)^d} \int_{\mathbb{R}^d} g(k)\, dk \qquad \text{instead of} \qquad \sum_{k \in B} g(k), \tag{2.47}$$

where $\text{vol}(\Omega)$ is the volume of the semiconductor. The advantage of the integral formulation is that integrals can be more easily computed than sums.

The answer to the second question is contained in the following lemma.

Lemma 2.12 *The mean number of electrons in a quantum state of energy E is given by the* Fermi-Dirac distribution function

$$f(E) = \frac{1}{1 + e^{(E - q\mu)/k_B T}}, \tag{2.48}$$

where k_B is the Boltzmann constant, T the (electron) temperature and μ the chemical potential.

The two parameters T and μ can be considered here as given by the corresponding quantum state. A deeper understanding is possible by means of thermodynamics, but we refer to [11, Ch. 5] for details since the significance of T and μ will become more transparent in subsequent chapters (see also Remark 2.13).

For the formal proof of Lemma 2.12 we notice that electrons are fermions, i.e. particles with half-integral spin, satisfying the following properties:

- Electrons cannot be distinguised from each other.
- The Pauli exclusion principle holds, i.e., each quantum state can be occupied by not more than two electrons with opposite spins.

Proof of Lemma 2.12: We proceed as in [11, p. 293 ff.]. We arrange M electrons into N bands each of which has g_n quantum states, $n = 1, \ldots, N$. Suppose that m_n electrons are occupying quantum states in the n-th band, where $m_n \leq g_n$. Notice that, by the Pauli exclusion principle, each electron is occupying exactly one quantum state. The number of different arrangements in the n-th band Q_n equals the number of all possible configurations,

$$g_n(g_n - 1) \cdot \ldots \cdot (g_n - m_n + 1),$$

divided by the number of all possible permutations of the m_n electrons (since they are indistinguishable),

$$m_n!,$$

hence

$$Q_n = \frac{g_n(g_n - 1) \cdot \ldots \cdot (g_n - m_n + 1)}{m_n!} = \frac{g_n!}{(g_n - m_n)! m_n!}.$$

The total number of configurations reads in the limit $N \to \infty$

$$Q(m_1, m_2, \ldots) = \prod_{n=1}^{\infty} Q_n = \prod_{n=1}^{\infty} \frac{g_n!}{(g_n - m_n)! m_n!}.$$

In order to manipulate this function, it is convenient to consider

$$\ln Q(m_1, m_2, \ldots) = \sum_{n=1}^{\infty} [\ln g_n! - \ln(g_n - m_n)! - \ln m_n!].$$

Using the Stirling formula

$$n! \sim \frac{n^n}{e^n} \qquad \text{or} \qquad \ln n! \sim n \ln n - n \quad (n \to \infty),$$

we obtain approximately

$$\ln Q(m_1, m_2, \ldots) = \sum_{n=1}^{\infty} [\ln g_n! - (g_n - m_n)\ln(g_n - m_n) - m_n \ln m_n + g_n].$$

This function is in some sense related to the thermodynamic entropy (see [11, p. 268]). The most probable configuration of $(m_1, m_2, \ldots)$ is that one which maximizes $\ln Q$, under the constraints that the particle number and the energy are conserved:

$$\max_{m_j} \ln Q(m_1, m_2, \ldots) \qquad \text{such that} \quad \sum_{n=1}^{\infty} m_n = M, \quad \sum_{n=1}^{\infty} E_n m_n = E.$$

We solve this constrained extremal problem with Lagrange multipliers, i.e., we extremize

$$F(\lambda_1, \lambda_2; m_1, m_2, \ldots) = \ln Q + \lambda_1 \left(\sum_{n=1}^{\infty} m_n - M \right) + \lambda_2 \left(\sum_{n=1}^{\infty} E_n m_n - E \right).$$

A necessary condition is

$$0 = \frac{\partial F}{\partial m_j} = \ln(g_j - m_j) + 1 - (\ln m_j + 1) + \lambda_1 + \lambda_2 E_j = \ln\left(\frac{g_j}{m_j} - 1\right) + \lambda_1 + \lambda_2 E_j.$$

Solving for m_j/g_j yields

$$\frac{m_j}{g_j} = \frac{1}{1 + e^{-\lambda_1 - \lambda_2 E_j}}.$$

Defining the temperature T and the chemical potential by

$$\lambda_1 = \frac{q\mu}{k_B T} \qquad \text{and} \qquad \lambda_2 = -\frac{1}{k_B T},$$

we obtain

$$\frac{m_j}{g_j} = \frac{1}{1 + e^{(E_j - q\mu)/k_B T}} = f(E_j).$$

Since the left-hand side is the mean number of electrons in a quantum state of energy E_j, the lemma is shown. □

Remark 2.13 The properties of the Fermi-Dirac distribution can be understood as follows (see also [11, p. 298 f.]). At zero temperature, this function becomes

$$f(E) = \begin{cases} 1 & \text{for } E < q\mu \\ 0 & \text{for } E > q\mu \end{cases} \qquad \text{and} \quad f(q\mu) = \frac{1}{2}$$

(see Figure 2.13). This means that all states which have an energy smaller than the chemical potential are occupied, and all states with an energy larger than $q\mu$ are empty. Physically, this behavior comes from the Pauli principle according to which two electrons do not occupy the same quantum state. Consequently, at zero temperature, the states with lowest energy are filled first. The energy of the state filled by the last particle is equal to the chemical potential. For non-zero temperature, there is a positive probability that some energy states above $q\mu$ will be occupied, i.e., some particles jump to higher energy levels due to thermal exertation.

We notice that the chemical potential $q\mu$ (more precisely, the product of elementary charge and chemical potential) in semiconductors is referred to as the *Fermi level* or *Fermi potential* and is usually denoted by E_F.

For energies much larger than the Fermi energy $E_F = q\mu$ in the sense of $E - E_F \gg k_B T$, we can approximate the Fermi-Dirac distribution by the *Maxwell-Boltzmann distribution*

$$F(E) = e^{-(E - E_F)/k_B T}$$

since $1/(1+e^x) \sim e^{-x}$ for $x \gg 1$ (Figure 2.13). Semiconductors whose electron distribution can be described by this distribution are called *non-degenerate*. Materials in which the Fermi-Dirac distribution has to be used are termed *degenerate*. □

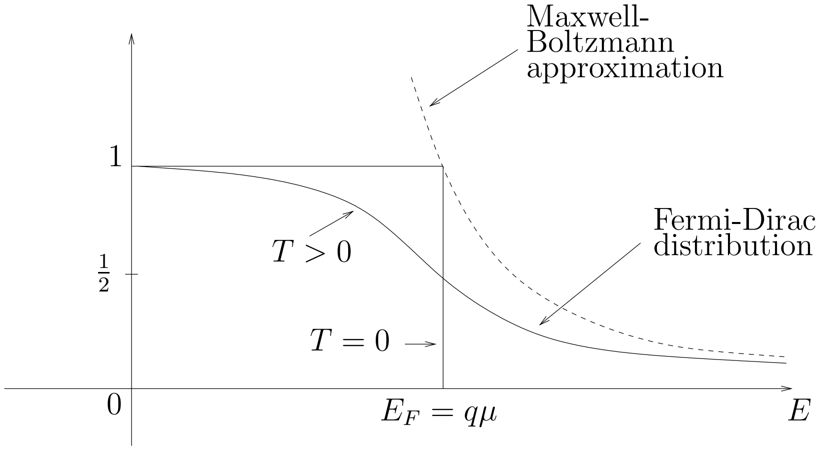

Figure 2.13: The Fermi-Dirac distribution at zero and non-zero temperature and the Maxwell-Boltzmann approximation.

Now we wish to determine the *density of states*, i.e. the number of quantum states (n, k) with energy E in the semiconductor Ω:

$$g(E) = \frac{2}{\mathrm{vol}(\Omega)} \sum_n \sum_{k \in B} \delta(E - E_n(k)), \tag{2.49}$$

where again the factor 2 comes from the two spin states of an electron and $\delta(z) = 1$ if $z = 0$ and $\delta(z) = 0$ if $z \neq 0$. We show:

Lemma 2.14 *The density of states in the continuum limit in $\mathbb{R}^d$ ($d \geq 1$) reads as*

$$g(E) = \frac{2}{(2\pi)^d} \sum_n \int_{\mathbb{R}^d} \delta(E - E_n(k))\, dk, \tag{2.50}$$

where δ is the delta distribution (see below). In the case $d > 1$ we can also write

$$g(E) = \frac{2}{(2\pi)^d} \sum_n \int_{E_n^{-1}(E)} \frac{dF_{d-1}}{|\nabla_k E_n(k)|},$$

where dF_{d-1} is the element of the surface $E_n^{-1}(E)$.

Before we can give a heuristic proof of this lemma, we need some preparations. First, we explain the delta distribution. Heuristically, δ is defined by

$$\delta(z) := \begin{cases} \infty & : \ z = 0 \\ 0 & : \ z \neq 0 \end{cases} \qquad \text{and} \qquad \int_{\mathbb{R}} \delta(z)\, dz = 1. \tag{2.51}$$

Clearly, there is no function satisfying these conditions. However, in the theory of distributions it is possible to give a sense to (2.51). More precisely, δ is defined as a functional,

$$\langle \delta, \phi \rangle = \phi(0) \qquad \text{for appropriate } \phi : \mathbb{R} \to \mathbb{R}.$$

Usually, the bracket is written as

$$\int_{\mathbb{R}} \delta(z)\phi(z)\,dz = \phi(0), \tag{2.52}$$

which is motivated by (2.51). We refer to [22] for details about distributions.

The second tool needed for the proof of Lemma 2.14 is the coarea formula (see, e.g., [30, 46]).

Theorem 2.15 (Coarea formula)
Let $f : B \subset \mathbb{R}^d \to \mathbb{R}$ $(d > 1)$ *be continuous and* $E : B \to \mathbb{R}$ *be continuously differentiable with* $\nabla_k E(k) \neq 0$ *for all* $k \in B$. *Then*

$$\int_B f(k)\,dk = \int_{\mathbb{R}} \int_{E^{-1}(\varepsilon)} f(k) \frac{dF_{d-1}(k)}{|\nabla_k E(k)|}\,d\varepsilon,$$

where $dF_{d-1}(k)$ *is the* $(d-1)$*-dimensional hypersurface element.*

Proof: In the following we motivate the coarea formula for the case $d = 3$. The idea of the proof is to make a transformation of the k- to the ε-variable. Let (u, v) be a parametrization of the surface $E^{-1}(\varepsilon) = \{k : E(k) = \varepsilon\} \subset \mathbb{R}^3$ (see Figure 2.14). Then $k = k(u, v, E)$ and $\varepsilon = E(k(u, v, \varepsilon))$. Taking partial derivatives gives

$$\begin{aligned} 0 &= \frac{\partial \varepsilon}{\partial u} &&= \nabla_k E(k) \cdot \frac{\partial k}{\partial u}, \\ 0 &= \frac{\partial \varepsilon}{\partial v} &&= \nabla_k E(k) \cdot \frac{\partial k}{\partial v}, \\ 1 &= \frac{\partial \varepsilon}{\partial \varepsilon} &&= \nabla_k E(k) \cdot \frac{\partial k}{\partial \varepsilon}. \end{aligned}$$

The first two equations imply that $\nabla_k E \perp \partial k/\partial u$ and $\nabla_k E \perp \partial k/\partial v$ and

$$\nabla_k E(k) \parallel \left(\frac{\partial k}{\partial u} \times \frac{\partial k}{\partial v}\right). \tag{2.53}$$

The third equation shows that $|\partial k/\partial \varepsilon| = 1/|\nabla_k E|$ and $\nabla_k E \parallel \partial k/\partial \varepsilon$. Then we infer from (2.53)

$$\frac{\partial k}{\partial \varepsilon} \parallel \left(\frac{\partial k}{\partial u} \times \frac{\partial k}{\partial v}\right). \tag{2.54}$$

We conclude from the transformation theorem and (2.54) that

$$\begin{aligned} dk &= \left|\det \frac{dk}{d(u, v, \varepsilon)}\right| d(u, v)\,d\varepsilon = \left|\det\left(\frac{\partial k}{\partial u}, \frac{\partial k}{\partial v}, \frac{\partial k}{\partial \varepsilon}\right)\right| d(u, v) d\varepsilon \\ &= \left|\frac{\partial k}{\partial \varepsilon} \cdot \left(\frac{\partial k}{\partial u} \times \frac{\partial k}{\partial v}\right)\right| d(u, v) d\varepsilon = \left|\frac{\partial k}{\partial \varepsilon}\right| \cdot \left|\frac{\partial k}{\partial u} \times \frac{\partial k}{\partial v}\right| d(u, v) d\varepsilon \\ &= \frac{1}{|\nabla_k E|} dF d\varepsilon \end{aligned}$$

and hence,

$$\int_B f(k)\,dk = \int_{\mathbb{R}} \int_{\{E(k)=\varepsilon\}} \frac{f(k)}{|\nabla_k E|} dF_{d-1} d\varepsilon,$$

which motivates the theorem. $\square$

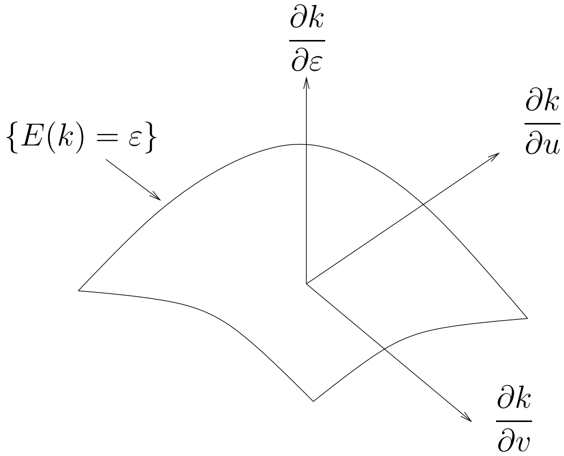

Figure 2.14: Illustration for the motivation of Theorem 2.15.

Proof of Lemma 2.14: In the continuum limit, the substitution (2.47) in the definition (2.49) of $g(E)$ immediately gives the first equality in (2.50). The second one follows from the definition (2.52) of the δ distribution and the coarea formula since for $E > 0$,

$$\begin{aligned}\int_{\mathbb{R}^d} \delta(E - E_n(k))\,dk &= \int_{\mathbb{R}} \left(\int_{\{E_n(k)=\varepsilon\}} \frac{dF_{d-1}}{|\nabla_k E_n|} \right) \delta(E - E_n(k))\,d\varepsilon \\ &= \int_{\mathbb{R}} \int_{\{E_n(k)=\varepsilon\}} \frac{dF_{d-1}}{|\nabla_k E_n|} \delta(E - \varepsilon)\,d\varepsilon \\ &= \int_{\{E_n(k)=E\}} \frac{dF_{d-1}}{|\nabla_k E_n|}.\end{aligned}$$ $\square$

The above lemmas allow us to determine the particle densities.

Lemma 2.16 *The electron and hole densities are given by*

$$n = \int_{\mathbb{R}} g_c(E) f(E)\,dE, \qquad p = \int_{\mathbb{R}} g_v(E)(1 - f(E))\,dE$$

where $g_c(E)$, $g_v(E)$ are the densities of states of the conduction or valence band, respectively,

$$g_c(E) = \frac{2}{(2\pi)^d} \int_{\mathbb{R}^d} \delta(E - E_c(k))\,dk, \qquad g_v(E) = \frac{2}{(2\pi)^d} \int_{\mathbb{R}^d} \delta(E - E_v(k))\,dk,$$

and $f(E)$ is the Fermi-Dirac distribution function (2.48).

Proof: The electron density is defined as the number of electrons N per volume $\mathrm{vol}(\Omega)$. Thus, by (2.46), (2.47) and (2.52), for a special band $E_c(k)$ and $B = \mathbb{R}^d$,

$$\begin{aligned} n &= \frac{N}{\mathrm{vol}(\Omega)} = \frac{2}{\mathrm{vol}(\Omega)} \sum_{k \in B} f(E_c(k)) = \frac{2}{(2\pi)^d} \int_{\mathbb{R}^d} f(E_c(k))\, dk \\ &= \frac{2}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \delta(E - E_c(k)) f(E)\, dE\, dk = \int_{\mathbb{R}} g_c(E) f(E) dE. \end{aligned}$$

In the last step we have used (2.50). The formula for p follows similarly, taking into account that the mean number of holes in a quantum state of energy E equals the mean number of *empty* states of energy E, $1 - f(E)$. □

In the parabolic band approximation, (2.31) and (2.32), the particle densities can be computed more explicitely. For this, we first compute the density of states.

Lemma 2.17 *In the parabolic band approximation $E_n(k) = E_0 + (\hbar^2/2m^*)|k|^2$ we obtain for $E \geq E_0$:*

$$\begin{aligned} g(E) &= \frac{m^*}{\pi\hbar^2} \frac{\sqrt{2m^*(E - E_0)}}{\pi\hbar} && \textit{for three-dimensional carriers,} \\ g(E) &= \frac{m^*}{\pi\hbar^2} && \textit{for two-dimensional carriers,} \\ g(E) &= \frac{m^*}{\pi\hbar^2} \frac{\hbar}{\sqrt{2m^*(E - E_0)}} && \textit{for one-dimensional carriers.} \end{aligned}$$

For $E < E_0$, we have $g(E) = 0$ in all three cases.

It is possible in modern quantum devices to confine carriers in one (or two) dimensions, i.e., the carriers are confined in the x-y-plane (or in the x-direction) and are free to move in the z-direction (or in the y-z-plane). Such structures can be constructed with so-called semiconductor heterostructures and are called quantum wires or quantum wells, respectively (see [45, Sec. 1.5.2]).

Proof: We start from the first equality in (2.50), use spherical coordinates (ϱ, θ, ϕ) in the three-dimensional case and substitute $z = \hbar^2 \varrho^2 / 2m^*$:

$$\begin{aligned} g(E) &= \frac{2}{(2\pi)^3} \int_{\mathbb{R}^3} \delta\left(E - E_0 - \frac{\hbar^2}{2m^*}|k|^2\right) dk \\ &= \frac{1}{4\pi^3} \int_0^{2\pi} \int_0^{\pi} \int_0^{\infty} \delta\left(E - E_0 - \frac{\hbar^2}{2m^*}\varrho^2\right) \varrho^2 \sin\theta\, d\varrho\, d\theta\, d\phi \\ &= \frac{4\pi}{4\pi^3} \frac{m^*}{\hbar^2} \frac{\sqrt{2m^*}}{\hbar} \int_0^{\infty} \delta(E - E_0 - z)\sqrt{z}\, dz. \end{aligned}$$

Introducing the Heaviside function H by $H(x) = 0$ for $x < 0$ and $H(x) = 1$ for $x > 0$ we obtain from (2.52):

$$g(E) = \frac{m^*}{\pi\hbar^2}\frac{\sqrt{2m^*}}{\pi\hbar}\int_{\mathbb{R}} \delta(E - E_0 - z)\sqrt{z}H(z)\,dz = \frac{m^*}{\pi\hbar^2}\frac{\sqrt{2m^*}}{\pi\hbar}\sqrt{E - E_0}\;H(E - E_0).$$

For the two-dimensional case, we start again from (2.50) and use polar coordinates (ϱ, ϕ) and the substitution $z = \hbar^2\varrho^2/2m^*$,

$$\begin{aligned} g(E) &= \frac{2}{(2\pi)^2}\int_0^{2\pi}\int_0^\infty \delta\left(E - E_0 - \frac{\hbar^2}{2m^*}\varrho^2\right)\varrho\,d\varrho\,d\phi \\ &= \frac{2\pi}{2\pi^2}\,\frac{m^*}{\hbar^2}\int_0^\infty \delta(E - E_0 - z)\cdot 1\,dz = \frac{m^*}{\pi\hbar^2}. \end{aligned}$$

Finally, for the one-dimensional case,

$$\begin{aligned} g(E) &= \frac{2}{(2\pi)}\int_{\mathbb{R}} \delta\left(E - E_0 - \frac{\hbar^2}{2m^*}k^2\right)\,dk = \frac{1}{\pi}\,\frac{\sqrt{2m^*}}{2\hbar}\int_0^\infty \delta(E - E_0 - z)\frac{dz}{\sqrt{z}} \\ &= \frac{m^*}{\pi\hbar^2}\,\frac{\hbar}{\sqrt{2m^*(E - E_0)}}. \end{aligned}$$

☐

Remark 2.18 In the non-parabolic band approximation

$$E_n(1 + \alpha E_n) = \frac{\hbar^2}{2m^*}|k|^2, \qquad \alpha > 0$$

(see (2.30)), the densities of states become in the three-dimensional case

$$g(E) = \frac{m^*}{\pi\hbar^2}\,\frac{\sqrt{2m^*E}}{\pi\hbar}\sqrt{1 + \alpha E}\,(1 + 2\alpha E)$$

and in the two-dimensional case

$$g(E) = \frac{m^*}{\pi\hbar^2}(1 + 2\alpha E)$$

(see [45, Problem 1.4]). ☐

Lemma 2.19 *Let the conduction and valence bands be given by the parabolic band approximations (2.31) and (2.32). Then, for three-dimensional particles,*

$$n = N_c F_{1/2}\left(\frac{q\mu - E_c}{k_B T}\right), \qquad p = N_v F_{1/2}\left(\frac{E_v - q\mu}{k_B T}\right),$$

where

$$N_c = 2\left(\frac{m_e^* k_B T}{2\pi\hbar^2}\right)^{3/2}, \qquad N_v = 2\left(\frac{m_h^* k_B T}{2\pi\hbar^2}\right)^{3/2} \tag{2.55}$$

are the effective densities of states *and*

$$F_{1/2}(z) = \frac{2}{\sqrt{\pi}}\int_0^\infty \frac{\sqrt{x}\,dx}{1 + e^{x-z}}, \qquad z \in \mathbb{R},$$

is the Fermi integral (of index 1/2). Furthermore, m_e^* and m_h^* denote the (isotropic) effective mass of the electrons and holes, respectively.

Proof: From Lemmas 2.16 and 2.17 and the substitution $x = (E - E_c)/k_BT$ we obtain

$$n = \frac{m_e^*}{\pi\hbar^2}\,\frac{\sqrt{2m_e^*}}{\pi\hbar}\int_{E_c}^{\infty}\frac{\sqrt{E-E_c}}{1+e^{(E-q\mu)/k_BT}}\,dE = \frac{4}{\sqrt{\pi}}\left(\frac{m_e^* k_BT}{2\pi\hbar^2}\right)^{3/2}\int_0^{\infty}\frac{\sqrt{x}\,dx}{1+e^{x-(q\mu-E_c)/k_BT}}$$

$$= N_c F_{1/2}\left(\frac{q\mu - E_c}{k_BT}\right).$$

In a similar way,

$$\begin{aligned} p &= \frac{m_h^*}{\pi\hbar^2}\,\frac{\sqrt{2m_h^*}}{\pi\hbar}\int_{-\infty}^{E_v}\sqrt{E_v - E}\,\frac{e^{(E-q\mu)/k_BT}}{1+e^{(E-q\mu)/k_BT}}\,dE \\ &= \frac{4}{\sqrt{\pi}}\left(\frac{m_h^*}{2\pi\hbar^2}\right)^{3/2}\int_{-\infty}^{E_v}\frac{\sqrt{E_v-E}\,dE}{1+e^{-(E-q\mu)/k_BT}} \\ &= \frac{4}{\sqrt{\pi}}\left(\frac{m_h^* k_BT}{2\pi\hbar^2}\right)^{3/2}\int_{-\infty}^{0}\frac{\sqrt{x}\,dx}{1+e^{x-(E_v-q\mu)/k_BT}} \\ &= N_v F_{1/2}\left(\frac{E_v - q\mu}{k_BT}\right). \end{aligned}$$

□

We compute the carrier densities in some special situations.

Lemma 2.20 *The electron density in a quantum well equals*

$$n = \frac{m_e^* k_BT}{\pi\hbar^2}\ln\left(1+e^{(q\mu-E_c)/k_BT}\right).$$

Proof: In a quantum well, electrons are confined in one direction. Therefore, using the density of states function for two-dimensional carriers (see Lemma 2.17),

$$\begin{aligned} n &= \frac{m_e^*}{\pi\hbar^2}\int_{E_c}^{\infty}\frac{dE}{1+e^{(E-q\mu)/k_BT}} = \frac{m_e^* k_BT}{\pi\hbar^2}\left[-\ln(1+e^{-(E-q\mu)/k_BT})\right]_{E_c}^{\infty} \\ &= \frac{m_e^* k_BT}{\pi\hbar^2}\ln(1+e^{-(E-q\mu)/k_BT}). \end{aligned}$$

□

Lemma 2.21 *The electron and hole densities in the three-dimensional parabolic band and Maxwell-Boltzmann approximation are*

$$n = N_c\mathrm{exp}\left(\frac{q\mu - E_c}{k_BT}\right), \qquad p = N_v\mathrm{exp}\left(\frac{E_v - q\mu}{k_BT}\right),$$

where N_c and N_v are the effective densities of states defined in (2.55).

Proof: For $z \to -\infty$ we can approximate

$$F_{1/2}(z) = e^z \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\sqrt{x}\,dx}{e^z + e^x} \sim e^z \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\sqrt{x}\,dx}{e^x} = e^z \frac{2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = e^z,$$

where $\Gamma(p)$ is the Γ function,

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x}\,dx,$$

with the properties $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\Gamma(p+1) = p\Gamma(p)$, $p > 0$. Thus, the result follows from Lemma 2.19. □

Finally, we discuss two notions needed in the subsequent chapters, the intrinsic density and the doping of semiconductors.

A pure semiconductor with no impurities is called an intrinsic semiconductor. In this case, electrons in the conduction band can only come from valence band levels leaving a vacancy behind them. Vacancies in the valence band are called holes (Figure 2.11). Therefore, the number of electrons in the conduction band is equal to the number of holes in the valence band,

$$n = p = n_i.$$

The quantity n_i is called *intrinsic density.* It can be computed in the non-degenerate parabolic band case from Lemma 2.21:

$$n_i = \sqrt{np} = \sqrt{N_c N_v}\,\exp\left(\frac{E_v - E_c}{2k_B T}\right) = \sqrt{N_c N_v}\,\exp\left(-\frac{E_g}{2k_B T}\right) \tag{2.56}$$

since the energy gap is $E_g = E_c - E_v$. This allows to determine the Fermi energy $E_F = q\mu$ of an intrinsic semiconductor [3, (28.22)]:

$$\begin{aligned} E_F &= E_c + k_B T \ln \frac{n}{N_c} = E_c + k_B T \ln \frac{n_i}{N_c} = E_c - \frac{E_g}{2} + \frac{k_B T}{2} \ln \frac{N_v}{N_c} \\ &= \frac{1}{2}(E_c + E_v) + \frac{3}{4} k_B T \ln \frac{m_h^*}{m_e^*}. \end{aligned}$$

This asserts that at zero temperature, the Fermi energy lies precisely in the middle of the energy gap. Furthermore, since $\ln(m_h^*/m_e^*)$ is of order one, the correction is only of order $k_B T$ for non-zero temperature. In most semiconductors at room temperature, the energy gap is much larger than $k_B T \approx 0.0259$ eV ($T = 300K$). This shows that the non-degeneracy assumptions

$$\begin{aligned} E - E_F \geq E_c - E_F &= \frac{E_g}{2} + \frac{3}{4} k_B T \ln \frac{m_h^*}{m_e^*} \gg k_B T, \\ E_F - E \geq E_F - E_v &= \frac{E_g}{2} + \frac{3}{4} k_B T \ln \frac{m_h^*}{m_e^*} \gg k_B T \end{aligned}$$

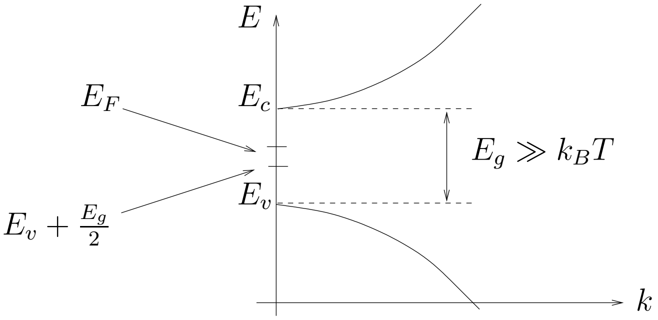

Figure 2.15: Illustration of the energy gap E_g in relation to the energies E_c, E_v, and E_F (see [3, Fig. 28.10]).

are satisfied and that the result is consistent with our assumptions (see Figure 2.15).

The intrinsic density is too small to result in a significant conductivity for non-zero temperature. For instance, in silicon we have $n_i \approx 6.93 \cdot 10^9$ cm^{-3} compared to N_c, $N_v \sim 10^{19}$ cm^{-3}. Replacing some atoms in the semiconductor crystal by atoms which provide free electrons in the conduction band or free holes in the valence band allows to increase the conductivity. Such a process is called *doping*. Impurities are called *donors* if they supply additional electrons to the conduction band, and *acceptors* if they supply additional holes to (i.e. capture electrons from) the valence band. A semiconductor which is doped with donors is termed *n-type semiconductor*, and a semiconductor doped with acceptors is called *p-type semiconductor*. For instance, when we dope a germanicum crystal, whose atoms have each 4 valence electrons, with arsenic, which has 5 valence electrons per atom, each arsenic atom provides one additional electron (see Figure 2.16). These additional electrons are only weakly bound to the arsenic atom. Indeed, the binding energy is about 0.013 eV (see [3, Table 28.2]) which is much smaller than the thermal energy $k_BT \approx 0.026$ eV at room temperature. More generally, denoting by E_d and E_a the energies of a donor electron and an acceptor hole, respectively, then $E_c - E_d$ and $E_a - E_v$ are small compared to k_BT (see Figure 2.17). This means that the additional carriers contribute at room temperature to the electron and hole density and increase the conductivity of the semiconductor.

Let $N_D(x)$, $N_A(x)$ denote the densities of the donor and acceptor impurities, respectively. Then the *doping profile* or doping concentration is $C(x) = N_A(x) - N_D(x)$ and the total space charge (see Chapter 1) is given by

$$\varrho = -qn + qp + qN_A(x) - qN_D(x) = -q(n - p - C(x)).$$

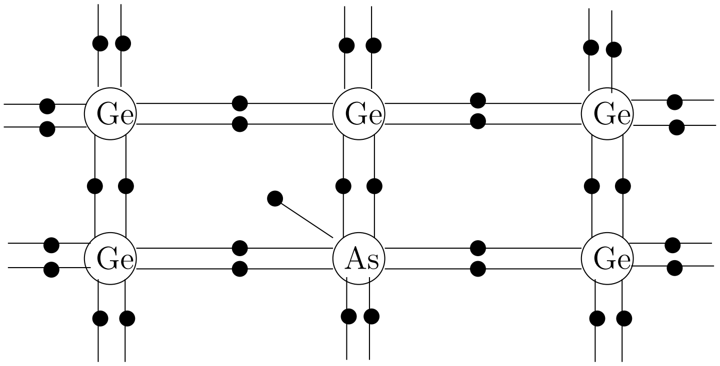

Figure 2.16: Germanium crystal with an arsenic impurity. The bullets • represent electrons bound to the atoms. The arsenic atom provides an additional electron which is not bound.

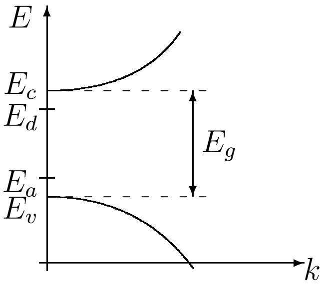

Figure 2.17: Illustration of the donor and acceptor energy levels E_d and E_a.

3 Classical Kinetic Models

3.1 The Liouville equation

We first analyze the motion of M particles with mass m in a vacuum under the action of a force. The particles are described as classical particles, i.e., we associate the position vector $x_i \in \mathbb{R}^d$ and the velocity vector $v_i \in \mathbb{R}^d$ with the i-th particle of the ensemble. The trajectories $(x_i(t), v_i(t))$ of the particles satisfy Newton's equations in the $(2d \cdot M)$-dimensional ensemble position-velocity space

$$\dot{x} = v, \qquad \dot{v} = \frac{1}{m} F(t), \quad t > 0, \tag{3.1}$$

with initial conditions

$$x(0) = x_0, \qquad v(0) = v_0. \tag{3.2}$$

Here, $x = (x_1, \ldots, x_M)$, $v = (v_1, \ldots, v_M)$, $F = (F_1, \ldots, F_M)$ is a force and the dot means differentiation with respect to t. For instance, the forces can be given by the electric field acting on the electron ensemble:

$$F_i = -qE(x,t), \qquad i = 1, \ldots, M,$$

where $q > 0$ is the elementary charge. In this case the forces are independent of the velocity. In semiconductors, M is a very large number (typically, $M \sim 10^4$) and therefore, the numerical solution of (3.1)–(3.2) is very expensive. It seems reasonable to use a statistical description. Instead of the precise initial conditions $x(0)$ and $v(0)$ we are given the probability density $f_I(x,v)$ of the initial position and velocity of the particles. The integral

$$\int_A f_I(x,v)\, dx\, dv$$

is the probability to find the particle ensemble at time $t = 0$ in the subset A of the (v,x)-phase space.

Let $f(x,v,t)$ be the probability density of the ensemble at time t. We wish to derive a differential equation for f. In Chapter 1 we have already argued that under the assumption that f is constant along the trajectories $(x(t), v(t))$, i.e.

$$f(x(t), v(t), t) = f_I(x_0, v_0), \tag{3.3}$$

differentiation with respect to t leads to the differential equation

$$\begin{aligned} 0 = \frac{d}{dt} f(x(t), v(t), t) &= \partial_t f + \dot{x} \cdot \nabla_x f + \dot{v} \cdot \nabla_v f \\ &= \partial_t f + v \cdot \nabla_x f + \frac{1}{m} F \cdot \nabla_v f. \end{aligned} \tag{3.4}$$

This equation is referred to as the *(classical) Liouville equation*. It has to be solved in $\mathbb{R}^{2dM} \times \mathbb{R}$ and is supplemented by the initial condition

$$f(x, v, 0) = f_I(x, v), \qquad (x, v) \in \mathbb{R}^{2dM}.$$

In the following we show under which conditions (3.3) is satisfied.

Theorem 3.1 (Liouville)
Assume that

$$\sum_{j=1}^{M} \left(\frac{\partial \dot{x}_i}{\partial x_i} + \frac{\partial \dot{v}_i}{\partial v_i} \right) = 0 \qquad \textit{for all } t > 0. \tag{3.5}$$

Then (3.3) holds.

Notice that (3.5) is trivially satisfied if Newton's laws (3.1) are assumed. However, in the semiconductor case, (3.1) has to be replaced by the semi-classical picture (2.33) for which (3.5) will be verified later. For details about the Liouville equation for gases, see [19].

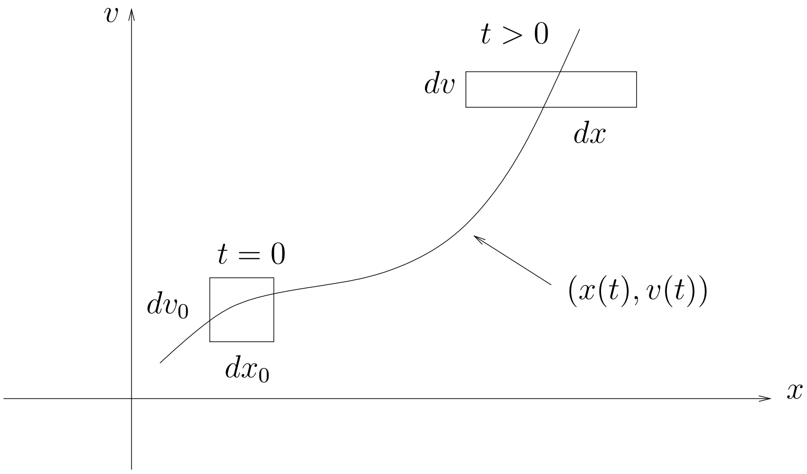

Figure 3.1: Infinitesimal volume elements $dx_0 dv_0$ at time $t = 0$ and $dxdv$ at time $t > 0$ in the phase space.

Proof: We will examine how the infinitesimal phase-space element $dx_0 dv_0$ will change during time. In particular, we show that [23, Sec. 1.2]

$$dxdv(t) = dx_0 dv_0. \tag{3.6}$$

This implies the assertion (3.3). Indeed, the number of particles at time t in the volume $dxdv(t)$ is given by

$$f(x(t), v(t), t) dxdv(t).$$

The conservation of particles implies that

$$f(x(t), v(t), t)dxdv(t) = f(x_0, v_0, 0)dx_0dv_0$$

and therefore, (3.6) gives (3.3).

In the following we show (3.6). Assuming sufficient regularity of the functions, the mapping $(x_0, v_0) \mapsto (x(t), v(t))$ is continuously differentiable and invertible with continuously differentiable inverse. Then

$$dxdv(t) = J(t)dx_0dv_0, \tag{3.7}$$

where

$$J(t) = \left|\det \frac{\partial(x(t), v(t))}{\partial(x_0, v_0)}\right| = \left|\det \begin{pmatrix} \frac{\partial x(t)}{\partial x_0} & \frac{\partial x(t)}{\partial v_0} \\ \frac{\partial v(t)}{\partial x_0} & \frac{\partial v(t)}{\partial v_0} \end{pmatrix}\right|.$$

We claim that (3.5) implies $\partial_t J(t) = 0$. In order to see this, we use the Taylor expansion

$$x(t+\varepsilon) = x(t) + \varepsilon\dot{x}(t) + O(\varepsilon^2), \qquad v(t+\varepsilon) = v(t) + \varepsilon\dot{v}(t) + O(\varepsilon^2).$$

This gives

$$\frac{\partial x(t+\varepsilon)}{\partial x(t)} = I + \varepsilon\frac{\partial \dot{x}(t)}{\partial x(t)} + O(\varepsilon^2), \qquad \frac{\partial v(t+\varepsilon)}{\partial v(t)} = I + \varepsilon\frac{\partial \dot{v}(t)}{\partial v(t)} + O(\varepsilon^2),$$

where I denotes the unit matrix in $(dM) \times (dM)$. Since $x(t)$ and $v(t)$ are independent variables, we have

$$\frac{\partial x(t)}{\partial v(t)} = 0, \qquad \frac{\partial v(t)}{\partial x(t)} = 0$$

and therefore,

$$\frac{\partial x(t+\varepsilon)}{\partial v(t)} = \varepsilon\frac{\partial \dot{x}(t)}{\partial v(t)} + O(\varepsilon^2), \qquad \frac{\partial v(t+\varepsilon)}{\partial x(t)} = \varepsilon\frac{\partial \dot{v}(t)}{\partial x(t)} + O(\varepsilon^2).$$

Thus we can compute

$$\begin{aligned}
\det \frac{\partial(x,v)(t+\varepsilon)}{\partial(x,v)(t)} &= \det \begin{pmatrix} \dfrac{\partial x(t+\varepsilon)}{\partial x(t)} & \dfrac{\partial x(t+\varepsilon)}{\partial v(t)} \\ \dfrac{\partial v(t+\varepsilon)}{\partial x(t)} & \dfrac{\partial v(t+\varepsilon)}{\partial v(t)} \end{pmatrix} \\
&= \det \begin{pmatrix} I + \varepsilon\dfrac{\partial \dot{x}(t)}{\partial x(t)} & \varepsilon\dfrac{\partial \dot{x}(t)}{\partial v(t)} \\ \varepsilon\dfrac{\partial \dot{v}(t)}{\partial x(t)} & I + \varepsilon\dfrac{\partial \dot{v}(t)}{\partial v(t)} \end{pmatrix} + O(\varepsilon^2) \\
&= 1 + \varepsilon\sum_i \left(\frac{\partial \dot{x}_i(t)}{\partial x_i(t)} + \frac{\partial \dot{v}_i(t)}{\partial v_i(t)}\right) + O(\varepsilon^2) \\
&= 1 + O(\varepsilon^2),
\end{aligned}$$

by using assumption (3.5). Hence, by the chain rule,

$$J(t+\varepsilon) = \left|\det\left(\frac{\partial(x,v)(t+\varepsilon)}{\partial(v,x)(t)}\ \frac{\partial(v,x)(t)}{\partial(x_0,v_0)}\right)\right| = \left|\det\frac{\partial(x,v)(t+\varepsilon)}{\partial(v,x)(t)}\ \det\frac{\partial(v,x)(t)}{\partial(x_0,v_0)}\right|$$

$$= \left|\det\frac{\partial(v,x)(t)}{\partial(x_0,v_0)}\right| + O(\varepsilon^2) = J(t) + O(\varepsilon^2),$$

and we obtain finally

$$\partial_t J(t) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}(J(t+\varepsilon)-J(t)) - \lim_{\varepsilon\to 0} O(\varepsilon) = 0.$$

The mapping $(x_0,v_0)\mapsto(x(t),v(t))$ is the identity for $t=0$ such that $J(0)=1$. This implies that $J(t)=1$ for all $t\geq 0$ and hence, (3.7) reduces to (3.6). The theorem is proved. □

The Liouville equation (3.4) is valid for an ensemble of particles moving in a vacuum according to the laws of classical mechanics. An ensemble of M electrons moving in a semiconductor crystal can be described semi-classically by the equations

$$\dot{x}_j = v_n(k) = \frac{1}{\hbar}\nabla_{k_j}E_n(k), \qquad \dot{k}_j = -\frac{q}{\hbar}\nabla_{x_j}V(x,t), \qquad j=1,\ldots,M \tag{3.8}$$

(see (2.33)), where $E_n(k_j)$ is the energy of the n-th band depending on the (pseudo-) wave vector k and $V(x,t)$ is the electrostatic potential. We denote as above $x=(x_1,\ldots,x_M)$ and $k=(k_1,\ldots,k_M)\in\mathbb{R}^{dM}$. We assume that the ensemble stays in the same energy band so that we can drop the index n in the equations (3.8).

We claim that Liouville's Theorem 3.1 can be applied to the semi-classical picture. Since the classical momentum $p=mv$ translates into the crystal momentum $p=\hbar k$, we have to verify

$$\sum_j\left(\frac{\partial\dot{x}_i}{\partial x_i}+\frac{\partial\dot{k}_i}{\partial k_i}\right)=0.$$

Actually, $\dot{x}$ depends on k but not on x, and $\dot{k}$ depends on x but not on k (see (3.8)) such that this condition is satisfied. Therefore, (3.3) holds and the Liouville equation (3.4) is valid here. Observing that Newton's law (3.1) can be written as

$$\dot{p}=F,$$

and that (3.8) can be reformulated as

$$\dot{p}=-q\nabla_x V,$$

the Liouville equation (3.4) becomes

$$\partial_t f + \frac{1}{\hbar}\nabla_k E(k)\cdot\nabla_x f - \frac{q}{\hbar}\nabla_x V\cdot\nabla_k f = 0. \tag{3.9}$$

Here, we have used that the classical term $\nabla_p = \frac{1}{m}\nabla_v$ translates into the semi-classical expression $\nabla_p = \frac{1}{\hbar}\nabla_k$. Equation (3.9) is referred to as the *semi-classical Liouville equation.* More precisely, the scalar products are defined as

$$\nabla_k E(k) \cdot \nabla_x f = \sum_{j=1}^{M} \nabla_{k_j} E(k_j) \cdot \nabla_{x_j} f, \qquad \nabla_x V \cdot \nabla_k f = \sum_{j=1}^{M} \nabla_{x_j} V(x) \cdot \nabla_{k_j} f.$$

Equation (3.9) has to be solved for $x \in \mathbb{R}^{dM}$, $k \in B^M$, $t > 0$, where B is the Brillouin zone (see Section 2.1). As the Brillouin zone is a bounded subset, we have to impose boundary conditions. Often, periodic boundary conditions

$$f(x, k_1, \ldots, k_j, \ldots, k_M, t) = f(x, k_1, \ldots, -k_j, \ldots, k_M, t), \qquad k_j \in \partial B,$$

for all $j = 1, \ldots, M$ are chosen [49, (1.2.47)]. This formulation makes sense since B is point symmetric to the origin, i.e., $k \in B$ if and only if $-k \in B$. We also impose the initial conditions

$$f(x, k, 0) = f_I(x, k), \qquad x \in \mathbb{R}^{dM},\ k \in B^M.$$

Before we show some properties of the solutions to (3.9), we introduce the electron-ensemble particle density and the electron-ensemble current density:

$$\begin{aligned} n(x,t) &= \int_{B^M} f(x,k,t)\,dk, \qquad (3.10) \\ J(x,t) &= -q \int_{B^M} f(x,k,t) v(k)\,dk, \end{aligned}$$

where $v(k) = (v(k_1), \ldots, v(k_M))$.

Remark 3.2 Definition (3.10) of the number density is slightly inexact. Indeed, similar to Lemma 2.16 and arguing as in [23, Sec. 1.2.2], the total number of electrons at time t in the volume $dx\,dk$ of the (conduction band) phase space is given by

$$dN(x,k) = f(x,k,t) g(x,k)\,dx\,dk,$$

where $g(x,k)$ is the momentum density of states (in the conduction band) and $f(x,k,t)$ can be interpreted as an occupation distribution, i.e, the ratio of the number of occupied states in the phase-space volume $dx\,dk$ and the total number of quantum states in this volume (in the conduction band). The volume in phase space occupied by a single quantum state is of the order of $(2\pi)^d$. Thus, we have $dx\,dk/(2\pi)^d$ quantum states in the volume $dx\,dk$. To be precise, we have to take into account the spin of the electrons. Each quantum state can be occupied by two electrons with opposite spin. Therefore, the number of quantum states is twice such that $g(x,k) = 2/(2\pi)^d$. The correct definition of the electron density becomes

$$n(x,t) = \int_{B^M} f(x,k,t)g(x,k)\,dk = \int_{B^M} f(x,k,t)\frac{2}{(2\pi)^d}\,dk$$

(see [23] for details). In the following we ignore the factor $2/(2\pi)^d$ which can be achieved, for instance, by scaling. □

Lemma 3.3 *The solution of the semi-classical Liouville equation (3.9) satisfies formally the following properties:*

(1) If $f_I(x,k) \geq 0$ for all $x \in \mathbb{R}^{dM}$, $k \in B^M$, then $f(x,k,t) \geq 0$ for all $x \in \mathbb{R}^{dM}$, $k \in B^M$.

(2) The following conservation law holds:

$$\partial_t n - \frac{1}{q}\mathrm{div}_x J = 0, \qquad x \in \mathbb{R}^{dM}, \quad t > 0.$$

(3) The number of particles is conserved:

$$\int_{\mathbb{R}^{dM}} n(x,t)\,dx = \int_{\mathbb{R}^{dM}}\int_{B^M} f_I(x,k)\,dk\,dx.$$

Proof: The first property follows immediately from (3.3). By formally integrating (3.9) over $k \in B^M$ we obtain from the divergence theorem,

$$\begin{aligned} \partial_t n &= \int_{B^M} \partial_t f\,dk = -\int_{B^M} \mathrm{div}_x(v(k)f)\,dk + \frac{q}{\hbar}\int_{B^M} \mathrm{div}_k(\nabla_x V f)\,dk \\ &= -\frac{1}{q}\mathrm{div}_x J. \end{aligned}$$

Finally, the last property follows from the second one after integrating over $x \in \mathbb{R}^{dM}$. □

We notice that in the parabolic band approximation

$$E(k_j) = \frac{\hbar^2}{2m^*}|k_j|^2, \quad j = 1,\ldots,M, \quad k_j \in \mathbb{R}^d,$$

the semi-classical Liouville equation (3.9) reduces to its classical counterpart (3.4) since $v(k_j) = \hbar k_j/m^* = p_j/m^*$, which is the classical momentum-velocity relation, and $\nabla_k f = (\hbar/m^*)\nabla_v f$. We obtain from (3.9)

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

3.2 The Vlasov equation

The main disadvantage of the (semi-)clasical Liouville equation is that it has to be solved in the very high-dimensional phase space $\mathbb{R}^{2dM}$. Typically, $M \sim 10^4$, $d = 3$, and the dimension becomes $6 \cdot 10^4$ which is prohibitive for numerical simulations. In this section we will formally derive a lower-dimensional equation, the so-called Vlasov equation. We proceed as in [49, Sec. 1.3].

The idea of the derivation is first to assume a certain structure of the force field, then to integrate the Liouville equation in a certain sub-phase space and finally to carry out the formal limit $M \to \infty$, where M is the number of particles. More precisely, we consider an ensemble of M electrons and denote by $x = (x_1 \ldots, x_M) \in \mathbb{R}^{dM}$, $k = (k_1, \ldots, k_m) \in B^M$ the position and wave vector coordinates of the particles, respectively. We impose the following assumptions:

(1) The motion is governed by an external electric field and by two-particle (long-range) interaction forces

$$F_j(x,t) = -qE_{\text{ext}}(x_j,t) - q \sum_{\ell=1, j\neq\ell}^{M} E_{\text{int}}(x_j,x_\ell), \qquad j = 1, \ldots, M,$$

such that

$$E_{\text{int}}(x_j,x_\ell) = -E_{\text{int}}(x_\ell,x_j). \tag{3.11}$$

(2) The interaction force $|E_{\text{int}}|$ is of order $1/M$.

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

$$f_I(x_1,\ldots,x_M,k_1,\ldots,k_M) = f_I(x_{\pi(1)},\ldots,x_{\pi(M)},k_{\pi(1)},\ldots,k_{\pi(M)}) \tag{3.12}$$

for all $x_j \in \mathbb{R}^d$, $k_j \in B$, $j = 1, \ldots, M$, and for all permutations π of $\{1, \ldots, M\}$.

(4) The sub-ensemble initial density

$$f_I^{(a)}(x_1,\ldots,x_a,k_1,\ldots,k_a) = \int_{(\mathbb{R}^d\times B)^{M-a}} f_I dx_{a+1}\cdots dx_M dk_{a+1}\cdots dk_M$$

can be factorized:

$$f_I^{(a)} = \prod_{j=1}^{a} P_I(x_j,k_j), \qquad a = 1, \ldots, M-1.$$

We discuss the above assumptions. The first assumption means that magnetic fields are ignored. In fact, this hypothesis can be discarded, see [49, p. 26ff]. This assumption, which is crucial for the derivation of the Vlasov equation, also means that the force field F_i

exerted on the i-th electron is given by the sum of an external electric field acting on the i-th electron and of the sum of $M-1$ two-particle interaction forces between the i-th electron and all other electrons. The interaction force E_{int} is independent of the electron indices which interprets the fact that the electrons are indistinguishable. The action-reaction law implies that the force exerted by the j-th electron on the ℓ-th electron is the negative force of the ℓ-th electron on the j-th electron, i.e. $E_{\text{int}}(x_j, x_\ell) = -E_{\text{int}}(x_\ell, x_j)$. This property and assumption (3) imply that also $f(x,k,t)$ is independent of the numbering of the particle! s for all $t > 0$. Finally, assumption (4) is needed for the limit $M \to \infty$ (in order to get a finite force F_j) and its meaning will become clear later.

We introduce the density $f^{(a)}$ of a subensemble consisting of $a < M$ electrons:

$$f^{(a)}(x_1, \ldots, x_a, k_1, \ldots, k_a, t) = \int_{(\mathbb{R}^d \times B)^{M-a}} f(x,k,t)\, dx_{(a+1)}\, dk_{(a+1)},$$

where $dx_{(a+1)} = dx_{a+1} \cdots dx_M$ and $dk_{(a+1)} = dk_{a+1} \cdots dk_M$.

Theorem 3.4 *Let the above assumptions (1)-(4) hold. Then the function $f(x,k,t)$ is a particular solution to the semi-classical Liouville equation (3.9) if $M \gg 1$,*

$$f^{(a)}(x_1, \ldots, x_a, k_1, \ldots, k_a, t) = \prod_{j=1}^{a} P(x_j, k_j, t) \tag{3.13}$$

and $F(x,k,t) := MP(x,k,t)$ is a solution to the semi-classical Vlasov equation

$$\begin{aligned} \partial_t F + v(k) \cdot \nabla_x F - \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k F &= 0, \qquad x \in \mathbb{R}^d,\ k \in B,\ t > 0, \\ F(x,k,0) &= F_I(x,k), \qquad x \in \mathbb{R}^d,\ k \in B, \end{aligned} \tag{3.14}$$

where

$$\begin{aligned} E_{\text{eff}}(x,t) &= E_{\text{ext}}(x,t) + \int_{\mathbb{R}^d \times B} F(x_2, k_2, t) E_{\text{int}}(x, x_2)\, dx_2 dk_2 \\ &= E_{\text{ext}}(x,t) + \int_{\mathbb{R}^d} n(x_*, t) E_{\text{int}}(x, x_*)\, dx_* \end{aligned} \tag{3.15}$$

and $F_I(x,k) = MP_I(x,k)$ (see assumption (4)). Furthermore,

$$n(x,t) = \int_B F(x,k,t)\, dk$$

represents the electron density, and we impose periodic boundary conditions:

$$F(x,k,t) = F(x,-k,t), \qquad x \in \mathbb{R}^d,\ k \in \partial B,\ t > 0.$$

Proof: We integrate the semi-classical Liouville equation

$$\partial_t f + \sum_{j=1}^{M} v(k_j)\cdot\nabla_{x_j} f - \frac{q}{\hbar}\sum_{j=1}^{M} E_{\text{ext}}(x_j,t)\cdot\nabla_{k_j} f - \frac{q}{\hbar}\sum_{j,\ell=1}^{M} E_{\text{int}}(x_j,x_\ell)\cdot\nabla_{k_j} f = 0 \quad (3.16)$$

with respect to $x_{a+1},\dots,x_M,k_{a+1},\dots,k_M$ in order to obtain an equation for $f^{(a)}$. We reformulate these integrals term by term.

Clearly, the first term on the left-hand side of (3.16) equals $\partial_t f^{(a)}$ after the integration. For the second term we compute, using the divergence theorem,

$$\begin{aligned}\sum_{j=1}^{M}\int_{(\mathbb{R}^d\times B)^{M-a}} v(k_j)\cdot\nabla_{x_j} f\,dx_{(a+1)}\,dk_{(a+1)} &= \sum_{j=1}^{a} v(k_j)\cdot\nabla_{x_j}\int_{(\mathbb{R}^d\times B)^{M-a}} f\,dx_{(a+1)}\,dk_{(a+1)}\\ &\quad+\sum_{j=a+1}^{M}\int_{(\mathbb{R}^d\times B)^{M-a}}\operatorname{div}_{x_j}(v(k_j)f)\,dx_{(a+1)}\,dk_{(a+1)}\\ &= \sum_{j=1}^{a} v(k_j)\cdot\nabla_{x_j} f^{(a)}.\end{aligned}$$

Similarly,

$$\sum_{j=1}^{M}\int_{(\mathbb{R}^d\times B)^{M-a}}\operatorname{div}_{k_j}\left(E_{\text{ext}}(x_j,t)f\right)\,dx_{(a+1)}\,dk_{(a+1)} = \sum_{j=1}^{a}\operatorname{div}_{k_j}\left(E_{\text{ext}}(x_j,t)f^{(a)}\right).$$

The last integral on the left-hand side becomes

$$\begin{aligned}&\sum_{j,\ell=1}^{M}\int_{(\mathbb{R}^d\times B)^{M-a}}\operatorname{div}_{k_j}\left(E_{\text{int}}(x_j,x_\ell)f\right)\,dx_{(a+1)}\,dk_{(a+1)}\\ &= \sum_{j,\ell=1}^{a} E_{\text{int}}(x_j,x_\ell)\cdot\nabla_{k_j} f^{(a)} + \sum_{j=a+1}^{M}\sum_{\ell=1}^{M}\int_{(\mathbb{R}^d\times B)^{M-a}}\operatorname{div}_{k_j}\left(E_{\text{int}} f\right)\,dx_{(a+1)}\,dk_{(a+1)}\\ &\quad+\sum_{j=1}^{a}\sum_{\ell=a+1}^{M}\int_{(\mathbb{R}^d\times B)^{M-a}}\operatorname{div}_{k_j}\left(E_{\text{int}}(x_j,x_\ell)f\right)\,dx_{(a+1)}\,dk_{(a+1)}.\end{aligned}$$

The second integral on the right-hand side of the above equation vanishes by the divergence theorem. For the last integral we use the assumptions (3.11) and (3.12). Indeed, it is possible to renumber the particles such that the last integral equals

$$\begin{aligned}&\sum_{j=1}^{a}(M-a)\int_{(\mathbb{R}^d\times B)^{M-a}}\operatorname{div}_{k_j}\left(E_{\text{int}}(x_j,x_{a+1})f\right)\,dx_{(a+1)}\,dk_{(a+1)}\\ &=\sum_{j=1}^{a}(M-a)\operatorname{div}_{k_j}\int_{\mathbb{R}^d\times B} E_{\text{int}}(x_j,x_*)f_*^{(a+1)}dx_*dk_*,\end{aligned}$$

where

$$f_*^{(a+1)} = f^{(a+1)}(x_1, \dots, x_a, x_*, k_1, \dots, k_a, k_*, t).$$

Thus, integration of (3.16) yields the system of equations

$$0 = \partial_t f^{(a)} + \sum_{j=1}^{a} v(k_j) \cdot \nabla_{x_j} f^{(a)} - \frac{q}{\hbar} \sum_{j=1}^{a} E_{\text{ext}}(x_j, t) \cdot \nabla_{k_j} f^{(a)} - \frac{q}{\hbar} \sum_{j,\ell=1}^{a} E_{\text{int}}(x_j, x_\ell) \cdot \nabla_{k_j} f^{(a)}$$

$$-\frac{q}{\hbar} \sum_{j=1}^{a} \operatorname{div}_{k_j} \int_{\mathbb{R}^d \times B} (M-a) E_{\text{int}}(x_j, x_*) f_*^{(a+1)} dx_* dk_*, \tag{3.17}$$

where $1 \le a \le M-1$. These equations are called the *BBGKY hierarchy* (from Bogoliubov [12], Born and Green [14], Kirkwood [39], and Yvon [67]). By assumption (2), $|E_{\text{int}}|$ is of the order of $1/M$ such that for $M \gg 1$, the fourth term on the right-hand side of (3.17) can be neglected. The term $(M-a)E_{\text{int}}$ however, stays finite and is approximately equal to ME_{int}. Therefore, we obtain for $M \gg 1$

$$\begin{aligned} 0 &= \partial_t f^{(a)} + \sum_{j=1}^{a} v(k_j) \cdot \nabla_{x_j} f - \frac{q}{\hbar} \sum_{j=1}^{a} E_{\text{ext}}(x_j, t) \nabla_{k_j} f^{(a)} \\ &\quad -\frac{q}{\hbar} \sum_{j=1}^{a} \operatorname{div}_{k_j} \int_{\mathbb{R}^d \times B} M f_*^{(a+1)} E_{\text{int}}(x_j, x_*) dx_* dk_*. \end{aligned} \tag{3.18}$$

Now we claim that this equation is solved by the ansatz (3.13). In order to see this we multiply (3.14) for $(x,k) = (x_j, k_j)$ by

$$Q_j = \frac{1}{M} \prod_{\ell \ne j} P(x_\ell, k_\ell, t)$$

and take the sum for $j = 1, \dots, a$. Then, for $F_j = MP(x_j, k_j, t)$,

$$\sum_{j=1}^{a} Q_j \partial_t F_j = \sum_{j=1}^{a} \prod_{\ell \ne j} P(x_\ell, k_\ell, t) \partial_t P(x_j, k_j, t) = \partial_t \prod_{j=1}^{a} P(x_j, k_j, t) = \partial_t f^{(a)}.$$

In a similar way,

$$\begin{aligned} \sum_{j=1}^{a} Q_j v(k_j) \cdot \nabla_{x_j} F_j &= \sum_{j=1}^{a} v(k_j) \cdot \nabla_{x_j} f^{(a)}, \\ \sum_{j=1}^{a} Q_j E_{\text{ext}}(x_j, t) \cdot \nabla_{k_j} F_j &= \sum_{j=1}^{a} E_{\text{ext}}(x_j, t) \cdot \nabla_{k_j} f^{(a)}. \end{aligned}$$

$$
\begin{aligned}
&\sum_{j=1}^{a} Q_j \left(\int_{\mathbb{R}^d \times B} MP(x_*, k_*, t) E_{\text{int}}(x_j, x_*)\, dx_*\, dk_* \right) \cdot \nabla_{k_j} F_j \\
&\quad = \sum_{j=1}^{a} \operatorname{div}_{k_j} \int_{\mathbb{R}^d \times B} M \prod_{\ell=1}^{a} P(x_\ell, k_\ell, t) P(x_*, k_*, t) E_{\text{int}}(x_j, x_*)\, dx_*\, dk_* \\
&\quad = \sum_{j=1}^{a} \operatorname{div}_{k_j} \int_{\mathbb{R}^d \times B} M f_{(*)}^{(a+1)} E_{\text{int}}(x_j, x_*)\, dx_*\, dk_*.
\end{aligned}
$$

Putting together the above computations, we see that the ansatz (3.13) indeed solves (3.18). This proves the theorem. ☐

The semi-classical Vlasov equation has the form of a Liouville equation for a single particle with the force $-qE_{\text{eff}}$. Many-particle physics is taken into account through the effective field E_{eff} which depends on the density n and hence on F. Thus, (3.14) is a nonlinear equation with a nonlocal quadratic nonlinearity. The Vlasov equation describes the macroscopic motion of many-particle systems with weak long-range forces. However, it does not provide a description of strong short-range forces such as scattering of particles. This case will be considered in Section 3.3.

As for the Liouville equation, the quantity $F(x, k, t)$ can be interpreted as the probability density of a particle to be in the state (x, k) at time t. Indeed, we obtain from the trajectory equations

$$
\dot{x} = v(k), \qquad \dot{k} = -\frac{q}{\hbar} E_{\text{eff}}, \qquad t > 0, \qquad x(0) = x_0, \qquad k(0) = k_0,
$$

and (3.14) the equation

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

and thus

$$
F(x(t), k(t), t) = F_I(x_0, k_0) \geq 0 \qquad \text{for all } t \geq 0.
$$

Finally, we wish to reformulate the nonlinear system (3.14)–(3.15) in the case of the *Coulomb force* in $\mathbb{R}^3$,

$$
E_{\text{int}}(x, y) = -\frac{q}{4\pi\varepsilon_s} \frac{x - y}{|x - y|^3}, \qquad x, y \in \mathbb{R}^3,\ x \neq y, \tag{3.19}
$$

which is the most important long-range force between two electrons. Here, the *permittivity* ε_s is a material constant. We assume that the external field is generated by doping atoms in the semiconductor crystal of charge $+q$:

$$
E_{\text{ext}}(x, t) = \frac{+q}{4\pi\varepsilon_s} \int_{\mathbb{R}^3} C(y) \frac{x - y}{|x - y|^3}\, dy, \tag{3.20}
$$

where $C(x)$ is the doping concentration (see Section 2.4).

Proposition 3.5 *In the case of the Coulomb force (3.19) and (3.20) the semi-classical Vlasov equation (3.14)–(3.15) can be written as the* Vlasov-Poisson system

$$\begin{aligned} \partial_t F + v(k)\cdot\nabla_x F - \frac{q}{\hbar}\nabla_x V_{\text{eff}}\cdot\nabla_k F &= 0, && (3.21)\\ \varepsilon_s \Delta_x V_{\text{eff}} &= q(n-C), \qquad x\in\mathbb{R}^3,\ k\in B,\ t>0. && (3.22) \end{aligned}$$

Proof: It is well known that the function

$$\phi(x) = -\frac{1}{4\pi}\int_{\mathbb{R}^3}\frac{f(y)}{|x-y|}\,dy, \qquad x\in\mathbb{R}^3,$$

solves the Poisson equation $\Delta\phi = f$ in $\mathbb{R}^3$ under some regularity assumptions on f. Therefore,

$$\begin{aligned} f(x) = \Delta\phi(x) &= \frac{1}{4\pi}\int_{\mathbb{R}^3} f(y)\text{div}_x\frac{x-y}{|x-y|^3}\,dy,\\ 0 = \text{curl}\,\nabla\phi(x) &= \frac{1}{4\pi}\text{curl}_{\,x}\int_{\mathbb{R}^3} f(y)\frac{x-y}{|x-y|^3}\,dy. \end{aligned}$$

This shows that

$$\begin{aligned} \text{div}E_{\text{eff}}(x,t) &= \text{div}E_{\text{ext}}(x,t) + \int_{\mathbb{R}^3} n(x_*,t)\text{div}E_{\text{int}}(x,x_*)\,dx_*\\ &= \text{div}E_{\text{ext}}(x,t) - \frac{q}{\varepsilon_s}n(x,t),\\ \text{curl}\,E_{\text{eff}}(x,t) &= \text{curl}\,E_{\text{ext}}(x,t) + \int_{\mathbb{R}^3} n(x_*,t)\text{curl}\,E_{\text{int}}(x,x_*)\,dx_*\\ &= \text{curl}\,E_{\text{ext}}(x,t), \qquad x\in\mathbb{R}^3,\ t>0. \end{aligned}$$

Furthermore,

$$\text{div}E_{\text{ext}} = \frac{q}{\varepsilon_s}C, \qquad \text{curl}\,E_{\text{ext}} = 0 \qquad \text{in } \mathbb{R}^3,$$

and hence,

$$\text{div}E_{\text{eff}} = -\frac{q}{\varepsilon_s}(n-C), \qquad \text{curl}\,E_{\text{eff}} = 0 \qquad \text{in } \mathbb{R}^3.$$

Since E_{eff} is vortex-free, there exists a potential V_{eff} such that $E_{\text{eff}} = -\nabla V_{\text{eff}}$. Thus,

$$\varepsilon_s\Delta V_{\text{eff}} = -\varepsilon_s\text{div}E_{\text{eff}} = q(n-C),$$

and the proposition follows. □

3.3 The Boltzmann equation

The Vlasov equation does not account for short-range particle interactions, like collisions of the particles with other particles or with the crystal lattice. We wish to extend the Vlasov equation to include scattering mechanisms which leads to the Boltzmann equation. We present only a phenomenological derivation; for details on more rigorous derivations, we refer to [4, Sec. 1.5.3], [15, Ch. 7], and [19]. The Boltzmann equation has been first formulated by Boltzmann in 1872 for the non-equilibrium transport of dilute gases [13].

The Vlasov equation along trajectories

$$\frac{dF}{dt} = 0$$

states that the probability F (of occupation of states) does not change in time. Scattering allows particles to jump to another trajectory. Our main assumption is that the rate of change of F due to convection and the effective field, dF/dt, and the rate of change of F due to collisions, $Q(F)$, balance:

$$\frac{dF}{dt} = Q(F).$$

Clearly, this equation has to be understood along trajectories. By (3.4) this equation equals (writing f instead of F)

$$\partial_t f + v(k) \cdot \nabla_x f - \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k f = Q(f), \qquad x \in \mathbb{R}^d,\ k \in B,\ t > 0, \tag{3.23}$$

where the effective field is given by (3.15).

It remains to derive an expression for $Q(f)$. We assume that scattering of particles occurs instantaneously and only changes the crystal momentum of the particles. The rate $P(x, k' \to k, t)$ at which a particle at (x, t) changes its Bloch state k' into another Bloch state k due to a scattering event is proportional to

- the occupation probability $f(x, k', t)$ and
- the probability $1 - f(x, k, t)$ that the state (x, k) is not occupied at time t.

Here, we used the Pauli exclusion principle. Thus,

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

where the proportionality constant $s(x, k', k)$ is called the *scattering rate*. Then the rate of change of f due to collisions is the sum of all in-scattering rates from k' to k minus the out-scattering rate from k to k',

$$P(x, k' \to k, t) - P(x, k \to k', t),$$

for all possible Bloch states k' in the volume element dk'. In the continuum limit, the sum becomes an integral and we obtain

$$\begin{aligned}(Q(f))(x,k,t) &= \int_B [P(x,k'\to k,t) - P(x,k\to k',t)]dk' \qquad (3.24)\\ &= \int_B [s(x,k',k)f'(1-f) - s(x,k,k')f(1-f')]dk',\end{aligned}$$

where $f = f(x,k,t)$, $f' = f(x,k',t)$. Equation (3.23), together with the effective field equation

$$E_{\text{eff}}(x,t) = E_{\text{ext}}(x,t) + \int_{\mathbb{R}^d} n(y,t)E_{\text{int}}(x,y)\,dy, \qquad (3.25)$$

where E_{ext} and E_{int} are given functions, and the collision operator (3.24), is called the *semi-classical Boltzmann equation.* When E_{ext} and E_{int} are given by the Coulomb forces (3.19) and (3.20), equations (3.23)–(3.25) are called the *Boltzmann-Poisson system* which can be written as (3.21)–(3.22) with f instead of F and with the right-hand side $Q(f)$ in (3.21).

Again we impose the initial and periodic boundary conditions

$$\begin{aligned}f(x,k,t) &= f(x,-k,t), \qquad x\in\mathbb{R}^d,\ k\in\partial B,\ t>0, \qquad (3.26)\\ f(x,k,0) &= f_I(x,k), \qquad x\in\mathbb{R}^d,\ k\in B. \qquad (3.27)\end{aligned}$$

The Boltzmann equation has two nonlinearities:

- a quadratic nonlocal nonlinearity in the position variable caused by the self-consistent field E_{eff} in (3.25) and
- another quadratic nonlocal nonlineariy in the wave vector caused by the collision integral (3.24).

These nonlinearities make the mathematical analysis of the initial-boundary-value problem (3.23)–(3.27) very difficult. Only at the end of the 1980ies, an existence proof for global-in-time solutions and large data in the field-free case $E_{\text{eff}} = 0$ has been obtained by DiPerna and Lions employing the theory of renormalized solutions [26, 27]. We refer to the reviews [44, 63] for details and more references.

We give now some examples of collision operators.

Example 3.6 In semiconductor crystals, scattering of electrons occurs by lattice defects, phonons, and other carriers. We consider only the following important collision events:

- electron-phonon scattering,
- ionized impurity scattering, and

- carrier-carrier scattering.

Extensive treatments of scattering mechanisms in semiconductors can be found in, e.g., [11, Ch. 9], [37, Ch. 7], [45, Ch. 2], [61, Ch. 6], and in the textbooks [33, 58, 68].

Phonon scattering. At nonzero temperature, the atoms in the crystal lattice vibrate about their fixed equilibrium. These vibrations are quantized and the quantum of lattice vibrations is called *phonon.* We can distinguish so-called *acoustic phonons* and *optical phonons.* Acoustic phonons arise from displacements of lattice atoms in the same direction like sound waves. Optical phonos describe displacements in the wave vector and are able to interact strongly with light. Denoting by $\hbar\omega_\alpha$ the energy of a phonon, the phonon occupation number N_α is given by Bose-Einstein statistics,

$$N_\alpha = \frac{1}{e^{\hbar\omega_\alpha/k_BT} - 1},$$

where the index α refers to either "op" for optical phonons or "ac" for accoustic phonons. Notice that Bose-Einstein statistics can be used for indistinguishable particles not obeying the Pauli exclusion principle and therefore also for phonons (see [11, p.307ff.] for a derivation).

An electron in the Bloch state k' with conduction-band energy $E(k')$ before the collision with a phonon can change to the state k after the collision if

$$E(k') - E(k) = \pm\hbar\omega_\alpha, \tag{3.28}$$

where the plus sign refers to the phonon emission and the minus sign for phonon absorption. The acoustic phonon energy $\hbar\omega_\alpha$ is usually a function of the difference $k - k'$ of the wave vectors before and after a scattering event [8, Sec. 2]. The transition rate $s(x, k, k')$ is non-zero only if (3.28) is satisfied. Therefore,

$$s_\alpha(x, k, k') = \sigma_\alpha(x, k, k')[(1 + N_\alpha)\delta(E(k') - E(k) + \hbar\omega_\alpha) + N_\alpha\delta(E(k') - E(k) - \hbar\omega)],$$

where $\sigma_\alpha(x, k, k')$ is assumed to be symmetric in k and k' and δ is the delta distribution used in Section 2.4. This transition rate can be derived more generally from *Fermi's golden rule* (see [11, Sec. 4.4], [45, Sec. 1.7.1], or [65, Appendix C]). The collision operator reads according to (3.24)

$$(Q_\alpha(f))(x, k, t) = \int_B [s_\alpha(x, k', k)f'(1 - f) - s_\alpha(x, k, k')f(1 - f')]\, dk'. \tag{3.29}$$

The above expression for s_α shows that the scattering rates can be highly non-smooth.

Ionized impurity scattering. A doping atom in the semiconductor material donates either an electron or a hole, leaving behind an ionized charged impurity center. This fixed charge may attract or repulse an electron propagating through the crystal lattice. The interaction of carriers with neutral impurities is another scattering possibility but we do not consider it here. Since the scattering is elastic, the electron energy $E(k')$ after the collision is the same as the energy $E(k)$ before the interaction, and the transition rate is [57]

$$s_{\mathrm{imp}}(x,k,k') = \sigma_{\mathrm{imp}}(x,k,k')\delta(E(k')-E(k)),$$

where σ_{imp} is symmetric in k and k'. Again, this expression can be derived from Fermi's golden rule (see above). The symmetry of σ_{imp} and δ implies that

$$s_{\mathrm{imp}}(x,k',k)f'f - s_{\mathrm{imp}}(x,k,k')ff' = \sigma_{\mathrm{imp}}(x,k,k')\delta(E(k')-E(k))(f'f-ff') = 0,$$

and hence, the collision operator becomes

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

Carrier-carrier scattering. We only consider binary electron-electron interactions. Also binary electron-hole scattering or collective carrier-carrier collisions (i.e. interactions of carriers with oscillations in the carrier density; see [45, Sec. 2.10.2]) are possible but we do not consider these mechanisms here. The influence of electron-electron interactions on the carrier dynamics is more pronounced in degenerate semiconductors in which Fermi-Dirac statistics instead of Maxwell-Boltzmann statistics has to be used (see Section 2.4). The transition rate that carrier in the Bloch states k' and k_1' collide and scatter to the states k and k_1 is given by

$$s_{\mathrm{ee}}(x,k,k',k_1,k_1') = \sigma_{\mathrm{ee}}(x,k,k',k_1,k_1')\delta(E(k')+E(k_1')-E(k)-E(k_1)),$$

since the collisions are elastic. Therefore, the collision operator becomes [9, (2.7)]:

$$(Q_{\mathrm{ee}}(f))(x,k,t) = \int_{B^3} s_{\mathrm{ee}}(x,k,k',k_1,k_1')[f'f_1'(1-f)(1-f')-ff_1(1-f')(1-f_1')]dk'\,dk_1\,dk_1', \tag{3.30}$$

where $f=f(x,k,t)$, $f'=f(x,k',t)$, $f_1=f(x,k_1,t)$, and $f_1'=f(x,k_1',t)$. Notice that this operator has a nonlocal nonlinearity of fourth order.

Generally, the collision operator in (3.23) can be written as the sum of the various collision operators considered above:

$$Q(f) = Q_{\mathrm{op}}(f) + Q_{\mathrm{ac}}(f) + Q_{\mathrm{imp}}(f) + Q_{\mathrm{ee}}(f). \qquad \square$$

In the following we show some properties of the collision operator (3.24). The transition rate $s(x,k,k')$ is assumed to satisfy

$$\frac{s(x,k,k')}{s(x,k',k)} = \exp\frac{E(k)-E(k')}{k_BT} \qquad \text{for } x\in\mathbb{R}^d,\ k,k'\in B. \tag{3.31}$$

We explain this hypothesis. The so-called *principle of detailed balance* [3, 37] asserts that in thermal equilibrium, the *local* scattering probabilities vanish,

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

and that the equilibrium occupation number density is given by the Fermi-Dirac distribution,

$$f_{\text{eq}}(x,k) = \frac{1}{1+e^{u(k)}}, \qquad u(k) = \frac{E(k)-E_F}{k_B T}$$

(see (2.48)). A calculation shows that

$$\frac{s(x,k,k')}{s(x,k',k)} = \frac{f'_{\text{eq}}(1-f_{\text{eq}})}{f_{\text{eq}}(1-f'_{\text{eq}})} = e^{u(k)-u(k')},$$

which is (3.31).

The following result is due to Poupand [52] (also see the review [54]).

Theorem 3.7 *Let (3.31) hold for some function $E(k)$ and let $s(x,k,k') > 0$ for all $x \in \mathbb{R}^d$, $k, k' \in B$.*

(1) For all (regular) functions f there holds:

$$\int_B (Q(f))(x,k,t)\,dk = 0 \qquad \textit{for} \quad x \in \mathbb{R}^d,\ t > 0.$$

(2) For all functions $f(x,k,t) \in (0,1)$ and non-decreasing functions $\chi : \mathbb{R} \to \mathbb{R}$ there holds:

$$\begin{aligned} \int_B (Q(f))(x,k,t)\chi\left(\frac{f(x,k,t)}{1-f(x,k,t)}e^{E(k)/k_BT}\right) dk &\leq 0, \qquad (3.32)\\ \int_B (Q(f))(x,k,t)\chi\left(\frac{1-f(x,k,t)}{f(x,k,t)}e^{-E(k)/k_BT}\right) dk &\geq 0. \end{aligned}$$

(3) The kernel of Q only consists of Fermi-Dirac distributions, i.e. $Q(f) = 0$ if and only if, for some $-\infty \leq E_F \leq \infty$,

$$f(k) = \frac{1}{1+\exp((E(k)-E_F)/k_BT)}, \qquad k \in B. \qquad (3.33)$$

The theorem can be physically interpreted as follows. Let f be a solution to the Boltzmann equation (3.23)–(3.24). Then, by Theorem 3.7 (1),

$$\begin{aligned} \partial_t \int_{\mathbb{R}^d} n(x,t)\,dx &= \int_{\mathbb{R}^d}\int_B \partial_t f(x,k,t)\,dk\,dx \\ &= \int_{\mathbb{R}^d}\int_B (-\text{div}_x(v(k)f) + \frac{q}{h}\text{div}_k(E_{\text{eff}}f) + Q(f))\,dk\,dx \\ &= 0, \end{aligned}$$

where we have used the divergence theorem. This implies that the total number of electrons is conserved in time:

$$\int_{\mathbb{R}^d} n(x,t)\,dx = \int_{\mathbb{R}^d} n(x,0)\,dx \qquad \text{for all} \quad t>0.$$

Physically, this is reasonable: collisions neither destroy nor generate particles. The statement of Theorem 3.7 (2) is also called *H-theorem.*

Proof of Theorem 3.7: (1) Changing k and k' in the second integral gives

$$\begin{aligned} \int_B Q(f)\,dk &= \int_{B^2} s(k',k)f'(1-f)\,dk'\,dk - \int_{B^2} s(k,k')f(1-f')\,dk'\,dk \\ &= \int_{B^2} s(k',k)f'(1-f)\,dk'\,dk - \int_{B^2} s(k',k)f'(1-f)\,dk\,dk' \\ &= 0. \end{aligned}$$

(2) We show only the second inequality. The proof of the first one is similar. Set

$$M(k) = e^{-E(k)/k_BT}, \qquad F(k) = \frac{1-f(k)}{f(k)}M(k). \tag{3.34}$$

The function $M(k)$ is called *Maxwellian.* Then assumption (3.31) is equivalent to

$$\frac{s(k',k)}{M(k)} = \frac{s(k,k')}{M(k')}, \tag{3.35}$$

and we obtain, with the notations $F=F(k)$, $F'=F(k')$, $M=M(k)$, $M'=M(k')$,

$$\begin{aligned} \int_B Q(f)\chi(F)\,dk &= \int_{B^2} s(k,k')\left[\frac{M}{M'}f'(1-f)\chi(F) - f(1-f')\chi(F)\right]dk'\,dk \\ &= \int_{B^2} \frac{s(k,k')}{M'}ff'(F-F')\chi(F)\,dk'\,dk, \end{aligned} \tag{3.36}$$

and, after changing k and k' and then using (3.35),

$$\begin{aligned} \int_B Q(f)\chi(F)\,dk &= \int_{B^2} \frac{s(k',k)}{M}f'f(F'-F)\chi(F')\,dk\,dk' \\ &= \int_{B^2} \frac{s(k,k')}{M'}f'f(F'-F)\chi(F')\,dk'\,dk. \end{aligned} \tag{3.37}$$

Adding (3.36) and (3.37) leads to

$$\begin{aligned} \int_B Q(f)\chi(F)\,dk &= \frac{1}{2}\int_{B^2} \frac{s(k,k')}{M'}ff'(F-F')(\chi(F)-\chi(F'))\,dk'\,dk \\ &\geq 0, \end{aligned} \tag{3.38}$$

since χ is non-decreasing and all the other factors are non-negative.

(3) We see from (2), with $\chi(x) = x$, taking into account (3.38), that $Q(f) = 0$ is equivalent to

$$ff'(F - F')^2 = 0 \qquad \text{for almost all } k, k' \in B.$$

This implies $f \equiv 0$ or $F = F'$ almost everywhere. The latter equation is equivalent to

$$\frac{1 - f(k)}{f(k)} M(k) = \frac{1 - f(k')}{f(k')} M(k') \qquad \text{for almost all } k, k' \in B.$$

We infer that both sides are constant and denote this constant by $\exp(-E_F/k_BT)$ with $E_F \in \mathbb{R}$. Notice that the constant is positive since $0 \leq f \leq 1$, except if $f \equiv 1$. But then

$$\frac{1 - f(k)}{f(k)} = \frac{e^{-E_F/k_BT}}{M(k)} = e^{(E(k)-E_F)/k_BT}$$

and solving for $f(k)$ yields (3.33). Finally, choosing $E_F = \pm\infty$ leads to the other two possibilities $F \equiv 0$ of $F \equiv 1$. $\square$

In the literature, two approximations of the collision operator are frequently used [49, p. 33ff]:

- the *low density approximation*

$$Q_0(f))(x, k, t) = \int_B \sigma(x, k', k)(Mf' - M'f)\, dk', \tag{3.39}$$

  where $\sigma(x, k', k) = s(x, k', k)/M(k)$ is called *collision cross-section*;

- the *relaxation time approximation*

$$(Q_\tau(f))(x, k, t) = -\frac{1}{\tau(x, k)} \left(f(x, k, t) - M(k)n(x)\right), \tag{3.40}$$

  where

$$\tau(x, k) = \left(\int_B s(x, k, k')\, dk'\right)^{-1}$$

  is called the *relaxation time* describing the average time between two consecutive scattering events at (x, k), and $n(x)$ is some given density such that

$$n(x) = \int_B f_I(x, k)\, dk.$$

These approximations can be derived as follows. The low-density collision operator is obtained by assuming that the distribution function is small:

$$0 \leq f(x, k, t) \ll 1.$$

Then $1 - f(x) \approx 1$ and we can approximate, using (3.35),

$$\begin{aligned} Q(f)(k) &= \int_B [s(k',k)f' - s(k,k')f]\,dk' \\ &= \int_B \frac{s(k',k)}{M(k)}[M(k)f' - M(k')f]\,dk', \end{aligned}$$

which is (3.39). Notice that (3.35) implies that $\sigma(x,k',k)$ is symmetric in k and k'.

When the initial distribution is close to a multiple of $M(k)$ it is reasonable to approximate f' in (3.39) by $n(x)M(k')$. Thus, (3.39) yields

$$\begin{aligned} (Q(f))(k) &= \int_B \sigma(k',k)(nMM' - M'f)\,dk' \\ &= \int_B s(k',k)\frac{M'}{M}(nM - f)\,dk' \\ &= \int_B s(k,k')(nM - f)\,dk' \\ &= (nM - f)\int_B s(k,k')\,dk', \end{aligned}$$

which is equal to (3.40).

Why is $\tau(x,k)$ called relaxation time? In order to see this, we argue as in [49, p. 34f.] and consider the Boltzmann equation with (3.40) along the trajectories $(x(t),k(t))$

$$\frac{df}{dt} = -\frac{1}{\tau}(f - nM), \qquad t > 0,$$

where $x(t), k(t)$ solves $\dot{x} = v(k)$, $\dot{k} = -qE_{\text{eff}}/\hbar$, $x(0) = x_0, k(0) = k_0$ for constant relaxation time τ. Multiplying this differential equation with $f - nM$ gives

$$\frac{1}{2}\frac{d}{dt}(f - nM)^2 = (f - nM)\frac{df}{dt} = -\frac{1}{\tau}(f - nM)^2.$$

Separation of variables leads to

$$f(x(t),k(t),t) - n(x(t))M(k(t)) = ce^{-t/\tau} \to 0 \qquad \text{as} \quad t \to \infty,$$

where $c \in \mathbb{R}$ is an integration constant. This means that the distribution function relaxes to the equilibrium density nM, starting from f_I, along the trajectory after a time of order τ.

Notice that the low-density and the relaxation-time approximations (3.39) and (3.40) coincide if the collision cross-section only depends on the position variable and if the Maxwellians are normalized such that $\int_B M\,dk = 1$. Indeed, we obtain

$$\begin{aligned} (Q_0(f))(x,k,t) &= \int_B \sigma(x)(Mf' - M'f)\,dk' \\ &= \sigma(x)M(k)\int_B f'\,dk' - \sigma(x)f(x)\int_B M'\,dk' \\ &= -\sigma(x)(f(x,k,t) - M(k)n(x,t)), \end{aligned}$$

which corresponds to (3.40) with $\tau(x) = 1/\sigma(k)$ and time-dependent density $n(x,t)$.

Lemma 3.8 *Let $\sigma(x,k',k)$ in (3.39) be symmetric in k and k'. Then the kernel of the low-density collision operator consists exactly of the Maxwellians, i.e. $Q_0(f) = 0$ if and only if*

$$f(x,k) = n(x)M(k) = n(x)e^{-E(k)/k_BT},$$

where $n(x)$ is any (non-negative) function.

Proof: The proof is very similar to the proof of Theorem 3.7, parts (2) and (3), using $\chi(x) = x$ and $F(k) = M(k)/f(k)$. Indeed, an analogous computation leads to $F(k) = F(k')$ or

$$\frac{M(k)}{f(k)} = \frac{M(k')}{f(k')} \qquad \text{for almost all } k, k' \in B.$$

This implies $M(k)/f(k) = \text{const.} = 1/n(x)$. $\square$

Remark 3.9 In the parabolic band approximation

$$E(k) = \frac{\hbar^2}{2m^*}|k-k_0|^2$$

we can characterize the kernel of the low-density operator by the family of functions

$$M_{\varrho,u,T}(v) = \varrho\left(\frac{m^*}{2\pi k_BT}\right)^{d/2}\exp\left(-\frac{m^*|v-u|^2}{2k_BT}\right) \tag{3.41}$$

with $\varrho, T > 0$ and $u \in \mathbb{R}$. Indeed, the velocity is

$$\frac{1}{\hbar}\nabla_k E(k) = \frac{\hbar}{m^*}(k-k_0) =: v-u$$

such that

$$M(v) = \exp\left(-\frac{E(k)}{k_BT}\right) = \exp\left(-\frac{m^*|v-u|^2}{2k_BT}\right).$$

Setting

$$n(x) = \varrho\left(\frac{m^*}{2\pi k_BT}\right)^{d/2}$$

then gives (3.41). The constant ϱ represents a density, u a velocity, and T the temperature. $\square$

The semi-classical Boltzmann equation (3.23) is fundamental in deriving simpler macroscopic models for semiconductors (see Chapter 3). It is the basis equation in (classical) semiconductor modeling and usually, other models are validated by numerical comparisons with the Boltzmann equation. Nevertheless, it is important to understand its limitations (see [45, Sec. 37]):

- The Boltzmann equation is only a single-particle description of many-particle systems of charged carriers. In particular, correlations between carriers are neglected.
- Quantum mechanical phenomena are only modeled in a semi-classical way. Electrons are considered as particles which obey Newton's laws.
- Collisions are assumed to be binary and to be instantaneous in time and local in space.

We can estimate the range of validity by using Heisenberg's uncertainty principles

$$\triangle p \triangle x \geq \hbar \qquad \text{and} \qquad \triangle E \triangle t \geq \hbar, \tag{3.42}$$

which means that it is impossible to determine both momentum and space at the same time and that the energy of a particle can be defined only if it stays in the same state for some time [45, p. 152]. For instance, writing

$$\frac{p^2}{2m^*} = E$$

for the momentum p and assuming that the energy spread is of the order of the thermal energy k_BT, we obtain

$$\triangle p = \sqrt{2m^* \triangle E} \sim \sqrt{2m^* k_B T}.$$

Introduce the *de Broglie length*

$$\lambda_B = \frac{h}{\sqrt{2m^* k_B T}}$$

which is the wave length of an electron with thermal energy. Then the first inequality in (3.42) leads to the requirement

$$\triangle x \geq \frac{\hbar}{\triangle p} \sim \frac{\lambda_B}{2\pi}.$$

Therefore, when treating electrons as particles, they cannot be localized sharper than $\triangle x$ which is of the order of λ_B. At room temperature and with the effective mass $m^* = 0.98 m_e$ (see Table 2.3) of silicon, we obtain $\triangle x \sim 8$ nm.

Taking $\triangle t$ to be the time between collisions and again assuming that $\triangle E \sim k_B T$, the second inequality in (3.42) gives

$$\triangle t \geq \frac{\hbar}{\triangle E} \sim \frac{\hbar}{k_B T}.$$

Supposing further that the electron has the velocity $v = p/m^* \sim \sqrt{2k_BT/m^*}$, corresponding to the thermal energy, the distance between two collisions is

$$\triangle L = v \triangle t \geq \sqrt{\frac{2k_BT}{m^*}} \frac{\hbar}{2\pi k_B T} = \frac{\lambda_B}{\pi}.$$

Thus, the *mean free path* $\triangle L$ should be (much) larger than the de Broglie length $\lambda_B \sim$ 8 nm.

Finally, in Figure 3.2 we present a summary of the models derived in this chapter and the relations between them.

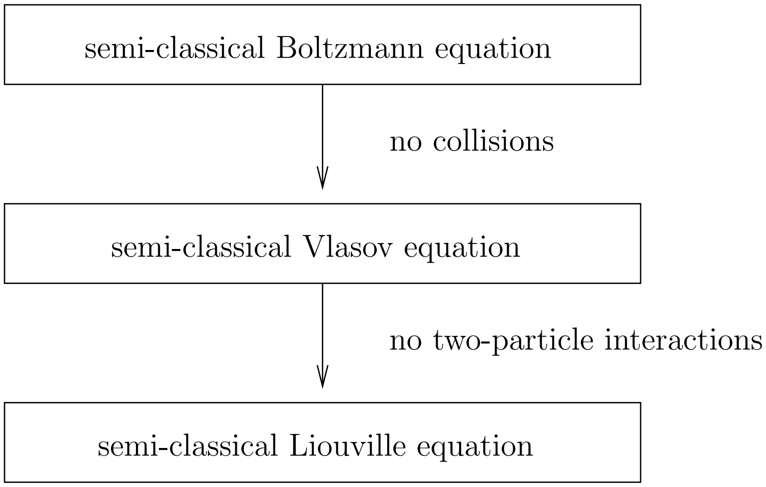

Figure 3.2: Relations between the models of Sections 3.1–3.3.

3.4 The bipolar Boltzmann model

So far we have only considered the transport of electrons in the conduction band. However, also holes in the valence band contribute to the carrier flow in semiconductors (see Section 2.2). It is possible that an electron moves from the valence band to the conduction band, leaving a hole in the valence band behind it. This process is called the *generation* of an electron-hole pair (see Figure 3.3). The electron has to overcome the energy gap, which is of the order of 1eV; on the other hand, the thermal energy of an electron is only of the order of $k_B T \sim 0.026$eV at room temperature. Therefore, a lot of absorption energy is necessary for such processes. The inverse process of an electron moving from the conduction to the valence band, occupying an empty state, is termed the *recombination* of an electron-hole pair. For such an event, energy is emitted. The basic mechanisms for generation-recombination processes are

- Auger/impact ionization generation-recombination,
- radiative generation-recombination,
- thermal generation-recombination.

An *Auger process* is defined as an electron-hole recombination followed by a transfer of energy to a free carrier which is excited to a state of higher energy. The inverse Auger process, i.e. the generation of an electron-hole pair, is called *impact ionization*. The energy for the pair generation comes from the collision of a high-energetic free carrier with the lattice or from electron-electron or hole-hole collisions.

When an electron from the conduction band recombines with a hole from the valence band and emittes a photon, we call this process *radiative* recombination. The energy of the photon is equal to the bandgap energy. Radiative generation occurs when a photon with energy larger than or equal to the gap energy is absorbed.

A third source of energy is coming from lattice vibrations or phonons. Thus, *thermal* recombination or generation arises from phonon emission or absorption, respectively.

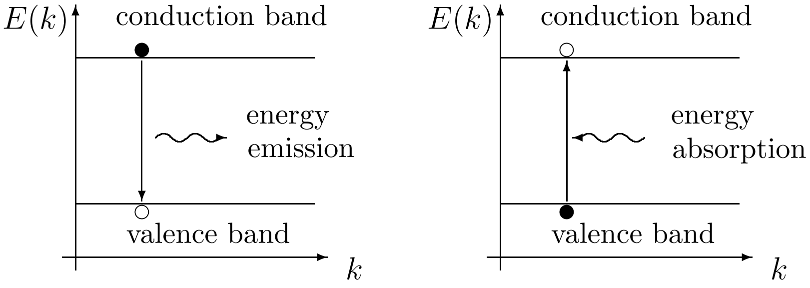

Figure 3.3: Recombination (left) and generation (right) of an electron-hole pair.

The recombination-generation operators can be derived as the collision operator of Section 3.3 from phenomenological considerations. The generation of an electron in the state k and a hole in the state k' is possible if both states are not occupied, and its rate is given by

$$g(x, k', k)(1 - f_n)(1 - f_p'),$$

where $g(x, k', k) \geq 0$ is the generation rate, $f_n = f_n(x, k, t)$ the electron distribution function and $f_p' = f_p(x, k', t)$ the hole distribution function. The rate of recombination of an electron at state k and a hole at state k' is

$$r(x, k, k')f_n f_p',$$

where $r(x, k, k') \geq 0$. The net rate is the difference of generation and recombination rates:

$$g(x, k', k)(1 - f_n)(1 - f_p') - r(x, k, k')f_n f_p',$$

and the recombination-generation operator for electrons in the conduction band is the integral over all states k' (see [49, Sec. 1.6]):

$$(I_n(f_n, f_p))(x, k, t) = \int_B [g(x, k', k)(1 - f_n)(1 - f'_p) - r(x, k, k') f_n f'_p] \, dk'. \tag{3.43}$$

In a similar way, the recombination-generation operator for holes in the valence band can be written as

$$(I_p(f_n, f_p))(x, k, t) = \int_B [g(x, k, k')(1 - f'_n)(1 - f_p) - r(x, k', k) f'_n f_p] \, dk'. \tag{3.44}$$

The recombination and generation rates are related by the equation

$$r(x, k, k') = \exp\left(\frac{E_n(k) - E_p(k')}{k_B T}\right) g(x, k', k) \tag{3.45}$$

which can be derived from the principle of detailed balance as in Section 3.3 (see (3.39)). This principle holds here since recombination and generation balance in thermal equilibrium.

The operators (3.43) and (3.44) are added to the electron and hole collision operators. Then, the evolution of the distribution functions f_n and f_p is given by the system of Boltzmann equations

$$\begin{aligned} \partial_t f_n + v_n(k) \cdot \nabla_x f_n - \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k f_n &= Q_n(f_n) + I_n(f_n, f_p) & (3.46) \\ \partial_t f_p + v_p(k) \cdot \nabla_x f_p + \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k f_p &= Q_p(f_p) + I_p(f_n, f_p), & (3.47) \end{aligned}$$

where

$$v_n(k) = \frac{1}{\hbar} \nabla_k E_n(k), \qquad v_p(k) = \frac{1}{\hbar} \nabla_k E_p(k)$$

and E_n and E_p are the conduction and valance band energies, respectively. Denoting by

$$n(x, t) = \int_B f_n(x, k, t) \, dk, \qquad p(x, t) = \int_B f_p(x, k, t) \, dk$$

the electron and hole densities, respectively, the effective-field equation (3.25) becomes

$$E_{\text{eff}}(x, t) = E_{\text{ext}}(x, t) + \int_{\mathbb{R}^d} (n(y, t) - p(y, t)) E_{\text{int}}(x, y) \, dy, \tag{3.48}$$

since electrons and holes have charges with opposite sign. The equations (3.46)–(3.48), together with the collision operators (3.24) and (3.43)–(3.44) are called the *semi-classical bipolar Boltzmann model.*

The bipolar model has an additional nonlinearity due to the coupling between the carrier densities n and p through (3.48). In particular, the total number of each type of particles is not conserved anymore since recombination-generation effects can take place.

However, the total space charge $n - p - C$ is conserved if the doping atoms are immobile, i.e., C is a function of the space variable x only. Indeed, taking the difference of the Boltzmann equations (3.46) and (3.47) and integrating over $(x, k) \in \mathbb{R}^d \times B$ yields

$$\partial_t \int_{\mathbb{R}^d} (n - p - C(x))\, dx = \int_{\mathbb{R}^d} \int_B (I_n(f_n, f_p) - I_p(f_n, f_P))\, dk\, dx = 0,$$

by arguing as in the proof of Theorem 3.7 (1).

Finally, we discuss two special cases. In the low density approximation $f_n, f_p \ll 1$ we can write the recombination-generation operators as

$$(I_n(f_n, f_p))(x, k, t) \;=\; \int_B g(x, k', k) \left(1 - e^{(E_n(k) - E_p(k'))/k_B T} f_n f_p'\right) dk', \qquad (3.49)$$

$$(I_p(f_n, f_p))(x, k, t) \;=\; \int_B g(x, k, k') \left(1 - e^{(E_n(k') - E_p(k))/k_B T} f_n' f_p\right) dk', \qquad (3.50)$$

using (3.45). In the case of Coulomb forces in $\mathbb{R}^3$, the effective field is given by

$$E_{\text{eff}}(x, t) = \frac{1}{4\pi\varepsilon_s} \int_{\mathbb{R}^3} \varrho(y, t) \frac{x - y}{|x - y|^3}\, dy,$$

where the total space charge ϱ is the sum of the electron density n, the hole density p, and the densities N_D, N_A of the implanted positively charged donor ions and the negatively charged acceptor ions, respectively, with which the semiconductor is doped (see Section 2.4), weighted by the corresponding charges $+q$ or $-q$:

$$\varrho = q(-n + p - N_A + N_D).$$

Defining the electrostatic potential $E_{\text{eff}} = -\nabla V$ and the doping profile $C = N_D - N_A$, we can replace the effective-field equation (3.48) by the Poisson equation

$$\varepsilon_s \Delta V = q(n - p - C) \qquad \text{in } \mathbb{R}^3. \qquad (3.51)$$

The Boltzmann equations (3.43)–(3.44) and the Poisson equation (3.51) constitute the so-called *bipolar Boltzmann-Poisson system.*

4 Classical Fluid Models

The mathematical and numerical solution of the Boltzmann equation of Chapter 3 is very difficult due to the complex structure of the collision operator and the high number of independent variables (three position plus three wave vector plus one time variable). In this chapter we derive simpler models with only four independent variables (three space and one time variable) from the Boltzmann equation. The main feature of these models is that they describe the evolution of averaged quantities, like the carrier density or the current density, rather than a distribution function. Thus the carrier ensemble is considered as a "fluid" consisting of charged particles.

4.1 Derivation of the drift-diffusion equations

The drift-diffusion equations are one of the simplest semiconductor models. It has been first derived by van Roosbroeck in 1950 [64]. We derive this model by employing the so-called *Hilbert method* which consists in expanding the distribution function in powers of the scaled mean free path. This procedure has been analyzed in [51]. We follow the presentation of [49, Sec. 22].

The starting point of the derivation is the bipolar Boltzmann model of Section 3.4 in three dimensions,

$$\begin{aligned} \partial_t f_n + v_n(k) \cdot \nabla_x f_n - \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k f_n &= Q_n(f_n) + I_n(f_n, f_p), && (4.1) \\ \partial_t f_p + v_p(k) \cdot \nabla_x f_p + \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k f_p &= Q_p(f_p) + I_p(f_n, f_p), && (4.2) \end{aligned}$$

together with the field equation

$$-\varepsilon_s \operatorname{div} E_{\text{eff}} = q(n - p - C(x)), \qquad x, k \in \mathbb{R}^3.$$

We assume low-density collision operators,

$$(Q_j(f_j))(x,k,t) = \int_B \sigma_j(x,k,k') \left(e^{-E_j(k)/k_B T} f_j' - e^{-E_j(k')/k_B T} f_j \right) dk', \tag{4.3}$$

where $j \in \{n, p\}$ (see (3.39)) and the low-density recombination-generation operators (3.49)–(3.50),

$$\begin{aligned} (I_n(f_n, f_p))(x,k,t) &= \int_B g(x,k',k) \left(1 - e^{(E_n(k) - E_p(k'))/k_B T} \right) f_n f_p' \, dk', && (4.4) \\ (I_p(f_n, f_p))(x,k,t) &= \int_B g(x,k,k') \left(1 - e^{(E_n(k') - E_p(k))/k_B T} \right) f_n f_p' \, dk'. && (4.5) \end{aligned}$$

As in Chapter 3, f_j' means evaluation at k', i.e., $f_j' = f_j(x,k',t)$, $j \in \{n,p\}$, and $B \subset \mathbb{R}^3$ denotes the Brillouin zone. Furthermore, we assume the parabolic band approximation

$$E_n(k) = E_c + \frac{\hbar^2}{2m_e^*}|k|^2, \qquad E_p(k) = E_v - \frac{\hbar^2}{2m_h^*}|k|^2,$$

where E_c denotes the conduction band minimum, E_v the valence band maximum, and m_e^* and m_h^* are the effective masses of electrons and holes, respectively. In particular, we can replace B by $\mathbb{R}^3$ in the above integrals. The velocities in (4.1)–(4.2) are given by

$$v_n(k) = \frac{1}{\hbar}\nabla_k E_n(k) = \frac{\hbar k}{m_e^*}, \qquad v_p(k) = -\frac{1}{\hbar}\nabla_k E_p(k) = \frac{\hbar k}{m_h^*}.$$

The drift-diffusion equations are derived under the (main) assumption that collisions occur on a much shorter time scale than recombination-generation events. Indeed, typically the time scale of collisions is 10^{-12}s whereas a typical time for recombination-generation effects is 10^{-9}s [49, p. 86]. In order to make this statement more precise, we scale the Boltzmann equations, i.e., we choose new variables which make the equation dimension free. We choose the following reference parameter:

- Define the reference velocity $v = \sqrt{k_B T/m_e^*}$. Behind this setting there are two assumptions: The effective masses of electrons and holes are of the same order and the thermal energy $k_B T$ is of the same order as the kinetic energy $m_e^* v^2/2$.

- Let τ_c, τ_R be characteristic times of collision and recombination-generation events. Then $\lambda_c = \tau_c v$ and $\lambda_R = \tau_R v$ are the so-called *mean free paths* between two consecutive scattering and recombination-generation events. Thus we have to scale

$$\begin{aligned} Q_n &= \frac{1}{\tau_c} Q_{n,s}, & Q_p &= \frac{1}{\tau_c} Q_{p,s}, \\ I_n &= \frac{1}{\tau_R} I_{n,s}, & I_p &= \frac{1}{\tau_R} I_{p,s}. \end{aligned}$$

- A reference length λ_0 is given by the geometric average of λ_c and λ_R:

$$\lambda_0 = \sqrt{\lambda_R \lambda_c}.$$

  Then $x = \lambda_0 x_s$.

- We use the reference time τ_R, the reference wave vector $m_e^* v/\hbar$ and the reference field strength $k_B T/q\lambda_0$:

$$t = \tau_R t_s, \qquad k = \frac{m_e^* v}{\hbar} k_s, \qquad E_{\text{eff}} = \frac{k_B T}{\lambda_0 q} E_{\text{eff},s}. \tag{4.6}$$

Our main assumption is

$$\tau_c \ll \tau_R.$$

Then the parameter $\alpha^2 = \lambda_c/\lambda_R = \tau_c/\tau_R$ satisfies $\alpha \ll 1$. With the above scaling we can rewrite (4.1) as

$$\frac{1}{\tau_R}\partial_{t_s} f_n + \frac{v}{\lambda_0} k_s \cdot \nabla_{x_s} f_n - \frac{k_B T}{\lambda_0 m_e^* v} E_{\text{eff},s} \cdot \nabla_{k_s} f_n = \frac{1}{\tau_c} Q_{n,s}(f_n) + \frac{1}{\tau_R} I_{n,s}(f_n, f_p).$$

Multiplying this equation by $\tau_c = \lambda_c/v$, using $\alpha = \lambda_c/\lambda_0$ and $k_B T = m_e^* v^2$, and omitting the index s, we obtain the *scaled* Boltzmann equation

$$\alpha^2 \partial_t f_n + \alpha \left(k \cdot \nabla_x f_n - E_{\text{eff}} \cdot \nabla_k f_n\right) = Q_n(f_n) + \alpha^2 I_n(f_n, f_p). \tag{4.7}$$

In a similar way, we have

$$\alpha^2 \partial_t f_p + \alpha \left(\frac{m_e^*}{m_h^*} k \cdot \nabla_x f_p + E_{\text{eff}} \cdot \nabla_k f_p\right) = Q_p(f_p) + \alpha^2 I_p(f_n, f_p). \tag{4.8}$$

The scaled operators $Q_{j,s}$ and $I_{j,s}$ have the same form as (4.3)–(4.5) but

$$e^{-E_n(k)/k_B T} \qquad \text{and} \qquad e^{-E_p(k)/k_B T}$$

have to be replaced by

$$e^{-E_c/k_B T - |k|^2/2} \qquad \text{and} \qquad e^{-E_v/k_B T - (m_e^*/m_h^*)|k|^2/2},$$

respectively, and the rates σ_n, σ_p, and g are multiplied by $(m_e^* v/\hbar)^d$.

We want to study the scaled Boltzmann equations for "‘small" α. First we analyze the collision operators Q_n and Q_p. We proceed as in [53]. It is convenient to write them in the form

$$Q_f(f_j) = \int_{\mathbb{R}^3} s_j(x, k, k')(M_j(k) f_j' - M_j(k') f_j)\, dk', \qquad j \in \{n, p\}, \tag{4.9}$$

where

$$\begin{aligned} s_n(x, k, k') &= N_n \sigma_n(x, k, k') e^{-E_c/k_B T}, \\ s_p(x, k, k') &= N_p \sigma_p(x, k, k') e^{-E_v/k_B T}, \end{aligned}$$

and

$$M_n(k) = \frac{1}{N_n} e^{-|k|^2/2}, \qquad M_p(k) = \frac{1}{N_n} e^{-(m_e^*/m_h^*)|k|^2/2}$$

are the scaled Maxwellians. The constants

$$N_n = (2\pi)^{3/2}, \qquad N_p = (2\pi)^{3/2} \left(\frac{m_h^*}{m_e^*}\right)^{3/2}$$

are chosen in such a way that the Maxwellians are normalized, i.e.

$$\int_{\mathbb{R}^3} M_n(k)\, dk = \int_{\mathbb{R}^3} M_p(k)\, dk = 1.$$

Here we have used that fact that

$$\int_{\mathbb{R}^d} e^{-|k|^2/2}\, dk = (2\pi)^{d/2}.$$

For the following analysis we introduce the functions

$$\lambda_j(k) = \int_{\mathbb{R}^3} s_j(x,k,k')M_j(k')\,dk', \qquad k \in \mathbb{R}^3, \quad j \in \{n,p\}, \tag{4.10}$$

for some fixed $x \in \mathbb{R}^d$, and the Banach spaces

$$\begin{aligned} X_j &= \{f : \mathbb{R}^3 \to \mathbb{R} \text{ measurable:} \quad \|f\|_{X_j} < \infty\}, \\ Y_j &= \{f : \mathbb{R}^3 \to \mathbb{R} \text{ measurable:} \quad \|f\|_{Y_j} < \infty\} \end{aligned}$$

with associated norms

$$\begin{aligned} \|f\|_{X_j} &= \left(\int_{\mathbb{R}^3} f(k)^2 \lambda_j(k) M_j(k)^{-1}\,dk\right)^{1/2}, \\ \|f\|_{Y_j} &= \left(\int_{\mathbb{R}^3} f(k)^2 \lambda_j(k)^{-1} M_j(k)^{-1}\,dk\right)^{1/2}. \end{aligned}$$

Lemma 4.1 *Let $j \in \{n,p\}$ and let $s_j > 0$ be symmetric in k and k'. Then the kernel of Q_j, $N(Q_j) = \{f \in X_j : Q_j(f) = 0\}$, only consists of Maxwellians:*

$$N(Q_j) = \{\sigma M_j : \sigma = \sigma(x) \in \mathbb{R}\}.$$

Proof: The proof is similar to the proofs of Theorem 3.7 and Lemma 3.8. In the following we omit the index j. First we symmetrize the collision operator by setting $f_s = (\lambda/M)^{1/2} f$ and $Q_s(f_s) = (\lambda M)^{-1/2} Q(f)$. Then

$$\begin{aligned} Q_s(f_s) &= (\lambda M)^{-1/2}\left(M\int_{\mathbb{R}^3} s(x,k,k')f'\,dk' - \lambda f\right) \\ &= \int_{\mathbb{R}^3} s(x,k,k')\left(\frac{MM'}{\lambda\lambda'}\right)^{1/2} f'_s\,dk' - f_s, \end{aligned}$$

where $M' = M(k')$ and $\lambda' = \lambda(k')$. Since s is symmetric in k and k' by assumption, the operator $Q_s : L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^3)$ is self-adjoint (see Section 2.1).

Now we analyze the kernel of Q_s. By definition of λ we have

$$\int_{\mathbb{R}^3} s(x,k,k')\frac{M'}{\lambda}\,dk' = 1$$

and therefore

$$\begin{aligned}
-\int_{\mathbb{R}^3} Q_s(f_s)f_s\,dk &= \int_{\mathbb{R}^3}\left[f_s^2 - \int_{\mathbb{R}^3} s(x,k,k')\left(\frac{MM'}{\lambda\lambda'}\right)^{1/2} f_s' f_s\,dk'\right] dk \\
&= \frac{1}{2}\int_{\mathbb{R}^3}\left(\int_{\mathbb{R}^3} s(x,k,k')\frac{M'}{\lambda}\,dk'\right) f_s^2\,dk \\
&\quad +\frac{1}{2}\int_{\mathbb{R}^3}\left(\int_{\mathbb{R}^3} s(x,k,k')\frac{M}{\lambda'}\,dk\right)(f_s')^2\,dk' \\
&\qquad -\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} s(x,k,k')\left(\frac{MM'}{\lambda\lambda'}\right)^{1/2} f_s' f_s\,dk'\,dk \\
&= \frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} s(x,k,k')MM'\left(\frac{f_s}{\sqrt{\lambda M}} - \frac{f_s'}{\sqrt{\lambda' M'}}\right)^2 dk\,dk' \\
&\geq 0. \qquad (4.11)
\end{aligned}$$

Thus, $Q_s(f_s) = 0$ implies that

$$\frac{f_s}{\sqrt{\lambda M}} = \frac{f_s'}{\sqrt{\lambda' M'}} \qquad \text{for } k, k' \in \mathbb{R}^3.$$

We conclude that both sides must be constant, $f_s/\sqrt{\lambda M} = \sigma = \text{const.}$, where $\sigma = \sigma(x)$ is a parameter. In the original variables, $Q(f) = 0$ implies $f = \sqrt{M/\lambda} f_s = \sqrt{M/\lambda} \cdot \sigma\sqrt{\lambda M} = \sigma M$ for all $\sigma \in \mathbb{R}$. Conversely, if $f = \sigma M$ for some $\sigma \in \mathbb{R}$, the formulation (4.9) immediately gives $Q(f) = 0$. This proves the lemma. □

Lemma 4.2 *Let the assumptions of Lemma 4.1 hold. Then*

(1) The equation $Q_j(f) = g \in Y_j$ has a solution $f \in X_j$ if and only if

$$\int_{\mathbb{R}^3} g(k)\,dk = 0.$$

In this situation, any solution of $Q_j(f) = g$ can be written as $f + \sigma M_j$, where $\sigma = \sigma(x)$ is a parameter.

(2) The solution $f \in X_j$ of $Q_j(f) = g$ is unique if the orthonogality relation

$$\int_{\mathbb{R}^3} f(k)\lambda_j(k)\,dk = 0 \qquad (4.12)$$

is satisfied.

This lemma is a consequence of the *Fredholm alternative* which we prove first.

Lemma 4.3 (Fredholm alternative)
Let X be a Hilbert space with scalar product $(\cdot,\cdot)$ and $Q : X \to X$ a linear, continuous and closed operator (i.e. $R(Q) = \{g \in X : \exists\, f \in X : Q(f) = g\}$ is closed). Then

$$Q(f) = g \text{ has a solution} \quad \Longleftrightarrow \quad g \in N(Q^*)^\perp.$$

Proof: We show first that $N(Q^*) = R(Q)^\perp$. Indeed,

$$\begin{aligned} f \in N(Q^*) \quad &\Longleftrightarrow \quad Q^*(f) = 0 \quad \Longleftrightarrow \quad (g, Q^*(f)) = 0 \quad \text{for all } g \in X \\ &\Longleftrightarrow \quad (Q(g), f) = 0 \text{ for all } g \in X \quad \Longleftrightarrow \quad f \perp R(Q) \\ &\Longleftrightarrow \quad f \in R(Q)^\perp. \end{aligned}$$

Thus, since Q is closed,

$$R(Q) = \overline{R(Q)} = \overline{R(Q)}^{\perp\perp} = N(Q^*)^\perp.$$

This identity is equivalent to the assertion. □

Proof of Lemma 4.2: Again, we omit the index j.

(1) It is possible to show that the operator $Q_s : L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^3)$ is linear, continuous, and closed. Moreover, we have shown in the proof of Lemma 4.1 that Q_s is self-adjoint. We conclude from Lemma 4.3 that $Q_s(f_s) = g_s$ has a solution if and only if $g_s \in N(Q_s^*)^\perp = N(Q_s)^\perp$ or

$$\int_{\mathbb{R}^3} g_s h \, dk = 0 \qquad \text{for all } h \in N(Q_s).$$

By Lemma 4.1, the kernel of Q_s is spanned by $\sqrt{\lambda M}$, such that this equivalent to

$$0 = \int_{\mathbb{R}^3} g_s \sqrt{\lambda M} \, dk = \int_{\mathbb{R}^d} g \, dk,$$

since in the original variables we have

$$g = Q(f) = \sqrt{\lambda M} Q_s(f_s) = \sqrt{\lambda M} g_s.$$

Let f_1 and f_2 be two solutions of $Q(f) = g$. Then, since Q is linear, $Q(f_1 - f_2) = Q(f_1) - Q(f_2) = 0$ and $f_1 - f_2 \in N(Q)$. This shows that $f_1 = f_2 + \sigma M$ for some $\sigma \in \mathbb{R}$.

(2) We only give a sketch of the proof and refer to [53] for details. It is possible to show that the operator $-Q$ is coercive in the following sense:

$$\int_{\mathbb{R}^3} (-Q(f)) f M^{-1} \, dk \geq c \|f\|_X^2 \tag{4.13}$$

for some $c > 0$ and for all $f \in X$ satisfying (4.12). Clearly, this implies that Q is one-to-one on the subset of functions satisfying (4.12), and the uniqueness property is shown. In order to prove (4.13) one needs to show that $\mathrm{Id} + Q$ is a Hilbert-Schmidt operator and to use general properties of these operators (see [53]). □

Lemma 4.4 *Let the assumptions of Lemma 4.1 hold and assume that*

$$s_j(x, Ak, Ak') = s_j(x, k, k') \qquad \text{for } x, k, k' \in \mathbb{R},\ j \in \{n, p\}, \tag{4.14}$$

and for all isometric matrices $A \in \mathbb{R}^{3\times 3}$. *Then the equations*

$$(Q_n(h_{n,i}))(k) = k_i M_n(k), \qquad (Q_p(h_{p,i}))(k) = \frac{m_e^*}{m_h^*} k_i M_p(k), \qquad i = 1, 2, 3,$$

have solutions $h_n(x,k) = (h_{n,1}(x,k), \ldots, h_{n,3}(x,k))$ *and* $h_p(x,k) = (h_{p,1}(x,k), \ldots, h_{p,3}(x,k))$ *with the property that there exist* $\mu_n(x)$, $\mu_p(x) \geq 0$ *satisfying*

$$\int_{\mathbb{R}^3} k \otimes h_n \, dk = -\mu_n(x) I, \tag{4.15}$$

$$\int_{\mathbb{R}^3} \frac{m_e^*}{m_h^*} k \otimes h_p \, dk = -\mu_p(x) I, \tag{4.16}$$

where $I \in \mathbb{R}^{3\times 3}$ *is the unit matrix and* $a \otimes b = a^\top b \in \mathbb{R}^{3\times 3}$ *for* $a, b \in \mathbb{R}^3$.

In the statement of the lemma we have omitted some technical assumptions on the scattering rates s_n and s_p (regularity conditions). Again, we refer to [53] for the precise hypotheses.

Proof: Again we omit the index $j \in \{n, p\}$. The existence of a solution $h(k) = (h_1(k), h_2(k), h_3(k))$ of $Q(h) = kM(k)$ follows from Lemma 4.3 (1) since

$$\int_{\mathbb{R}^3} k_i M(k) \, dk = 0 \qquad \text{for all } i = 1, 2, 3.$$

We only show (4.15) since the proof of (4.16) is similar.

Let A be the matrix of a rotation with axis k_1. Then $(Ak)_1 = k_1$ for all $k = (k_1, k_2, k_3)^\top \in \mathbb{R}^3$. Since A is isometric, i.e., $|Ak| = |k|$, we obtain $M(Ak) = N^{-1}e^{-|Ak|^2/2} = N^{-1}e^{-|k|^2/2} = M(k)$ and $(Ak)_1 M(Ak) = k_1 M(k)$. This implies, together with the assumption (4.14) and the transformation $w = Ak'$ with $dw = |\det A| dk' = dk'$

$$\begin{aligned}
(Q(h_1 \circ A))(k) &= \int_{\mathbb{R}^3} s(x, k, k')[M(k)h_1(Ak') - M(k')h_1(Ak)] \, dk' \\
&= \int_{\mathbb{R}^3} s(x, Ak, Ak')[M(Ak)h_1(Ak') - M(Ak')h_1(Ak)] \, dk' \\
&= \int_{\mathbb{R}^3} s(x, Ak, w)[M(Ak)h_1(w) - M(w)h_1(Ak)] \, dw \\
&= (Q(h_1))(Ak) = (Ak)_1 M(Ak) = k_1 M(k) = (Q(h_1))(k),
\end{aligned}$$

and thus $Q(h_1 \circ A - h_1) = 0$. Another computation yields (see (4.10))

$$\begin{aligned}
\int_{\mathbb{R}^3} h_1(Ak)\lambda(k)\,dk &= \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} s(x,k,k')h_1(Ak)M(k')\,dk'\,dk \\
&= \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} s(x,Ak,Ak')h_1(Ak)M(Ak')\,dk'\,dk \\
&= \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} s(x,v,w)h_1(v)M(w)\,dw\,dv \\
&= \int_{\mathbb{R}^3} h_1(v)\lambda(v)\,dv
\end{aligned}$$

and

$$\int_{\mathbb{R}^3} (h_1 \circ A - h_1)\lambda\,dk = 0.$$

This is the orthogonality condition (4.12) which ensures the uniqueness of the solution of $Q(h_1 \circ A - h_1) = 0$. Therefore, $h_1 \circ A - h_1 = 0$. We conclude that h_1 remains invariant under a rotation with axis k_1. In particular, we can write

$$h_1(k) = H_1(k_1, |k|^2 - k_1^2).$$

Now, let A be the isometric matrix of the linear mapping $k \mapsto (-k_1, k_2, k_3)$. Since $k \mapsto k_1 M(k)$ is odd, a similar computation as above gives $Q(h_1 \circ A) = -Q(h_1)$ and

$$\int_{\mathbb{R}^3} (h_1 \circ A + h_1)\lambda\,dk = 0.$$

This implies as above that $h_1 \circ A + h_1 = 0$. Thus, h_1 is an odd function with respect to k_1.

In a similar way, we can show that

$$h_i(k) = H_i(k_i, |k|^2 - k_i^2), \quad i = 2, 3,$$

and H_i are odd functions with respect to k_i. In fact, all the functions H_i equal H since, for instance, exchanging k_1 and k_2 in

$$Q(H_1(k_1, k_2^2 + k_3^2)) = k_1 M(k_1^2 + k_2^2 + k_3^2)$$

(with a slight abuse of notation) leads to

$$Q(H_1(k_2, k_1^2 + k_3^2)) = k_2 M(k_1^2 + k_2^2 + k_3^2) = Q(H_2(k_2, |k|^2 - k_2^2))$$

or $Q(H_1 - H_2) = 0$, and a similar argument as above shows that $H_1 = H_2$. We set $H := H_1$.

Since H is odd with respect to the first argument and $|k|^2 - k_j^2$ does not depend on k_j, we obtain for all $i \neq j$,

$$\int_{\mathbb{R}^3} k_i h_j(k)\,dk = \int_{\mathbb{R}^3} k_i H(k_j, |k|^2 - k_j^2)\,dk = 0. \tag{4.17}$$

Furthermore,

$$\begin{aligned}\int_{\mathbb{R}^3} k_i h_i(k)\,dk &= \int_{\mathbb{R}^3} k_i H(k_i, |k|^2 - k_i^2)\,dk = \int_{\mathbb{R}^3} k_j H(k_j, |k|^2 - k_j^2)\,dk \\ &= \int_{\mathbb{R}^3} k_j h_j(k)\,dk\end{aligned}$$

for all i, j. This means that the integral is independent of i, and we can set

$$\mu := -\int_{\mathbb{R}^3} k_1 h_1(k)\,dk = -\int_{\mathbb{R}^3} Q(h_1) h_1 M^{-1}\,dk.$$

The parameter μ depends on x since h_1 depends on x through Q. Moreover, by (4.11),

$$\mu(x) = \int_{\mathbb{R}^3} (-Q(h_1)) h_1 M^{-1}\,dk \geq 0$$

and by (4.17)

$$\int_{\mathbb{R}^3} k \otimes h(k)\,dk = -\mu(x) I,$$

which shows (4.15). □

Now, we are able to analyze the scaled Boltzmann equations (4.7)–(4.8). Setting formally $\alpha = 0$ gives

$$Q_n(f_n) = 0, \qquad Q_p(f_p) = 0.$$

By Lemma 4.1, these equations possess the solutions

$$f_{n0} = n(x,t) M_n, \qquad f_{p0} = p(x,t) M_p, \tag{4.18}$$

respectively, where $n(x,t)$ and $p(x,t)$ are some parameters. Since

$$\int_{\mathbb{R}^3} f_{n0}\,dk = n(x,t), \qquad \int_{\mathbb{R}^3} f_{p0}\,dk = p(x,t),$$

we can interpet n and p as scaled number densities of electrons and holes, repectively. In order to obtain more information from (4.7)–(4.8) we use the *Hilbert method.* The idea is to expand f_n and f_p in terms of powers of α,

$$f_n = f_{n0} + \alpha f_{n1} + \alpha^2 f_{n2} + \cdots, \tag{4.19}$$

$$f_p = f_{p0} + \alpha f_{p1} + \alpha^2 f_{p2} + \cdots, \tag{4.20}$$

and to derive equations for f_{ni} and f_{pi}. We substitute this ansatz into (4.7)–(4.8) and equate coefficients of equal powers of α, noticing that Q_j and I_j are linear operators,

- terms α^0:

$$Q_n(f_{n0}) = 0, \qquad Q_p(f_{p0}) = 0; \tag{4.21}$$

- terms α^1:

$$\begin{aligned} k \cdot \nabla_x f_{n0} - E_{\text{eff}} \cdot \nabla_k f_{n0} &= Q_n(f_{n1}), & (4.22)\\ k \cdot \nabla_x f_{p0} + E_{\text{eff}} \cdot \nabla_k f_{p0} &= Q_p(f_{p1}); & (4.23) \end{aligned}$$

- terms α^2:

$$\begin{aligned} \partial_t f_{n0} + k \cdot \nabla_x f_{n1} - E_{\text{eff}} \cdot \nabla_k f_{n1} &= Q_n(f_{n2}) + I_n(f_{n0}, f_{p0}), & (4.24)\\ \partial_t f_{p0} + k \cdot \nabla_x f_{p1} + E_{\text{eff}} \cdot \nabla_k f_{p1} &= Q_p(f_{p2}) + I_p(f_{n0}, f_{p0}). & (4.25) \end{aligned}$$

Theorem 4.5 *The Hilbert expansions (4.19)–(4.20) are solvable up to second order if and only if*

$$n = \int_{\mathbb{R}^3} f_{n0}\, dk, \qquad p = \int_{\mathbb{R}^3} f_{p0}\, dk \tag{4.26}$$

solve the (bipolar) drift-diffusion equations

$$\begin{aligned} \partial_t n - \operatorname{div}_x J_n &= -R(n,p), & J_n &= \mu_n(\nabla_x n + nE_{\text{eff}}), & (4.27)\\ \partial_t p - \operatorname{div}_x J_p &= -R(n,p), & J_p &= -\mu_p(\nabla_x p - pE_{\text{eff}}), & (4.28)\\ R(n,p) &= A(x)(np - n_i^2), & & & (4.29) \end{aligned}$$

where

$$A(x) = \frac{1}{n_i^2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} g(x,k,k')\, dk'\, dk, \qquad n_i = \sqrt{N_n N_p}\, e^{(E_v - E_c)/2k_B T}.$$

Proof: We already solved (4.21). With the solutions (4.18) we can rewrite (4.22)–(4.23), using $\nabla_k M_n(k) = -kM_n(k), \nabla_k M_p(k) = -(m_e^*/m_h^*)kM_p(k)$,

$$\begin{aligned} Q_n(f_{n1}) &= M_n k \cdot (\nabla_x n + nE_{\text{eff}}) = M_n k \cdot \frac{J_n}{\mu_n},\\ Q_p(f_{p1}) &= \frac{m_e^*}{m_h^*} M_p k \cdot (\nabla_x p - pE_{\text{eff}}) = -\frac{m_e^*}{m_h^*} M_p k \cdot \frac{J_p}{\mu_p}. \end{aligned}$$

By Lemma 4.4, these equations have solutions and moreover, any solution can be expressed as

$$f_{n1} = \frac{J_n}{\mu_n} \cdot h_n + \sigma_n M_n, \qquad f_{p1} = -\frac{J_p}{\mu_p} \cdot h_p + \sigma_p M_p$$

for some unspecified parameters $\sigma_n(x,t)$, $\sigma_p(x,t)$.

Equation (4.24) is solvable by Lemma 4.2 (1) if and only if

$$\begin{aligned} 0 &= \int_{\mathbb{R}^3} \left(\partial_t f_{n0} + k \cdot \nabla_x f_{n1} - \operatorname{div}_k(E_{\text{eff}} f_{n1}) - I_n(f_{n0}, f_{p0})\right) dk \\ &= \partial_t n + \int_{\mathbb{R}^3} k \cdot \nabla_x f_{n1}\, dk - \int_{\mathbb{R}^3} I_n(f_{n0}, f_{p0})\, dk. \end{aligned}$$

We infer from the definition (4.15) of μ_n that

$$\begin{aligned} \int_{\mathbb{R}^3} k \cdot \nabla_x f_{n1}\, dk &= \int_{\mathbb{R}^3} k \cdot \nabla_x \left(\frac{J_n}{\mu_n} \cdot h_n + \sigma_n M_n\right) dk \\ &= \sum_{i,j=1}^{3} \frac{\partial}{\partial x_i} \left(\frac{J_{nj}}{\mu_n} \int_{\mathbb{R}^3} k_i h_{nj}\, dk\right) + \nabla_x \sigma_n \cdot \int_{\mathbb{R}^3} k M_n\, dk \\ &= -\sum_{i,j=1}^{3} \frac{\partial}{\partial x_i} J_{nj} \delta_{ij} = -\operatorname{div}_x J_n. \end{aligned}$$

Furthermore,

$$\begin{aligned} -\int_{\mathbb{R}^3} I_n(f_{n0}, f_{p0})\, dk &= \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} g(x, k', k) \Bigg[\exp\left(\frac{E_c - E_v}{k_B T} + \frac{|k|^2}{2} + \frac{m_e^*}{m_h^*} \frac{|k'|^2}{2}\right) \\ &\qquad \times n M_n(k) p M_p(k') - 1\Bigg] dk'\, dk \\ &= \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} g(x, k, k') \left[\exp\left(\frac{E_c - E_v}{k_B T}\right) \frac{np}{N_n N_p} - 1\right] dk'\, dk \\ &= \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} g(x, k, k') \left[\frac{np}{n_i^2} - 1\right] dk',dk \\ &= A(x)(np - n_i^2). \end{aligned}$$

This shows (4.27) and (4.29). The proof of (4.28) is analogous. □

The first equations in (4.27) and (4.28) show that the quantities J_n and J_p can be interpreted as particle current densities. They are the sum of the drift currents $\mu_n n E_{\text{eff}}$, $\mu_p p E_{\text{eff}}$ and the diffusion currents $\mu_n \nabla_x n$, $-\mu_p \nabla_x p$, respectively, which explains the name of the model.

Equations (4.27)–(4.29) are in scaled form. In order to scale back to the physical variables we notice that the scaled number densities, now called n_s and p_s, read

$$\begin{aligned} n_s &= \int_{\mathbb{R}^3} f_{n0}\, dk_s = \left(\frac{\hbar}{m_e^* v}\right)^3 \int_{\mathbb{R}^d} f_{n0}\, dk = \left(\frac{\hbar^2}{m_e^* k_B T}\right)^{3/2} n, \\ p_s &= \int_{\mathbb{R}^3} f_{n0}\, dk_s = \left(\frac{\hbar^2}{m_e^* k_B T}\right)^{3/2} p, \end{aligned}$$

where n and p are the unscaled variables. Thus, using

$$t_s = \frac{t}{\tau_R}, \quad E_{\text{eff},s} = \frac{\iota_0}{U_T} E_{\text{eff}}, \quad x_s = \frac{x}{\iota_0}, \quad \mu_{n,s} = \frac{m_e^*}{q\tau_c}\mu_n,$$

where $U_T = k_B T/q$ is the *thermal voltage*, we obtain after some computations the unscaled equations

$$\partial_t n - \frac{1}{q}\mathrm{div}_x J_n = -R(n,p), \qquad J_n = q\mu_n(U_T \nabla_x n + n E_{\text{eff}}), \tag{4.30}$$

$$\partial_t p + \frac{1}{q}\mathrm{div}_x J_p = -R(n,p), \qquad J_p = -q\mu_p(U_T \nabla_x p - p E_{\text{eff}}), \tag{4.31}$$

the unscaled recombination-generation rate equals

$$R(n,p) = A(x)(np - n_i^2)$$

for some unscaled $A(x)$ and with the *intrinsic density*

$$n_i = \left(\frac{2\pi k_B T \sqrt{m_e^* m_h^*}}{\hbar^2} \right)^{3/2} \exp\left(\frac{E_v - E_c}{2k_B T} \right)$$

which equals the intrinsic density (2.56) up to a factor $2/(2\pi)^{3/2}$. The reason for this difference is that we ignored the moment density of states in (4.26) (see Section 2.4).

Equations (4.30)–(4.31) are solved in $x \in \mathbb{R}^3$, $t > 0$, together with the initial conditions

$$n(x,0) = n_I(x), \qquad p(x,0) = p_I(x), \quad x \in \mathbb{R}^3.$$

For a self-consistent treatment of the electric field $E_{\text{eff}} = -\nabla V$, equations (4.30)–(4.31) are supplemented by the Poisson equations

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

where ε_s is the semiconductor permittivity and $C(x)$ the doping concentration (see Proposition 3.5).

Mathematically, (4.30)–(4.31) are parabolic convection-diffusion equations with diffusion coefficients

$$D_n := \mu_n U_T, \qquad D_p := \mu_p U_T$$

and so-called *mobilities* μ_n and μ_p. It is remarkable that the quotients of diffusivities and mobilities are constant; the equations

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = U_T$$

are called *Einstein relations.*

The derivation of the drift-diffusion model is mainly based on the following assumptions:

- The mean free path λ_c between two consecutive scattering events is much smaller than the mean free path λ_R between two recombination-generation events (typically, $\lambda_c \sim 10^{-8}$m and $\lambda_R \sim 10^{-5}$m).
- The device diameter is of the order of $\lambda_0 = \sqrt{\lambda_R \lambda_c}$ (typically, $\lambda_0 \sim 10^{-6}$m).
- The electrostatic potential is of the order of $U_T = k_B T/q$ (at room temperature: $U_T \approx 0.026$V).

This implies that the drift-diffusion model is appropriate for semiconductor devices with typical length scales not much smaller than 10^{-6}m and applied voltages much smaller than 1V. However, in applications, this model is used also for voltages much larger than U_T, sometimes together with some correction forms.

4.2 Derivation of the hydrodynamic equations

In the previous section we have derived a fluiddynamical model from the Boltzmann equation by using the Hilbert expansion method. A second approach for the derivation of fluid models is the *moment method* which consists in deriving evolution equations for averaged quantities, like particle density, current density, and energy.

We consider the semi-classical Boltzmann equation in the parabolic-band approximation

$$\partial_t f + \frac{\hbar}{m_e^*} k \cdot \nabla_x f - \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k f = Q(f), \qquad x, k \in \mathbb{R}^d, \ t > 0. \tag{4.32}$$

We assume that the collision operator splits into two parts

$$Q(f) = Q_1(f) + Q_2(f),$$

with the mean free paths λ_c for the first part and λ for the second part. Similarly as in Section 4.1 we choose the reference length λ, the reference velocity $v = \sqrt{k_B T_L / m_e^*}$, the reference time $\tau = \lambda / v$, the reference wave vector $m_e^* v/\hbar$, and the reference field U_T/λ, where $U_T = k_B T_L / q$, and T_L is the lattice temperature. Then, defining

$$x = \lambda x_s, \quad t = \tau t_s, \quad k = \frac{m_e^* v}{\hbar} k_s, \quad E_{\text{eff}} = \frac{U_T}{\lambda} E_{\text{eff},s}$$

and

$$Q_1(f) = \frac{1}{\tau_c} Q_{1,s}(f), \qquad Q_2(f) = \frac{1}{\tau} Q_{2,n}(f),$$

where $\tau_c = \lambda_c / v$, inserting this scaling into (4.32), multiplying this equation by τ, and omitting the index s, gives the scaled equation

$$\partial_t f_\alpha + k \cdot \nabla_x f_\alpha - E_{\text{eff}} \cdot \nabla_k f_\alpha = \frac{1}{\alpha} Q_1(f_\alpha) + Q_2(f_\alpha), \tag{4.33}$$

where $\alpha = \lambda_c/\lambda$.

For the following we set

$$\langle g(k)\rangle = \int_{\mathbb{R}^d} g(k)\,dk$$

for any function g depending on k. We call the averaged quantities $\langle \chi(k) f\rangle$ *moments* of f, In particular, $\langle f\rangle$, $\langle k_i f\rangle$, and $\langle \frac{1}{2}|k|^2 f\rangle$ are called the zeroth, first, and second moments of f, respectively.

The collision operators are assumed to satisfy

$$\begin{aligned} \langle \alpha^{-1}Q_1(f) + Q_2(f)\rangle &= 0, && (4.34)\\ \langle k(\alpha^{-1}Q_1(f) + Q_2(f))\rangle &= -\langle kf\rangle, && (4.35)\\ \langle \tfrac{1}{2}|k|^2(\alpha^{-1}Q_1(f) + Q_2(f))\rangle &= \tfrac{d}{2}\langle f\rangle - \langle c\tfrac{1}{2}|k|^2 f\rangle && (4.36) \end{aligned}$$

for all functions f. The following example shows that the relaxation-time operator (3.40) fulfills these conditions.

Example 4.6 Let

$$(Q_1(f))(x,k,t) = \alpha\langle f(x,\cdot,t)\rangle M(k), \qquad (Q_2(f))(x,k,t) = -f(x,k,t),$$

where $M(k) = (2\pi)^{-d/2}e^{-|k|^2/2}$. Then $\alpha^{-1}Q_1 + Q_2$ satisfies (4.34)–(4.36). Indeed, since

$$\begin{aligned} \langle M\rangle &= \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d} e^{-|k|^2/2}\,dk = 1,\\ \langle k_i M\rangle &= \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d} k_i e^{-k_i^2/2}\,dk_i \prod_{j\neq i}\int_{\mathbb{R}} e^{-k_j^2/2}\,dk_j = 0,\\ \langle \frac{1}{2}|k|^2 M\rangle &= \frac{1}{2(2\pi)^{d/2}}\sum_{i=1}^d \int_{\mathbb{R}^d} k_i^2 e^{-k_i^2/2}\,dk_i \prod_{j\neq i}\int_{\mathbb{R}} e^{-k_j^2/2}\,dk_j,\\ &= \frac{1}{2(2\pi)^{d/2}}\sum_{i=1}^d (2\pi)^{1/2}(2\pi)^{(d-1)/2} = \frac{d}{2}, \end{aligned}$$

we obtain

$$\begin{aligned} \langle \alpha^{-1}Q_1(f) + Q_2(f)\rangle &= \langle f\rangle\langle M\rangle - \langle f\rangle = 0,\\ \langle k(\alpha^{-1}Q_1(f) + Q_2(f))\rangle &= \langle f\rangle\langle kM\rangle - \langle kf\rangle = -\langle kf\rangle,\\ \langle \tfrac{1}{2}|k|^2(\alpha^{-1}Q_1(f) + Q_2(f))\rangle &= \langle f\rangle\langle \tfrac{1}{2}|k|^2 M\rangle - \langle \tfrac{1}{2}|k|^2 f\rangle = \tfrac{d}{2}\langle f\rangle - \langle \tfrac{1}{2}|k|^2 f\rangle. \qquad \square \end{aligned}$$

Multiplying the scaled Boltzmann equation (4.33) by 1, k, $\frac{1}{2}|k|^2$, respectively, and integrating over $k \in \mathbb{R}^d$ yields the so-called *moment equations*

$$\partial_t\langle f_\alpha\rangle + \operatorname{div}_x\langle kf_\alpha\rangle = 0 \qquad (4.37)$$

$$\partial_t\langle kf_\alpha\rangle + \operatorname{div}_x\langle k\otimes kf_\alpha\rangle + E_{\text{eff}}\langle f_\alpha\rangle = -\langle kf_\alpha\rangle \qquad (4.38)$$

$$\partial_t\langle \tfrac{1}{2}|k|^2 f_\alpha\rangle + \operatorname{div}_x\langle \tfrac{1}{2}k|k|^2 f_\alpha\rangle + E_{\text{eff}}\langle kf_\alpha\rangle = \tfrac{d}{2}\langle kf_\alpha\rangle - \langle \tfrac{1}{2}|k|^2 f_\alpha\rangle, \qquad (4.39)$$

where we have used the divergence theorem to compute the force terms:

$$\begin{aligned}\langle E_{\mathrm{eff}}\cdot\nabla_k f_\alpha\rangle &= \langle \mathrm{div}_k(E_{\mathrm{eff}} f_\alpha)\rangle = 0,\\ \langle k_i E_{\mathrm{eff}}\cdot\nabla_k f_\alpha\rangle &= -\langle \nabla_k k_i\cdot E_{\mathrm{eff}} f_\alpha\rangle = -\langle E_{\mathrm{eff},i} f_\alpha\rangle,\\ \langle \tfrac{1}{2}|k|^2 E_{\mathrm{eff}}\cdot\nabla_k f_\alpha\rangle &= -\langle \tfrac{1}{2}\nabla_k |k|^2\cdot E_{\mathrm{eff}} f_\alpha\rangle = -\langle k\cdot E_{\mathrm{eff}} f_\alpha\rangle.\end{aligned}$$

The goal of the moment method is to express all integrals in the moment equations in terms of the first moments $\langle f_\alpha\rangle$,$\langle kf_\alpha\rangle$ and $\langle\frac{1}{2}|k|^2 f_\alpha\rangle$. However, it is not clear how to express the fluxes $\langle k\otimes kf_\alpha\rangle$ and $\langle\frac{1}{2}k|k|^2 f_\alpha\rangle$ in terms of the first moments. Moreover, the third-order moment flux $\langle\frac{1}{2}k|k|^2 f_\alpha\rangle$ cannot be written with the help of the moments up to second order. Therefore, an additional assumption is necessary. We present two approaches. The first one consists in assuming that the mean free path λ_c of the first part of the collision operator is much smaller than the typical device length λ such that $\alpha = \lambda_c/\lambda \ll 1$ and to perform the formal limit $\alpha \to 0$. This idea is due to Bardos, Golse and Levermore [6, 7] (also see the review [36]). The second approach closes the moment hierarchy by choosing a special! function f_α which gives the minimum of the entropy functional. This approach is called the *entropy minimization principle* and has been used by Levermore [42, 43].

For the first approach we impose the following assumptions. Let f_α be a solution to (4.33) such that, as $\alpha \to 0$,

$$f_\alpha \to f \qquad \text{pointwise almost everywhere,} \tag{4.40}$$

$$\langle\chi(k)f_\alpha\rangle \to \langle\chi(k)f\rangle \qquad \text{for } \chi(k) = 1, k_1, \ldots, k_d, |k|^2, \tag{4.41}$$

and assume that

$$Q_1(f) = 0 \qquad \text{if and only if} \qquad f = \exp(\alpha + \beta\cdot k + \gamma|k|^2), \tag{4.42}$$

where $\alpha, \gamma \in \mathbb{R}$ and $\beta = (\beta_1, \ldots, \beta_d)^\top \in \mathbb{R}^d$.

The last condition implies that there exist functions $n(x,t) > 0$, $u(x,t) \in \mathbb{R}$, and $T(x,t) > 0$ such that $f \in N(Q_1)$ can be written equivalently as a so-called *shifted Maxwellian* (see (3.41))

$$M(x,k,t) = \frac{n(x,t)}{(2\pi T(x,t))^{d/2}}\exp\left(-\frac{|k-u(x,t)|^2}{2T(x,t)}\right). \tag{4.43}$$

The functions n, u, and T can be interpreted physically as the particle density, mean velocity, and particle temperature, respectively.

Lemma 4.7 *The Maxwellian (4.43) satisfies*

$$\langle M\rangle = n, \quad \langle kM\rangle = nu, \quad \langle k\otimes kM\rangle = n(u\otimes u) + nTI,$$

$$\langle \tfrac{1}{2}|k|^2 M\rangle = n(\tfrac{1}{2}|u|^2 + \tfrac{d}{2}T), \qquad \langle k|k|^2 M\rangle = 2nu(\tfrac{1}{2}|u|^2 + \tfrac{d+2}{2}T),$$

where I *is the identity matrix in* $\mathbb{R}^{d\times d}$ *and* $a \otimes b = a^\top b$ *for* $a, b \in \mathbb{R}^d$.

The quantity $J = -nu$ *is called the particle current density, and* $e = \frac{1}{2}|u|^2 + \frac{d}{2}T$ *is the total energy consisting of the sum of kinetic and thermal energy.*

Proof: We use the identities

$$\int_{\mathbb{R}} x^{2m} e^{-x^2/2}\,dx = (2m-1)(2m-3)\cdot\ldots\cdot 5\cdot 3\cdot 1\sqrt{2\pi}, \qquad \int_{\mathbb{R}} x^{2m+1} e^{-x^2/2}\,dx = 0 \quad (4.44)$$

for $m \in \mathbb{N}_0$ and the transformation $z = (k-u)/\sqrt{T}$ to compute

$$\begin{aligned}
\langle M\rangle &= n\int_{\mathbb{R}^d} e^{-|z|^2/2}\,dz = n,\\
\langle kM\rangle &= \frac{n}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}(u+\sqrt{T}z)e^{-|z|^2/2}\,dz = nu\\
\langle k\otimes kM\rangle &= \frac{n}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}(u+\sqrt{T}z)\otimes(u+\sqrt{T}z)e^{-|z|^2/2}\,dz\\
&= n(u\otimes u) + nT(2\pi)^{-d/2}\int_{\mathbb{R}^d} z\otimes z e^{-|z|^2/2}\,dz\\
&= n(u\otimes u) + nTI.
\end{aligned}$$

From the last identity follows

$$\langle \tfrac{1}{2}|k|^2 M\rangle = \tfrac{1}{2}\textstyle\sum_{i=1}^d (nu_i^2 + nT) = n(\tfrac{1}{2}|u|^2 + \tfrac{d}{2}T),$$

and finally,

$$\begin{aligned}
\langle k_i|k|^2 M\rangle &= \frac{n}{(2\pi)^{d/2}}\sum_{j=1}^d \int_{\mathbb{R}^d}(u_i+\sqrt{T}z_i)(u_j+\sqrt{T}z_j)^2 e^{-|z|^2/2}\,dz\\
&= nu_i|u|^2 + \frac{u_i nT}{(2\pi)^{d/2}}\sum_{j=1}^d z_j^2 e^{-|z|^2/2}\,dz\\
&\quad +2\frac{nT}{(2\pi)^{d/2}}\sum_{j=1}^d u_j\int_{\mathbb{R}^d} z_i z_j e^{-|z|^2/2}\,dz\\
&\quad +\frac{nT^{3/2}}{(2\pi)^{d/2}}\int_{\mathbb{R}^d} z_i|z|^2 e^{-|z|^2/2}\,dz\\
&= nu_i|u|^2 + nT\sum_{j=1}^d (u_i + 2u_j\delta_{ij})\\
&= 2nu_i\left(\frac{1}{2}|u|^2 + \frac{d+2}{2}T\right).
\end{aligned}$$

□

Theorem 4.8 *Let the assumptions (4.40)–(4.42) hold. Then, as* $\alpha \to 0$*, the moment equations (4.37)–(4.38) become*

$$\begin{aligned}
\partial_t n - \mathrm{div} J &= 0, & (4.45)\\
\partial_t J - \mathrm{div}\left(\frac{J \otimes J}{n}\right) - \nabla(nT) + nE_{\mathrm{eff}} &= -J, & (4.46)\\
\partial_t(ne) - \mathrm{div}[J(e+T)] + J \cdot E_{\mathrm{eff}} &= -n(e - \frac{d}{2}), & (4.47)
\end{aligned}$$

where $J = -nu$ *and* $e = \frac{1}{2}\,|u|^2 + \frac{d}{2}T$.

Equations (4.45)–(4.47) are called the (scaled) *hydrodynamic equations* for semiconductors. They have the same form as the Euler equations of gas dynamics subject to the electric force E_{eff} and the momentum and energy relaxation terms $-J$, $-n(e - \frac{d}{2})$, respectively. Mathematically, they are nonlinear equations of hyperbolic type. For constant particle temperature $T = 1$ (or $T = T_L$ in unscaled variable) equations (4.45)–(4.46) are referred to as the *isothermal hydrodynamic model.* For Coulomb forces $E_{\mathrm{eff}} = -\nabla V$, they are supplemented with the scaled Poisson equation

$$\lambda_D^2 \Delta V = n - C(x),$$

where the so-called *Debye length* $\lambda_D^2 = \varepsilon_s U_T/(qL^2 \max C)$ comes from the scaling. Equations (4.45)–(4.46) are solved for $x \in \mathbb{R}^d$, $t > 0$, together with the initial conditions

$$n(x,0) = n_I(x), \qquad J(x,0) = J_I(x), \qquad T(x,0) = T_I(x), \qquad x \in \mathbb{R}^d,$$

which determine the initial function $e(x,0)$.

Proof: Assumption (4.40) implies that for all smooth functions ϕ,

$$\int_{\mathbb{R}^{2d+1}} (\partial_t f_\alpha + k \cdot \nabla_x f_\alpha - E_{\mathrm{eff}} \cdot \nabla_k f_\alpha)\phi \, dk \, dx = -\int_{\mathbb{R}^{2d+1}} f_\alpha(\partial_t \phi + k \cdot \nabla_x \phi - E_{\mathrm{eff}} \cdot \nabla_k \phi) \, dk \, dx$$

is bounded; therefore, in the weak sense,

$$Q_1(f) = \lim_{\alpha \to 0} Q_1(f_\alpha) = 0$$

and by assumption (4.42), $f = M$. Assumption (4.41) allows us to pass to the limit $\alpha \to 0$ in the moment equations (4.37)–(4.38) yielding

$$\begin{aligned}
\partial_t \langle M \rangle + \mathrm{div}\langle kM \rangle &= 0,\\
\partial_t \langle kM \rangle + \mathrm{div}\langle k \otimes kM \rangle + E_{\mathrm{eff}}\langle M \rangle &= -\langle kM \rangle,\\
\partial_t \langle \tfrac{1}{2}|k|^2 M \rangle + \mathrm{div}\langle \tfrac{1}{2} k|k|^2 M \rangle + E_{\mathrm{eff}}\langle kM \rangle &= \tfrac{d}{2}\langle M \rangle - \langle \tfrac{1}{2}|k|^2 M \rangle.
\end{aligned}$$

Then Lemma 4.7 gives the assertion. □

Remark 4.9 The unscaled hydrodynamic equations can be written as follows:

$$\begin{aligned} \partial_t n - \frac{1}{q}\mathrm{div} J &= 0 && (4.48)\\ \partial_t J - \frac{1}{q}\mathrm{div}\left(\frac{J\otimes J}{n}\right) + \frac{qk_B}{m_e^*}\nabla(nT) - \frac{q^2}{m_e^*} nE_{\mathrm{eff}} &= -\frac{J}{\tau} && (4.49)\\ \partial_t(ne) - \frac{1}{q}\mathrm{div}[J(e+k_BT)] - J\cdot E_{\mathrm{eff}} &= -\frac{n}{\tau}\left(e - \frac{d}{2}k_BT_L\right), && (4.50) \end{aligned}$$

where the energy is given by

$$e = \frac{m_e^*}{2q^2}\frac{|J|^2}{n^2} + \frac{d}{2}k_BT. \qquad \square$$

We now turn to the second approach of deriving the hydrodynamic model. The main hypotheses of the first approach are $\alpha \ll 1$ and that the kernel of Q_1 consists of the shifted Maxwellians such that they can be used as a closure function in the moment equations. The idea of the second approach is to choose a closure function f^* in the moment equations, which minimizes the entropy functional

$$(H(f))(x,t) = \int_{\mathbb{R}^d} f(\log f - 1)\, dk. \tag{4.51}$$

We will see that f^* is in fact equal to the shifted Maxwellian. We present this entropy minimization method in a slightly more general framework.

Let $\chi_i(k)$ be some monomials in the variables $k_1, \ldots, k_d$ for $i = 0, \ldots, N$ and set $\alpha = 1$. The moment equations are now

$$\partial_t\langle\chi_i(k)f\rangle + \mathrm{div}_x\langle k\chi_i(k)f\rangle - E_{\mathrm{eff}}\cdot\langle\chi_i(k)\nabla_k f\rangle = \langle\chi_i(k)(Q_1+Q_2)\rangle \tag{4.52}$$

for $x \in \mathbb{R}^d$, $t > 0$, $i = 0, \ldots, N$. In order to close this system of equations we choose the distribution function f^* which minimizes the entropy (4.51) among those functions whose moments $m(x) = (m_0(x), \ldots, m_N(x))$ are given. In the first approach, we have used the moments $m_0 = n$, $m_i = -nu_i$ $(i = 1, \ldots, d)$, and $m_{d+1} = e$. More precisely, let f^* be the solution to the so-called *Gibbs problem*

$$H(f^*) = \min\{H(f) : \langle\chi_i(k)f\rangle = m_i(x) \text{ for all } x, i\}. \tag{4.53}$$

Notice that the *physical* entropy is $-H(f)$, i.e., the physical entropy has to be maximized.

Lemma 4.10 *The formal solution to (4.53) is*

$$f^*(x,k) = e^{-\lambda(x)\cdot\chi(k)}, \tag{4.54}$$

where $\lambda(x)$ *are Lagrange multipliers and* $\chi(k) = (\chi_0(k), \ldots, \chi_N(k))^\top$.

Proof: Using Lagrange multipliers, we have to analyze the functional

$$G(f,\lambda) = \int_{\mathbb{R}^d} f(\log f - 1)\,dk + \sum_{i=0}^{N} \lambda_i(\langle \chi_i f\rangle - m_i).$$

We claim that

$$\frac{\partial G(f,\lambda)}{\partial f}(g) = \int_{\mathbb{R}^d} g \log f\,dk + \lambda \cdot \langle \chi g\rangle. \tag{4.55}$$

Indeed, we compute formally

$$\frac{1}{\varepsilon}[G(f+\varepsilon g,\lambda) - G(f,\lambda)] = \int_{\mathbb{R}^d} \left[\frac{1}{\varepsilon} f \log\left(1 + \varepsilon\frac{g}{f}\right) + g(\log(f+\varepsilon g) - 1)\right] dk + \frac{1}{\varepsilon}\lambda \cdot \langle \chi \varepsilon g\rangle.$$

Since $\log(1+\varepsilon g/f) \sim \varepsilon g/f$ $(\varepsilon \to 0)$, we obtain

$$\begin{aligned} \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[G(f+\varepsilon g,\lambda) - G(f,\lambda)\right] &= \lim_{\varepsilon\to 0} \int_{\mathbb{R}^d} g \log(f+\varepsilon g)\,dk + \lambda\cdot\langle \chi g\rangle \\ &= \int_{\mathbb{R}^d} g\log f\,dk + \lambda\cdot\langle\chi g\rangle, \end{aligned}$$

which proves (4.55).

Now, the necessary condition

$$\frac{\partial G(f^*,\lambda)}{\partial f}(g) = 0 \qquad \text{for all functions } g$$

for an extremal f leads to

$$\int_{\mathbb{R}^d} g(\log f^* + \lambda\cdot\chi)\,dk = 0 \qquad \text{for all functions } g$$

and hence, $\log f^* = -\lambda\cdot\chi$. ☐

The Lagrange multipliers $\lambda_i(x)$ can be derived by inserting (4.54) into the constraints (4.53):

$$m(x) = \int_{\mathbb{R}^d} \chi_i(k) e^{-\lambda(x)\cdot\chi(k)}\,dk \qquad \text{for } x \in \mathbb{R}^d,\ i = 0,\ldots,N. \tag{4.56}$$

These are implicit equations for $\lambda_0,\ldots,\lambda_N$ which, generally, cannot be directly inverted to give explicit expressions for $\lambda_0,\ldots,\lambda_N$. However, when choosing

$$m = (n, nu, ne) \qquad \text{and} \qquad \chi(k) = \left(1, k, \frac{1}{2}|k|^2\right), \tag{4.57}$$

we can solve (4.56). To see this, we observe that

$$f^* = \exp(-\lambda\cdot\chi) = \exp\left(-\lambda_0 - \sum_{j=1}^{d} \lambda_j k_j - \frac{1}{2}\lambda_{d+1}|k|^2\right)$$

can be equivalently written as the shifted Maxwellians

$$f^* = \frac{n}{(2\pi T)^{d/2}}\exp\left(-\frac{|k-u|^2}{2T}\right),$$

where now the parameters n, u, T play the role of the Lagrange multipliers. The notation is chosen correctly since the moments of this Maxwellian yield the moments (n, nu, ne). This shows that (4.56) is satisfied. Then Lemma 4.7 and (4.34)–(4.36) show that the choice $f = f^*$ in the moment equations (4.52) yields the hydrodynamic model (4.45)–(4.47). We have shown the following result.

Theorem 4.11 *Assume that (4.34)–(4.36) hold and let* f^* *be the solution to the Gibbs problem (4.53). Then, choosing* $m_i(x)$ *and* $\chi_i(k)$ *according to (4.57), the moments* (n, nu, e) *solve the hydrodynamic equations (4.45)–(4.47).*

The hydrodynamic equations have been first introduced by Bløtekjær [10] and Baccarani and Wordeman [5] in the context of semiconductor devices. In their model, sometimes called the BBW equations, the heat conduction term $-\mathrm{div}(\kappa(n,T)\nabla T)$ is added to the left-hand side of (4.50). The heat conductivity $\kappa(n,T)$ is usually modeled in $\mathbb{R}^3$ by

$$\kappa(n,T) = \frac{3}{2}\frac{k_B^2\tau}{m_e^*}nT$$

(see [5]). The heat conduction term makes equation (4.50) parabolic since we can write it as

$$\frac{d}{2}k_B\partial_t T - \mathrm{div}(\kappa(n,T)\nabla T) = \text{ function of } n, J, T \text{ and first derivatives.}$$

In particular, the heat conduction term is often used in numerical simulations. However, it is not clear how to justify this term.

4.3 Derivation of the spherical harmonics expansion equations

In Section 4.1 we have derived the drift-diffusion model from the Boltzmann equation by the Hilbert expansion method under the assumption that the mean free path between two consecutive collision events is much smaller than the typical time. In this section we perform a Hilbert expansion using the same hypothesis but a different collision operator. We show that this approximation leads to the Spherical Harmonics Expansion (SHE) model.

We consider the semi-classical Boltzmann equation

$$\partial_t f + v(k)\cdot\nabla_k f - \frac{q}{\hbar}E_{\mathrm{eff}}\cdot\nabla_k f = Q_{\mathrm{el}}(f) + Q_{\mathrm{inel}}(f), \qquad x\in\mathbb{R}^3,\ k\in B,\ t>0, \quad (4.58)$$

where $v(k) = \hbar^{-1}\nabla_k E(k)$, $Q_{\rm el}(f)$ models elastic collisions and is given by

$$(Q_{\rm el}(f))(x,k,t) = \int_B \sigma(x,k,k')\delta(E(k') - E(k))(f' - f)\,dk', \tag{4.59}$$

with the scattering rate $\sigma(x,k,k')$ and $f = f(x,k,t)$, $f' = f(x,k',t)$ (see Example 3.6). Furthermore, $Q_{\rm inel}(f)$ models inelastic collisions (e.g., inelastic corrections to electron-phonon and electron-electron scattering) and will not specified in this section. We impose the following assumptions:

- The operator $Q_{\rm inel}$ is linear.
- The scattering rate σ is positive and symmetric in k and k':

$$\sigma(x,k,k') > 0, \qquad \sigma(x,k,k') = \sigma(x,k',k) \qquad \text{for all } x \in \mathbb{R}^3,\ k,k' \in B. \tag{4.60}$$

We scale the Boltzmann equation (4.58) by using the same scaling as in Section 4.1, i.e., we fix the reference velocity $v_0 = \sqrt{k_B T/m_e^*}$, the characteristic times $\tau_{\rm el}$, $\tau_{\rm inel}$ of elastic and inelastic collisions, respectively, with corresonding mean free paths $\lambda_{\rm el} = \tau_{\rm el}v_0$ and $\lambda_{\rm inel} = \tau_{\rm inel}v_0$, the reference length $\lambda_0 = \sqrt{\lambda_{\rm el}\lambda_{\rm inel}}$ and time $\tau_0 = \lambda_0/v_0$. Then, with the scaling

$$x = \lambda_0 x_s, \quad t = \tau_{\rm inel}t_s, \quad k = \frac{m_e^* v_0}{\hbar}k_s, \quad E_{\rm eff} = \frac{U_T}{\lambda_0}E_{\rm eff,s},$$

$$E(k) = m_e^* v_0^2 E_s(k_s), \quad Q_{\rm el}(f) = \frac{1}{\tau_{\rm el}}Q_{{\rm el},s}(f), \quad Q_{\rm inel}(f) = \frac{1}{\tau_{\rm inel}}Q_{\rm inel}(f),$$

we obtain, omitting the index s and introducing $\alpha^2 = \lambda_{\rm el}/\lambda_{\rm inel}$,

$$\alpha^2\partial_t f + \alpha\left(\nabla_k E(k)\cdot\nabla_x f - E_{\rm eff}\cdot\nabla_k f\right) = Q_{\rm el}(f) + \alpha^2 Q_{\rm inel}(f), \tag{4.61}$$

since $v_0\tau_{\rm el}/\lambda_0 = \alpha$. We assume that $\alpha \ll 1$.

We now analyze the collision operator $Q_{\rm el}$. In [8] the following results are shown.

Lemma 4.12 *Assume that (4.60) holds. Then*

(1) $-Q_{\rm el}$ is self-adjoint and non-negative on $L^2(B)$, i.e.

$$-\int_B Q_{\rm el}(f)f\,dk \geq 0 \qquad \textit{for all } f \in L^2(B).$$

(2) The kernel of $Q_{\rm el}$ is given by

$$N(Q_{\rm el}) = \{f \in L^2(B) : f \textit{ only depends on } E(k)\}.$$

(3) The equation $Q_{\mathrm{el}}(f) = h$ has a solution if and only if

$$\int_B h(k)\delta(e - E(k))\, dk = 0 \qquad \textit{for all } e \in R(E). \tag{4.62}$$

(4) For all functions g only depending on $E(k)$ and all $f \in L^2(B)$ there holds

$$Q_{\mathrm{el}}(gf) = gQ_{\mathrm{el}}(f). \tag{4.63}$$

In the statement of the lemma we have omitted some technical assumptions on the regularity of $\sigma(x, k, k')$ and $g(E(k))$; we refer to [8] for details.

Proof: (1) Let $x \in \mathbb{R}^3$, $f, g \in L^2(B)$, and set $f = f(k)$, $f' = f(k')$, $g = g(k)$, $g' = g(k')$, $E = E(k)$, and $E' = E(k')$. Then, using the property

$$\int_B \delta(E' - E)\psi(k, k')\, dk' = \int_B \delta(E - E')\psi(k, k')\, dk'$$

for any (smooth) function ψ and employing the symmetry of σ, we compute

$$\begin{aligned}
\int_B Q_{\mathrm{el}}(f)g\, dk &= \frac{1}{2}\int_{B^2} \sigma(x, k, k')\delta(E' - E)(f' - f)g\, dk'\, dk \\
&\qquad + \frac{1}{2}\int_{B^2} \sigma(x, k', k)\delta(E - E')(f - f')g'\, dk\, dk' \\
&= -\frac{1}{2}\int_{B^2} \sigma(x, k, k')\delta(E' - E)(f' - f)(g' - g)\, dk'\, dk \\
&= \int_B Q_{\mathrm{el}}(g)f\, dk.
\end{aligned}$$

This shows that Q_{el} is self-adjoint. Taking $f = g$ gives

$$-\int_B Q_{\mathrm{el}}(f)f\, dk = \frac{1}{2}\int_{B^2} \sigma(x, k, k')\delta(E' - E)(f' - f)^2\, dk'\, dk \geq 0 \tag{4.64}$$

and thus, $-Q_{\mathrm{el}}$ is non-negative.

(2) Let $f \in N(Q_{\mathrm{el}})$. Then (4.64) implies that

$$\delta(E(k') - E(k))(f(k') - f(k))^2 = 0 \qquad \text{for } k, k' \in B.$$

Heuristically, $\delta(x) = 0$ for $x \neq 0$ and $\delta(x) \neq 0$ for $x = 0$. This motivates

$$E(k') = E(k) \qquad \text{and} \qquad f(k') = f(k) \qquad \text{for } k, k' \in B.$$

Hence, f must be constant on each energy surface $\{k \in B : E(k) = e\}$. We infer that f is a function of $E(k)$ only.

(3) We show that $N(Q_{\mathrm{el}})^\perp$ only consists of functions satisfying (4.62). Then the self-adjointness of Q_{el} and Fredholm's alternative give the conclusion. Let $h \in N(Q_{\mathrm{el}})^\perp$ and $f \in N(Q_{\mathrm{el}})$. By part (2), we can write $f(k) = g(E(k))$ for some function g and

$$\begin{aligned} 0 &= \int_B h(k)f(k)\,dk = \int_B h(k)g(E(k))\,dk = \int_B h(k)\int_{R(E)} g(e)\delta(E(k)-e)\,de\,dk \\ &= \int_{R(E)} \left(\int_B h(k)\delta(E(k)-e)\,dk\right) g(e)\,de. \end{aligned}$$

This equation holds for all $f \in N(Q_{\mathrm{el}})$ and thus for any function g. Hence

$$\int_B h(k)\delta(E(k)-e)\,dk = 0 \qquad \text{for } e \in R(E).$$

Thus, functions of $N(Q_{\mathrm{el}})^\perp$ are satisfying (4.62). On the other hand, if h fulfills (4.62) the same arguments as above show that $h \in N(Q_{\mathrm{el}})^\perp$.

(4) We claim that

$$\int_B \sigma(x,k,k')\delta(E'-E)f'g(E')\,dk' = g(E(k))\int_B \sigma(x,k,k')\delta(E'-E)f'\,dk'. \tag{4.65}$$

Indeed, by the coarea formula (Theorem 2.15) we can write formally

$$\begin{aligned} &\int_B \sigma(x,k,k')\delta(E'-E)f'g(E')\,dk' \\ &= \int_{R(E)}\int_{E^{-1}(e)} \sigma(x,k,k')\delta(e-E(k))f(k')g(e)\frac{dF(k')}{|\nabla_k E(k')|}\,de \\ &= \int_{R(E)}\int_{E^{-1}(e)} \sigma(x,k,k')\delta(e-E(k))f(k')g(E(k))\frac{dF(k')}{|\nabla_k E(k')|}\,de \\ &= g(E(k))\int_B \sigma(x,k,k')\delta(E'-E)f'\,dk'. \end{aligned}$$

This gives

$$\begin{aligned} (Q_{\mathrm{el}}(gf))(k) &= \int_B \sigma(x,k,k')\delta(E'-E)f'g(E')\,dk' - \int_B \sigma(x,k,k')\delta(E'-E)fg(E)\,dk' \\ &= g(E(k))\int_B \sigma(x,k,k')\delta(E'-E)f'\,dk' \\ &\qquad - g(E(k))\int_B \sigma(x,k,k')\delta(E'-E)f\,dk' \\ &= g(E(k))(Q_{\mathrm{el}}(f))(k). \end{aligned}$$

This finishes the proof of the lemma. □

Remark 4.13 From the proof of Lemma 4.12 (4) we conclude the following general result: For all functions $\psi(k,k')$ and $g(E(k))$ it holds

$$\int_B \psi(k,k')\delta(E'-E)g(E')\,dk' = g(E(k))\int_B \psi(k,k')\delta(E'-E)\,dk'. \tag{4.66}$$

Now we insert the Hilbert expansion.

$$f = f_0 + \alpha f_1 + \alpha^2 f + \cdots \tag{4.67}$$

into the scaled Boltzmann equation (4.61), use the linearity of the collision operators Q_{el} and Q_{inel} and identify terms of equal order in α to find

- terms α^0:
$$Q_{\mathrm{el}}(f_0) = 0; \tag{4.68}$$
- terms α^1:
$$Q_{\mathrm{el}}(f_1) = \nabla_k E(k)\cdot\nabla_x f_0 - E_{\mathrm{eff}}\cdot\nabla_k f_0; \tag{4.69}$$
- terms α^2:
$$Q_{\mathrm{el}}(f_2) = \partial_t f_0 + \nabla_k E(k)\cdot\nabla_x f_1 - E_{\mathrm{eff}}\cdot\nabla_k f_1 - Q_{\mathrm{inel}}(f_0). \tag{4.70}$$

These equations can be solved using the properties of the operator Q_{el}.

Theorem 4.14 *The Hilbert expansion (4.67) is solvable up to second order if and only if* $f_0(x,k,t) = F(x,E(k),t)$ *and* F *is a solution of*

$$N(\varepsilon)\partial_t F - \mathrm{div}_x J + E_{\mathrm{eff}}\cdot\frac{\partial J}{\partial\varepsilon} = S(F), \qquad x\in\mathbb{R}^3,\ \varepsilon\in R(E),\ t>0, \tag{4.71}$$

where

$$J(x,\varepsilon,t) = D(x,\varepsilon)\left(\nabla_x F - E_{\mathrm{eff}}\frac{\partial F}{\partial\varepsilon}\right) \tag{4.72}$$

is the current density,

$$N(\varepsilon) = \int_B \delta(\varepsilon - E(k))\,dk \tag{4.73}$$

is called the density of states of energy ε *(compare with Lemma 2.14),*

$$D(x,\varepsilon) = \int_B \nabla_k E(k)\otimes\mu(x,k)\delta(\varepsilon-E(k))\,dk \in \mathbb{R}^{3\times 3} \tag{4.74}$$

is the diffusion matrix, where μ *is a solution of* $Q_{\mathrm{el}}(\mu) = -\nabla_k E(k)$*, and*

$$(S(F))(x,\varepsilon,t) = \int_B (Q_{\mathrm{inel}}(F))(x,k,t)\delta(\varepsilon-E(k))\,dk \tag{4.75}$$

is the collision operator, averaged over the energy surface $\varepsilon = E(k)$.

Equations (4.71)–(4.72) are called the *Spherical Harmonic Expansion (SHE) model.* It has been first derived in the physical literature for spherically symmetric band diagrams and rotationally invariant collision operators as a numerical approximation of the Boltzmann equation [34, 35] or through a macroscopic limit [28] which explains the name. Notice, however, that the above SHE model is derived without the assumption of spherical symmetric energy bands. The presented expansion is no more "spherical" but rather on surfaces of constant energy. The SHE model is supplemented by the initial conditions

$$F(x, \varepsilon, 0) = F_I(x, \varepsilon), \qquad x \in \mathbb{R}^3,\ \varepsilon \in R(E),$$

and appropriate boundary conditions for ε which we do not detail here.

Proof: By Lemma 4.12 (2), any solution to (4.68) can be written as

$$f_0(x, k, t) = F(x, E(k), t)$$

for some function F. Thus we can reformulate (4.69) as

$$Q_{\text{el}}(f_1) = \nabla_k E \cdot \left(\nabla_x F - E_{\text{eff}} \frac{\partial F}{\partial \varepsilon} \right).$$

We claim that any solution of this equation can be written as

$$f_1(x, k, t) = -\mu(x, k) \cdot \left(\nabla_x F - E_{\text{eff}} \frac{\partial F}{\partial \varepsilon} \right) + F_1(x, E(k), t), \tag{4.76}$$

where $F_1 \in N(Q_{\text{el}})$ and $\mu(x, k)$ is a solution of

$$Q_{\text{el}}(\mu) = -\nabla_k E. \tag{4.77}$$

Notice that μ depends on the parameter x since σ in the definition of Q_{el} also depends on x. By Lemma 4.12 (3), (4.77) is solvable if and only if

$$\int_B \nabla_k E \delta(e - E(k))\, dk = 0 \qquad \text{for all } e \in R(E).$$

To see this we introduce the Heaviside function H, defined by $H(x) = 0$ for $x < 0$ and $H(x) = 1$ for $x > 0$. This function has the property that $H'(x) = \delta(x)$. Hence

$$\int_B \nabla_k E \delta(e - E(k))\, dk = -\int_B \nabla_k H(e - E(k))\, dk = 0, \tag{4.78}$$

since $E(k)$ is periodic on B. Thus (4.77) is solvable. Observing that $\nabla_x F - E_{\text{eff}}(\partial F / \partial \varepsilon)$ only depends on $E(k)$ (and on the parameters x, t), Lemma 4.12 (4) shows that

$$Q_{\text{el}}(f_1) = \left(\nabla_x F - E_{\text{eff}} \frac{\partial F}{\partial \varepsilon} \right) \cdot \nabla_k E = Q_{\text{el}} \left(-\mu \left(\nabla_x F - E_{\text{eff}} \frac{\partial F}{\partial \varepsilon} \right) \right).$$

The linearity of Q_{el} implies that

$$Q_{\mathrm{el}}\left(f_1 + \mu\left(\nabla_x F - E_{\mathrm{eff}}\frac{\partial F}{\partial \varepsilon}\right)\right) = 0$$

and therefore, (4.76) follows. In the following, we take a particular solution to (4.69) by choosing $F_1 = 0$.

It remains to solve (4.70). Again, by Lemma 4.12 (3), this operator equation is solvable if and only if

$$\int_B \left(\partial_t f_0 + \nabla_k E \cdot \nabla_x f_1 - E_{\mathrm{eff}} \cdot \nabla_k f_1 - Q_{\mathrm{inel}}(f_0)\right) \delta(e - E(k))\, dk = 0 \tag{4.79}$$

for all $e \in R(E)$. We claim that this condition is equivalent to (4.71). To show this, we compute the terms separately. By (4.66) we obtain for the first term

$$\begin{aligned}\int_B \partial_t f_0 \delta(e - E(k))\, dk &= \int_B \partial_t F(x, E(k), t)\delta(e - E(k))\, dk \\ &= \partial_t F(x, e, t) \int_B \delta(e - E(k))\, dk = N(e)\partial_t F(x, e, t).\end{aligned}$$

We introduce the electron current density by

$$J(x, e, t) = -\int_B \nabla_k E(k) f_1(x, k, t)\delta(e - E(k))\, dk.$$

Then the second term in (4.79) becomes

$$\int_B \nabla_k E \cdot \nabla_x f_1 \delta(e - E(k))\, dk = -\mathrm{div}_x J(x, e, t).$$

The current density can be reformulated in terms of the function F by using the expression (4.76) for f_1 and the property (4.66):

$$\begin{aligned}J(x, e, t) &= \int_B \nabla_k E(k)\mu(x, k) \cdot \left(\nabla_x F - E_{\mathrm{eff}}\frac{\partial F}{\partial \varepsilon}\right)(x, E(k), t)\delta(e - E(k))\, dk \\ &= D(x, e)\left(\nabla_x F - E_{\mathrm{eff}}\frac{\partial F}{\partial \varepsilon}\right)(x, e, t).\end{aligned}$$

For the computation of the third term in (4.79) we choose a smooth function ψ with compact support on R(E) and use the definition of the δ distribution and integration by parts to obtain

$$\begin{aligned}&\int_{R(E)} \psi(e) \int_B \nabla_k f_1(x, k, t)\delta(e - E(k))\, dk\, de \\ &= \int_B \psi(E(k))\nabla_k f_1(x, k, t)\, dk = -\int_B f_1(x, k, t)\psi'(E(k))\nabla_k E(k)\, dk \\ &= -\int_{R(E)} \psi'(e) \int_B \nabla_k E(k) f_1(x, k, t)\delta(e - E(k))\, dk\, de \\ &= \int_{R(E)} \psi'(e) J(x, e, t)\, de = -\int_{R(E)} \psi(e)\frac{\partial J}{\partial \varepsilon}(x, e, t)\, de.\end{aligned}$$

Since ψ is arbitrary, it follows

$$\int_B \nabla_k f_1(x,k,t)\delta(e-E(k))\,dk = -\frac{\partial J}{\partial \varepsilon}(x,e,t).$$

We have shown that (4.79) is equivalent to (4.71)–(4.72). ☐

The advantage of the SHE model is twofold:

- The SHE equations have to be solved in $3+1$ dimensions of the position-energy space instead of the $(3+3)$-dimensional phase space of the Boltzmann equation.
- The SHE model is of parabolic type which simplifies its numerical solution.

In order to see that the SHE model is of parabolic type, we assume that the electric field E_{eff} is a gradient, i.e, $E_{\text{eff}} = -\nabla_x V$, and we introduce the total energy variable

$$u = \varepsilon - V(x,t).$$

Then, with the change of unknowns

$$\begin{aligned} F(x,\varepsilon,t) &= f(x,\varepsilon - V(x,t),t) = f(x,u,t), \\ D(x,\varepsilon) &= d(x,\varepsilon - V(x,t),t) = d(x,u,t), \\ N(\varepsilon) &= N(u+V(x,t)) = \varrho(x,u,t), \end{aligned}$$

we obtain

$$\nabla_x F = \nabla_x f - \nabla_x V \frac{\partial f}{\partial u} = \nabla_x f - \nabla_x V \frac{\partial F}{\partial \varepsilon}$$

or

$$\nabla_* F := \left(\nabla_x + \nabla_x V \frac{\partial}{\partial \varepsilon}\right) F = \nabla_x f$$

and

$$\partial_t F = \partial_t f - \partial_t V \frac{\partial f}{\partial u}.$$

Since we can write (4.71)–(4.72) equivalently as

$$N(\varepsilon)\partial_t F - \nabla_* \cdot J = S(F), \qquad J = D\nabla_* F,$$

we obtain, in the total energy variable,

$$\varrho(x,u,t)\partial_t f - \operatorname{div}_x(d(x,u,t)\nabla_x f) = S(F) + \varrho(x,u,t)\partial_t V \frac{\partial f}{\partial u}.$$

The parabolicity now follows from the positive (semi-) definiteness of the diffusion matrix d or D, respectively, as stated in the following lemma.

Lemma 4.15 *The diffusion matrix $D(x,\varepsilon) \in \mathbb{R}^{3\times 3}$ is symmetric and positive semi-definite.*

In [8, Prop. 3.6] a stronger property for $D(x,\varepsilon)$ is shown: There exists $K>0$ such that

$$D_{ij}(x,\varepsilon) \geq \frac{K}{N(\varepsilon)} \int_B \frac{\partial E}{\partial k_i}(k)\frac{\partial E}{\partial k_j}(k)\delta(\varepsilon - E(k))\,dk, \qquad i,j=1,2,3. \tag{4.80}$$

In particular, $\det D = 0$ at critical points of E, (i.e. $\nabla_k E(k) = 0$ and $\varepsilon = E(k)$).

Proof: We use the symmetry of Q_{el} (Lemma 4.12 (1)) and the property (4.63) of Lemma 4.12 (4) to show the symmetry of $D(x,\varepsilon)$. Let ψ be a (smooth) function. Then

$$\begin{aligned}\int_{\mathbb{R}} D_{ij}(x,\varepsilon)\psi(\varepsilon)\,d\varepsilon &= \int_{\mathbb{R}}\int_B \frac{\partial E}{\partial k_i}\mu_j\delta(\varepsilon - E(k))\,dk\,\psi(\varepsilon)\,d\varepsilon \\ &= -\int_B Q_{\mathrm{el}}(\mu_i)\mu_j\psi(E(k))\,dk = -\int_B Q_{\mathrm{el}}(\psi(E(k))\mu_i)\mu_j\,dk \\ &= -\int_B Q_{\mathrm{el}}(\psi(E(k))\mu_j)\mu_i\,dk = \int_{\mathbb{R}} D_{ji}(x,\varepsilon)\psi(\varepsilon)\,d\varepsilon.\end{aligned}$$

since ψ is arbitrary, $D_{ij}(x,\varepsilon) = D_{ji}(x,\varepsilon)$ for all $x \in \mathbb{R}^3$, $\varepsilon \in \mathbb{R}$.

In order to show that $D(x,\varepsilon)$ is positive semi-definite, let $z \in \mathbb{R}^3$. Then, by definition of Q_{el},

$$\begin{aligned} z^\top D(x,\varepsilon)z &= -\sum_{i,j=1}^3 \int_B z_i Q_{\mathrm{el}}(\mu_i)\mu_j z_j \delta(\varepsilon - E(k))\,dk \\ &= -\sum_{i,j=1}^3 \int_B\int_B \sigma(x,k,k')\delta(E(k')-E(k))z_i(\mu_i(k')-\mu_i(k))z_j\mu_j(k) \\ &\qquad \times\delta(\varepsilon - E(k))\,dk'\,dk.\end{aligned}$$

Similar as in the proof of Lemma 4.12 (1), we can write

$$\begin{aligned} z^\top D(x,\varepsilon)z &= \frac{1}{2}\sum_{i,j=1}^3 \int_B\int_B \sigma(x,k,k')\delta(E(k')-E(k))\delta(\varepsilon-E(k)) \\ &\qquad \times z_i(\mu_i(k')-\mu_i(k))z_j(\mu_j(k')-\mu_j(k))\,dk'\,dk \\ &= \frac{1}{2}\int_B\int_B \sigma(x,k,k')\delta(E(k')-E(k))\delta(\varepsilon-E(k)) \\ &\qquad \times\left(\sum_{i=1}^3 z_i(\mu_i(k')-\mu_j(k))\right)^2 dk'\,dk \\ &\geq 0.\end{aligned}$$

For the last equality we have used the elementary identity

$$\sum_{i,j=1}^n a_i a_j = \left(\sum_{i=1}^n a_i\right)^2 \qquad \text{for all } a_n,\ldots,a_n \in \mathbb{R}.$$

The lemma is shown. $\square$

More explicit expressions for the density of states $N(\varepsilon)$ and the diffusion matrix $D(x,\varepsilon)$ can be derived for special energy bands (see [8, Sec. 3.4]).

Example 4.16 (Spherical symmetric energy bands)
Assume that the scattering rate only depends on $E(k)$, i.e.

$$\sigma(x,k,k') = s(x,E(k)) \qquad \text{for all } k,k' \text{ with } E(k)=E(k'),$$

and that $E(k)$ is spherical symmetric, i.e. $E=E(|k|)$, and strictly monotone in $|k|$, i.e., there exists a function $\gamma:\mathbb{R}\to\mathbb{R}$ such that

$$|k|^2 = \gamma(E(|k|)).$$

Notice that the first assumption makes sense since due to the term $\delta(E(k')-E(k))$ in the definition of Q_{el}, the scattering rate needs to be defined only on the surface $\{k'\in B : E(k')=E(k)\}$ of energy $E(k)$. The second assumption implies that we can choose $B=\mathbb{R}^3$.

We claim that

$$\begin{aligned} N(\varepsilon) &= 2\pi\sqrt{\gamma(\varepsilon)}\gamma'(\varepsilon), && (4.81)\\ D(x,\varepsilon) &= \frac{4}{3}\,\frac{\gamma(\varepsilon)}{s(x,\varepsilon)\gamma'(\varepsilon)^2}I, && (4.82)\end{aligned}$$

where $I\in\mathbb{R}^{3\times 3}$ is the identity matrix.

First we reformulate $D(x,\varepsilon)$. From the first assumption and the definition (4.73) of the density of states follows, with $E=E(k)$, $E'=E(k')$,

$$\begin{aligned} (Q_{\mathrm{el}}(f))(x,k,t) &= s(x,E(k))\int_{\mathbb{R}^3}\delta(E'-E)f(k')\,dk' - s(x,E(k))N(E(k))f(k)\\ &= \frac{([f]-f)(x,k,t)}{\tau(x,E(k))}, && (4.83)\end{aligned}$$

where $\tau(x,\varepsilon)=1/s(x,\varepsilon)N(\varepsilon)$ is called relaxation time and

$$[f](k) = \frac{1}{N(E(k))}\int_{\mathbb{R}^3}\delta(E'-E)f(k')\,dk'$$

is the average of f on the energy surface $\{k' : E(k')=E(k)\}$. The expression (4.83) is called *relaxation-time operator* (compare with (3.40)).

We claim that the solution of $Q_{\mathrm{el}}(\mu)=-\nabla_k E$ can now be written explicitly as

$$\mu(x,k) = \tau(x,E(k))\nabla_k E(k).$$

Indeed, by (4.64) and (4.78),

$$\begin{aligned} [\mu](k) &= \frac{1}{N(E(k))}\int_{\mathbb{R}^3}\delta(E'-E)\tau(x,E')\nabla_k E(k')\,dk' \\ &= \frac{\tau(x,E(k))}{N(E(k))}\int_{\mathbb{R}^3}\delta(E'-E)\nabla_k E(k')\,dk' = 0, \end{aligned}$$

and hence

$$Q_{\text{el}}(\mu) = \frac{[\mu]-\mu}{\tau(x,E(k))} = -\nabla_k E(k).$$

Thus, we can write

$$\begin{aligned} D(x,\varepsilon) &= \int_{\mathbb{R}^3}\nabla_k E(k)\otimes\nabla_k E(k)\tau(x,E(k))\delta(\varepsilon-E(k))\,dk \\ &= \tau(x,\varepsilon)\int_{\mathbb{R}^3}\nabla_k E(k)\otimes\nabla_k E(k)\delta(\varepsilon-E(k))\,dk. \end{aligned}$$

This shows that the estimate (4.80) is sharp.

The above expression for $D(x,\varepsilon)$ can be further simplified under the second assumption. As E only depends on $|k|$, it is convenient to use spherical coordinates $k=\varrho\omega$, where $\varrho>0$ and

$$\omega = \begin{pmatrix}\sin\theta\ \cos\phi\\ \sin\theta\ \sin\phi\\ \cos\theta\end{pmatrix}, \qquad 0\le\phi<2\pi,\ 0\le\theta<\pi.$$

Then, transforming $\eta=E(\varrho)$ with $d\eta=E'(\varrho)\,d\varrho$,

$$\begin{aligned} D(x,\varepsilon) &= \tau(x,\varepsilon)\int_0^\infty\int_0^{2\pi}\int_0^\pi E'(\varrho)\omega\otimes E'(\varrho)\omega\delta(\varepsilon-E(\varrho))\varrho^2\sin\theta\,d\theta\,d\phi\,d\varrho \\ &= \tau(x,\varepsilon)\int_{\mathbb{R}}E'(\varrho)\delta(\varepsilon-\eta)\varrho^2\,d\eta\cdot\int_0^{2\pi}\int_0^\pi\omega\otimes\omega\sin\theta\,d\theta\,d\phi. \end{aligned}$$

By the definition of the δ distribution, the first integral is

$$\int_{\mathbb{R}}E'(\varrho)\delta(\varepsilon-\eta)\varrho^2\,d\eta = E'(|k|)|k|^2 \qquad \text{with } \varepsilon=E(|k|).$$

For the second integral an elementary computation gives

$$\int_0^{2\pi}\int_0^\pi\omega\otimes\omega\sin\theta\,d\theta\,d\phi = \frac{4\pi}{3}I,$$

where I is the identity matrix. Therefore

$$D(x,\varepsilon) = \frac{4\pi}{3}\tau(x,\varepsilon)E'(|k|)|k|^2 \qquad \text{with } \varepsilon=E(|k|).$$

Differentiation of $|k|^2 = \gamma(E(|k|))$ with respect to $|k|$ yields $2|k| = \gamma'(E(|k|))E'(|k|)$ and therefore

$$E'(|k|) = \frac{2\sqrt{\gamma(\varepsilon)}}{\gamma'(\varepsilon)} \qquad \text{with } \varepsilon = E(|k|). \tag{4.84}$$

We conclude that

$$D(x,\varepsilon) = \frac{8\pi}{3}\,\frac{\tau(x,\varepsilon)\gamma(\varepsilon)^{3/2}}{\gamma'(\varepsilon)} I = \frac{8\pi}{3}\,\frac{\gamma(\varepsilon)^{3/2}}{s(x,\varepsilon)N(\varepsilon)\gamma'(\varepsilon)} I.$$

The density of states can be computed as

$$\begin{aligned} N(\varepsilon) &= \int_{\mathbb{R}^3} \delta(\varepsilon - E(|k|))\,dk = \int_0^\infty \int_0^{2\pi} \int_0^\pi \delta(\varepsilon - E(\varrho))\varrho^2 \sin\theta\,d\theta\,d\phi\,d\varrho \\ &= 4\pi \int_{\mathbb{R}} \delta(\varepsilon - \eta)E'(\varrho)^{-1}\varrho^2\,d\eta = \frac{4\pi|k|^2}{E'(|k|)} \qquad \text{with } \varepsilon = E(|k|). \end{aligned}$$

Employing (4.84) we infer that

$$N(\varepsilon) = 2\pi\sqrt{\gamma(\varepsilon)}\gamma'(\varepsilon)$$

and

$$D(x,\varepsilon) = \frac{4}{3}\,\frac{\gamma(\varepsilon)}{s(x,\varepsilon)\gamma'(\varepsilon)^2} I.$$

This shows (4.81) and (4.82). □

Example 4.17 (Parabolic band approximation)
We assume that $\sigma(x,k,k') = s(x,E(k))$ for all $k,k' \in \mathbb{R}^3$ such that $E(k) = E(k')$ and

$$E(k) = \frac{1}{2}|k|^2.$$

Then $\gamma(\varepsilon) = 2\varepsilon$ and

$$N(\varepsilon) = 4\pi\sqrt{2\varepsilon}, \qquad D(x,\varepsilon) = \frac{2\varepsilon}{3s(x,\varepsilon)} I.$$

With the specification (see [24, (2.6)] or [45, (3.63)])

$$\tau(x,\varepsilon) = \frac{1}{s_0\varepsilon^\beta N(\varepsilon)}, \qquad \beta > -1, \tag{4.85}$$

we obtain $s(x,\varepsilon) = 1/\tau(x,\varepsilon)N(\varepsilon) = s_0\varepsilon^\beta$ and hence

$$D(x,\varepsilon) = \frac{2}{3s_0}\varepsilon^{1-\beta} I.$$ □

It remains to determine the average collision operator $S(F)$ (see (4.75)). The precise structure of this term depends on the assumptions on the inelastic collision operator Q_{inel}. Here we only notice that a simplified expression is given by the Fokker-Planck approximation [60]

$$S(F) = \frac{\partial}{\partial \varepsilon}\left[s_1 \varepsilon^{\beta} N(\varepsilon)^2 \left(F + \frac{\partial F}{\partial \varepsilon}\right)\right]. \tag{4.86}$$

Remark 4.18 The SHE equations read in unscaled form

$$N(\varepsilon)\partial_t F - \frac{1}{q}\mathrm{div}_x J - E_{\text{eff}} \cdot \frac{\partial J}{\partial \varepsilon} = S(F),$$

$$J(x,\varepsilon,t) = D(x,\varepsilon)\left(\nabla_x F - qE_{\text{eff}}\frac{\partial F}{\partial \varepsilon}\right),$$

where

$$\begin{aligned} N(\varepsilon) &= \int_B \delta(\varepsilon - E(k))\,dk, \\ D(x,\varepsilon) &= \frac{q}{\hbar}\int_B \nabla_k E(k) \otimes \mu(x,k)\delta(\varepsilon - E(k))\,dk. \end{aligned}$$

□

4.4 Derivation of the energy-transport equations

The SHE model of Section 4.3 is much simpler than the semi-classical Boltzmann equation since it has a parabolic structure and has to be solved in the one-dimensional energy variable instead of the three-dimensional wave vector. However, it is numerically more expensive than usual macroscopic models due to the additional energy variable. In this section we derive a macroscopic energy-transport model from the SHE model following [8].

We start with the scaled SHE equations of Theorem 4.14

$$\begin{aligned} N(\varepsilon)\partial_t F - \mathrm{div} J - \nabla V \cdot \frac{\partial J}{\partial \varepsilon} &= S(F), \qquad x \in \mathbb{R}^3,\ \varepsilon \in \mathbb{R},\ t > 0, && (4.87) \\ J(x,\varepsilon,t) &= D(x,\varepsilon)\left(\nabla F + \nabla V \cdot \frac{\partial F}{\partial \varepsilon}\right), && (4.88) \end{aligned}$$

for the distribution function $F(x,\varepsilon,t)$ depending on the position x, energy ε and time t. We have assumed that the effective field in Theorem 4.14 can be written as $E_{\text{eff}} = -\nabla V$ with the electrostatic potential $V(x,t)$ which is assumed to be a given function. The density of states $N(\varepsilon)$, the diffusion matrix $D(x,\varepsilon)$, and the averaged inelastic collision operator $S(F)$ are defined in (4.73), (4.74), and (4.75), respectively. We recall that the inelastic collision operator contains the inelastic corrections to phonon scattering and electron-electron collisions.

Our main assumption is that the electron-electron collision operator is dominant in the sense

$$S(F) = \frac{1}{\alpha} S_{\text{ee}}(F) + S_{\text{ph}}(F),$$

where S_{ee} and $S_{\text{ph}}(F)$ are given by

$$\begin{aligned}
(S_{\text{ee}}(F))(x,\varepsilon,t) &= \int_B (Q_{\text{ee}}(F))(x,k,t)\delta(\varepsilon - E(k))\,dk, \\
(S_{\text{ph}}(F))(x,\varepsilon,t) &= \int_B (Q_{\text{ac}}(F) + Q_{\text{op}}(F))(x,k,t)\delta(\varepsilon - E(k))\,dk,
\end{aligned}$$

and the collision operators Q_{ee}, Q_{ac} and Q_{op} model electron-electron scattering and scattering with acoustic or optical phonons and are defined by (3.30) and (3.29), respectively. Our assumption means that the energy loss due to phonon scattering occurs on a longer time scale than carrier-carrier collisions. Note that the rescaled SHE equation

$$N(\varepsilon)\partial_t F - \text{div} J - \nabla V \cdot \frac{\partial J}{\partial \varepsilon} = \frac{1}{\alpha} S_{\text{ee}}(F) + S_{\text{ph}}(F) \tag{4.89}$$

corresponds to the hyperbolic scaling of Section 4.2 (see (4.33)).

We need the following properties of the operators

$$\begin{aligned}
(S_{\text{ee}}(F))(x,\varepsilon,t) &= \int_{B^4} s_{\text{ee}}(k,k',k_1,k_1')[F'F_1'(1-F)(1-F_1) - FF_1(1-F')(1-F_1')] \\
&\qquad \times \delta(\varepsilon - E(k))\,dk\,dk'\,dk_1\,dk_1', \qquad\qquad (4.90) \\
(S_{\text{ph}}(F))(x,\varepsilon,t) &= \int_{B^2} [s_{\text{ph}}(k',k)F'(1-F) - s_{\text{ph}}(k,k')F(1-F')]\delta(\varepsilon - E(k))\,dk\,dk',
\end{aligned}$$

where $F = F(E(k))$, $F' = F(E(k'))$, $F_1 = F(E(k_1))$, $F_1' = F(E(k_1'))$, and

$$s_{\text{ee}}(k,k',k_1,k_1') = \sigma_{\text{ee}}(x,k,k',k_1,k_1')\delta(E(k) + E(k_1) - E(k') - E(k_1'))$$

has the properties

$$s_{\text{ee}}(k,k',k_1,k_1') = s_{\text{ee}}(k_1,k',k,k_1') = s_{\text{ee}}(k',k,k_1',k_1). \tag{4.91}$$

Lemma 4.19 *Let $S_{\text{ee}}(F)$ and $S_{\text{ph}}(F)$ be given as above satisfying (4.91). Then*

$$\int_{\mathbb{R}} S_{\text{ee}}(F)\varepsilon^i d\varepsilon = 0 \quad (i = 0,1), \qquad \int_{\mathbb{R}} S_{\text{ph}}(F) d\varepsilon = 0.$$

Proof: The second assertion follows directly by integrating over $\varepsilon \in \mathbb{R}$ and using the definition of the delta distribution:

$$\begin{aligned}
\int_{\mathbb{R}} S_{\text{ph}}(F) d\varepsilon &= \int_{B^2} [s_{\text{ph}}(k',k)F'(1-F) - s_{\text{ph}}(k,k')F(1-F')])\,dk\,dk' \\
&= \int_{B^2} s_{\text{ph}}(k',k)F'(1-F)\,dk\,dk' - \int_{B^2} s_{\text{ph}}(k',k)F'(1-F)\,dk'\,dk \\
&= 0.
\end{aligned}$$

To prove the first assertion we show a more general result. It holds for all F and G:

$$\begin{aligned}\int_{\mathbb{R}} S_{\mathrm{ee}}(F)G\,d\varepsilon &= -\frac{1}{4}\int_{B^4} s(k,k',k_1,k_1')[F'F_1'(1-F)(1-F_1)-FF_1(1-F')(1-F_1')] \\ &\qquad \times(G'+G_1'-G-G_1)\,d^4k, \end{aligned} \tag{4.92}$$

where $d^4k = dk\,dk'\,dk_1\,dk_1'$. Indeed, we obtain from (4.91)

$$\begin{aligned}&\int_{\mathbb{R}} S_{\mathrm{ee}}(F)G\,d\varepsilon \\ &= \int_{B^4} Q_{\mathrm{ee}}(F(E(k))G(E(k))\,dk \\ &= \frac{1}{4}\int_{B^4} s_{\mathrm{ee}}(k,k',k_1,k_1')[F'F_1'(1-F)(1-F_1)-FF_1(1-F')(1-F_1')]G\,d^4k \\ &\quad +\frac{1}{4}\int_{B^4} s_{\mathrm{ee}}(k',k,k_1',k_1)[FF_1(1-F')(1-F_1')-F'F_1'(1-F)(1-F_1)]G'\,d^4k \\ &\quad +\frac{1}{4}\int_{B^4} s_{\mathrm{ee}}(k_1,k',k,k_1')[F'F_1'(1-F_1)(1-F)-F_1F(1-F')(1-F_1')]G_1\,d^4k \\ &\quad +\frac{1}{4}\int_{B^4} s_{\mathrm{ee}}(k_1',k_1,k',k)[F_1F(1-F_1')(1-F')-F_1'F'(1-F_1)(1-F)]G_1'\,d^4k \\ &= -\frac{1}{4}\int_{B^4} s_{\mathrm{ee}}(k,k',k_1,k_1')[F'F_1'(1-F)(1-F_1)-FF_1(1-F')(1-F_1')] \\ &\qquad \times(G-G'+G_1-G_1')\,d^4k. \end{aligned}$$

□

It remains to show that the first assertion follows from (4.92). Taking $G(\varepsilon)=1$ it is clear that

$$\int_{\mathbb{R}} S_{\mathrm{ee}}(F)\,d\varepsilon = 0.$$

Choosing $G(\varepsilon)=\varepsilon$, we obtain an integral over B^4 involving the product

$$\delta(E+E_1-E'-E_1')(E'+E_1'-E-E_1)$$

which vanishes in view of the properties of the δ distribution. This shows that

$$\int_{\mathbb{R}} S_{\mathrm{ee}}(F)\varepsilon\,d\varepsilon = 0.$$

Lemma 4.20 *(1) The kernel of S_{ee} is given by*

$$N(S_{\mathrm{ee}}) = \{F:\mathbb{R}\to\mathbb{R} : \exists \mu\in\mathbb{R},\, T>0 : F = F_{\mu,T}\},$$

where

$$F_{\mu,T}(\varepsilon) = \frac{1}{1+e^{(\varepsilon-\mu)/T}}.$$

(2) For given F_0, the equation $DS_{\mathrm{ee}}(F_0)F = G$ is solvable if and only if

$$\int_{\mathbb{R}} G(\varepsilon)\, d\varepsilon = 0 \qquad \textit{and} \qquad \int_{\mathbb{R}} G(\varepsilon)\varepsilon\, d\varepsilon = 0,$$

where $DS_{\mathrm{ee}}(F_0)$ is the first derivative of S_{ee} with respect to F_0.

The variable μ is called *chemical potential* and T the *electron temperature.* The function $F_{\mu,T}$ is termed *Fermi-Dirac distribution* and has already been derived in Theorem 3.7 under different assumptions on the collision operator.

Proof: We only give a sketch of the proof and refer to [8, Sec. 4.2] for details.

(1) Choosing $G = H(F) = \ln F - \ln(1-F)$ in (4.92) yields

$$\begin{aligned}
&\int_{\mathbb{R}} S_{\mathrm{ee}}(F)H(F)\, d\varepsilon \\
&= \; -\frac{1}{4}\int_{B^4} s_{\mathrm{ee}}(k,k',k_1,k_1')[F'F_1'(1-F)(1-F_1) - FF_1(1-F')(1-F_1')] \\
&\qquad \times[\ln(F'F_1'(1-F)(1-F_1)) - \ln(FF_1(1-F')(1-F_1'))]\, d^4k \\
&\leq \; 0
\end{aligned}$$

since $(x-y)(\ln x - \ln y) \geq 0$ for all $x, y\, 0$. Thus, taking $F \in N(S_{\mathrm{ee}})$, we obtain

$$\ln(F'F_1'(1-F)(1-F_1)) = \ln(FF_1(1-F')(1-F_1'))$$

or

$$H(F') + H(F_1') = H(F) + H(F_1)$$

when $E(k') + E(k_1') = E(k) + E(k_1)$ (which follows from the δ distribution in (4.90)) and $k' + k_1' = k + k_1$ modulo B (which follows from the so-called umklapp processes, see [8, (2.5)]). It can be motivated that

$$H(F(\varepsilon - c)) + H(F(\varepsilon + c)) = H(F(\varepsilon)) + H(F(\varepsilon))$$

for some $c > 0$. This means that the function $x \mapsto g(x) := H(F(x))$ fulfills the functional equation

$$g(x-c) + g(x+c) = 2g(x) \qquad \text{for all } x \in \mathbb{R} \text{ and some } c = c(x).$$

The only functions solving such an identiy are affine, i.e. $g(x) = ax + b$. Then, introducing formally

$$T = -1/a \qquad \text{and} \qquad \mu = -b/a,$$

we obtain from

$$\ln\frac{F}{1-F} = H(F) = a\varepsilon + b = -\frac{\varepsilon - \mu}{T}$$

the equation

$$F(\varepsilon) = \frac{1}{1+e^{(\varepsilon-\mu)/T}}.$$

(2) The assertion is a consequence of the Fredholm alternative Lemma 4.3; we refer to [8, Prop. 4.2]. □

We now employ the Hilbert expansion

$$F = F_0 + \alpha F_1 + \alpha^2 F_2 + \cdots, \qquad J = J_0 + \alpha J_1 + \alpha^2 J_2 + \cdots \tag{4.93}$$

and Taylor approximation for the collision operators in (4.89):

$$\begin{aligned} N(\varepsilon)\partial_t F_0 - \mathrm{div} J_0 - \nabla V \cdot \frac{\partial J_0}{\partial \varepsilon} &= \frac{1}{\alpha}(S_{\mathrm{ee}}(F_0) + \alpha D S_{\mathrm{ee}}(F_0) F_1) \\ &\qquad + S_{\mathrm{ph}}(F_0) + \alpha D S_{\mathrm{ph}}(F_0) F_1 + O(\varepsilon^2). \end{aligned}$$

Identifying equal powers of α leads to

$$S_{\mathrm{ee}}(F_0) = 0, \quad N(\varepsilon) = \partial_t F_0 - \mathrm{div} J_0 - \nabla V \cdot \frac{\partial J_0}{\partial \varepsilon} = D S_{\mathrm{ee}}(F_0) F_1 + S_{\mathrm{ph}}(F_0). \tag{4.94}$$

Theorem 4.21 *The Hilbert expansion (4.93) is solvable up to first order if and only if* $F_0(x,\varepsilon,t) = F_{\mu(x,t),T(x,t)}(\varepsilon)$ *and* $\mu(x,t), T(x,t)$ *satisfy formally the equations*

$$\begin{aligned} \partial_t n(\mu,t) - \mathrm{div} J_n &= 0, \\ \partial_t (ne)(\mu,t) - \mathrm{div} J_e + \nabla V \cdot J_n &= W(\mu,T), \\ J_n &= D_{11} \nabla \left(\frac{\mu}{T}\right) + D_{12} \nabla \left(\frac{-1}{T}\right) - D_{11} \frac{\nabla V}{T}, \\ J_e &= D_{21} \nabla \left(\frac{\mu}{T}\right) + D_{22} \nabla \left(\frac{-1}{T}\right) - D_{21} \frac{\nabla V}{T}, \end{aligned}$$

where

$$n(\mu,T) = \int_{\mathbb{R}} F_{\mu,T}(\varepsilon) N(\varepsilon)\, d\varepsilon, \tag{4.95}$$

$$ne(\mu,T) = \int_{\mathbb{R}} F_{\mu,T}(\varepsilon) \varepsilon N(\varepsilon)\, d\varepsilon, \tag{4.96}$$

are the electron density and energy density, respectively,

$$W(\mu,T) = \int_{\mathbb{R}} S_{\mathrm{ph}}(F_{\mu,T}) \varepsilon \, d\varepsilon$$

is the relaxation term, and

$$D_{ij}(x,\mu,T) = \int_{\mathbb{R}} D(x,\varepsilon) F_{\mu,T}(1-F_{\mu,T}) \varepsilon^{i+j-2}\, d\varepsilon, \qquad i,j = 1,2, \tag{4.97}$$

are diffusion matrices in $\mathbb{R}^{3\times 3}$.

The variables J_n *and* J_e *are called the macroscopic partial and energy current densities, respectively, and* μ/T *and* $-1/T$ *are the so-called entropy variables.*

Proof: From Lemma 4.20 (1) follows that the solutions of the first equation in (4.94) are given by

$$F_0(x,\varepsilon,t) = F_{\mu,T}(\varepsilon),$$

where μ and T also depend on the parameters x and t. Lemma 4.20 (2) implies that the second equation in (4.94) is solvable if and only if

$$\int_{\mathbb{R}} \left(N(\varepsilon)\partial_t F_{\mu,T} - \mathrm{div} J_0 - \nabla V \cdot \frac{\partial J_0}{\partial \varepsilon} - S_{\mathrm{ph}}(F_{\mu,T}) \right) \varepsilon^i \, d\varepsilon = 0, \qquad i = 0, 1. \tag{4.98}$$

Defining the macroscopic particle and energy current densities by

$$J_n(x,t) = \int_{\mathbb{R}} J_0(x,\varepsilon,t)\, d\varepsilon, \qquad J_e(x,t) = \int_{\mathbb{R}} J_0(x,\varepsilon,t)\varepsilon \, d\varepsilon,$$

we can reformulate (4.98) as

$$\begin{aligned}
\partial_t n - \mathrm{div} J_n &= \nabla V \cdot \int_{\mathbb{R}} \frac{\partial J_0}{\partial \varepsilon}\, d\varepsilon + \int_{\mathbb{R}} S_{\mathrm{ph}}(F_{\mu,T})\, d\varepsilon = 0, \\
\partial_t (ne) - \mathrm{div} J_e &= \nabla V \cdot \int_{\mathbb{R}} \frac{\partial J_0}{\partial \varepsilon}\varepsilon \, d\varepsilon + \int_{\mathbb{R}} S_{\mathrm{ph}}(F_{\mu,T})\varepsilon \, d\varepsilon \\
&= -\nabla V \cdot J + W(\mu, T),
\end{aligned}$$

using Lemma 4.19.

It remains to compute the fluxes J_n and J_e. We interpret $F_{\mu,T}(\varepsilon)$ as a function of the entropy variables $u_1 = \mu/T$, $u_2 = -1/T$, and ε,

$$F(u_1, u_2, \varepsilon) = F_{\mu,T}(\varepsilon).$$

Then

$$F(u_1, u_2, \varepsilon) = \frac{1}{1 + e^{-(u_1 + \varepsilon u_2)}}$$

and

$$\begin{aligned}
\frac{\partial F}{\partial u_1} &= \frac{e^{-(u_1+\varepsilon u_2)}}{(1 + e^{-(u_1+\varepsilon u_2)})^2} = F_{\mu,T}(1 - F_{\mu,T}), \\
\frac{\partial F}{\partial u_2} &= \frac{\varepsilon e^{-(u_1+\varepsilon u_2)}}{(1 + e^{-(u_1+\varepsilon u_2)})^2} = \varepsilon F_{\mu,T}(1 - F_{\mu,T}), \\
\frac{\partial F}{\partial \varepsilon} &= \frac{u_2 e^{-(u_1+\varepsilon u_2)}}{(1 + e^{-(u_1+\varepsilon u_2)})^2} = -\frac{1}{T} F_{\mu,T}(1 - F_{\mu,T}).
\end{aligned}$$

This yields

$$J_0 = D(x,\varepsilon)\left(\nabla F_{\mu,T} + \nabla V \cdot \frac{\partial F_{\mu,T}}{\partial \varepsilon} \right) = D(x,\varepsilon)\left(\frac{\partial F}{\partial u_1}\nabla u_1 + \frac{\partial F}{\partial u_2}\nabla u_2 + \nabla V \cdot \frac{\partial F}{\partial \varepsilon} \right)$$

and

$$\begin{aligned} J_n &= D_{11}\nabla u_1 + D_{12}\nabla u_2 - D_{11}\frac{\nabla V}{T}, \\ J_e &= D_{21}\nabla u_2 + D_{22}\nabla u_2 - D_{21}\frac{\nabla V}{T} \end{aligned}$$

and completes the proof. $\square$

The matrix $\mathcal{D} = (D_{ij}) \in \mathbb{R}^{6\times 6}$ has the following properties.

Proposition 4.22 *It holds:*

(1) The matrix $\mathcal{D} = (D_{ij})$ *is symmetric. Moreover,* $D_{12} = D_{21}$ *and* $D_{ij}^\top = D_{ji}$ *for* $i, j = 1, 2$.

(2) If the functions

$$\nabla_k E, \; E\nabla_k E \tag{4.99}$$

are linearly independent then $\mathcal{D}(x,\mu,T)$ *is symmetric, positive definite for any* $\mu \in \mathbb{R}$ *and* $T > 0$.

The hypothesis (4.99) is a geometric assumption on the band structure. It requires that the energy band has a real three-dimensional structure, excluding bands depending only on one or two variables, for instance.

Proof: Part (1) follows from the definition (4.97) of the (3×3)-matrices D_{ij} and Lemma 4.15. For the proof of part (2) we choose $z = (\xi, \eta) \in \mathbb{R}^6$ with $\xi, \eta \in \mathbb{R}^3$ such that $\xi \neq 0$. Then, by the symmetry of $D(x,\varepsilon)$ and the property (4.80), we compute

$$\begin{aligned} z^\top \mathcal{D} z &= \int_{\mathbb{R}} z^\top \begin{pmatrix} D(x,\varepsilon) & \varepsilon D(x,\varepsilon) \\ \varepsilon D(x,\varepsilon) & \varepsilon^2 D(x,\varepsilon) \end{pmatrix} z F_{\mu,T}(1-F_{\mu,T})\, d\varepsilon \\ &= \int_{\mathbb{R}} (\xi + \varepsilon\eta)^\top D(x,\varepsilon)(\xi+\varepsilon\eta) F_{\mu,T}(1-F_{\mu,T})\, d\varepsilon \\ &\geq \int_{\mathbb{R}} \frac{K}{N(\varepsilon)} \int_B \sum_{i,j=1}^3 (\xi_i + \varepsilon\eta_i)\frac{\partial E}{\partial k_i}\frac{\partial E}{\partial k_j}(\xi_j + \varepsilon\eta_j)\delta(\varepsilon - E(k)) F_{\mu,T}(1-F_{\mu,T})\, dk\, d\varepsilon \\ &= \int_B \frac{K}{N(E(k))}\left(\sum_{j=1}^3 (\xi_j + E(k)\eta_j)\frac{\partial E}{\partial k_j}\right)^2 F_{\mu,T}(1-F_{\mu,T})\, dk \\ &= \int_B \frac{K}{N(E(k))}\left| \begin{pmatrix} \nabla_k E \\ E\nabla_k E \end{pmatrix} \cdot z \right|^2 F_{\mu,T}(1-F_{\mu,T})\, dk. \end{aligned}$$

The last integral is positive since otherwise,

$$\begin{pmatrix} \nabla_k E \\ E\nabla_k E \end{pmatrix} \cdot z = 0, \qquad z \neq 0,$$

would imply that $(\nabla_k E, E\nabla_k E)$ is linearly dependent which is excluded. $\square$

The following result shows that the name “relaxation term” for $W(\mu, T)$ is justified. In fact, the temperature of the particles is expected to relax to the constant lattice temperature $T_L = 1$ if there are no other forces.

Proposition 4.23 *The relaxation term $W(\mu, T)$ with phonon collision operator (4.75) and (3.29) is monotone in the sense of operator theory, i.e.*

$$W(\mu, T)(T - 1) \leq 0.$$

Proof: We recall the definition of $W(\mu, T)$,

$$\begin{aligned} W(\mu, T) &= \int_{\mathbb{R}} S_{\text{ph}}(\varepsilon)\varepsilon \, d\varepsilon \\ &= \int_{\mathbb{R}} \int_{B^2} [s_{\text{ph}}(k', k)F'(1 - F) - s_{\text{ph}}(k, k')F(1 - F')]\delta(\varepsilon - E(k))\varepsilon \, dk \, dk' \, d\varepsilon, \end{aligned}$$

where $F = F_{\mu,T}$,

$$s_{\text{ph}}(k, k') = \sigma_{\text{ph}}(x, k, k')[(1 + N_{\text{ph}})\delta(E' - E + E_{\text{ph}}) + N_{\text{ph}}\delta(E' - E + E_{\text{ph}})], \qquad (4.100)$$

and $E = E(k)$, $E' = E(k')$, $N_{\text{ph}} = (e^{E_{\text{ph}}} - 1)^{-1}$, and $E_{\text{ph}} > 0$ is the phonon energy. Since

$$F = (1 - F)M, \qquad M = e^{-(\varepsilon-\mu)/T},$$

we obtain

$$W(\mu, T) = \int_{B^2} (1 - F')(1 - F)[s_{\text{ph}}(k', k)M' - s_{\text{ph}}(k, k')M]E(k) \, dk \, dk'.$$

Definition (4.100) and the properties $\delta(x) = \delta(-x)$ and $\delta(x - x_0)x = \delta(x - x_0)x_0$ yield

$$\begin{aligned} &[s_{\text{ph}}(k', k)M' - s_{\text{ph}}(k, k')M]E(k) \\ &\quad = \sigma_{\text{ph}}(1 + N_{\text{ph}})\delta(E - E' + E_{\text{ph}})M'E + N_{\text{ph}}\delta(E - E' - E_{\text{ph}})M'E \\ &\qquad\qquad -\sigma_{\text{ph}}(1 + N_{\text{ph}})\delta(E - E' - E_{\text{ph}})ME - N_{\text{ph}}\delta(E - E' + E_{\text{ph}})ME \\ &\quad = \sigma_{\text{ph}}\delta(E - E' + E_{\text{ph}})[(1 + N_{\text{ph}})M' - N_{\text{ph}}M]E \\ &\qquad\qquad +\sigma_{\text{ph}}\delta(E - E' - E_{\text{ph}})[N_{\text{ph}}M' - (1 + N_{\text{ph}})M](E' + E_{\text{ph}}). \end{aligned}$$

Thus, exchanging k and k' in the second integral and using $\delta(x) = \delta(-x)$ again, some terms cancel and we end up with

$$\begin{aligned} W(\mu, T) &= \int_{B^2} (1-F')(1-F)\delta(E - E' + E_{\text{ph}})E_{\text{ph}}[N_{\text{ph}}M - (1+N_{\text{ph}})M']\,dk\,dk' \\ &= \int_{B^2} (1-F')(1-F)\delta(E - E' + E_{\text{ph}})E_{\text{ph}}N_{\text{ph}}[M - e^{E_{\text{ph}}}M']\,dk\,dk', \end{aligned}$$

since $1 + N_{\text{ph}} = e^{E_{\text{ph}}}N_{\text{ph}}$. The δ distribution allows to substitute E in M to $E' - E_{\text{ph}}$ such that

$$M - e^{E_{\text{ph}}}M' = e^{-(E-\mu)/T} - e^{E_{\text{ph}}}e^{-(E'-\mu)/T} = e^{-(E'-\mu)/T}\left(e^{E_{\text{ph}}/T} - e^{E_{\text{ph}}}\right)$$

for all E and E' satisfying $E = E' - E_{\text{ph}}$. We conclude that

$$\begin{aligned} W(\mu, T)(T-1) &= \int_{B^2} (1-F')(1-F)\delta(E - E' + E_{\text{ph}})E_{\text{ph}}N_{\text{ph}}M' \\ &\qquad \times \left(e^{E_{\text{ph}}/T} - e^{E_{\text{ph}}}\right)(T-1)\,dk\,dk' \\ &\le 0, \end{aligned}$$

since the function $x \mapsto e^{E_{\text{ph}}/x}$ is decreasing and hence,

$$\left(e^{E_{\text{ph}}/T} - e^{E_{\text{ph}}}\right)(T-1) \le 0 \qquad \text{for all } T > 0. \qquad \square$$

More explicit expressions for the particle and energy densities and for the diffusion matrix can be derived for spherically symmetric energy bands and Maxwell-Boltzmann statistics.

Example 4.24 (Spherically symmetric energy bands)
We impose the following simplifying hypotheses:

- The distribution function $F_{\mu,T}$ is approximated by $e^{-(\varepsilon-\mu)/T}$. This is possible if $\varepsilon - \mu \gg T$.
- The scattering rate σ of the elastic collision operator (4.59) depends only on x and $E(k)$.
- The energy band is spherically symmetric and strictly monotone in $|k|$.

From Example 4.16 and definitions (4.95)–(4.96) of the particle and energy densities follows

$$\begin{aligned} n(\mu, T) &= \int_{\mathbb{R}} e^{-(\varepsilon-\mu)/T}N(\varepsilon)\,d\varepsilon = 2\pi e^{\mu/T}\int_{\mathbb{R}} \sqrt{\gamma(\varepsilon)}\gamma'(\varepsilon)e^{-\varepsilon/T}\,d\varepsilon, \\ ne(\mu, T) &= \int_{\mathbb{R}} e^{-(\varepsilon-\mu)/T}N(\varepsilon)\varepsilon\,d\varepsilon = 2\pi e^{\mu/T}\int_{\mathbb{R}} \sqrt{\gamma(\varepsilon)}\gamma'(\varepsilon)\varepsilon e^{-\varepsilon/T}\,d\varepsilon, \end{aligned}$$

where we have used the expression (4.81) for $N(\varepsilon)$ and $|k|^2 = \gamma(E(k))$. From (4.82) the diffusion matrices become

$$D_{ij}(\mu, T) = \frac{4}{3} e^{\mu/T} \int_{\mathbb{R}} \frac{\gamma(\varepsilon)\varepsilon^{i+j-2}}{s(x,\varepsilon)\gamma'(\varepsilon)^2} e^{-\varepsilon/T}\, d\varepsilon \cdot I,$$

where $I \in \mathbb{R}^{3\times 3}$ is the identity matrix. $\square$

Example 4.25 (Parabolic band approximation)
We impose the same assumptions as in the previous example and, additionally,

- The energy band is parabolic: $E(k) = |k|^2/2$, $k \in \mathbb{R}^3$.
- The scattering rate is given by the so-called *Chen model* [21, 24]: $s(x,\varepsilon) = s_0(x)\sqrt{\varepsilon}$.

The first assumption implies that $\gamma(\varepsilon) = 2\varepsilon$, $\varepsilon \ge 0$. The second hypothesis means that the relaxation time (4.85) in the formulation (4.83) of the elastic collision operator,

$$\tau(x,\varepsilon) = \frac{1}{s(x,\varepsilon)N(\varepsilon)} = \frac{1}{s_0(x)\varepsilon^\beta N(\varepsilon)},$$

holds with $\beta = 1/2$. Recall that $N(\varepsilon) = 4\pi\sqrt{2\varepsilon}$ by Example 4.17. Using

$$\int_0^\infty e^{-z^2/2} z^2\, dz = \frac{1}{2}\int_{\mathbb{R}} e^{-z^2/2} z^2\, dz = \frac{1}{2}\int_{\mathbb{R}} e^{-z^2/2}\, dz = \sqrt{\frac{\pi}{2}}$$

gives

$$\begin{aligned} n(\mu, T) &= 4\sqrt{2}\pi e^{\mu/T} \int_0^\infty e^{-\varepsilon/T}\sqrt{\varepsilon}\, d\varepsilon = 4\sqrt{2}\pi e^{\mu/T} \frac{T^{3/2}}{\sqrt{2}} \int_0^\infty e^{-z^2/2} z^2\, dz \\ &= (2\pi T)^{3/2} e^{\mu/T}. \end{aligned} \tag{4.101}$$

Furthermore, since

$$\int_0^\infty e^{-z^2/2} z^4\, dz = \frac{1}{2}\int_{\mathbb{R}} e^{-z^2/2} z^4\, dz = \frac{3}{2}\sqrt{2\pi}$$

(see (4.44)), we obtain

$$\begin{aligned} (ne)(\mu, T) &= 4\sqrt{2}\pi e^{\mu/T} \int_0^\infty e^{-\varepsilon/T}\varepsilon^{3/2}\, d\varepsilon = 4\sqrt{2}\pi e^{\mu/T} \frac{T^{5/2}}{2^{3/2}} \int_0^\infty e^{-z^2/2} z^4\, dz \\ &= \frac{3}{2}(2\pi)^{3/2} T^{5/2} e^{\mu/T} = \frac{3}{2} nT. \end{aligned}$$

The diffusion matrices become (see Example 4.24)

$$\begin{aligned} D_{ij}(x, \mu, T) &= \frac{2}{3}\frac{e^{\mu/T}}{s_0(x)} \int_{\mathbb{R}} \varepsilon^{i+j-3/2} e^{-\varepsilon/T}\, d\varepsilon \\ &= \frac{2}{3}\frac{e^{\mu/T}}{s_0(x)} \left(\frac{T}{2}\right)^{i+j-3/2} T \int_0^\infty z^{2(i+j-1)} e^{-z^2/2}\, dz. \end{aligned}$$

Therefore, since

$$\int_0^\infty e^{-z^2/2} z^6\,dz = \frac{1}{2}\int_{\mathbb{R}} e^{-z^2/2} z^6\,dz = \frac{5\cdot 3}{2}\sqrt{2\pi}$$

(see (4.44)), we infer that

$$\begin{aligned} D(x,\mu,T) &= \frac{\sqrt{\pi}}{3s_0(x)} T^{3/2} e^{\mu/T} \begin{pmatrix} I & \frac{3}{2}TI \\ \frac{3}{2}TI & \frac{15}{4}T^2 I \end{pmatrix} \\ &= \frac{n(\mu,T)}{6\sqrt{2}\pi s_0(x)} \begin{pmatrix} I & \frac{3}{2}TI \\ \frac{3}{2}TI & \frac{15}{4}T^2 I \end{pmatrix} \in \mathbb{R}^{6\times 6}. \end{aligned}$$

Notice that the matrix $D(x,\mu,T)$ can be identified by the (2×2)-matrix

$$D(x,\mu,T) = \frac{n(\mu,T)}{6\sqrt{2}\pi s_0(x)} \begin{pmatrix} 1 & \frac{3}{2}T \\ \frac{3}{2}T & \frac{15}{4}T^2 \end{pmatrix}.$$

Example 4.26 (Fokker-Planck relaxation term)
The relaxation term $W(\mu,T)$ can be written explicitely in terms of μ and T using the Fokker-Planck approximation (4.86) and the assumptions of the previous example. Then, by integrating by parts,

$$\begin{aligned} W(\mu,T) &= \int_0^\infty \frac{\partial}{\partial\varepsilon}\left[s_1\varepsilon^{1/2}N(\varepsilon)^2\left(F_{\mu,T} + \frac{\partial F_{\mu,T}}{\partial\varepsilon}\right)\right]\varepsilon\,d\varepsilon \\ &= 2(4\pi)^2 s_1 \int_0^\infty \frac{\partial}{\partial\varepsilon}\left[\varepsilon^{3/2} e^{-(\varepsilon-\mu)/T}\left(1-\frac{1}{T}\right)\right]\varepsilon\,d\varepsilon \\ &= -2(4\pi)^2 s_1 \int_0^\infty \varepsilon^{3/2} e^{-(\varepsilon-\mu)/T}\left(1-\frac{1}{T}\right)d\varepsilon \\ &= -2(4\pi)^2 s_1 \cdot \frac{3}{4}\sqrt{\pi}T^{5/2}\cdot e^{-\mu/T}\left(1-\frac{1}{T}\right) \\ &= \frac{3}{2}\,\frac{n(\mu,T)(1-T)}{\tau_0}, \end{aligned}$$

where

$$\tau_0 = \frac{1}{4\sqrt{2}\pi s_1}$$

is called the *energy relaxation time.* $\square$

Remark 4.27 The energy-transport model of Theorem 4.21 ist written with physical parameters as

$$\begin{aligned}
\partial_t n(\mu,T) - \frac{1}{q}\mathrm{div} J_n &= 0,\\
\partial_t (ne)(\mu,T) - \frac{1}{q}\mathrm{div} J_e + \nabla V \cdot J_n &= W(\mu,T),\\
J_n &= D_{11}\nabla\left(\frac{q\mu}{k_BT}\right) + D_{12}\nabla\left(\frac{-1}{k_BT}\right) - D_{11}\frac{q\nabla V}{k_BT},\\
J_e &= D_{21}\nabla\left(\frac{q\mu}{k_BT}\right) + D_{22}\nabla\left(\frac{-1}{k_BT}\right) - D_{21}\frac{q\nabla V}{k_BT}.
\end{aligned}$$

In the situation of Examples 4.25 and 4.26 we call this system of equations the *Chen model*. In particular,

$$\begin{aligned}
n(\mu,T) &= \left(\frac{2\pi m_e^* k_BT}{\hbar^2}\right)^{3/2} e^{q\mu/k_BT},\\
(ne)(\mu,T) &= \frac{3}{2} n(\mu,T)\cdot k_BT,\\
D(x,\mu,T) &= \frac{q\hbar^3}{6\sqrt{2}\pi S_0(x)} n(\mu,T)\begin{pmatrix} 1 & \frac{3}{2}k_BT \\ \frac{3}{2}k_BT & \frac{15}{4}(k_BT)^2 \end{pmatrix},\\
W(\mu,T) &= \frac{3}{2}\,\frac{n(\mu,T)k_B(T-T_L)}{\tau_0},
\end{aligned}$$

where T_L is the lattice temperature and $\tau_0 = (\hbar^2/2m_e^*)^{3/2}/2\pi S_1$. □

4.5 Relaxation-time limits

In this chapter we have derived four fluid-type models: the drift-diffusion, hydrodynamic, SHE, and energy-transport equations (see Figure 4.1 for a summary). In this section we show that the drift-diffusion and energy-transport equations can be formally derived from the hydrodynamic model by performing so-called relaxation-time limits.

We recall the unscaled hydrodynamic model with electrostatic potential:

$$\partial_t n - \frac{1}{q}\mathrm{div} J = 0 \tag{4.102}$$

$$\partial_t J - \frac{1}{q}\mathrm{div}\left(\frac{J\otimes J}{n}\right) - \frac{qk_B}{m_e^*}\nabla(nT) + \frac{q^2}{m_e^*}n\nabla V = -\frac{J}{\tau_p}, \tag{4.103}$$

$$\begin{aligned}
&\partial_t(ne) - \frac{1}{q}\mathrm{div}[J(e+k_BT)] + J\cdot\nabla V - \mathrm{div}\left(\frac{\tau_p}{m_e^*}\kappa_0 n k_BT\nabla(k_BT)\right) \qquad (4.104)\\
&\quad = -\frac{n}{\tau_W}\left(e - \frac{d}{2}k_BT_L\right), \qquad x\in\Omega\subset\mathbb{R}^d;\ t>0,
\end{aligned}$$

where the energy density is

$$ne = \frac{m_e^*}{2q^2}\,\frac{|J|^2}{n} + \frac{d}{2}nk_BT.$$

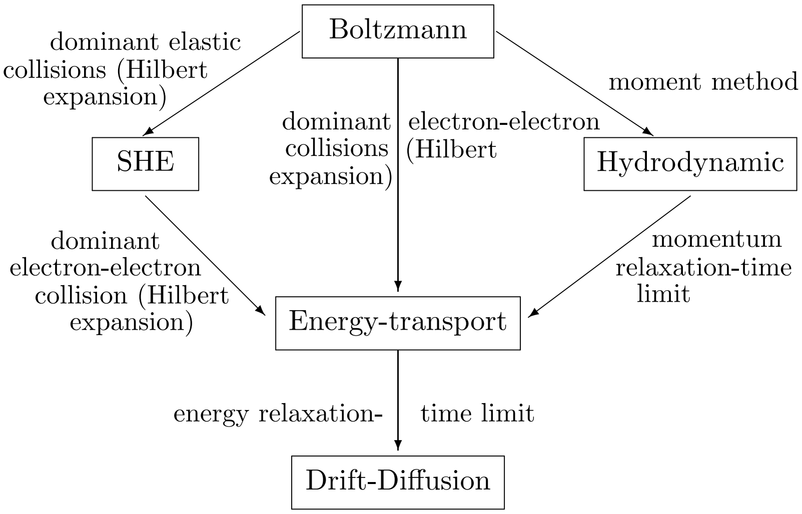

Figure 4.1: Hierarchy of classical semiconductor models. The drift-diffusion model can be also directly derived from the Boltzmann equation via a moment method (see [9]).

These equations deviate from (4.102)–(4.104) by the inclusion of the heat conduction term with the heat conductivity κ_0 (see the end of Section 4.2) and by the introduction of two different relaxation times; the momentum relaxation time τ_p and the energy relaxation time τ_W. The reason is that we wish to consider two different time scales for momentum and energy relaxation.

In order to scale (4.102)–(4.104) we choose the following reference values:

- reference length L,
- reference particle density N,
- reference temperature T_L (lattice temperature),
- reference time τ_0,
- reference potential $U_T = k_B T_L / q$,
- reference particle current density $J_0 = qNL/\tau_0$.

We can choose, for instance, L as the diameter of the semiconductor domain Ω and N as the maximal value of the doping concentration. The reference time τ_0 is given by the assumption that the thermal energy is of the same order as the geometric average of the kinetic energies to cross the domain in time τ_0, τ_p, respectively:

$$k_B T_0 = \sqrt{m_e^* \left(\frac{L}{\tau_0}\right)^2} \sqrt{m_e^* \left(\frac{L}{\tau_p}\right)^2}.$$

With these reference values we define the scaled variables:

$$\begin{aligned} x &= Lx_s, & t &= \tau_0 t_s, & n &= Nn_s, \\ T &= T_L T_s, & V &= U_T V_S, & J &= J_0 J_s. \end{aligned}$$

Replacing the dimensional variables in (4.102)–(4.104) by the scaled ones and omitting the index s we obtain

$$\partial_t n - \mathrm{div} J = 0, \tag{4.105}$$

$$\alpha \partial_t J - \alpha \mathrm{div}\left(\frac{J \otimes J}{n}\right) - \nabla(nT) + n\nabla V = -J, \tag{4.106}$$

$$\partial_t(ne) - \mathrm{div}[J(e+T)] + J \cdot \nabla V - \mathrm{div}\,(\kappa_0 nT\nabla T) = -\frac{n}{\beta}\left(e - \frac{d}{2}\right), \tag{4.107}$$

where

$$\alpha = \frac{\tau_p}{\tau_0}, \qquad \beta = \frac{\tau_W}{\tau_0},$$

and the scaled energy is given by

$$e = \alpha \frac{|J|^2}{2n^2} + \frac{d}{2}T. \tag{4.108}$$

We consider the following limits:

(1) $\alpha \to 0$ and $\beta \to 0$,

(2) $\alpha \to 0$ and β fixed,

(3) α fixed and $\beta \to 0$.

(1) The combined limit $\alpha \to 0$ and $\beta \to 0$ corresponds to the physical situation when the kinetic energy needed to cross the domain in time τ_0 is assumed to be much smaller than the thermal energy, and when the momentum and energy relaxation times are assumed to be of the same order. This means

$$1 \gg \frac{m_e^*(L/\tau_0)^2}{k_B T_0} = \frac{\tau_p}{\tau_0} = \alpha \qquad \text{and} \qquad O(1) = \frac{\tau_p}{\tau_W} = \frac{\alpha}{\beta}. \tag{4.109}$$

The second condition implies that also $\beta \ll 1$. Equations (4.105)–(4.107) become in the limit $\alpha \to 0$ and $\beta \to 0$

$$\partial_t n - \mathrm{div} J = 0, \qquad J = \nabla(nT) - n\nabla V, \qquad e = \frac{d}{2}. \tag{4.110}$$

Moreover, by (4.108), $e = (d/2)T$, and the temperature is $T = 1$. Then the limit equations (4.110) are the drift-diffusion equations.

(2) The limit $\alpha \to 0$ corresponds to the physical situation expressed by the first equation in (4.109). Then (4.105)–(4.107) yield in the limit $\alpha \to 0$

$$\partial_t n - \mathrm{div} J = 0, \qquad J = \nabla(nT) - n\nabla V \tag{4.111}$$

and

$$\partial_t \left(\frac{d}{2} nT\right) - \mathrm{div}\left(\frac{d+2}{2} JT + \kappa_0 nT\nabla T\right) + J\cdot\nabla V = -\frac{d}{2}\frac{n}{\beta}(T-1). \tag{4.112}$$

In the entropic variables $u_1 = \mu/T$ and $u_2 = -1/T$ we can write in view of the expression $n = (2\pi T)^{d/2} e^{\mu/T}$ (see (4.95) for the case $d = 3$),

$$J = \frac{\partial(nT)}{\partial u_1}\nabla u_1 + \frac{\partial(nT)}{\partial u_2}\nabla u_2 - n\nabla V = n\nabla\left(\frac{\mu}{T}\right) + \frac{d+2}{2} nT^2 \nabla\left(-\frac{1}{T}\right) - n\nabla V,$$

and

$$\begin{aligned}
&\frac{d+2}{2} JT + \kappa_0 nT\nabla T \\
&\quad= \frac{d+2}{2} T \left[n\nabla\left(\frac{\mu}{T}\right) + \frac{d+2}{2} nT^2\nabla\left(-\frac{1}{T}\right) - n\nabla V\right] + \kappa_0 nT^3 \nabla\left(-\frac{1}{T}\right) \\
&\quad= \frac{d+2}{2} nT\nabla\left(\frac{\mu}{T}\right) + \left[\left(\frac{d+2}{2}\right)^2 + \kappa_0\right] nT^3\nabla\left(-\frac{1}{T}\right) - \frac{d+2}{2} nT\nabla V.
\end{aligned}$$

Thus we can write (4.111)–(4.112) as the energy-transport model in the entropic variables with the diffusion matrix

$$D = nT \begin{pmatrix} 1 & \frac{d+2}{2} T \\ \frac{d+2}{2} T & \left(\left(\frac{d+2}{2}\right)^2 + \kappa_0\right) T^2 \end{pmatrix}.$$

The determinant of this matrix is positive (for $n, T > 0$) if and only if $\kappa_0 > 0$. This shows that the heat conduction term is necessary to obtain wellposedness of the energy-transport equations. We recall that this term is introduced heuristically, and no justification has been given.

(3) The formal limit $\beta \to 0$ only gives the equation $e = d/2$ which means that the sum of kinetic and thermal energy is constant in space and time. More interesting is the limit $\beta \to 0$ in the energy-transport equations (4.111)–(4.112). This yields formally $T = 1$ and

$$\partial_t n - \mathrm{div} J = 0, \qquad J = \nabla n - n\nabla V,$$

which are the drift-diffusion euqations.

The three relaxation-time limits are summarized in Figure 4.2.

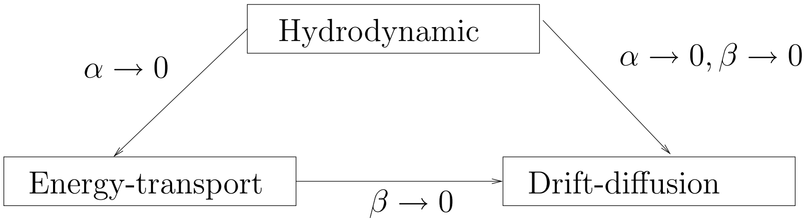

Figure 4.2: Relaxation-time limits in the hydrodynamic and energy-transport model.

5 Quantum Kinetic Models

5.1 The Schrödinger equation

As the dimensions of a semiconductor device decrease, quantum mechanical effects have to be included in the modeling of the transport phenomena. It is of great importance to devise models which, on the one hand, are capable of describing quantum effects and, on the other hand, are sufficiently simple to allow for efficient numerical simulations. We start with the many-particle Schrödinger equation and follow [49, Sec. 1.4].

The motion of an electron ensemble consisting of M particles in a vacuum under the influence of a (real-valued) electrostatic potential V is described by the many-particle Schrödinger equation,

$$\begin{aligned} i\hbar\partial_t\psi &= -\frac{\hbar^2}{2m}\sum_{j=1}^{M}\Delta_{x_j}\psi - qV(x,t)\psi, \qquad x\in\mathbb{R}^{dM},\ t>0, && (5.1)\\ \psi(x,0) &= \psi_I(x), && (5.2)\end{aligned}$$

where $x=(x_1,\ldots,x_M)^\top\in\mathbb{R}^{dM}$, $i^2=-1$, $\hbar=h/2\pi$ is the reduced Planck constant, m the electron mass, q the elementary charge, and $\psi=\psi(x,t)$ is called the *wave function* of the electron ensemble. The Laplace operator Δ_{x_j} only acts on the position variable x_j,

$$\Delta_{x_j} = \sum_{\ell=1}^{d}\frac{\partial}{\partial x_{j,\ell}}, \qquad \text{where } x_j=(x_{j,1},\ldots,x_{j,d})^\top\in\mathbb{R}^d.$$

The variable $x_j\in\mathbb{R}^d$ denotes the position vector of the j-th electron of the ensemble. The ensemble position density n and the ensemble current density J are defined by

$$n=|\psi|^2, \qquad J=-\frac{\hbar q}{m}\operatorname{Im}(\overline{\psi}\nabla_x\psi),$$

where $\overline{\psi}$ is the complex conjugate of ψ.

There are two alternative formulations of the motion of the electron ensemble:

- the density-matrix formulation and
- the kinetic (Wigner) formulation.

We introduce the density matrix in this section and explain the kinetic formulation in the following section.

Define the *ensemble Hamilton operator*

$$H = H_x = -\frac{\hbar^2}{2m}\Delta_x - qV(x,t) = -\frac{\hbar^2}{2m}\sum_{j=1}^{M}\Delta_{x_j} - qV(x,t)$$

and the *density matrix*

$$\varrho(r,s,t) = \overline{\psi(r,t)}\psi(s,t), \qquad r,s \in \mathbb{R}^{dM},\ t > 0. \tag{5.3}$$

Then the ensemble densities can be written in terms of ϱ as

$$\begin{aligned} n(x,t) &= \varrho(x,x,t), && (5.4)\\ J(x,t) &= \frac{i\hbar q}{2m}(\overline{\psi}\nabla_x\psi - \psi\nabla_x\overline{\psi})(x,t) = \frac{i\hbar q}{2m}(\nabla_s - \nabla_r)\varrho(x,x,t). && (5.5) \end{aligned}$$

The evolution equation for ϱ is obtained by differentiating (5.3) and using the Schrödinger equation

$$i\hbar\partial_t\psi = H\psi \tag{5.6}$$

which gives

$$\begin{aligned} i\hbar\partial_t\varrho(r,s,t) &= i\hbar(\overline{\partial_t\psi(r,t)}\psi(s,t) + \overline{\psi(r,t)}\partial_t\psi(s,t)) \\ &= (H_s - H_r)\varrho(r,s,t), && (5.7) \end{aligned}$$

since $V(x,t)$ is real-valued. This equation is referred to as the *Liouville-von Neumann* or *Heisenberg* equation. With the initial condition

$$\varrho(r,s,0) = \psi_I(r)\psi_I(s), \qquad r,s \in \mathbb{R}^{dM}, \tag{5.8}$$

the Schrödinger problem (5.1)–(5.2) and the Heisenberg Problem (5.7)–(5.8) are formally equivalent.

Is there any relation between the Schrödinger and the Heisenberg picture in the case of general initial conditions

$$\varrho(r,s,0) = \varrho_I(r,s), \qquad r,s \in \mathbb{R}^{dM}? \tag{5.9}$$

In order to derive a relation let (ϕ_j) be an orthonormal basis of $L^2(\mathbb{R}^{dM})$. Then $(\overline{\phi_j}\phi_k)$ is an orthonormal basis of $L^2(\mathbb{R}^{dM} \times \mathbb{R}^{dM})$ and we can develop

$$\varrho_I(r,s) = \sum_{j,k=1}^{\infty} \varrho_{jk}\overline{\phi_j(r)}\phi_k(s),$$

if $\varrho_I \in L^2(\mathbb{R}^{dM} \times \mathbb{R}^{dM})$ is assumed. Now let ψ_j be the solution of the Schrödinger equation (5.6) with initial datum ϕ_j. A computation as above shows that the product $\overline{\psi_j(r,t)}\psi_k(s,t)$ solves the Heisenberg equation (5.7), and since this equation is linear, also

$$\varrho(r,s,t) = \sum_{j,k=1}^{\infty} \varrho_{jk}\overline{\psi_j(r,t)}\psi_k(s,t) \tag{5.10}$$

solves the Heisenberg equation with initial datum ϱ_I. Thus, the solution of the Heisenberg problem (5.7), (5.9) can be written as an infinite sum of wave functions. Such a situation is called *mixed quantum states*, whereas the formulation (5.3) refers to a *single quantum state.*

In the mixed state case (5.10), the ensemble position density is written as

$$n(x,t) = \sum_{j=1}^{\infty} \varrho_{jj} |\psi_j(x,t)|^2,$$

and the weights ϱ_{jj} can be interpreted as occupation probabilites of the j-th quantum state.

5.2 The quantum Liouville equation

The Wigner equation is the quantum equivalent of the Liouville equation and can be derived from the Heisenberg equation by Fourier transformation. We recall that the Fourier transform F is defined by

$$\widehat{g}(\eta) = (Fg)(\eta) = (2\pi)^{-dM/2} \int_{\mathbb{R}^{dM}} g(v) e^{-i\eta\cdot v}\, dv$$

for (sufficiently smooth) functions $g : \mathbb{R}^{dM} \to \mathbb{C}$ with inverse

$$\check{h}(v) = (F^{-1}h)(v) = (2\pi)^{-dM/2} \int_{\mathbb{R}^{dM}} h(\eta) e^{i\eta\cdot v}\, d\eta$$

for functions $h : \mathbb{R}^{dM} \to \mathbb{C}$.

Definition 5.1 *The inverse Fourier transform of the density matrix,*

$$w(x,v,t) = (2\pi)^{-dM} \int_{\mathbb{R}^{dM}} \varrho\left(x + \frac{\hbar}{2m}\eta, x - \frac{\hbar}{2m}\eta, t\right) e^{i\eta\cdot v}\, d\eta, \tag{5.11}$$

is called Wigner function.

The Wigner function has been introduced by Wigner in 1932 [66]. Setting

$$u(x,\eta,t) = \varrho\left(x + \frac{\hbar}{2m}\eta, x - \frac{\hbar}{2m}\eta, t\right), \tag{5.12}$$

the expression

$$w = (2\pi)^{-dM/2} \check{u}$$

shows that w is indeed related to the inverse Fourier transform of ϱ. Since $(\hbar/2m)\eta$ has the dimension of length and $\hbar/2m$ has the dimension of m^2/s, η has the unit of inverse velocity s/m. Thus, the variable v in the Fourier transform has the dimension of velocity m/s.

Lemma 5.2 *Let ϱ be a solution of the Heisenberg equation (5.7), (5.9). Then the Wigner function w, defined in (5.11), is formally a solution of*

$$\begin{aligned} \partial_t w + v \cdot \nabla_x w - \frac{q}{m}\theta[V]w &= 0, \qquad x, v \in \mathbb{R}^{dM},\ t > 0, \\ w(x,v,0) &= w_I(x,v), \qquad x, v \in \mathbb{R}^{dM}, \end{aligned} \tag{5.13}$$

where

$$w_I(x,v) = (2\pi)^{-dM} \int_{\mathbb{R}^{dM}} \varrho_I\left(x + \frac{\hbar}{2m}\eta, x - \frac{\hbar}{2m}\eta\right) e^{i\eta\cdot v}\, d\eta$$

and $\theta[V]$ is a so-called pseudo-differential operator*, defined by*

$$(\widehat{\theta[V]w})(x,\eta,t) = \delta V(x,\eta,t)\widehat{w}(x,\eta,t), \tag{5.14}$$

where

$$\delta V(x,\eta,t) = \frac{im}{\hbar}\left[V\left(x + \frac{\hbar}{2m}\eta, t\right) - V\left(x - \frac{\hbar}{2m}\eta, t\right)\right].$$

Equation (5.13) is called the *many-particle Wigner equation* or *quantum Liouville equation.* Generally an operator whose Fourier transform acts as a multiplication operator on the Fourier transform of its argument, is called a (linear) pseudo-differential operator. We refer to [62] for the mathematical theory of pseudo-differential operators. The function δV is a kind of discrete directional derivative. Indeed, in the formal limit "$\hbar \to 0$" it holds

$$\delta V(x,\eta,t) \to i\nabla_x V(x,t)\cdot\eta = i\frac{\partial V}{\partial \eta}(x,t). \tag{5.15}$$

Clearly, the limit "$\hbar \to 0$" makes sense only after an appropriate scaling (see below).

The operator $\theta[V]w$ is written more explicitly as

$$\begin{aligned} (\theta[V]w)(x,v,t) &= ((\delta V)\widehat{w})^{\vee}(x,v,t) \ = \ (2\pi)^{-dM/2}((\delta V)u)^{\vee}(x,v,t) && (5.16) \\ &= \frac{1}{(2\pi)^{dM}} \int_{\mathbb{R}^{dM}} \delta V(x,\eta,t)u(x,\eta,t)e^{i\eta\cdot v}\, d\eta \\ &= \frac{1}{(2\pi)^{dM}} \int_{\mathbb{R}^{dM}} \int_{\mathbb{R}^{dM}} \delta V(x,\eta,t)w(x,v',t)e^{i\eta\cdot(v-v')}\, dv'\, d\eta. && (5.17) \end{aligned}$$

Proof: Since, by (5.12),

$$\begin{aligned} \mathrm{div}_\eta(\nabla_x u)(x,\eta,t) &= \mathrm{div}_\eta(\nabla_r\varrho + \nabla_s\varrho)\left(x + \frac{\hbar}{2m}\eta, x - \frac{\hbar}{2m}\eta, t\right) \\ &= \frac{\hbar}{2m}(\Delta_r\varrho - \Delta_s\varrho)\left(x + \frac{\hbar}{2m}\eta, x - \frac{\hbar}{2m}\eta, t\right), \end{aligned}$$

the transformed Heisenberg equation for u becomes, for $r = x + \hbar\eta/2m$, $s = x - \hbar\eta/2m$,

$$\begin{aligned} i\hbar\partial_t u(x,\eta,t) &= i\hbar\partial_t \varrho(r,s,t) \\ &= \left(-\frac{\hbar^2}{2m}(\Delta_s - \Delta_r) - qV(s,t) + qV(r,t)\right)\varrho(r,s,t) \\ &= \hbar \operatorname{div}_\eta(\nabla_x u)(x,\eta,t) + \frac{i\hbar q}{m}\delta V(x,\eta,t)u(x,\eta,t) \end{aligned}$$

or

$$\partial_t u + i\operatorname{div}_\eta(\nabla_x u) - \frac{q}{m}(\delta V)u = 0, \qquad x,\eta \in \mathbb{R}^{dM},\ t > 0.$$

Taking the inverse Fourier transform gives

$$\partial_t \check{u} + i(\operatorname{div}_\eta \nabla_x u)^\vee - \frac{q}{m}((\delta V)u)^\vee = 0. \tag{5.18}$$

We compute the second and the third term of this equation:

$$\begin{aligned} i(\operatorname{div}_\eta \nabla_x u)^\vee(x,v,t) &= \frac{i}{(2\pi)^{dM/2}}\int_{\mathbb{R}^{dM}} \operatorname{div}_\eta \nabla_x u(x,\eta,t)e^{i\eta\cdot v}\,d\eta \\ &= \frac{v}{(2\pi)^{dM/2}}\int_{\mathbb{R}^{dM}} \nabla_x u(x,\eta,t)e^{i\eta\cdot v}\,d\eta \\ &= v\cdot\nabla_x\check{u}(x,v,t) \ = \ (2\pi)^{dM/2}v\cdot\nabla_x w(x,v,t), \end{aligned}$$

where we have employed integration by parts, and, by (5.16),

$$(\theta[V]w)(x,v,t) = (2\pi)^{-dM/2}((\delta V)u)^\vee(x,v,t).$$

Therefore, (5.18) equals the Wigner equation (5.13). □

Lemma 5.3 *The ensemble particle density n and the ensemble current density J can be expressed in terms of the Wigner function as*

$$n(x,t) = \int_{\mathbb{R}^{dM}} w(x,v,t)\,dv, \qquad J(x,t) = -q\int_{\mathbb{R}^{dM}} w(x,v,t)v\,dv.$$

The above integrals are called the zeroth and first moments of the Wigner function, respectively, in analogy to the classical situation (see Section 3.1).

Proof: The first identity follows from (5.4):

$$n(x,t) = \varrho(x,x,t) = u(x,0,t) = (2\pi)^{dM/2}\widehat{w}(x,0,t) = \int_{\mathbb{R}^{dM}} w(x,v,t)\,dv.$$

For the proof of the second identity we use the fact that the Fourier transform translates differential operators into a multiplication:

$$i\nabla_\eta\widehat{w}(x,\eta,t) = \frac{1}{(2\pi)^{dM/2}}\int_{\mathbb{R}^{dM}} w(x,v,t)ve^{-i\eta\cdot v}\,dv = \widehat{vw}(x,\eta,t).$$

Therefore, by (5.5),

$$\begin{aligned} J(x,t) &= \frac{i\hbar q}{2m}(\nabla_s - \nabla_r)\varrho(x,x,t) = -iq\nabla_\eta u(x,0,t) \\ &= -iq(2\pi)^{dM/2}\nabla_\eta \widehat{w}(x,0,t) = -q(2\pi)^{dM/2}\widehat{vw}(x,0,t) \\ &= -q\int_{\mathbb{R}^{dM}} w(x,v,t)v\,dv. \end{aligned}$$

$\square$

We discuss two questions related to the quantum Liouville equation:

- How can we formalize the classical limit "$\hbar \to 0$" and which is the limit equation?
- Are the solutions of the quantum Liouville equation nonnegative if this property holds true initially?

The limit "$\hbar \to 0$" can be formalized in an appropriate scaling. We choose the reference length λ, the reference time τ, the reference velocity λ/τ, and the reference voltage k_BT/q, where k_B is the Boltzmann constant and T the temperature. We assume that the reference wave energy $\hbar/\tau$ is much smaller than the thermal and kinetic energies,

$$\frac{\hbar/\tau}{k_BT} = \varepsilon \qquad \text{and} \qquad \frac{\hbar/\tau}{m(\lambda/\tau)^2} = \varepsilon \qquad \text{with } \varepsilon \ll 1_X$$

(this fixes λ for given τ and vice versa). Thus, introducing the scaling

$$x = \lambda x_s, \quad t = \tau t_s, \quad v = \frac{\lambda}{\tau}v_s, \quad V = \frac{k_BT}{q}V_s,$$

we obtain, after omitting the index s,

$$\partial_t w + v\cdot\nabla_x w - \theta[V]w = 0, \tag{5.19}$$

where $\theta[V]w$ is given by (5.17) with

$$\delta V(x,\eta,t) = \frac{i}{\varepsilon}\left[V(x+\frac{\varepsilon}{2}\eta,t) - V(x-\frac{\varepsilon}{2}\eta,t)\right].$$

The limit $\varepsilon \to 0$ leads to (see (5.15))

$$\delta V(x,\eta,t) \to i\nabla_x V(x,t)\cdot\eta$$

and hence, using $i(\eta u)^\vee = \nabla_v \check{u}$,

$$\begin{aligned} \theta[V]w &= (2\pi)^{-dM/2}((\delta V)u)^\vee \to i(2\pi)^{-dM/2}\left((\nabla_x V\cdot\eta)u\right)^\vee \\ &= (2\pi)^{-dM/2}\nabla_x V\cdot\nabla_v\check{u} = \nabla_x V\cdot\nabla_v w. \end{aligned}$$

Thus, (5.19) becomes in the limit $\varepsilon \to 0$

$$\partial_t w + v \cdot \nabla_x w - \nabla_x V \cdot \nabla_v w = 0,$$

which is the classical Liouville equation (see Section 3.1). The above consideration makes clear the limit performed in (5.15). The classical limit $\varepsilon \to 0$ has been made rigorous in [47, 48] for smooth potentials.

The solution of the classical Liouville equation stays nonnegative for all time if the initial distribution function is nonnegative. Unfortunately, this property does *not* hold for the quantum Liouville equation. Thus, a (fully) probabilistic interpretation of the Wigner function as a distribution function is not possible. In the case of a pure quantum state it is possible to characterize the states for which the Wigner function is nonnegative exactly. It is shown in [38] that

$$w(x, v, t) = \frac{1}{(2\pi)^{dM}} \int_{\mathbb{R}^{dM}} \overline{\psi}\left(x + \frac{\hbar}{2m}\eta, t\right) \psi\left(x - \frac{\hbar}{2m}\eta, t\right) e^{i\eta\cdot v}\, d\eta$$

is nonnegative if and only if either $\psi \equiv 0$ or

$$\psi(x, t) = \exp\left(-x^\top A(t)x - a(t) \cdot x - b(t)\right), \qquad x \in \mathbb{R}^{dM},\ t > 0,$$

where $A(t) \in \mathbb{C}^{dM\times dM}$ is a matrix with symmetric positive definite real part and $a(t) \in \mathbb{C}^{dM}$, $b(t) \in \mathbb{C}$. Inserting this ansatz into the Schrödinger equation (5.1) shows that the potential has to be quadratic in x, i.e.

$$V(x, t) = x^\top \widetilde{A}(t)x + \widetilde{a}(t) \cdot x + \widetilde{b}(t)$$

for some $\widetilde{A}(t) \in \mathbb{C}^{dM\times dM}$, $\widetilde{a}(t) \in \mathbb{C}^{dM}$, $\widetilde{b}(t) \in \mathbb{C}$, in order to obtain a nonnegative Wigner solution.

The case of mixed quantum states, i.e. for arbitrary initial data $w_I \in L^2(\mathbb{R}^{dM} \times \mathbb{R}^{dM})$, is more involved. In fact, a necessary condition for the nonnegativity of w is not known.

5.3 The quantum Vlasov and Boltzmann equation

The quantum Liouville equation has the same disadvantage as its classical analogue, namely, the equation needs to be solved in a very high-dimensional phase space which makes its numerical solution almost unfeasible. In this section we derive the quantum analogue of the classical Vlasov equation, the quantum Vlasov equation, which acts on a $2d$-dimensional phase space. We proceed similarly as in [49, Sec. 1.5] (also see Section 3.2).

Consider as in the proceeding section an ensemble of M electrons with mass m in a vacuum under the action of a (real-valued) electrostatic potential $V(x, t)$. The motion of the particle ensemble is described by the density matrix as a solution of the Heisenberg equation (5.7). We impose the following assumptions:

(1) The potential can be decomposed into a sum of external potentials acting on one particle and of two-particle interaction potentials,

$$V(x_1,\ldots,x_M,t)=\sum_{j=1}^{M}V_{\text{ext}}(x_j,t)+\frac{1}{2}\sum_{i,j=1}^{M}V_{\text{int}}(x_i,x_j), \tag{5.20}$$

where V_{int} is symmetric, i.e. $V_{\text{int}}(x_i,x_j)=V_{\text{int}}(x_j,x_i)$ for all $i,j=1,\ldots,M$, and is of the order $1/M$ as $M\to\infty$.

(2) The electrons of the ensemble are initially indistinguishable in the sense

$$\varrho(r_1,\ldots,r_M,s_1,\ldots,s_M,0)=\varrho(r_{\pi(1)},\ldots,r_{\pi(M)},s_{\pi(1)},\ldots,s_{\pi(M)},0) \tag{5.21}$$

for all permutations π of $\{1,\ldots,M\}$ and all $r_j,s_j\in\mathbb{R}^d$.

(3) The initial subsensemble density matrices can be factorized

$$\varrho_I^{(a)}(r^{(a)},s^{(a)})=\varrho^{(a)}(r^{(a)},s^{(a)},0)=\prod_{j=1}^{a}R_I(r_j,s_j),\qquad 1\le a<M-1.$$

We discuss now these assumptions. The factor $\frac{1}{2}$ in (5.20) is necessary since each electron-electron pair in the sum of two-particle interactions is counted twice. The symmetry of the interaction potentials implies that

$$V(x_1,\ldots,x_M,t)=V(x_{\pi(1)},\ldots,x_{\pi(M)},t)\qquad\text{for all } t\ge 0$$

and for all permutations π. This property and (5.21) has the consequence that

$$\varrho(r_1,\ldots,r_M,s_1,\ldots,s_M,t)=\varrho(r_{\pi(1)},\ldots,r_{\pi(M)},s_{\pi(1)},\ldots,s_{\pi(M)},t) \tag{5.22}$$

holds for all $t>0$. Physically, this means that the electrons are indistinguishable for all time.

In fact, (5.22) is not enough to describe the behavior of the electron ensemble. Indeed, in the pure quantum state,

$$\varrho(r_1,\ldots,r_M,s_1,\ldots,s_M,t)=\overline{\psi(r_1,\ldots,r_M,t)}\psi(s_1,\ldots,r_M,t),$$

where ψ is the electron-ensemble wave function, the condition (5.22) is satisfied if either the wave function is antisymmetric,

$$\psi(x_1,\ldots,x_M,t)=\text{sign}(\pi)\psi(x_{\pi(1)},\ldots,x_{\pi(M)},t)$$

for all permutations π, where $\text{sign}(\pi)\in\{-1,+1\}$ is the sign of π, or if it is symmetric,

$$\psi(x_1,\ldots,x_M,t)=\psi(x_{\pi(1)},\ldots,x_{\pi(M)},t)$$

for all π. The first case corresponds to ensembles consisting of Fermions (and in particular, of electrons) and the latter case to ensembles of Bosons. In the first case (which is the interesting case for us) we obtain

$$\psi(x_1, \ldots, x_M, t) = 0 \qquad \text{if } x_i = x_j \text{ for some } i \neq j.$$

This expresses the Pauli exclusion principle which is *not* contained in condition (5.22).

We wish to derive the evolution equation for a subensemble density matrix. The density matrix of a subensemble consisting of a electrons is defined by

$$\varrho^{(a)}(r^{(a)}, s^{(a)}, t) = \int_{\mathbb{R}^{d(M-a)}} \varrho(r^{(a)}, u_{a+1}, \ldots, u_M, s^{(a)}, u_{a+1}, \ldots, u_M, t)\, du_{a+1} \ldots du_M,$$

where

$$r^{(a)} = (r_1, \ldots, r_a), \qquad s^{(a)} = (s_1, \ldots, s_a) \in \mathbb{R}^{da}.$$

Clearly, in view of the indistinguishable property (5.22), the subensemble density matrices satisfy

$$\varrho^{(a)}(r_1, \ldots, r_a, s_1, \ldots, s_a, t) = \varrho^{(a)}(r_{\pi(1)}, \ldots, r_{\pi(a)}, s_{\pi(1)}, \ldots, s_{\pi(a)}, t) \tag{5.23}$$

for all permutations π of $\{1, \ldots, a\}$ and all $r_j, s_j \in \mathbb{R}^d$, $t \geq 0$.

We recall that the evolution of the complete electron ensemble is governed by the Heisenberg equation (5.7), rewritten as

$$\begin{aligned} i\hbar\partial_t \varrho \;&=\; -\frac{\hbar^2}{2m}\sum_{j=1}^{M}(\Delta_{s_j} - \Delta_{r_j})\varrho - q\sum_{j=1}^{M}\left(V_{\text{ext}}(s_j, t) - V_{\text{ext}}(r_j, t)\right)\varrho \\ &\qquad -\frac{q}{2}\sum_{j,\ell=1}^{M}\left(V_{\text{int}}(s_j, s_\ell) - V_{\text{int}}(r_j, r_\ell)\right)\varrho. \end{aligned} \tag{5.24}$$

Theorem 5.4 *Let the above hypotheses (1)–(3) hold. Then the density matrix* $\varrho(r, s, t)$ *is a particular solution to the Heisenberg equation (5.23) if* $M \gg 1$,

$$\varrho^{(a)}(r^{(a)}, s^{(a)}, t) = \prod_{j=1}^{a} R(r_j, s_j, t), \tag{5.25}$$

and the function $W = (2\pi)^{-d/2}M\check{R}$ *is a solution to the* quantum Vlasov equation

$$\begin{aligned} \partial_t W + v\cdot\nabla_x W - \frac{q}{m}\theta[V_{\text{eff}}]W \;&=\; 0, \qquad x, v \in \mathbb{R}^d,\ t > 0 \\ W(x, v, 0) \;&=\; W_I(x, v), \qquad x, v \in \mathbb{R}^d, \end{aligned} \tag{5.26}$$

where the pseudo-differential operator $\theta[V_{\text{eff}}]$ *is defined as in (5.14), the effective potential* V_{eff} *is*

$$V_{\text{eff}}(x, t) = V_{\text{ext}}(x, t) + \int_{\mathbb{R}^d} n(z, t)V_{\text{int}}(x, z)\, dz, \tag{5.27}$$

the quantum electron density is

$$n(x,t) = \int_{\mathbb{R}} W(x,v,t)\,dv = MR(x,x,t), \tag{5.28}$$

and $W_I = (2\pi)^{-d/2} M\check{R}_I$*.*

The expression (5.25) is also called *Hartree ansatz*. As the effective potential depends on the function W through (5.28), we obtain a *nonlinear* pseudo-differential equation.

Proof: We set $u_j = s_j = r_j$ for $j = a+1, \ldots, M$ in (5.24), integrate over $(u_{a+1}, \ldots, u_M) \in \mathbb{R}^{d(M-a)}$ and use the property (5.23) to obtain, after an analogous computation as in Section 3.2, a quantum equivalent of the BBGKY hierarchy,

$$\begin{aligned} i\hbar\partial_t \varrho^{(a)} &= -\frac{\hbar^2}{2m}\sum_{j=1}^{M}(\Delta_{s_j} - \Delta_{r_j})\varrho^{(a)} - q\sum_{j=1}^{M}\left(V_{\text{ext}}(s_j,t) - V_{\text{ext}}(r_j,t)\right)\varrho^{(a)} \\ &\quad -q(M-a)\sum_{j=1}^{a}\int_{\mathbb{R}^d}\left(V_{\text{int}}(s_j,u_*) - V_{\text{int}}(r_j,u_*)\right)\varrho_*^{(a+1)}\,du_* \end{aligned}$$

for all $1 \le a \le M-1$, where

$$\varrho_*^{(a+1)} = \varrho^{(a+1)}(r^{(a)}, u_*, s^{(a)}, u_*, t).$$

Since we have assumed that V_{int} is of the order of $1/M$ as $M \to \infty$, we can approximate the above equation, by neglecting terms of order $1/M$, by

$$\begin{aligned} i\hbar\partial_t \varrho^{(a)} &= -\frac{\hbar^2}{2m}\sum_{j=1}^{a}(\Delta_{s_j} - \Delta_{r_j})\varrho^{(a)} - q\sum_{j=1}^{a}\left(V_{\text{ext}}(s_j,t) - V_{\text{ext}}(r_j,t)\right)\varrho^{(a)} \\ &\quad -q\sum_{j=1}^{a}\int_{\mathbb{R}^d}\left(V_{\text{int}}(s_j,u_*) - V_{\text{int}}(r_j,u_*)\right)M\varrho_*^{(a+1)}\,du_*. \end{aligned}$$

Similar to the classical case, it can be seen that this equation is satisfied by the ansatz (5.25) if R solves the equation

$$i\hbar\partial_t R = -\frac{\hbar^2}{2m}(\Delta_s - \Delta_r)R - q\left(V_{\text{eff}}(s,t) - V_{\text{eff}}(r,t)\right)R, \qquad r, s, \in \mathbb{R}^d,\ t > 0, \tag{5.29}$$

where the effective potential V_{eff} is defined in (5.27) and (5.28).

The kinetic formulation of (5.29) is derived as in Section 5.2. Introducing the change of coordinates

$$r = x + \frac{\hbar}{2m}\eta, \qquad s = x - \frac{\hbar}{2m}\eta,$$

the function $U(x,\eta,t) = MR(r,s,t)$ solves the equation

$$\partial_t U + i\,\mathrm{div}_\eta \nabla_x U + \frac{iq}{\hbar}\left[V_{\text{eff}}\left(x + \frac{\hbar}{2m}\eta, t\right) - V_{\text{eff}}\left(x - \frac{\hbar}{2m}\eta, t\right)\right]U = 0.$$

Finally, the inverse Fourier transform $W = (2\pi)^{-d/2}\check{U}$ solves the quantum Vlasov equation (5.26). $\square$

Contrary to the classical Vlasov equation, the quantum Vlasov equation does not preserve the nonnegativity of the solution; cf. the discussion in Section 5.2. However, if the initial single-particle density matrix $R(r, s, 0)$ is positive semi-definite, the electron density n, defined in (5.28), remains nonnegative for all times.

Similar to the quantum Liouville equation, the solution to the quantum Vlasov equation converges (at least formally) as "$\hbar \to 0$" to a solution to the classical Vlasov equation

$$\partial_t W + v \cdot \nabla_x W - \frac{q}{m}\nabla_x V_{\text{eff}} \cdot \nabla_v W = 0.$$

The limit "$\hbar \to 0$" has to be understood in the sense explained in Section 5.2.

As in Section 3.2, a usual choice for the two-particle interaction potential is the Coulomb potential which reads in three space dimensions as

$$V_{\text{int}}(x, y) = -\frac{q}{4\pi\varepsilon_s}\frac{1}{|x-y|}, \qquad x, y \in \mathbb{R}^3,\ x \neq y,$$

where ε_s denotes the permittivity of the semiconductor material (see (3.19)). In Section 3.2 it has been shown that the effective potential

$$V_{\text{eff}}(x, t) = V_{\text{ext}}(x, t) - \frac{q}{4\pi\varepsilon_s}\int_{\mathbb{R}^3}\frac{n(z, t)}{|z-x|}\,dz \tag{5.30}$$

solves the Poisson equation

$$\varepsilon_s \Delta V_{\text{eff}} = q(n - C(x)),$$

where $C(x) = -(\varepsilon_s/q)V_{\text{ext}}(x)$ is the doping concentration if V_{ext} is generated by ions of charge $+q$ in the semiconductor. This concept can be generalized to any space dimension. The initial-value problem (5.26), (5.30) (together with (5.28)) is called the *quantum Vlasov-Poisson problem.*

The presented quantum Vlasov equation models the motion of an electron ensemble in a vacuum under the influence of long-range interactions. However, the electrons are moving in a crystal whose influence needs to be taken into account. Including a lattice potential into the Schrödinger equation, making a Bloch decomposition (see Theorem 2.5), and letting the length scale of the Brillouin zone to infinity, it can be shown [2, 55] that the resulting equation equals

$$\partial_t w + v(k) \cdot \nabla_x w - \frac{q}{m}\theta[V_{\text{eff}}]w = 0, \qquad x, k \in \mathbb{R}^d,\ t > 0, \tag{5.31}$$

together with equations (5.27) and (5.28) and initial conditions for w. Here, $v(k) = \nabla_k E(k)/\hbar$ is the group velocity, $E(k)$ the band structure, and k the wave vector which has been extended to the whole space by the limiting procedure.

A second generalization of the quantum Vlasov equation is given by the inclusion of short-range interactions modeled by scattering events of particles. The quantum mechanical modeling of collisions of electrons (with phonons, for instance) is a very difficult task, and there is no complete theory up to now. The phenomenological approach used in Section 3.3 cannot be used here since the notion of particle trajectories or characteristics does not make sense in this quantum mechanical framework. On the other hand, most of the collision models derived in the literature are highly complicated and numerically too expensive [18, 31]. A simple approach is to formulate the so-called *quantum Boltzmann equation*

$$\partial_t w + v \cdot \nabla_x w - \frac{q}{m}\theta[V_{\text{eff}}]w = Q(w), \qquad x, v \in \mathbb{R}^d,\ t > 0, \tag{5.32}$$

by adding a heuristic collision term to the right-hande side of the quantum Vlasov equation (5.26) or of its energy-band version (5.31). In numerical studies, often one of the following collision operators are used (assuming the energy-band version of the quantum Boltzmann equation)

- the *relaxation-time* or *BGK model* [59]

  $$Q(w) = \frac{1}{\tau}\left(\frac{n}{n_0}w_0 - w\right),$$

  where

  $$n(x,t) = \int_{\mathbb{R}^d} w(x,k,t)\,dk, \qquad n_0(x,t) = \int_{\mathbb{R}^d} w_0(x,k,t)\,dk,$$

  and w_0 is the distribution function of the quantum mechanical thermal equilibrium, defined, for instance in the mixed state, by the thermal equilibrium density matrix

  $$\varrho_0(r,s) = \sum_j f(\lambda_j)\overline{\psi_j(r)}\psi_j(s),$$

  where (λ_j, ψ_j) are eigenvalue-eigenfunction pairs of the quantum Hamiltonian, and $f(\lambda)$ is the Fermi-Dirac distribution;

- the *Caldeira-Leggett model* [16]

  $$Q(w) = \frac{1}{\tau_0}\operatorname{div}_k\left(\frac{m^* k_B T}{\hbar^2}\nabla_k w + kw\right), \tag{5.33}$$

  where τ_0 denotes the relaxation time, T the lattice temperature and m^* the effective electron mass.

The Caldeira-Leggett model has the disadvantage that the positivity of the density matrix is not preserved under temporal evolution. In [17] the approach of Caldeira and Leggett has been improved by deriving the Fokker-Planck operator

$$Q(w) = \frac{1}{\tau_0}\operatorname{div}_k\left(\frac{m^* k_B T}{\hbar^2}\nabla_k w + kw\right) + \frac{1}{\tau_0}\operatorname{div}_x\left(\frac{\Omega\hbar}{6\pi k_B T}\nabla_k w + \frac{\hbar^2}{12 m^* k_B T}\nabla_x w\right),$$

where Ω is the cut-off frequency modeling the interaction of the electrons with the phonons of the crystal (see [17] for details). This collision operator preserves the posivitiy of the density matrix.

A summary of the kinetic models derived in Chapter 3 and this chapter is presented in Figure 5.1. Notice that for each model, there is an energy-band version in the (x, k, t) variables, which reduces to a model in the (x, v, t) variables in the parabolic band approximation via $v = v(k) = \hbar k/m^*$.

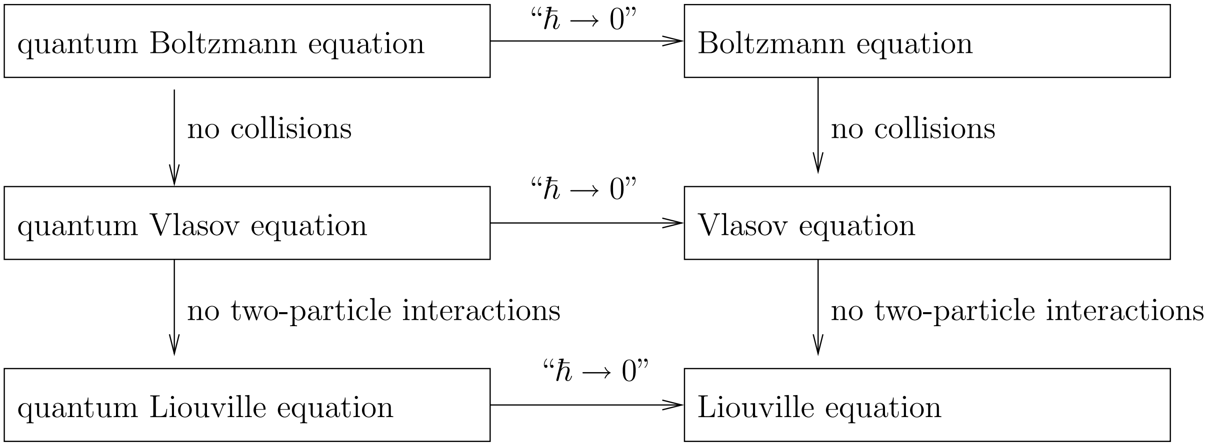

Figure 5.1: Hierarchy of classical and quantum kinetic models.

5.4 Open quantum systems

Up to now, we have only considered *closed quantum systems*. In a closed system, all elements and their interactions are completely known. On the other hand, *open quantum systems* are characterized by the fact that elements of the system interact with an environment but not all interactions are known precisely. As an example of an open system we consider the one-dimensional stationary Schrödinger equation with some potential $V(x)$,

$$-\frac{\hbar^2}{2m^*}\psi'' - qV(x)\psi = E\psi \qquad \text{in } (0, L), \tag{5.34}$$

where m^* is the effective mass which is assumed, for simplicity, to be constant. It is convenient to scale the Schrödinger equation. Choosing the reference length L, the reference potential k_BT/q, and the reference energy k_BT and introducing the scaled variables

$$x = Lx_s, \qquad V = \frac{k_BT}{q}V_s, \qquad E = (k_BT)E_s,$$

we obtain from (5.34), after omitting the index s,

$$-\frac{\varepsilon^2}{2}\psi'' - V(x)\psi = E\psi, \qquad x \in (0, 1), \tag{5.35}$$

where

$$\varepsilon = \frac{\hbar}{\sqrt{m^* L^2 k_B T}}$$

is the scaled Planck constant.

Equation (5.35) is an eigenvector-eigenvalue problem. For its solution we need to specify boundary conditions. Since we do not know the wave function at the boundary or outside of the interval, this constitutes an open system. In order to solve it, we have to specify what happens with the electron wave outside of the interval. We assume: Electron waves are injected at $x = 0$; they exit the interval at $x = L$ or they are reflected by the potential at $x = 0$ (Figure 5.2). We solve the problem (5.35) by extending the potential to the whole line,

$$V(x) = V(0) \quad \text{for } x < 0, \qquad V(x) = V(1) \quad \text{for } x > 1, \tag{5.36}$$

and solve (5.35) for $x \in \mathbb{R}$. Since the potential is constant in $\{x < 0\}$ and $\{x > 1\}$, respectively, we expect that the solutions are plane waves in these intervals. This motivates the following ansatz. Let first $p > 0$. Then we define

$$\psi_p(x) = \begin{cases} e^{ipx/\varepsilon} + r(p)e^{-ipx/\varepsilon} & : x < 0 \\ t(p)e^{ip_+(p)(x-1)/\varepsilon} & : x > 1, \end{cases} \tag{5.37}$$

where $p_+(p)$ has to be determined. This ansatz means that a wave with amplitude one is coming from $-\infty$ (since $p > 0$) and is either transmitted to $+\infty$ with amplitude $t(p)$ or it is reflected by the potential and travels back to $-\infty$ with amplitude $r(p)$. Inserting (5.37) into the scaled Schrödinger equation (5.35) yields for $x < 0$

$$(E + V(0))\psi_p = -\frac{\varepsilon^2}{2}\psi_p'' = \frac{p^2}{2}e^{ipx/\varepsilon} + \frac{p^2}{2}r(p)e^{-ipx/\varepsilon} = \frac{p^2}{2}\psi_p$$

and thus

$$E_p = E = \frac{p^2}{2} - V(0);$$

for $x > 1$ we obtain

$$(E_p + V(1))\psi_p = -\frac{\varepsilon^2}{2}\psi_p'' = \frac{p_+^2(p)}{2}\psi_p$$

and

$$p_+(p) = \sqrt{2(E_p + V(1))} = \sqrt{p^2 + 2(V(1) - V(0))}.$$

We take the positive root since the wave travels to $+\infty$ and hence, $p_+(p) > 0$ is required.

For the case $p < 0$ we make an analogous ansatz:

$$\psi_p(x) = \begin{cases} t(p)e^{-ip_-(p)x/\varepsilon} & : x < 0 \\ e^{-ip(x-1)/\varepsilon} + r(p)e^{ip(x-1)/\varepsilon} & : x > 1, \end{cases}$$

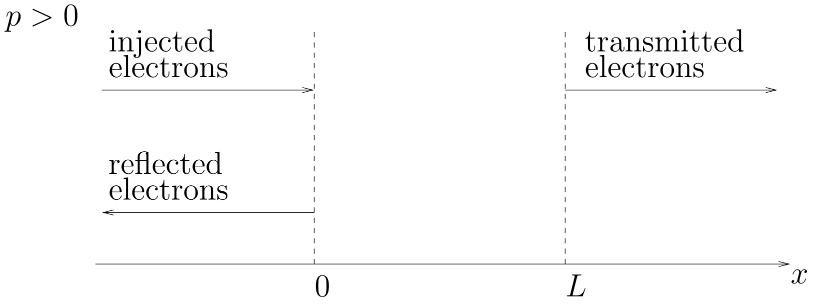

Figure 5.2: Electrons with $p > 0$ are injected at $x = 0$ and are reflected at $x = 0$ or transmitted at $x = L$.

where again $p_-(p)$ is unknown. This ansatz models a wave coming from $+\infty$ and being either transmitted to $-\infty$ or reflected at $x = 1$ and traveling back to $+\infty$. Inserting this ansatz into (5.35) gives, after a similar computation as above,

$$E_p = E = \frac{p^2}{2} - V(1), \qquad p_-(p) = \sqrt{p^2 - 2(V(1) - V(0))}.$$

Assuming that the wave function ψ_p is continuous in $\mathbb{R}$, this allows to specify the boundary conditions for ψ_p at $x = 0$ and $x = 1$. Indeed, for $p > 0$ and $x < 0, x \to 0$ there holds:

$$\varepsilon\psi_p'(0) = ip(1 - r(p)) \qquad \text{and} \qquad ip\psi_p(0) = ip(1 + r(p)),$$

so

$$\varepsilon\psi_p'(0) + ip\psi_p(0) = 2ip.$$

At $x > 1$, $x \to 1$, we infer

$$\varepsilon\psi_p'(1) = ip_+(p)t(p) = ip_+(p)\psi_p(1).$$

For $p < 0$ we obtain

$$\varepsilon\psi_p'(1) - ip\psi_p(1) = -ip(1 - r(p)) - ip(1 + r(p)) = -2ip,$$
$$\varepsilon\psi_p'(0) = -ip_-(p)t(p) = -ip_-(p)\psi_p(0).$$

We have shown the following result.

Lemma 5.5 *Let $V(x)$, $x \in (0, 1)$, be a given potential extended to $\mathbb{R}$ by (5.36). Then the solution of the eigenvalue problem*

$$-\frac{\varepsilon^2}{2}\psi'' - V(x)\psi = E\psi, \qquad x \in \mathbb{R},$$

can be written as

$$p > 0: \quad E_p = \tfrac{p^2}{2} - V(0), \quad \psi_p(x) = \begin{cases} e^{ipx/\varepsilon} + r(p)e^{-ipx/\varepsilon} & : x \leq 0 \\ t(p)e^{ip_+(p)(x-1)/\varepsilon} & : x \geq 1, \end{cases} \tag{5.38}$$

$$p < 0: \quad E_p = \tfrac{p^2}{2} - V(1), \quad \psi_p(x) = \begin{cases} t(p)e^{-ip_-(p)x/\varepsilon} & : x \leq 0 \\ e^{-ip(x-1)/\varepsilon} + r(p)e^{ip(x-1)/\varepsilon} & : x \geq 1, \end{cases} \tag{5.39}$$

where

$$p_\pm(p) = \sqrt{p^2 \pm 2(V(1) - V(0))},$$

and in the interval $(0,1)$ ψ_p *is the solution of*

$$-\frac{\varepsilon^2}{2}\psi_p'' - V(x)\psi_p = E_p\psi_p, \qquad x \in (0,1)$$

$$\varepsilon\psi_p'(0) + ip\psi_p(0) = 2ip, \quad \varepsilon\psi_p'(1) = ip_+(p)\psi_p(1) \quad \textit{for}\, p > 0, \tag{5.40}$$

$$-\varepsilon\psi_p'(1) + ip\psi_p(1) = 2ip, \quad \varepsilon\psi_p'(0) = -ip_-(p)\psi_p(0) \quad \textit{for}\, p < 0, \tag{5.41}$$

where E_p *is given as above. The reflection and transmission amplitudes* $r(p)$ *and* $t(p)$ *are determined by*

$$p > 0: \quad r(p) = \tfrac{1}{2}\left(\psi_p(0) + \tfrac{i\varepsilon}{p}\psi_p'(0)\right), \quad t(p) = \psi_p(1), \tag{5.42}$$

$$p < 0: \quad r(p) = \tfrac{1}{2}\left(\psi_p(1) - \tfrac{i\varepsilon}{p}\psi_p'(1)\right), \quad t(p) = \psi_p(0). \tag{5.43}$$

The boundary conditions (5.40)–(5.41) are called the *Lent-Kirkner boundary conditions* [41].

Proof: It remains to show the formulas (5.42)–(5.43). In fact, they follow immediately from (5.40) and (5.41). For instance, for $p > 0$, we have from the first equation in (5.40)

$$\frac{i\varepsilon}{p}\psi_p'(0) = \psi_p(0) - 2$$

and hence, by (5.38),

$$\frac{1}{2}\left(\psi_p(0) + \frac{i\varepsilon}{p}\psi_p'(0)\right) = \psi_p(0) - 1 = r(p), \qquad \psi_p(1) = t(p).$$

In a similar way, (5.41) and (5.39) imply (5.43). □

The unscaled macroscopic electron and current densities are given by

$$\begin{aligned} n(x) &= \int_{\mathbb{R}} f(p)|\psi_p(x)|^2\, dp \\ J(x) &= \frac{q\hbar}{m^*}\int_{\mathbb{R}} f(p)\mathrm{Im}\,(\overline{\psi_p(x)}\psi_p(x))\, dp, \end{aligned}$$

where $f(p)$ describes the statistics of the mixed states. In Lemma 2.20 we have derived the following formula:

$$f(p) = \frac{m^* k_B T}{2\pi^2 \hbar^3} \ln \left(1 + e^{(-p^2/2m^* + E_F)/k_B T} \right),$$

where E_F is the Fermi energy.

6 Quantum Fluid Models

6.1 Zero-temperature quantum hydrodynamic equations

We consider a single electron in a vacuum. The evolution of the particle is described by the Schrödinger equation

$$\begin{aligned} i\hbar\partial_t\psi &= -\frac{\hbar^2}{2m}\Delta\psi - qV(x,t)\psi, \qquad x\in\mathbb{R}^d,\ t>0, \\ \psi(x,0) &= \psi_I(x), \qquad x\in\mathbb{R}^d. \end{aligned} \tag{6.1}$$

First we scale the equations by introducing reference values for the time, length, and potential,

$$t=\tau t_s, \qquad x = Lx_s, \qquad V = UV_s.$$

We assume that the kinetic energy is of the same order as the electric energy,

$$m\left(\frac{L}{\tau}\right)^2 = qU.$$

Then the scaled Schrödinger equation becomes (omitting the index "s")

$$i\varepsilon\partial_t\psi = -\frac{\varepsilon^2}{2}\Delta\psi - V\psi, \qquad x\in\mathbb{R}^d,\ t>0, \tag{6.2}$$

where

$$\varepsilon = \frac{\hbar/\tau}{m(L/\tau)^2} = \frac{\hbar\tau}{mL^2}$$

is the scaled Planck constant.

In order to derive a fluid-type equation, we asume that the initial wave function is given by the WKB (Wentzel, Kramers, Brillouin) state

$$\psi_I = \sqrt{n_I}\exp(iS_I/\varepsilon), \tag{6.3}$$

where $n_I(x) \geq 0$, $S_I(x)\in\mathbb{R}$ are some functions. We claim that the *zero-temperature quantum hydrodynamic equations*

$$\partial_t n - \mathrm{div} J = 0, \tag{6.4}$$

$$\partial_t J - \mathrm{div}\left(\frac{J\otimes J}{n}\right) + n\nabla V + \frac{\varepsilon^2}{2}n\nabla\left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) = 0, \qquad x\in\mathbb{R}^d,\ t>0 \tag{6.5}$$

$$n(x,0) = n_I(x), \qquad J(x,0) = J_I(x), \qquad x\in\mathbb{R}^d, \tag{6.6}$$

are formally equivalent to the Schrödinger equation (6.2) in the sense of the following theorem.

Theorem 6.1 *(1) Let ψ be a solution of (6.1)–(6.2) with initial datum (6.3). Then $n = |\psi|^2$, $J = -\varepsilon \mathrm{Im}\,(\overline{\psi}\nabla\psi)$ solve (6.4)–(6.6) with initial data*

$$n_I = |\psi_I|^2, \qquad J_I = -n_I \nabla S_I \tag{6.7}$$

as long as $n > 0$ in $\mathbb{R}^d$.

(2) Let (n, S) be a solution of

$$\partial_t n + \mathrm{div}(n\nabla S) = 0, \qquad \partial_t S + \frac{1}{2}|\nabla S|^2 - V - \frac{\varepsilon^2}{2}\frac{\Delta\sqrt{n}}{\sqrt{n}} = 0, \qquad x \in \mathbb{R}^d,\ t > 0, \tag{6.8}$$

$$n(x,0) = n_I(x), \qquad S(x,0) = S_I(x), \qquad x \in \mathbb{R}^d,$$

such that $n > 0$ in $\mathbb{R}^d$, $t > 0$. then $\psi = \sqrt{n}\exp(iS/\varepsilon)$ solves (6.1)–(6.2) with initial datum (6.3).

Notice that (6.8) implies (6.4)–(6.5). Indeed, setting $J = -n\nabla S$, we obtain immediately (6.4), and differentiation of the second equation in (6.8) with respect to x and multiplication with n yields

$$\begin{aligned} n\nabla V + \frac{\varepsilon^2}{2} n\nabla \frac{\Delta\sqrt{n}}{\sqrt{n}} &= n\partial_t \nabla S + \frac{1}{2} n\nabla|\nabla S|^2 \\ &= \partial_t(n\nabla S) - \mathrm{div} J \cdot \nabla S + \frac{1}{2} n\nabla|\nabla S|^2 \\ &= -\partial_t J + \mathrm{div}\left(\frac{J\otimes J}{n}\right), \end{aligned}$$

since, with the Hessian D^2S,

$$\begin{aligned} \frac{1}{2} n\nabla|\nabla S|^2 &= n(D^2 S)\nabla S = \mathrm{div}(n\nabla S\otimes\nabla S) - \mathrm{div}(n\nabla S)\nabla S \\ &= \mathrm{div}\left(\frac{J\otimes J}{n}\right) + \mathrm{div} J\cdot\nabla S. \end{aligned}$$

Proof of Theorem 6.1: (1) We can write the complex-valued wave function ψ as $\psi = \sqrt{n}\exp(iS/\varepsilon)$, where $n = |\psi|^2$ and S is some phase function. Then

$$J = -\varepsilon \mathrm{Im}\,(\overline{\psi}\nabla\psi) = -\varepsilon\ \mathrm{Im}\left(\sqrt{n}\nabla\sqrt{n} + \frac{i}{\varepsilon} n\nabla S\right) = -n\nabla S.$$

Thus, n and S satisfy the initial conditions (6.7). It remains to show that (n, J) solves (6.8) since this implies (6.4)–(6.5).

Inserting the expression $\psi = \sqrt{n}\exp(iS/\varepsilon)$ into (6.2) gives, after division of $\exp(iS/\varepsilon)$,

$$\frac{i\varepsilon}{2}\frac{\partial_t n}{\sqrt{n}} - \sqrt{n}\partial_t S = -\frac{\varepsilon^2}{2}\left(\Delta\sqrt{n} + \frac{2i}{\varepsilon}\nabla\sqrt{n}\cdot\nabla S + \frac{i}{\varepsilon}\sqrt{n}\Delta S - \frac{\sqrt{n}}{\varepsilon^2}|\nabla S|^2\right) - \sqrt{n}V.$$

The imaginary part of this equation equals

$$\partial_t n = -2\sqrt{n}\nabla\sqrt{n}\cdot\nabla S - n\Delta S = -\mathrm{div}(n\nabla S),$$

which is the first equation in (6.8). The real part of the above equation reads

$$\partial_t S = \frac{\varepsilon^2}{2}\frac{\Delta\sqrt{n}}{\sqrt{n}} - \frac{1}{2}|\nabla S|^2 + V,$$

which is the second equation in (6.8).

(2) Differentiation of $\psi = \sqrt{n}\exp(iS/\varepsilon)$ gives, using (6.8),

$$\begin{aligned} i\varepsilon\partial_t\psi + \frac{\varepsilon^2}{2}\Delta\psi &= e^{iS/\varepsilon}\bigg(i\varepsilon\frac{\partial_t n}{2\sqrt{n}} - \sqrt{n}\partial_t S + \frac{\varepsilon^2}{2}\Delta\sqrt{n} + i\varepsilon\nabla\sqrt{n}\cdot\nabla S \\ &\qquad + \frac{i\varepsilon}{2}\sqrt{n}\Delta S - \frac{\sqrt{n}}{2}|\nabla S|^2\bigg) \\ &= e^{iS/\varepsilon}\left(-\frac{i\varepsilon}{2}\frac{\mathrm{div}(n\nabla S)}{\sqrt{n}} + i\varepsilon\nabla\sqrt{n}\cdot\nabla S + \frac{i\varepsilon}{2}\sqrt{n}\Delta S - \sqrt{n}V\right) \\ &= -\sqrt{n}\,e^{iS/\varepsilon}V \;=\; -V\psi. \end{aligned}$$

□

The system (6.4)–(6.5) is the quantum analogue of the classical pressureless Euler equations of gas dynamics. Notice that the above derivation requires an irrational initial velocity J/n (since $\mathrm{curl}\,(J/n) = -\mathrm{curl}\,(\nabla S) = 0$).

The quantum term can be interpreted either as a quantum self-potential term with the so-called *Bohm potential* $\Delta\sqrt{n}/\sqrt{n}$ or as a non-diagonal pressure tensor

$$P = \frac{\varepsilon^2}{4}n(\nabla\otimes\nabla)\log n$$

since $\mathrm{div}P = (\varepsilon^2/2)n\nabla(\Delta\sqrt{n}/\sqrt{n})$.

6.2 Quantum hydrodynamics and the Schrödinger equation

The quantum hydrodynamic model of the previous section is derived for a single particle and therefore, it does not contain a temperature term. In order to include temperature, many-particle systems need to be studied. We assume that the considered electron ensemble is represented by a *mixed state* (see Section 5.1). A mixed quantum state consists of a sequence of single states with occupation probability $\lambda_k \geq 0$, $k \in \mathbb{N}$, for the k-th state described by the Schrödinger equations

$$\begin{aligned} i\varepsilon\partial_t\psi_k &= -\frac{\varepsilon^2}{2}\Delta\psi_k - V\psi_k, \qquad x\in\mathbb{R}^d,\ t>0, &(6.9)\\ \psi_k(x,0) &= \psi_{I,k}(x), \qquad x\in\mathbb{R}^d. &(6.10)\end{aligned}$$

The occupation probabilities clearly satisfy $\sum_{k=1}^{\infty} \lambda_k = 1$, i.e, the probability of finding the electron ensemble in any of the quantum states is one.

Define the single-particle particle and current densities of the k-th state as in the previous subsection as

$$n_k = |\psi_k|^2, \qquad J_k = -\varepsilon \, \mathrm{Im}\,(\overline{\psi}_k \nabla \psi_k), \qquad k \in \mathbb{N}.$$

We claim that the total carrier density n and current density J of the mixed state, given by

$$n = \sum_{k=1}^{\infty} \lambda_k n_k, \qquad J = \sum_{k=1}^{\infty} \lambda_k J_k, \tag{6.11}$$

is a solution of the quantum hydrodynamic equations with temperature tensor:

$$\begin{aligned} \partial_t n - \mathrm{div} J &= 0, & (6.12)\\ \partial_t J - \mathrm{div}\left(\frac{J \otimes J}{n} + n\theta\right) + n\nabla V + \frac{\varepsilon^2}{2} n \nabla \left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) &= 0, \quad x \in \mathbb{R}^d,\ t > 0, & (6.13)\end{aligned}$$

with initial conditions

$$n(\cdot,0) = \sum_{k=1}^{\infty} \lambda_k |\psi_{I,k}|^2, \qquad J(\cdot,0) = -\varepsilon \sum_{k=1}^{\infty} \lambda_k \, \mathrm{Im}\,(\overline{\psi}_{I,k} \nabla \psi_{I,k}), \qquad \text{in } \mathbb{R}^d, \tag{6.14}$$

where $\theta \in \mathbb{R}^{d\times d}$ is the temperature tensor defined below (see (6.15)).

Theorem 6.2 *Let ψ_k be a solution of (6.9)–(6.10) for all $k \in \mathbb{N}$. Then (n, J), defined in (6.11), solves (6.12)–(6.14), where*

$$\theta = \theta_c + \theta_{os} \tag{6.15}$$

and the current temperature θ_c and osmotic temperature θ_{os} are given by

$$\begin{aligned} \theta_c &= \sum_{k=1}^{\infty} \lambda_k \frac{n_k}{n} (u_{c,k} - u_c) \otimes (u_{c,k} - u_c), & (6.16)\\ \theta_{os} &= \sum_{k=1}^{\infty} \lambda_k \frac{n_k}{n} (u_{os,k} - u_{os}) \otimes (u_{os,k} - u_{os}), & (6.17)\end{aligned}$$

and

$$u_c = -\frac{J}{n}, \quad u_{c,k} = -\frac{J_k}{n_k}, \qquad u_{os} = \frac{\varepsilon}{2}\nabla \ln n, \quad u_{os,k} = \frac{\varepsilon}{2}\nabla \ln n_k \tag{6.18}$$

are the "current" and "osmotic" velocities, respectively.

The notion "osmotic" comes from the fact that the quantum term can be written as

$$\frac{\varepsilon^2}{2} n \nabla \left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) = \mathrm{div} P \qquad \text{with } P = \frac{\varepsilon^2}{4} n (\nabla \otimes \nabla) \ln n,$$

and P can be interpreted as a non-diagonal pressure tensor.

Proof: Theorem 6.1 shows that (n_k, J_k) solves the system of equations

$$\begin{gathered}\partial_t n_k - \mathrm{div} J_k = 0,\\ \partial_t J_k - \mathrm{div}\left(\frac{J_k \otimes J_k}{n_k}\right) + n_k \nabla V + \frac{\varepsilon^2}{2} n_k \nabla \left(\frac{\Delta\sqrt{n_k}}{\sqrt{n_k}}\right) = 0, \qquad x \in \mathbb{R}^d\\ n_k(\cdot, 0) = |\psi_{I,k}|^2, \qquad J_k(\cdot, 0) = -\varepsilon\ \mathrm{Im}\,(\overline{\psi}_{I,k} \nabla \psi_{I,k}).\end{gathered}$$

Multiplication of the above equations by λ_k and summation over k yields

$$\begin{gathered}\partial_t n - \mathrm{div} J = 0,\\ \partial_t J - \sum_{k=1}^{\infty} \lambda_k \mathrm{div}\left(\frac{J_k \otimes J_k}{n_k}\right) + n \nabla V + \frac{\varepsilon^2}{2} \sum_{k=1}^{\infty} \lambda_k n_k \nabla \left(\frac{\Delta\sqrt{n_k}}{\sqrt{n_k}}\right) = 0. \qquad (6.19)\end{gathered}$$

We rewrite the second and fourth term of the second equation. Using (6.16) and (6.18) we obtain

$$\begin{aligned}&\sum_{k=1}^{\infty} \lambda_k \mathrm{div}\left(\frac{J_k \otimes J_k}{n_k}\right) \qquad (6.20)\\ &= \sum_{k=1}^{\infty} \lambda_k \mathrm{div}(n_k u_{c,k} \otimes u_{c,k})\\ &= \sum_{k=1}^{\infty} \lambda_k \mathrm{div}\left(n_k(u_{c,k} - u_c) \otimes (u_{c,k} - u_c) + 2 n_k u_{c,k} \otimes u_c\right) - \mathrm{div}(n u_c \otimes u_c)\\ &= \mathrm{div}(n\theta_c) + \sum_{k=1}^{\infty} \mathrm{div}\left(2\lambda_k \frac{J_k \otimes J}{n}\right) - \mathrm{div}\left(\frac{J \otimes J}{n}\right)\\ &= \mathrm{div}(n\theta_c) + \mathrm{div}\left(\frac{J \otimes J}{n}\right). \qquad (6.21)\end{aligned}$$

Furthermore, employing (6.17) and (6.18),

$$\begin{aligned}\sum_{k=1}^{\infty} \lambda_k n_k \nabla \left(\frac{\Delta\sqrt{n_k}}{\sqrt{n_k}}\right) &= \frac{1}{2} \sum_{k=1}^{\infty} \lambda_k \mathrm{div}\left((\nabla \otimes \nabla) n_k - \frac{\nabla n_k \otimes \nabla n_k}{n_k}\right)\\ &= \frac{1}{2} \sum_{k=1}^{\infty} \lambda_k \mathrm{div}\Bigg[(\nabla \otimes \nabla) n_k + \frac{\nabla n \otimes \nabla n}{n^2} - 2\frac{\nabla n \otimes \nabla n_k}{n}\\ &\qquad - n_k \left(\frac{\nabla n_k}{n_k} - \frac{\nabla n}{n}\right) \otimes \left(\frac{\nabla n_k}{n_k} - \frac{\nabla n}{n}\right)\Bigg]\\ &= \frac{1}{2} \mathrm{div}\left((\nabla \otimes \nabla) n - \frac{\nabla n \otimes \nabla n}{n}\right) + \mathrm{div}(n\theta_{os})\\ &= n \nabla \left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) + \mathrm{div}(n\theta_{os}). \qquad (6.22)\end{aligned}$$

Putting (6.21) and (6.22) into (6.19) gives (6.13). □

The temperature tensor cannot be expressed in terms of the particle and current densities without further assumptions, and as in the derivation of the classical hydrodynamic model (see Section 4.2), we need a closure condition to obtain a closed set of equations. In the literature, the following conditions have been used:

- Derive an evolution equation for the temperature tensor and close the corresponding equation (we do not consider this approach here since we derive a temperature equation in Section 6.3).
- Assume that the temperature tensor is diagonal with equal entries on the diagonal,
  $$\theta = T \cdot \mathrm{Id}\,,$$
  where T is a scalar temperature and $\mathrm{Id} \in \mathbb{R}^{d\times d}$ is the identity matrix. Usually, the temperature T is a constant (*isothermal case*) but also a nonlinear temperature $T(n) = T_0 n^{\alpha-1}$ for some $\alpha > 1$ can be considered (*isentropic case*).

The second choice implies that the (classical) pressure $n\theta$ becomes

$$n\theta = T_0 n^\alpha \cdot \mathrm{Id}\,,$$

with $\alpha = 1$ in the isothermal case and $\alpha > 1$ in the isentropic case.

6.3 Quantum hydrodynamics and the Wigner equation

The quantum hydrodynamic model of the previous section does not include any collisional effects. In fact, the Schrödinger equation only models ballistic quantum transport. In order to include scattering phenomena, we use the Wigner equation of Section 5.3 with the Caldeira-Leggett collision operator (see (5.32)–(5.33)):

$$\begin{aligned} \partial_t w + \frac{\hbar k}{m^*}\cdot\nabla_x w + \frac{q}{m^*}\theta[V]w &= \frac{1}{\tau_0}\mathrm{div}_k\left(kw + \frac{m^* k_B T_0}{\hbar^2}\nabla_k w\right), \quad t>0, \qquad (6.23)\\ w(x,k,0) &= w_I(x,k), \qquad x,k\in\mathbb{R}^d, \end{aligned}$$

where we recall the definition of the pseudodifferential operator $\theta[V]$,

$$\begin{aligned} (\theta[V]w)(x,k,t) &= \frac{i}{(2\pi)^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{m^*}{\hbar}\left[V\left(x+\frac{\eta}{2},t\right) - V\left(x-\frac{\eta}{2},t\right)\right]\\ &\qquad \times w(x,k',t)e^{i(k-k')\cdot\eta}\,dk'\,d\eta, \end{aligned}$$

and m^* is the effective electron mass. According to (5.31) we have assumed a parabolic energy band $E(k) = \hbar^2|k|^2/2m^*$, implying

$$v(k) = \frac{1}{\hbar}\nabla_k E(k) = \frac{\hbar k}{m^*}.$$

First we scale the so-called *Wigner-Fokker-Planck equation* (6.23). For this we introduce reference values for the time τ, the length L, and the potential U by requiring that

$$k_B T_0 = m^* \left(\frac{L}{\tau}\right)^2 \qquad \text{and} \qquad m^* \left(\frac{L}{\tau}\right)^2 = qU.$$

This means that we consider a scaling in which thermal, kinetic, and electric energy are of the same order. Then the scaling

$$t_s = \frac{t}{\tau}, \quad x - s = \frac{x}{L}, \quad k_s = \frac{\hbar\tau}{m^* L} k, \quad V_s = (x_s, t_s) = \frac{V(x,t)}{U}$$

leads, after some elementary computations, to the following dimensionless equations (omitting the index "s"):

$$\begin{aligned} \partial_t w + k \cdot \nabla_x w + \theta[V]w &= \operatorname{div}_k(kw) + \Delta_k w, \qquad t > 0, \\ w(x,k,0) &= w_I(x,k) \qquad x, k \in \mathbb{R}^d, \end{aligned} \tag{6.24}$$

with the scaled operator

$$\begin{aligned} (\theta[V]w)(x,k,t) &= \frac{i}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{1}{\varepsilon} \left[V\left(x + \frac{\varepsilon}{2}\eta, t\right) - V\left(x - \frac{\varepsilon}{2}\eta, t\right) \right] \\ &\quad \times w(x,k',t) e^{i(k-k')\cdot\eta} \, dk' \, d\eta. \end{aligned}$$

The parameter

$$\varepsilon = \frac{\hbar/\tau}{m^*(L/\tau)^2}$$

denotes the scaled Planck constant.

In order to derive macroscopic equations from (6.24) we apply the moment method as in Section 4.2. We define similarly as in the classical case the electron and current density by

$$n(x,t) = \int_{\mathbb{R}^d} w(x,k,t) \, dk, \qquad J(x,t) = -\int_{\mathbb{R}^d} w(x,k,t) k \, dk,$$

and we set

$$\langle f(k) \rangle = \int_{\mathbb{R}^d} w(x,k,t) f(k) \, dk$$

for any function f. Then $n = \langle 1 \rangle$ and $J = -\langle k \rangle$.

Lemma 6.3 *The first moment equations corresponding to (6.24) are as follows:*

$$\begin{aligned}
\partial_t n - \mathrm{div}_x J &= 0, && (6.25)\\
\partial_t J - \mathrm{div}_x \langle k \otimes k \rangle + n \nabla_x V &= -J, && (6.26)\\
\partial_t \langle \frac{1}{2}|k|^2 \rangle + \mathrm{div}_x \langle \frac{1}{2} k |k|^2 \rangle + \nabla_x V \cdot J &= -\langle |k|^2 \rangle + dn, && (6.27)
\end{aligned}$$

where $x \in \mathbb{R}^d$, $t > 0$. The initial conditions are given by

$$\begin{aligned}
n(x,0) &= \int_{\mathbb{R}^d} w_I(x,k)\, dk,\\
J(x,0) &= -\int_{\mathbb{R}^d} w_I(x,k) k\, dk,\\
\langle \frac{1}{2}|k|^2 \rangle (x,0) &= \frac{1}{2} \int_{\mathbb{R}^d} w_I(x,k) |k|^2\, dk, \qquad x \in \mathbb{R}^d.
\end{aligned}$$

Proof: First we integrate (6.24) over $\mathbb{R}^d$ with respect to k:

$$\partial_t \langle 1 \rangle + \mathrm{div}_x \langle k \rangle + \int_{\mathbb{R}^d} \theta[V] w\, dk = \int_{\mathbb{R}^d} \mathrm{div}_k (kw + \nabla_k w)\, dk = 0. \qquad (6.28)$$

For the evaluation of the integral on the left-hand side we recall the Fourier transform

$$\widehat{\phi}(x,\eta,t) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \phi(x,k,t) e^{-ik\cdot\eta}\, dk$$

with inverse

$$\phi(x,k,t) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \widehat{\phi}(x,\eta,t) e^{ik\cdot\eta}\, d\eta.$$

Integrating the last equation over $\mathbb{R}^d$ with respect to k gives the formula

$$\widehat{\phi}(x,0,t) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \widehat{\phi}(x,\eta,t) e^{ik\cdot\eta}\, d\eta\, dk. \qquad (6.29)$$

With this expression we compute

$$\begin{aligned}
\int_{\mathbb{R}^d} \theta[V] w\, dk &= \frac{i}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{1}{\varepsilon} \Big[V(x + \frac{\varepsilon}{2}\eta, t) - V(x - \frac{\varepsilon}{2}\eta, t) \Big] && (6.30)\\
&\qquad \times \widehat{w}(x,\eta,t) e^{ik\cdot\eta}\, d\eta\, dk\\
&= (2\pi)^{d/2} \frac{i}{\varepsilon} \Big[V(x + \frac{\varepsilon}{2}\eta, t) - V(x - \frac{\varepsilon}{2}\eta, t) \Big]_{\eta=0} \widehat{w}(x,0,t)\\
&= 0. && (6.31)
\end{aligned}$$

Therefore, (6.28) equals (6.25).

Now we multiply (6.24) with k_j and integrate over $\mathbb{R}^d$ with respect to k. We compute first, integrating by parts,

$$\begin{aligned}\int_{\mathbb{R}^d} \mathrm{div}_k(kw+\nabla_k w)k_j\,dk &= -\int_{\mathbb{R}^d}(kw+\nabla_k w)\cdot\nabla_k k_j\,dk \\ &= -\int_{\mathbb{R}^d}\left(k_j w+\frac{\partial w}{\partial k_j}\right)dk \;=\; -\langle k_j\rangle \;=\; J_j\end{aligned}$$

and, using (6.29),

$$\begin{aligned}&\int_{\mathbb{R}^d}\theta[V]wk_j\,dk \qquad (6.32)\\ &= \frac{i}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{1}{\varepsilon}\left[V(x+\frac{\varepsilon}{2}\eta,t)-V(x-\frac{\varepsilon}{2}\eta,t)\right]\widehat{w}(x,\eta,t)(-i)\frac{\partial}{\partial\eta_j}e^{ik\cdot\eta}\,d\eta\,dk\\ &= -(2\pi)^{d/2}\frac{1}{\varepsilon}\frac{\partial}{\partial\eta_j}\left[V(x+\frac{\varepsilon}{2}\eta,t)-V(x-\frac{\varepsilon}{2}\eta,t)\right]_{\eta=0}\widehat{w}(x,0,t)\\ &\qquad -(2\pi)^{d/2}\frac{1}{\varepsilon}\left[V(x+\frac{\varepsilon}{2}\eta,t)-V(x-\frac{\varepsilon}{2}\eta,t)\right]_{\eta=0}\frac{\partial\widehat{w}}{\partial\eta_j}(x,0,t)\\ &= -\frac{1}{2}\left[\frac{\partial V}{\partial x_j}(x+\frac{\varepsilon}{2}\eta,t)+\frac{\partial V}{\partial x_j}(x-\frac{\varepsilon}{2}\eta,t)\right]_{\eta=0}\int_{\mathbb{R}^d}w(x,k,t)\,dk\\ &= -\frac{\partial V}{\partial x_j}n(x,t). \qquad (6.33)\end{aligned}$$

Thus

$$\partial_t\langle k_j\rangle+\mathrm{div}_x\langle k_j k\rangle - n\frac{\partial V}{\partial x_j} = J_j.$$

This equals (6.26).

Finally, we multiply (6.24) with $\frac{1}{2}|k|^2$ and integrate over $k\in\mathbb{R}^d$. Again, we first compute the collision and electric-force terms. We obtain, integrating by parts,

$$\begin{aligned}\frac{1}{2}\int_{\mathbb{R}^d}\mathrm{div}_k(kw+\nabla_k w)|k|^2\,dw &= -\int_{\mathbb{R}^d}(kw+\nabla_k w)\cdot k\,dk\\ &= \int_{\mathbb{R}^d}(-w|k|^2+w\mathrm{div}_k k)\,dk \;=\; -\langle|k|^2\rangle+dn\end{aligned}$$

and, with the notation $\delta V(x,\eta,t) = [V(x+\varepsilon\eta/2,t) - V(x-\varepsilon\eta/2,t)]/\varepsilon$ and using (6.29),

$$\begin{aligned}
\frac{1}{2}\int_{\mathbb{R}^d} \theta[V]w|k|^2\,dk &= \frac{i}{2(2\pi)^{d/2}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \delta V(x,\eta,t)\widehat{w}(x,\eta,t)(-\Delta_\eta)e^{ik\cdot\eta}\,d\eta\,dk \\
&= \frac{-i}{2(2\pi)^{d/2}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Delta_\eta(\delta V\cdot\widehat{w})(x,\eta,t)e^{ik\cdot\eta}\,d\eta\,dk \\
&= -\frac{i}{2}(2\pi)^{d/2}\Delta_\eta(\delta V\cdot\widehat{w})(x,0,t) \\
&= -\frac{i}{2}(2\pi)^{d/2}\Big(\Delta_\eta\delta V(x,0,t)\widehat{w}(x,0,t) + 2\nabla_\eta\delta V(x,0,t)\cdot\nabla_\eta\widehat{w}(x,0,t) \\
&\qquad + \delta V(x,0,t)\Delta_\eta\widehat{w}(x,0,t)\Big) \\
&= -\nabla_x V\cdot\langle k\rangle \;=\; \nabla_x V\cdot J,
\end{aligned}$$

since

$$\begin{aligned}
\delta V(x,0,t) &= 0, \\
\nabla_\eta\delta V(x,0,t) &= \frac{1}{2}\Big[\nabla_x V(x+\frac{\varepsilon}{2}\eta,t) + \nabla_x V(x-\frac{\varepsilon}{2}\eta,t)\Big]_{\eta=0} \;=\; \nabla_x V, \\
5\Delta_\eta\delta V(x,0,t) &= \frac{\varepsilon}{4}\Big[\Delta_x V(x+\frac{\varepsilon}{2}\eta,t) - \Delta_x V(x-\frac{\varepsilon}{2}\eta,t)\Big]_{\eta=0} \;=\; 0, \\
\nabla_\eta\widehat{w}(x,0,t) &= \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d} w(x,k,t)\nabla_\eta e^{-i\eta\cdot k}\,dk\Big|_{\eta=0} \;=\; -\frac{i}{(2\pi)^{d/2}}\langle k\rangle.
\end{aligned}$$

Thus, (6.27) is shown. □

The system of moment equations (6.25)–(6.27) has to be closed. As in the case of classical kinetic theory, we achieve the closure by assuming that the Wigner function w is close to a wave vector displaced equilibrium density,

$$w(x,k,t) = w_{eq}(x,k-u(x,t),t),$$

similar to (4.43), where w_{eq} is assumed to be given by

$$\begin{aligned}
w_{eq}(x,k,t) = A(x,t)e^{(-|k|^2/2+V)/T(x,t)}\Big[1 &+ \frac{\varepsilon^2}{8T(x,t)^2}\Delta_x V + \frac{\varepsilon^2}{24T(x,t)^3}|\nabla_x V|^2 \\
&-\frac{\varepsilon^2}{24T(x,t)^3}\sum_{j,\ell=1}^{d} k_j k_\ell \frac{\partial^2 V}{\partial x_j\partial x_\ell} + O(\varepsilon^4)\Big],
\end{aligned} \tag{6.34}$$

and the functions $u(x,t)$ and $T(x,t)$, later interpreted as the electron group velocity and the electron temperature, have to be determined. Notice that the equilibrium density w_{eq} reduces in the classical case $\varepsilon \to 0$ to the classical Maxwellian,

$$w_{eq}^{cl}(x,k,t) = A(x,t)ne^{-|k|^2/2T},$$

since $n = e^{V/T}$ in classical thermal equilibrium. The first moments with respect to w_{eq} are computed in the following lemma.

Lemma 6.4 *The first moments corresponding to (6.34) are*

$$
\begin{aligned}
n = \langle 1 \rangle &= B(x,t)e^{V/T}\left(1 + \frac{\varepsilon^2}{12T^2}\Delta_x V + \frac{\varepsilon^2}{24T^3}|\nabla_x V|^2\right) + O(\varepsilon^4), && (6.35)\\
J = -\langle k \rangle &= -nu, && (6.36)\\
\langle k \otimes k \rangle &= nu \otimes u + nT\,\mathrm{Id} - \frac{\varepsilon^2}{12T} n(\nabla \otimes \nabla)V + O(\varepsilon^4), && (6.37)\\
\langle k|k|^2 \rangle &= nu|u|^2 + \left(dT - \frac{\varepsilon^2}{12T}\Delta_x V\right) nu + 2\left(T\,\mathrm{Id} - \frac{\varepsilon^2}{12T}(\nabla \otimes \nabla)V\right) nu + O(\varepsilon^4), && (6.38)
\end{aligned}
$$

where

$$B(x,t) = A(x,t)(2\pi T)^{d/2},$$

Id *is the identiy matrix of* $\mathbb{R}^{d\times d}$ *and* $(\nabla \otimes \nabla)V$ *the Hessian of* V*.*

Proof: The symmetry of w_{eq} with respect to k implies that all odd order moments of w_{eq} vanish. Therefore,

$$
\begin{aligned}
\langle 1 \rangle &= \int_{\mathbb{R}^d} w_{eq}(x, k - u(x,t), t)\, dk = \int_{\mathbb{R}^d} w_{eq}(x,\eta,t)\, d\eta,\\
\langle k \rangle &= \int_{\mathbb{R}^d} (\eta + u) w_{eq}(x,\eta,t)\, d\eta = nu,\\
\langle k \otimes k \rangle &= \int_{\mathbb{R}^d} (\eta+u)\otimes(\eta+u) w_{eq}(x,\eta,t)\, d\eta = nu\otimes u + \int_{\mathbb{R}^d} \eta\otimes\eta\, w_{eq}\, d\eta, && (6.39)\\
\langle k|k|^2 \rangle &= nu|u|^2 + u\int_{\mathbb{R}^d} |\eta|^2 w_{eq}\, d\eta + 2\left(\int_{\mathbb{R}^d} \eta\otimes\eta\, w_{eq} d\eta\right) u. && (6.40)
\end{aligned}
$$

We derive now an $O(\varepsilon^2)$ approximation for n. Since

$$
\begin{aligned}
\int_{\mathbb{R}^d} e^{(-|\eta|^2/2+V)/T}\, d\eta &= T^{d/2}e^{V/T}\int_{\mathbb{R}^d} e^{-|z|^2/2}\, dz = (2\pi T)^{d/2}e^{V/T},\\
\int_{\mathbb{R}^d} \eta_j\eta_\ell e^{(-|\eta|^2/2+V)/T}\, d\eta &= T^{d/2+1}e^{V/T}\int_{\mathbb{R}^d} z_j z_\ell e^{-|z|^2/2}\, dz\\
&= (2\pi T)^{d/2} T e^{V/T}\,\mathrm{Id}\,,
\end{aligned}
$$

we obtain

$$
\begin{aligned}
n &= Be^{V/T}\left(1 + \frac{\varepsilon^2}{8T^2}\Delta_x V + \frac{\varepsilon^2}{24T^3}|\nabla_x V|^2 - \frac{\varepsilon^2}{24T^3}\sum_{j,\ell=1}^{d}\frac{\partial^2 V}{\partial x_j \partial x_\ell}\delta_{j\ell}\right) + O(\varepsilon^4)\\
&= Be^{V/T}\left(1 + \frac{\varepsilon^2}{12T^2}\Delta_x V + \frac{\varepsilon^2}{24T^3}|\nabla_x V|^2\right) + O(\varepsilon^4). && (6.41)
\end{aligned}
$$

The computation of the integrals on the right-hand sides of (6.39)–(6.40) requires the evaluation of

$$\begin{aligned}\int_{\mathbb{R}^d} \eta_j \eta_\ell \eta_m \eta_n e^{-|\eta|^2/2T}\, d\eta &= T^{d/2+2} \int_{\mathbb{R}^d} z_j z_\ell z_m z_n e^{-|z|^2/2}\, dz \\ &= T^{d/2+2} \alpha(j,\ell,m,n),\end{aligned}$$

where

$$\alpha(j,\ell,m,n) = (2\pi)^{d/2} \begin{cases} 3 & : \; j = \ell = m = n \\ 1 & : \; j = \ell \neq m = n \text{ or } j = m \neq \ell = n \text{ or } j = n \neq \ell = m \\ 0 & : \; \text{else.} \end{cases}$$

This yields

$$\begin{aligned}\int_{\mathbb{R}^d} \eta_j \eta_\ell w_{eq} d\eta &= Be^{V/T} \Bigg[T \left(1 + \frac{\varepsilon^2}{8T^2} \Delta_x V + \frac{\varepsilon^2}{24T^2} |\nabla_x V|^2 \right) d_{j\ell} \\ &\qquad - \frac{\varepsilon^2}{24T^3} \sum_{m,n=1}^{d} T^2 \alpha(j,\ell,m,n) \frac{\partial^2 V}{\partial x_m \partial x_n} \Bigg] + O(\varepsilon^4).\end{aligned}$$

A simple computation shows that

$$\sum_{m,n=1}^{d} \alpha(j,\ell,m,n) \frac{\partial^2 V}{\partial x_m \partial x_n} = 2 \frac{\partial^2 V}{\partial x_j \partial x_\ell} + \Delta_x V \delta_{j\ell}.$$

Hence

$$\begin{aligned}\int_{\mathbb{R}^d} \eta_j \eta_\ell w_{eq} d\eta &= Be^{V/T} \Bigg[\left(T + \frac{\varepsilon^2}{12T} \Delta_x V + \frac{\varepsilon^2}{24T^2} |\nabla_x V|^2 \right) d_{j\ell} \\ &\qquad - \frac{\varepsilon^2}{12T} \frac{\partial^2 V}{\partial x_j \partial x_\ell} \Bigg] + O(\varepsilon^4). \qquad (6.42)\end{aligned}$$

We substitute the following expressions, derived from (6.41),

$$\begin{aligned} Be^{V/T} \left(T + \frac{\varepsilon^2}{12T} \Delta_x V + \frac{\varepsilon^2}{24T^2} |\nabla_x V|^2 \right) &= nT + O(\varepsilon^4), \\ Be^{V/T} &= n + O(\varepsilon^2),\end{aligned}$$

into (6.42) yielding

$$\int_{\mathbb{R}^d} \eta_j \eta_\ell w_{eq} d\eta = nT \delta_{j\ell} - \frac{\varepsilon^2}{12T} n \frac{\partial^2 V}{\partial x_j \partial x_\ell} + O(\varepsilon^4).$$

Putting this into (6.39) and (6.40) gives the remaining formulas (6.37) and (6.38). □

We wish to replace the second derivatives of the electrostatic potential in (6.37) and (6.38) by (derivatives of) the electron density. From the Taylor expansion $\log(x+\delta) = \log x + O(\delta)$ and the formula (6.35), written as

$$n = B(x,t)e^{V/T} + O(\varepsilon^2),$$

we obtain

$$\ln n = \ln(B(x,t)e^{V/T}) + O(\varepsilon^2)$$

and therefore

$$\frac{\partial^2 \ln n}{\partial x_j \partial x_\ell} = \frac{\partial^2}{\partial x_j \partial x_\ell}\left(\ln B(x,t) + \frac{V}{T}\right) + O(\varepsilon^2).$$

Our main assumption is that $B(x,t)$ and $T(x,t)$ are slowly varying with respect to x, compared to spatial variations of $V(x,t)$ such that we can write

$$\frac{\partial^2 \ln n}{\partial x_j \partial x_\ell} = \frac{1}{T}\frac{\partial^2 V}{\partial x_j \partial x_\ell} + O(\varepsilon^2). \tag{6.43}$$

Lemma 6.5 *Under the assumption (6.43) the second and third moments corresponding to (6.34) are*

$$\begin{aligned}
\langle k \otimes k \rangle &= \frac{J \otimes J}{n} + nT\ \mathrm{Id} - \frac{\varepsilon^2}{12} n (\nabla \otimes \nabla) \ln n + O(\varepsilon^4), \\
\langle k |k|^2 \rangle &= -\frac{J|J|^2}{n^2} - \left(dT - \frac{\varepsilon^2}{12}\Delta_x \ln n\right) J - 2\left(T\ \mathrm{Id} - \frac{\varepsilon^2}{12}(\nabla \otimes \nabla)\ln n\right) J + O(\varepsilon^4).
\end{aligned}$$

Proof: The lemma immediately follows from (6.37)–(6.38) and (6.43) using $J = -nu$. □

Now we can combine Lemmas 6.3 and 6.5 to achieve our main result.

Theorem 6.6 *The first moment equations corresponding to (6.24) with the closure condition (6.34) and the assumption (6.43) are given by, up to order $O(\varepsilon^4)$,*

$$\partial_t n - \mathrm{div} J = 0, \tag{6.44}$$

$$\partial_t J - \mathrm{div}\frac{J \otimes J}{n} - \nabla(nT) + n\nabla V + \frac{\varepsilon^2}{6} n \nabla\left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) = -J, \tag{6.45}$$

$$\partial_t e - \mathrm{div}\left((e - P)\frac{J}{n}\right) + \nabla V \cdot J = -2\left(e - \frac{d}{2}n\right), \tag{6.46}$$

where

$$\begin{aligned}
e &= \frac{|J|^2}{2n} + \frac{d}{2}nT - \frac{\varepsilon^2}{24} n \Delta \ln n, \\
P &= -nT\ \mathrm{Id} + \frac{\varepsilon^2}{12} n (\nabla \otimes \nabla) \ln n
\end{aligned}$$

are the energy density *and the* stress tensor*, respectively.*

Equations (6.44)–(6.46) are solved for $x \in \mathbb{R}^d$, $t > 0$, supplemented with the initial condtions

$$n(\cdot,0) = n_I, \qquad J(\cdot,0) = J_I, \qquad e(\cdot,0) = e_I \qquad \text{in } \mathbb{R}^d.$$

Proof: A simple computation shows that

$$\operatorname{div}\left(n(\nabla \otimes \nabla)\ln n\right) = 2n\nabla\left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right),$$

which proves (6.45). Equation (6.46) follows from (6.27) and Lemma 6.5. □

Equations (6.44)–(6.46) are called the *quantum hydrodynamic equations.* The above energy density is the sum of the kinetic energy, $|J|^2/2n = n|u|^2/2$, thermal energy, $(d/2)nT$, and quantum energy, $-(\varepsilon^2/24)n\Delta \ln n$. The stress tensor consists of a classical part, $-nT\,\mathrm{Id}$, and a quantum part, $(\varepsilon^2/12)n(\nabla\otimes\nabla)\ln n$. It allows to write (6.45) in the more compact form

$$\partial_t J - \operatorname{div}\left(\frac{J\otimes J}{n} + P\right) + n\nabla V = -J.$$

We have used the formulation (6.45) since it allows to interpret the quantum term as a force term with the self-interacting potential $\Delta\sqrt{n}/\sqrt{n}$, which is also called the *Bohm potential.*

The terms on the right-hand sides of (6.45) and (6.46) are called the *momentum* and *energy relaxation* terms, respectively. Neglecting all forces in the equations, we obtain

$$\partial_t J = -J, \qquad \partial_t e = -2\left(e - \frac{d}{2}n\right).$$

The solution to this system satisfies

$$J(t) \to 0, \qquad e(t) \to \frac{d}{2}n,$$

i.e., the current density vanishes and the energy relaxes to the thermal energy (with constant temperature $T = 1$). We notice that for constant temperature, equations (6.44)–(6.45) give the quantum hydrodynamic model from Section 6.2 if $\theta = T$.

Finally, we rewrite the quantum hydrodynamic equations in unscaled form:

$$\partial_t n - \frac{1}{q}\operatorname{div} J = 0,$$

$$\partial_t J - \frac{1}{q}\operatorname{div}\left(\frac{J\otimes J}{n}\right) - \frac{qk_B}{m^*}\nabla(nT) + \frac{q^2}{m^*}n\nabla V + \frac{q\hbar^2}{6(m^*)^2}n\nabla\left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) = -\frac{J}{\tau},$$

$$\partial_t e - \frac{1}{q}\operatorname{div}\left((e-P)\frac{J}{n}\right) + J\cdot\nabla V = -\frac{2}{\tau}\left(e - \frac{d}{2}nT_0\right),$$

where now

$$\begin{aligned} e &= m^* \frac{|J|^2}{2n} + \frac{d}{2} n k_B T - \frac{\hbar^2}{24 m^*} \Delta \ln n, \\ P &= -n k_B T \ \mathrm{Id} + \frac{\hbar^2}{12 m^*} n (\nabla \otimes \nabla) \ln n. \end{aligned}$$

6.4 Derivation of the quantum drift-diffusion equations

In this section we derive a quantum analogue of the classical drift-diffusion equations from a Wigner-Boltzmann model (i.e. the Wigner equation including a collision operator). Our main assumption is that collisions are dominant, similar to the situation of Section 4.1. We proceed as in [25] with the scaled Wigner equation

$$\begin{aligned} \partial_t w + k \cdot \nabla_x w + \theta[V] w &= Q(w), \qquad t > 0, && (6.47) \\ w(x, k, 0) &= w_I(x, k), \qquad x, k \in \mathbb{R}^d && (6.48) \end{aligned}$$

(see (6.24)), recalling the definition of the pseudodifferential operator $\theta[V]$,

$$(\theta[V] w)(x, k, t) = \frac{i}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \delta V(x, \eta, t) w(x, k', t) e^{i(k-k') \cdot \eta} \, dk' \, d\eta,$$

where

$$\delta V(x, \eta, t) = \frac{1}{\varepsilon} \left[V(x + \frac{\varepsilon}{2} \eta, t) - V(x - \frac{\varepsilon}{2} \eta, t) \right].$$

The collision operator $Q(w)$ is assumed to be of relaxation type, i.e.

$$Q(w) = M[w] - w, \qquad (6.49)$$

where $M[w]$ is a quantum analogue of the Maxwellian $\exp(A(x,t) - |k|^2/2)$ for some function $A(x,t)$. The quantum Maxwellian $M[w]$ is assumed to minimize the quantum entropy, under the constraint of given particle density.

More precisely, similar to Section 4.2 we introduce the quantum entropy

$$H(w) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} w \left(\mathrm{Ln}\, w - 1 + \frac{|k|^2}{2} - V(x) \right) dx \, dk,$$

where here and in the following we omit the time variable which only acts as a parameter and where the "quantum logarithm" is defined by

$$\mathrm{Ln}\, w = W(\ln W^{-1}(w)).$$

In this definition, W is the Wigner transform

$$(W(\varrho))(x, k) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varrho \left(x + \frac{\varepsilon}{2} \eta, x - \frac{\varepsilon}{2} \eta \right) e^{i \eta \cdot k} \, d\eta$$

and W^{-1} its inverse, the so-called *Weyl quantization*,

$$(W^{-1}(w))\phi(x) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} w\left(\frac{x+y}{2}, k\right)\phi(y)e^{ik\cdot(x-y)/\varepsilon}\,dk\,dy,$$

where $\varrho(x, x')$ is a density matrix and $w(x, k)$ a function. The logarithm ln is defined in the sense of operator theory.

The quantum logarithm has the following properties.

Lemma 6.7 *(1) The inverse of the quantum logarithm* Ln *is given by the so-called quantum exponential* Exp*, defined by*

$$\text{Exp}\, w = W(\exp W^{-1}(w)) \qquad \textit{for suitable functions } w(x,k).$$

(2) The functionals Ln *and* Exp *are (Fréchet) differentiable and*

$$\frac{d}{dw}\text{Ln}\, w = \frac{1}{w}, \qquad \frac{d}{dw}\text{Exp}\, w = \text{Exp}\, w.$$

Proof: (1) Formally, it holds

$$\begin{aligned}\text{Ln}(\text{Exp}\, w) &= W\left(\exp W^{-1}\Big(W\big(\ln W^{-1}(w)\big)\Big)\right) = W\left(\exp(\ln W^{-1}(w))\right) \\ &= W(W^{-1}(w)) = w.\end{aligned}$$

(2) Left to the reader. □

Lemma 6.8 *The solution of the constrained minimization problem*

$$H(w^*) = \min\left\{H(w) : \int w(x,k)\,dk = n(x) \textit{ for all } x\right\},$$

for given electron density $n(x)$, if it exists, is

$$w^* = \text{Exp}(A(x) - |k|^2/2),$$

where the constant $A(x)$ is fixed by the constraint

$$\int_{\mathbb{R}^d} w^*(x,k)\,dk = n(x).$$

We assume that the above constraint fixes $A(x)$ uniquely.

Proof: We use Lagrange multipliers and define

$$F(w,\lambda) = H(w) + \int_{\mathbb{R}^d}\lambda(x)\left(\int_{\mathbb{R}^d} w(x,k)\,dk - n(x)\right)dx.$$

By Lemma 6.7 (2),

$$\left(\frac{d}{dw}H(w)\right)(u) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left(\mathrm{Ln}\, w + \frac{|k|^2}{2} - V(x)\right)u(x,k)\,dx\,dk.$$

Furthermore,

$$\left(\frac{d}{dw}\int_{\mathbb{R}^d} w\,dk\right)(g) = \int_{\mathbb{R}^d} g(k)\,dk.$$

Therefore, we obtain the necessary condition for an extremum

$$0 = \left(\frac{d}{dw}F(w^*,\lambda^*)\right)(u) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left(\mathrm{Ln}\, w^* + \frac{|k|^2}{2} - V(x) + \lambda^*(x)\right)u(x,k)\,dx\,dk$$

for all functions $u(x,k)$. This implies that

$$\mathrm{Ln}\, w^* + \frac{|k|^2}{2} - V(x) + \lambda^*(x) = 0$$

or

$$w^* = \mathrm{Exp}(A(x) - |k|^2/2) \qquad \text{for } A(x) = V(x) - \lambda^*(x),$$

which shows the lemma. ☐

Definition 6.9 *The* quantum Maxwellian *is defined by*

$$M[w] = \mathrm{Exp}(A(x) - |k|^2/2)$$

and $A(x)$ is determined through

$$\int_{\mathbb{R}^d} M[w]\,dk = \int_{\mathbb{R}^d} w\,dk.$$

With this definition, the relaxation-time Wigner model (6.47)–(6.49) is well defined. In order to derive the quantum drift-diffusion model, we consider the case of dominant collisions and large time, i.e., we rescale the equation (6.47) by replacing

$$t \quad \text{by} \quad t/\delta \qquad \text{and} \qquad Q(w) \quad \text{by} \quad Q(w)/\delta.$$

This yields

$$\begin{aligned} \delta\,\partial_t w_\delta + k\cdot\nabla_x w_\delta + \theta[V]w_\delta &= Q(w_\delta) = \frac{1}{\delta}(M[w_\delta] - w_\delta), \qquad t > 0, && (6.50)\\ w_\delta(x,k,0) &= w_I(x,k), \qquad x,k\in\mathbb{R}^d. && (6.51)\end{aligned}$$

Now we perform the limit $\delta \to 0$.

Theorem 6.10 *Let w_δ be a solution of (6.50)–(6.51). Then, formally, $w_\delta \to w_0$ as $\delta \to 0$, where $w_0 = \mathrm{Exp}(A - |k|^2/2)$ and $A(x,t)$ is a solution of*

$$\begin{aligned} \partial_t n - \mathrm{div} J &= 0, \qquad J \;=\; \mathrm{div} P - n\nabla V, \qquad t > 0 \qquad (6.52)\\ n(x,0) &= n_I(x), \qquad x \in \mathbb{R}^d, \qquad (6.53) \end{aligned}$$

where

$$\begin{aligned} n &= \int_{\mathbb{R}^d} \mathrm{Exp}\left(A - \frac{|k|^2}{2}\right) dk, \qquad (6.54)\\ P &= \int_{\mathbb{R}^d} k \otimes k \; \mathrm{Exp}\left(A - \frac{|k|^2}{2}\right) dk, \qquad (6.55) \end{aligned}$$

and the initial datum is given by

$$n_I = \int_{\mathbb{R}^d} w_I \, dk.$$

Proof: Performing the (formal) limit $\delta \to 0$ in (6.50) directly, we obtain $w_0 = M[w_0]$. In order to derive higher-order corrections, we apply the so-called *Chapman-Enskog method* first used in classical kinetic theory [20, 29]. For this, we define the first-order correction

$$w_\delta^{(1)} = \frac{1}{\delta}(w_\delta - M[w_\delta]) \qquad \text{or} \qquad w_\delta = M[w_\delta] + \delta w_\delta^{(1)}.$$

Denote by $w^{(1)}$ the limit of $w_\delta^{(1)}$. Since, as $\delta \to 0$,

$$w_\delta^{(1)} = -Q(w_\delta) = -\delta\partial_t w_\delta - (k \cdot \nabla_x w_\delta + \theta[V]w_\delta) \to -(k \cdot \nabla_x w_0 + \theta[V]w_0),$$

we infer

$$w^{(1)} = -(k \cdot \nabla_x w_0 + \theta[V]w_0). \qquad (6.56)$$

We claim that in the (formal) limit $\delta \to 0$,

$$\partial_t \int_{\mathbb{R}^d} w_0 \, dk + \mathrm{div}_x \int_{\mathbb{R}^d} w^{(1)} k \, dk + \int_{\mathbb{R}^d} \theta[V] w^{(1)} \, dk = 0. \qquad (6.57)$$

To see this we integrate the rescaled Wigner equation (6.50) with respect to $k \in \mathbb{R}^d$,

$$\delta\partial_t \int_{\mathbb{R}^d} w_\delta \, dk + \mathrm{div}_x \int_{\mathbb{R}^d} w_\delta k \, dk + \int_{\mathbb{R}^d} \theta[V] w_\delta \, dk = \int_{\mathbb{R}^d} Q(w_\delta) \, dk, \qquad (6.58)$$

and manipulate this equation term by term. By definition, $M[w_\delta]$ and w_δ have the same (zeroth-order) moments, yielding

$$\int_{\mathbb{R}^d} Q(w_\delta) \, dk = \frac{1}{\delta} \int_{\mathbb{R}^d} (M[w_\delta] - w_\delta) \, dk = 0.$$

The function $k \mapsto (\theta[V]M[w_\delta])(x,k)$ is odd since, transforming $\eta \mapsto -\eta$, $k' \mapsto -k'$,

$$\begin{aligned}(\theta[V]M[w_\delta])(x,-k) &= \frac{i}{(2\pi)^{d/2}} \int_{\mathbb{R}^{2d}} \delta V(x,\eta) \mathrm{Exp}\left(A - \frac{|k'|^2}{2}\right) e^{i\eta\cdot(-k-k')}\, d\eta\, dk' \\ &= \frac{-i}{(2\pi)^{d/2}} \int_{\mathbb{R}^{2d}} \delta V(x,\eta) \mathrm{Exp}\left(A - \frac{|k'|^2}{2}\right) e^{i\eta\cdot(k-k')}\, d\eta\, dk' \\ &= -(\theta[V]M[w_\delta])(x,k),\end{aligned}$$

using $\delta V(x,-\eta) = -\delta V(x,\eta)$. This implies that the integral over $\theta[V]M[w_\delta]$ vanishes and therefore,

$$\int_{\mathbb{R}^d} \theta[V]w_\delta\, dk = \int_{\mathbb{R}^d} \theta[V](M[w_\delta] + \delta w_\delta^{(1)})\, dk = \delta \int_{\mathbb{R}^d} \theta[V]w_\delta^{(1)}\, dk.$$

Since also $k \mapsto M[w_\delta](x,k)k$ is odd,

$$\int_{\mathbb{R}^d} w_\delta k\, dk = \delta \int_{\mathbb{R}^d} w_\delta^{(1)} k\, dk.$$

Thus (6.57) follows from (6.58) after division by δ.

Setting $n = \int w_0\, dk$ and using the expression (6.56) in (6.57) we obtain

$$\partial_t n - \mathrm{div}_x \int_{\mathbb{R}^d} (k \cdot \nabla_x w_0 + \theta[V]w_0) k\, dk + \int_{\mathbb{R}^d} \theta[V] w^{(1)}\, dk = 0.$$

In Section 6.3 we have derived the following formulas (see (6.31) and (6.33)):

$$\begin{aligned}\int_{\mathbb{R}^d} \theta[V]w\, dk &= 0 \qquad \text{for all functions } w, \\ \int_{\mathbb{R}^d} \theta[V]w_0 k\, dk &= -n\nabla_x V \qquad \text{for all functions } w_0 \text{ such that } n = \textstyle\int w_0\, dk.\end{aligned}$$

This gives

$$\partial_t n - \mathrm{div}_x \int_{\mathbb{R}^d} \mathrm{div}_x(k \otimes k) w_0\, dk + \mathrm{div}_x(n \nabla V_x) = 0,$$

and hence the conclusion. $\square$

The initial-value problem (6.52)–(6.55) can be considered as a generalized quantum drift-diffusion equation. The drift current is given by $n\nabla V$ and P is a quantum pressure tensor. In the classical case, $P = n\delta_{ij}$ yielding the drift-diffusion equations of Section 4.1. In general, P is a nonlocal operator involving $A(x)$ which relates to n through (6.54)–(6.55). The usual quantum drift-diffusion model is obtained from (6.52)–(6.55) as a $O(\varepsilon^4)$ approximation.

Theorem 6.11 *Let (n, J) be a solution of the generalized quantum drift-diffusion equations (6.52)–(6.55). Then, formally, $J = J_0 + O(\varepsilon^4)$ and (n, J_0) solves*

$$\partial_t n - \operatorname{div} J_0 = 0, \qquad J_0 = \nabla n + n\nabla V + \frac{\varepsilon^2}{6} n \nabla \left(\frac{\Delta\sqrt{n}}{\sqrt{n}} \right), \qquad t > 0, \tag{6.59}$$

$$n(x, 0) = n_I(x), \qquad x \in \mathbb{R}^d. \tag{6.60}$$

The system (6.59)–(6.60) is called the *quantum drift-diffusion model*. For the proof of Theorem 6.11 we need the following results. It holds

$$\int_{\mathbb{R}^d} e^{-|k|^2/2}\, dk = (2\pi)^{d/2}, \tag{6.61}$$

$$\int_{\mathbb{R}^d} k_i k_j e^{-|k|^2/2}\, dk = (2\pi)^{d/2} \delta_{ij}, \tag{6.62}$$

$$\int_{\mathbb{R}^d} k_i k_j k_\ell k_m e^{-|k|^2/2}\, dk = (2\pi)^{d/2} \left(\delta_{ij}\delta_{\ell m} + \delta_{i\ell}\delta_{jm} + \delta_{im}\delta_{j\ell} \right). \tag{6.63}$$

Lemma 6.12 *The following (formal) expansion holds:*

$$\operatorname{Exp}\left(A - \frac{|k|^2}{2} \right) = \exp\left(A - \frac{|k|^2}{2} \right) \left[1 + \frac{\varepsilon^2}{8} \left(\Delta A + \frac{1}{3}|\nabla A|^2 - \frac{1}{3} k^\top D^2 A k \right) \right] + O(\varepsilon^4),$$

where $D^2 A$ denotes the Hessian of A.

The technical proof uses heavily pseudo-differential calculus and is therefore omitted. We refer to [25, Lemma 5.6] for details of the proof.

Proof of Theorem 6.11: First we develop the electron density in powers of ε^2. By (6.54), Lemma 6.12, and (6.61)–(6.62),

$$\begin{aligned} n &= e^A \left[1 + \frac{\varepsilon^2}{8} \left(\Delta A + \frac{1}{3}|\nabla A|^2 \right) \right] \int_{\mathbb{R}^d} e^{-|k|^2/2}\, dk - \frac{\varepsilon^2}{24} e^A \sum_{i,j} \partial^2_{ij} A \int_{\mathbb{R}^d} k_i k_j e^{-|k|^2/2}\, dk + O(\varepsilon^4) \\ &= (2\pi)^{d/2} e^A \left[1 + \frac{\varepsilon^2}{8} \left(\Delta A + \frac{1}{3}|\nabla A|^2 \right) - \frac{\varepsilon^2}{24} \Delta A \right] + O(\varepsilon^4) \\ &= (2\pi)^{d/2} e^A \left[1 + \frac{\varepsilon^2}{12} \left(\Delta A + \frac{1}{2}|\nabla A|^2 \right) \right] + O(\varepsilon^4), \end{aligned} \tag{6.64}$$

where ∂_i abbreviates the partial derivative $\partial/\partial x_i$. In particular,

$$n = (2\pi)^{d/2} e^A + O(\varepsilon^2). \tag{6.65}$$

Next we develop the quantum pressure in powers of ε^2. Using (6.55), Lemma 6.12, and (6.62)–(6.63),

$$\begin{aligned} P_{ij} &= e^A \left[1 + \frac{\varepsilon^2}{8}\left(\Delta A + \frac{1}{3}|\nabla A|^2\right)\right] \int_{\mathbb{R}^d} k_i k_j e^{-|k|^2/2}\, dk \\ &\qquad -\frac{\varepsilon^2}{24} e^A \sum_{\ell,m} \partial^2_{\ell m} A \int_{\mathbb{R}^d} k_i k_j k_\ell k_m e^{-|k|^2/2}\, dk + O(\varepsilon^4) \\ &= (2\pi)^{d/2} e^A \left[1 + \frac{\varepsilon^2}{8}\left(\Delta A + \frac{1}{3}|\nabla A|^2\right)\right] \delta_{ij} - \frac{\varepsilon^2}{24}(2\pi)^{d/2} e^A \left(\Delta A \delta_{ij} + 2\partial^2_{ij} A\right) + O(\varepsilon^4) \\ &= (2\pi)^{d/2} e^A \left[1 + \frac{\varepsilon^2}{12}\left(\Delta A + \frac{1}{2}|\nabla A|^2\right)\right] \delta_{ij} - \frac{\varepsilon^2}{12}(2\pi)^{d/2} e^A \partial^2_{ij} A + O(\varepsilon^4). \end{aligned}$$

We employ the developments (6.64) and (6.65) to obtain

$$P_{ij} = n\delta_{ij} - \frac{\varepsilon^2 n}{12}\partial^2_{ij} A + O(\varepsilon^4).$$

Thus, using $\nabla n = n\nabla A + +O(\varepsilon^2)$, which follows immediately from (6.65),

$$\begin{aligned} (\mathrm{div} P)_j = \sum_i \partial_i P_{ij} &= \partial_j n - \frac{\varepsilon^2}{12}\sum_i \left(\partial_i n \partial^2_{ij} A + n\partial^3_{iij} A\right) + O(\varepsilon^4) \\ &= \partial_j n - \frac{\varepsilon^2}{12}\sum_i n\partial_j \left(\frac{1}{2}(\partial_i A)^2 + \partial_{ii} A\right) + O(\varepsilon^4). \end{aligned}$$

Hence,

$$\mathrm{div} P = \nabla n - \frac{\varepsilon^2}{12} n\nabla \left(\Delta A + \frac{1}{2}|\nabla A|^2\right) + O(\varepsilon^4).$$

Finally, we relate A and n using (6.65) or

$$\nabla A = \frac{\nabla n}{n} + O(\varepsilon^2),$$

from which we conclude

$$\begin{aligned} \Delta A + \frac{1}{2}|\nabla A|^2 &= \frac{\Delta n}{n} - \frac{|\nabla n|^2}{n^2} + \frac{1}{2}\left|\frac{\nabla n}{n}\right|^2 + O(\varepsilon^2) = \frac{\Delta n}{n} - \frac{|\nabla n|^2}{2n^2} + O(\varepsilon^2) \\ &= 2\frac{\Delta\sqrt{n}}{\sqrt{n}} + O(\varepsilon^2). \end{aligned}$$

Therefore,

$$J = \mathrm{div} P - n\nabla V = \nabla n - n\nabla V - \frac{\varepsilon^2}{2} n\nabla \left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) + O(\varepsilon^4),$$

and the conclusion follows. □

References

[1] R. Adler, A. Smith, and R. Longini. *Introduction to Semiconductor Physics.* John Wiley & Sons, New York, 1964.

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

[3] N. Ashcroft and N. Mermin. *Solid State Physics.* Sanners College, Philadelphia, 1976.

[4] H. Babovsky. *Die Boltzmann-Gleichung.* Teubner, Stuttgart, 1998.

[5] G. Baccarani and M. Wordeman. An investigation on steady-state velocity overshoot in silicon. *Solid-State Electr.* **29** (1982), 970-977.

[6] C. Bardos, F. Golse, and C. Levermore. Fluid dynamic limits of kinetic equations I: formal derivations. *J. Stat. Phys.* **63** (1991), 323-344.

[7] C. Bardos, F. Golse, and C. Levermore. Fluid dynamic limits of kinetic equations II: convergence proofs for the Boltzmann equation. *Comm. Pure Appl. Math.* **46** (1993), 667-753.

[8] N. Ben Abdallah and P. Degond. On a hierarchy of macroscopic models for semiconductors. *J. Math. Phys.* **37** (1996), 3308-3333.

[9] N. Ben Abdallah, P. Degond, and S. Génieys. An energy-transport model for semiconductors derived from the Boltzmann equation. *J. Stat. Phys.* **84** (1996), 205-231.

[10] K. Bløtekjær. Transport equations for electrons in two-valley semiconductors. *IEEE Trans. Electr. Devices* **17** (1970), 38-47.

[11] K. Brennan. *The Physics of Semiconductors.* Cambridge University Press, Cambridge, 1999.

[12] N. Bogoliubov. Problems of a dynamical theory in statistical physics. In: J. de Boer, U. Uhlenbeck (eds.). *Studies in Statistical Mechanics, Vol. I.* North-Holland, Amsterdam (1962), 5.

[13] L. Boltzmann. Weitere Studien über das Wärmegleichgewicht unter Gasmoleküle. *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. Pergamon Press, Oxford (1966), 88-174.

[14] M. Born and H. Green. *A General Kinetic Theory of Fluids.* Cambridge University Press, Cambridge, 1949.

[15] F. Bufler and A. Schenk. *Halbleitertransporttheorie und Monte-Carlo-Bauelementsimulation.* Lecture Notes, ETH Zürich, Switzerland, 2002.

[16] A. Caldeira and A. Leggett. Path integral approach to quantum Brownian motion. Physica A **121A** (1983), 587-616.

[17] A. Caldeira, L. Erdös, F. Frommlet, and P. Markowich. Fokker-Planck equations as scaling limits of reversible quantum systems. J. Stat. Phys. **100** (2000), 543-601.

[18] P. Carruthers and F. Zachariasen. Quantum collision theory with phase-space distributions. Rev. Modern. Phys. **55** (1983), 245-285.

[19] C. Cercignani. *The Boltzmann Equation and Its Applications.* Springer, Berlin, 1988.

[20] S. Chapman. The kinetic theory of simple and composite gases: viscosity, thermal conduction and diffusion. Proc. Roy. Soc. London A **93** (1916/17), 1-20.

[21] D. Chen, E. Kan, U. Ravaioli, C. Shu, and R. Dutton. An improved energy transport model including nonparabolicity and non-Maxwellian distribution effects. *IEEE Electr. Dev. Lett.* **13** (1992), 26-28.

[22] R. Dautray and J.L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology.* Springer, Berlin, 1992.

[23] P. Degond. Mathematical modelling of microelectronics semiconductor devices. In: L. Hsiao (ed.). *Some Current Topics on Nonlinear Conservation Laws.* Amer. Math. Soc., Providence, USA (2000), 77-100.

[24] P. Degond, A. Jüngel, and P. Pietra. Numerical discretization of energy-transport models for semiconductors with nonparabolic band structure. *SIAM J. Sci. Comp.* **22** (2000), 986-1007.

[25] P. Degond, F. Méhats, and C. Ringhofer. Quantum energy-transport and drift-diffusion models. J. Stat. Phys. **118** (2005), 625-665.

[26] R. DiPerna, P.L. Lions. On the Cauchy problem for Boltzmann equations: global existence and weak stability. Ann. Math. **130** (1989), 321-366.

[27] R. DiPerna, P.L. Lions. Global solutions of Boltzmann's equation and the entropy inequality. Arch. Ration. Mech. Anal. **114** (1991), 47-55.

[28] P. Dmitruk, A. Saul, and L. Reyna. High electric-field approximation in semiconductor devices. *Appl. Math. Letters* **5** (1992), 99-102.

[29] D. Enskog. *Kinetische Theorie der Vorgänge in mäßig verdünnten Gasen. 1. Allgemeiner Teil.* Almquist & Wiksell, Uppsala, 1917.

[30] H. Federer. *Geometric Measure Theory.* Springer, New York, 1969.

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

[32] H. Föll. *Semiconductors.* Lecture Notes, Universität Kiel, Germany, 2004. (www.tf.uni-kiel.de/matwis/amat/semi_en/index.html).

[33] V. Gantmakher and Y. Levinson. *Carrier Scattering in Metals and Semiconductors.* North Holland, New York, 1987.

[34] A. Gnudi, D. Ventura, G. Baccarani, and F. Odeh. Two-dimensional MOSFET simulations by means of a multidimensional spherical harmonic expansion of the Boltzmann transport equation. *Solid State Electr.* **36** (1993), 575-581.

[35] N. Goldsman, L. Henrickson, and J. Frey. A physics based analytical numerical solution to the Boltzmann transport equation for use in device simulation. *Solid State Electr.* **34** (1991), 389-396.

[36] F. Golse. *The Boltzmann Equation and Its Hydrodynamic Limits.* Lecture Notes for the Ecole thématique "Les modèles cinétiques non linéaire et leurs limites fluides", Université de Franche-Comté, Besanǫn, 1997.

[37] H. Grahn. *Introduction to Semiconductor Physics.* World Scientific, Singpore, 1999.

[38] R. Hudson. When is the Wigner quasiprobability nonnegative?, Reports Math. Phys. **6** (1974), 249-252.

[39] J. Kirkwood. Statistical mechanical theory of transport processes I. General theory. *J. Chem. Phys.* **14** (1946), 180-201. Errata: **14**, 347.

[40] C. Kittel. *Introduction to Solid State Physcis.* 7th edition, John Wiley & Sons, New York, 1996.

[41] C. Lent and K. Kirkner. The quantum transmitting method. J. appl. Phys. **67** (1990), 6353-6359.

[42] C. Levermore. Moment closure hierarchies for kinetic theories. *J. Stat. Phys.* **83** (1996), 1021-1065.

[43] C. Levermore. Entropy-based moment closures for kinetic equations. *Transp. Theory Stat. Phys.* **26** (1997), 591-606.

[44] P. L. Lions. Global solutions of kinetic models and related problems. In: C. Cercignani and M. Pulvirenti (eds.). *Nonequilibrium Problems in Many-Particle Systems.* Lecture Notes in Math., Vol. 1551. Springer (1992), 58-86.

[45] M. Lundstrom. *Fundamentals of Carrier Transport.* 2nd edition, Cambridge University Press, Cambridge, 2000.

[46] J. Malý, D. Swanson, and W. Ziemer. The coarea formula for Sobolev mappings. *Trans. Amer. Math. Soc.* **355** (2003), 477-492.

[47] P. Markowich. On the equivalence of the Schrödinger and the quantum Liouville equations. Math. Meth. Appl. Sci. **11** (1989), 459-469.

[48] P. Markowich and C. Ringhofer. An analysis of the quantum Liouville equation. Z. Angew. Math. Mech. **69** (1989), 121-127.

[49] P. Markowich, C. Ringhofer, and C. Schmeiser. *Semiconductor Equations.* Springer, Vienna, 1990.

[50] D. Neamen. *Semiconductor Physics and Devices.* 3rd edition, McGraw Hill, 2002.

[51] F. Poupaud. *Etude mathématique et simulations numériques de quelques équations de Boltzmann.* PhD thesis, Université de Nice, France, 1988.

[52] F. Poupaud. On a system of nonlinear Boltzmann equations of semiconductor physics. *SIAM J. Appl. Math.* **50** (1990), 1593-1606.

[53] F. Poupaud. Diffusion approximation of the linear semiconductor Boltzmann equation: analysis of boundary layers. *Asympt. Anal.* **4** (1991), 293-317.

[54] F. Poupaud. Mathematical theory of kinetic equations for transport modelling in semiconductors. In: B. Perthame (ed.). *Advances in Kinetic Theory and Computing: Selected Papers.* World Scientific, Singapore (1994), 141-168.

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

[56] M. Reed and B. Simon. *Methods of Modern Mathematical Physics IV: Analysis of Operators.* Academic Press, New York, 1978.

[57] L. Reggiani (ed.). *Hot Electron Transport in Semiconductors.* Springer, Berlin, 1985.

[58] B. Ridley. *Quantum Processes in Semiconductors.* Clarendon Press, Oxford, 1982.

[59] C. Ringhofer. Computational methods for semiclassical and quantum transport in semiconductor devices. Acta Numerica (1997), 485-521.

[60] C. Schmeiser and A. Zwirchmayr. Elastic and drift-diffusion limits of electron-phonon interaction in semiconductor. *Math. Models Meth. Appl. Sci.* **8** (1998), 37-53.

[61] K. Seeger. *Semiconductor Physics. An Introduction.* 6th edition, Springer, Berlin, 1977.

[62] M. Taylor. *Pseudodifferential Operators.* Princeton University Press, Princeton, 1981.

[63] C. Villani. A review of mathematical topics in collisional kinetic theory. In: S. Friedlander and D. Serre. *Handbook of Mathematical Fluid Dynamics*, Vol. 1. Elsevier, Amsterdam (2002), 71-305.

[64] W. Van Roosbroeck. Theory of flow of electron and holes in germanium and other semiconductors. *Bell Syst. Techn. J.* **29** (1950), 560-607.

[65] W. Wenckebach. *Essentials of Semiconductor Physics.* John Wiley & Sons, Chichester, 1999.

[66] E. Wigner. On the quantum correction for the thermodynamic equilibrium. Phy. Rev. **40** (1932), 749-759.

[67] J. Yvon. *La théorie statistique des fluides.* Hermann, Paris, 1935.

[68] W. Zawadzki. Mechanics of electron scattering in semiconductors. in: T. Moss (ed.). *Handbook of Semiconductors.* Vol. 1. North-Holland, New York (1982), Chapter 12.