Mathematical Modeling
of Semiconductor Devices

Prof. Dr. Ansgar Jüngel
Fachbereich Mathematik und Statistik
Universität Konstanz

Preliminary version

Contents

1   Introduction

The modern computer and telecommunication industry relies heavily on the use and development of semiconductor devices. Since the first semiconductor device (a germanium transistor) has been built by Bardeen, Brattain and Shockley in 1947, a lot of different devices for special applications have been invented in the following decades. A very important fact of the success of semiconductor devices is that the device length is very small compared to 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. Thanks to the progressive miniaturization of semiconductor devices, the transistors in a modern Pentium IV processor have a characteristic length of 0.18 μm. The device length of tunneling diodes, produced in laboratories, is only of the order of 0.075 μm.

Nowadays, semiconductor materials are contained in almost all electronic devices. Some examples of semiconductor devices and their use are described in the following.

- *Photonic devices* capture light (photons) and convert it into an electronic signal. They are used in camcorders, solar cells, and light-wave communication systems as optical fibers.

- *Optoelectronic emitters* convert an electronic signal into light. Examples are light-emitting diodes (LED) used in displays and indication lambs and semiconductor lasers used in compact disk systems, laser printers, and eye surgery.

- *Flat-panel displays* create an image by controlling light that either passes through the device or is reflected off of it. They are made, for instance, of liquid crystals (liquid-crystal displays, LCD) or of thin semiconductor films (electroluminescent displays).

- In *field-effect devices* the conductivity is modulated by applying an electric field to a gate contact on the surface of the device. The most important field-effect device is the MOSFET (metal-oxide semiconductor field-effect transistor), used as a switch or an amplifier. Integrated circuits are mainly made of MOSFETs.

- *Quantum devices* are based on quantum mechanical phenomena, like tunneling of electrons through potential barriers which are impenetrable classically. Examples are resonant tunneling diodes, superlattices (multi-quantum-well structures), quantum wires in which the motion of carriers is restricted to one space dimension and confined quantum mechanically in the other two directions, and quantum dots.

Clearly, there are many other semiconductor devices which are not mentioned (for instance, bipolar transistors, Schottky barrier diodes, thyristors). Other new developments are, for instance, nanostructure devices (heterostructures) and solar cells made of amorphous silicon or organic semiconductor materials (see [14, 41]).

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 outside world by contacts at which a voltage (potential difference) is applied. We are mainly interested in devices which produce an electronic signal, for instance the macroscopically measurable electric current (electron flow), generated by the applied bias. In this situation, the input parameter is the applied voltage and the output parameter is the electric current through one contact. The relation between these two physical quantities is called *current-voltage characteristic*. It is a curve in the two-dimensional current-voltage space. The current-voltage characteristic does not need to be a monotone function and it does not need to be a function (but a relation).

The main objective of this book 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 of the electrons may be very different, for instance, due to drift, diffusion, convection, or quantum mechanical effects. For this reason, we have to devise *different* mathematical models which are able to describe the main physical phenomena for a particular situation or for a particular device.

This leads to a *hierarchy* of semiconductor models. Roughly speaking, we can divide semiconductor models in three classes: quantum models, kinetic models and fluiddynamical (macroscopic) models. In order to give some flavor of these models and the methods used to derive them, we explain these three view-points: quantum, kinetic and fluiddynamic in a simplified situation.

The quantum view. Consider a single electron of mass m and elementary charge q moving in a vacuum under the action of an electric field $E = E(x,t)$. The motion of the electron in space $x \in \mathbb{R}^d$ and time $t > 0$ is governed by the single-particle *Schrödinger equation*

$$i\hbar\partial_t\Psi = -\frac{\hbar^2}{2m}\Delta\Psi - qV(x,t)\Psi, \quad x \in \mathbb{R}^d, \ t > 0, \tag{1.1}$$

with some initial condition

$$\Psi(x,0) = \Psi_0(x), \quad x \in \mathbb{R}^d. \tag{1.2}$$

Here $\hbar = h/2\pi$ is the reduced Planck constant, $V(x,t)$ the real-valued electrostatic potential related to the electric field by $E = -\nabla V$, and $i^2 = -1$. Any solution Ψ of (1.1)-(1.2) is called *wave function*. Macroscopic variables are ob-

tained by the definitions:

electron density: $\quad\quad\quad n = |\Psi|^2,$

electron current density: $\quad J = -\dfrac{\hbar q}{m}\mathrm{Im}(\overline{\Psi}\nabla\Psi),$

where $\overline{\Psi}$ is the complex conjugate of Ψ. More precisely, $n(x,t)$ is the position probability density of the particle, i.e.

$$\int_A n(x,t)\, dx$$

is the probability to find the electron in the subset $A \subset \mathbb{R}^d$ at the time t.

Without too much exageration we can say that all other descriptions for the evolution of the electron, and in particular the fluiddynamical and kinetic models, can be derived from the Schrödinger equation (1.1).

The fluiddynamical view. In order to derive fluiddynamical models, for instance, for the evolution of the particle density n and the current density J, we assume that the wave function can be decomposed in its amplitude $\sqrt{n}(x,t) > 0$ and phase $S(x,t) \in \mathbb{R}$:

$$\Psi = \sqrt{n}e^{imS/\hbar}. \tag{1.3}$$

This expression makes sense as long as the electron density stays positive. We call this change of unknowns $\Psi \mapsto (n, S)$ *Madelung transform*. The current density now reads

$$J = -\frac{\hbar q}{m}\mathrm{Im}\left[\sqrt{n}\left(\frac{\nabla n}{2\sqrt{n}} + \frac{im}{\hbar}\sqrt{n}\nabla S\right)\right] = -qn\nabla S,$$

and we can interpret S also as a velocity potential. Setting (1.3) into the Schrödinger equation (1.1) and taking the imaginary and real part of the resulting equation, gives, after some computations (see Chapter 8):

$$\partial_t n - \frac{1}{q}\mathrm{div}\, J \;=\; 0, \tag{1.4}$$

$$\partial_t J - \frac{1}{q}\mathrm{div}\left(\frac{J \otimes J}{n}\right) + \frac{\hbar^2 q}{2m^2}n\nabla\left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) - \frac{q^2}{m}nE \;=\; 0. \tag{1.5}$$

Here, $J \otimes J$ is a tensor (or matrix) with components $J_j J_k$, where $j, k = 1, \ldots, d$. Equation (1.4) expresses the conversation of mass. Indeed,

$$\frac{d}{dt}\int_{\mathbb{R}^d} n\, dx = \frac{1}{q}\int_{\mathbb{R}^d}\mathrm{div}\, J\, dx = 0.$$

The equations (1.4)-(1.5) are referred to as the *quantum hydrodynamic equations*. They are equivalent to the Schrödinger equation (1.1) as long as $n > 0$. If they are studied in the whole space, the initial conditions

$$n(\cdot, 0) = |\Psi_0|^2, \quad J(\cdot, 0) = -\frac{\hbar q}{m}\mathrm{Im}(\overline{\Psi}_0\nabla\Psi_0) \quad \text{in } \mathbb{R}^d$$

have to be prescribed.

In the semi-classical limit "$\hbar \to 0$" the equations (1.4)-(1.5) formally reduce to the *hydrodynamic equations*

$$\partial_t n - \frac{1}{q} \operatorname{div} J = 0, \tag{1.6}$$

$$\partial_t J - \frac{1}{q} \operatorname{div} \left(\frac{J \otimes J}{n} \right) - \frac{q^2}{m} nE = 0, \tag{1.7}$$

which are the Euler equations for a gas of charged particles with zero temperature in a given electric field. Clearly, the limit "$\hbar \to 0$" does not make sense since $\hbar$ is a constant. After a scaling which makes the variables non-dimensional, a parameter (the so-called *scaled* Planck constant) appears in front of the scaled quantum term in (1.5). This parameter is small compared to one in certain physical ranges, for instance, when the characteristic length is "large" (compared to some reference length). Thus, the limit "$\hbar \to 0$" has always to be understood as the limit of vanishing *scaled* Planck constant.

The fluiddynamical formulation has some advantages:

- The equations are already formulated for macroscopic quantities.

- If the equations are considered in bounded domains (which is natural when dealing with semiconductor devices), boundary conditions can be more easily prescribed compared to the Schrödinger formulation.

- In two or three space dimensions, fluiddynamical models are usually numerically cheaper than the Schrödinger equation.

The kinetic view. We start again with the Schrödinger equation (1.1). First, introduce the so-called *density matrix*

$$\rho(r, s, t) = \overline{\Psi(r, t)} \Psi(s, t), \quad r, s \in \mathbb{R}^d, \ t > 0,$$

and make the change of coordinates

$$r = x + \frac{\hbar}{2m} \eta, \quad s = x - \frac{\hbar}{2m} \eta$$

in the density matrix defining

$$u(x, \eta, t) = \rho \left(x + \frac{\hbar}{2m} \eta, x - \frac{\hbar}{2m} \eta, t \right).$$

Since $(\hbar/2m)\eta$ has the dimension of length and $\hbar/2m$ has the dimension of m^2/s, η has the dimension of inverse velocity. We prefer to work with the variables space, velocity and time instead of space, inverse velocity and time and employ

therefore the Fourier transformation to u. We recall that the Fourier transform F is defined by

$$Fg(\eta) = \hat{g}(\eta) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} g(\eta)e^{-i\eta \cdot v} \, dv$$

for (sufficiently smooth) functions $g : \mathbb{R}^d \to \mathbb{C}$ with inverse

$$F^{-1}h(v) = \check{h}(v) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} h(v)e^{i\eta \cdot v} \, d\eta$$

for $h : \mathbb{R}^d \to \mathbb{C}$. The inverse Fourier transform of u is called *Wigner function w*:

$$w = (2\pi)^{-d/2}F^{-1}u = (2\pi)^{-d/2}\check{u}.$$

This function was introduced by Wigner in 1932 [43]. It can be shown (see Chapter 4) that the macroscopic electron density n and current density J defined above in terms of the wave function can now be written in terms of the Wigner function as

$$n(x,t) \;\; = \;\; \int_{\mathbb{R}^d} w(x,v,t) \, dv, \tag{1.8}$$

$$J(x,t) \;\; = \;\; -q \int_{\mathbb{R}^d} v \, w(x,v,t) \, dv. \tag{1.9}$$

The integrals are called the zeroth-order and first-order moments.

The transport equation for w is obtained by first deriving the evolution equation for u from the Schrödinger equation (1.1) and then by taking the inverse Fourier transform of the equation for u. A computation leads to the so-called *Wigner equation* (see Chapter 4 for details):

$$\partial_t w + v \cdot \nabla_x w + \frac{q}{m}\theta[V]w = 0, \quad x,v \in \mathbb{R}^d, \; t > 0, \tag{1.10}$$

where

$$(\theta[V]w)(x,v,t) \;\; = \;\; \frac{im}{\hbar(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left[V\left(x + \frac{\hbar}{2m}\eta, t\right) - V\left(x - \frac{\hbar}{2m}\eta, t\right) \right]$$
$$\times w(x,v',t)e^{i\eta \cdot (v-v')} \, dv' \, d\eta.$$

The initial condition for w reads

$$w(x,v,0) \;\; = \;\; (2\pi)^{-d/2}\check{u}(x,v,0)$$
$$= \;\; (2\pi)^{-d} \int_{\mathbb{R}^d} \overline{\Psi_0\left(x + \frac{\hbar}{2m}\eta\right)} \Psi_0\left(x - \frac{\hbar}{2m}\eta\right) e^{iv\cdot\eta} \, d\eta, \quad x,v \in \mathbb{R}^d.$$

An operator, whose Fourier transform acts as a multiplication operator on the Fourier transform of the function, is called a linear pseudo-differential operator [42]. Since

$$\widehat{\theta[V]w}(x, \eta, t) = \frac{im}{\hbar} \left[V\left(x + \frac{\hbar}{2m}\eta, t\right) - V\left(x - \frac{\hbar}{2m}\eta, t\right) \right] \hat{w}(x, \eta, t),$$

the operator $\theta[V]$ is a pseudo-differential operator and the Wigner equation (1.10) is a linear pseudo-differential equation.

In the formal limit "$\hbar \to 0$" it holds

$$\frac{m}{\hbar} \left[V\left(x + \frac{\hbar}{2m}\eta, t\right) - V\left(x - \frac{\hbar}{2m}\eta, t\right) \right] \to \nabla_x V(x, t) \cdot \eta,$$

and using the identity $i(\eta u)^\vee = \nabla_v \check{u}$, we obtain, as "$\hbar \to 0$",

$$\theta[v]w \to \frac{i}{(2\pi)^{d/2}} (\nabla_x V \cdot \eta u)^\vee = \frac{1}{(2\pi)^{d/2}} \nabla_x V \cdot \nabla_v \check{u} = \nabla_x V \cdot \nabla_v w.$$

Hence the Wigner equation becomes in the semi-classical limit

$$\partial_t w + v \cdot \nabla_x w + \frac{q}{m} \nabla_x V \cdot \nabla_v w = 0. \tag{1.11}$$

This equation is called *Liouville equation.* It has to be solved in the position-velocity space $(x, v) \in \mathbb{R}^d \times \mathbb{R}^d$ for times $t > 0$, supplemented with the initial condition

$$
\begin{aligned}
w(x, v, 0) &= (2\pi)^{-d/2} \lim_{\hbar \to 0} \check{u}(x, v, 0) \\
&= \frac{1}{(2\pi)^d} \lim_{\hbar \to 0} \int_{\mathbb{R}^d} \overline{\Psi_0\left(x + \frac{\hbar}{2m}\eta\right)} \Psi_0\left(x - \frac{\hbar}{2m}\eta\right) e^{iv \cdot \eta} \, d\eta, \quad x, v \in \mathbb{R}^d.
\end{aligned}
$$

Clearly, the above limits have to be understood in a formal way.

From the solution w of the kinetic Liouville equation, the macroscopic electron density n and current density J are computed as in the quantum case by the formulas (1.8)-(1.9).

The kinetic formulation of the motion of the electron seems much more complicate than the Schrödinger formulation since we pass from $d + 1$ variables (d for the space and one for the time) to $2d + 1$ variables (d for the space, d for the velocity and one for the time). However, this formulation has the following advantages:

- In the case of an electron ensemble with many particles, an approximate formulation in the position-velocity space is possible (leading to the Vlasov equation; see Chapter 3). This means that the kinetic equation has to be solved in $2d + 1$ variables instead of $Md + 1$ variables in the Schrödinger formulation, where $M \gg 1$ is the number of particles.

- Short-range interactions of the particles (collisions) can be easily included in the classical kinetic models.

We have described how fluiddynamical and kinetic models can be derived formally from the Schrödinger equation. Is there any relation between fluiddynamical and kinetic models? The answer is yes, and the method to see this is called *moment method.* In the following, we describe the idea of this method.

We start with the Liouville equation (1.11). Integrating (1.11) over the velocity space and using the definitions of the first two moments (1.8)-(1.9) and the divergence theorem gives for $E = -\nabla V$

$$
\begin{aligned}
0 &= \partial_t \int_{\mathbb{R}^d} w \, dv + \sum_{k=1}^d \int_{\mathbb{R}^d} v_k \frac{\partial w}{\partial x_k} \, dv - \frac{q}{m} \sum_{k=1}^d E_k \int_{\mathbb{R}^d} \frac{\partial w}{\partial v_k} \, dv \\
&= \partial_t n - \frac{1}{q} \text{div}_x J,
\end{aligned}
$$

which equals (1.6) expressing the conservation of mass. Now we multiply (1.11) with $-qv_j$ and integrate over the velocity space:

$$
\begin{aligned}
0 &= \partial_t J_j - q \sum_{k=1}^d \int_{\mathbb{R}^d} v_j v_k \frac{\partial w}{\partial x_k} \, dv + \frac{q^2}{m} \sum_{k=1}^d E_k \int_{\mathbb{R}^d} v_j \frac{\partial w}{\partial v_k} \, dv \\
&= \partial_t J_j - q \sum_{k=1}^d \frac{\partial}{\partial x_k} \int_{\mathbb{R}^d} v_j v_k w \, dv - \frac{q^2}{m} \sum_{k=1}^d E_k \int_{\mathbb{R}^d} \frac{\partial v_j}{\partial v_k} w \, dv.
\end{aligned}
$$

Since $\partial v_j / \partial v_k = \delta_{jk}$ with the Kronecker symbol δ_{jk}, this yields

$$
\partial_t J - q \text{div}_x \int_{\mathbb{R}^d} (v \otimes v) w \, dv - \frac{q^2}{m} E n = 0. \tag{1.12}
$$

The term

$$
\int_{\mathbb{R}^d} (v \otimes v) w \, dv
$$

is the matrix of all second-order moments and is called energy tensor. We make now the complete heuristic assumption that the velocity v can be approximated by an average velocity

$$
\bar{v} = \frac{\int vw \, dv}{\int w \, dv} = -\frac{J}{qn},
$$

that means,

$$
\int_{\mathbb{R}} (v \otimes v) w \, dv \approx (\bar{v} \otimes \bar{v}) \int_{\mathbb{R}^d} w \, dv = \frac{1}{q^2} \frac{J \otimes J}{n},
$$

9

and (1.12) becomes

$$\partial_t J - \frac{1}{q}\mathrm{div}_x\left(\frac{J \otimes J}{n}\right) - \frac{q^2}{m}nE = 0,$$

which equals (1.7). Under the above approximation, we have derived the hydrodynamic model (1.6)-(1.7). For a more general derivation (including thermal effects), see Chapter 5.

In a similar way, we can apply the moment method to the Wigner equation (1.10) in order to derive the quantum hydrodynamic model (1.4)-(1.5) (see Chapter 8).

All the models explained in this chapter and the relations between them are summarized in Figure 1.1.
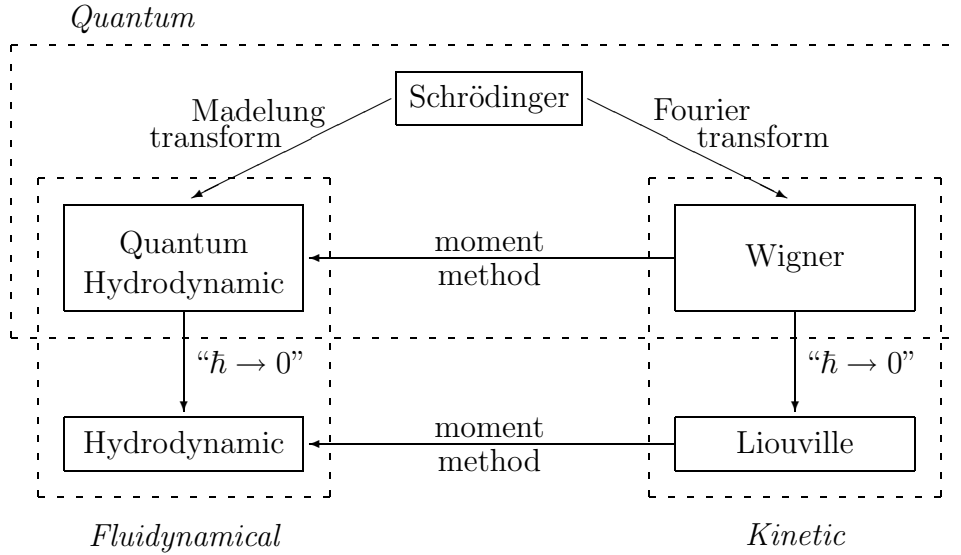
Figure 1.1: Models for the motion of an electron in a vaccum.

The above models are all derived in a very simplified situation. For the modeling of the electron flow in semiconductors we have to take into account:

- the description of the motion of many particles;

- the influence of the semiconductor crystal lattice;

- two-particle long-range interactions (Coulomb force between the electrons);

- two-particle short-range interactions (collisions of the particles with the crystal lattice or with other particles).

These features will be included in our models step by step in the following chapters.

2 Some Semiconductor Physics

In this chapter we present a short summary of the physics and main properties of semiconductors. Only these subjects relevant to the subsequent chapters are included here. We refer to [1, 8, 14, 28, 41] for a detailed description of solid-state and semiconductor physics.

2.1 Semiconductor crystals

What is a semiconductor? Historically, the term semiconductor has been used to denote solid materials with a much higher conductivity than insulators, but a much lower conductivity than metals measured at room temperature. A more precise definition is that a semiconductor is a solid with an *energy gap* larger than zero and smaller than about $4\,\text{eV}$. Metals have no energy gap, whereas this gap is usually larger than $4\,\text{eV}$ in insulators. In order to understand what an energy gap is, we have to explain the crystal structure of solids.

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

$$L = \{\ell \vec{a_1} + m \vec{a_2} + n \vec{a_3} : \; \ell, m, n \in \mathbb{Z}\} \subset \mathbb{R}^3,$$

where $\vec{a_1}, \vec{a_2}, \vec{a_3}$ are the basis vectors of L, usually called *primitive vectors* of the lattice. The lattice atoms (or ions) generate a periodic electrostatic potential V_L:

$$V_L(x + y) = V_L(x) \quad \forall x \in \mathbb{R}^3, \; y \in L.$$

The state of an electron moving in this periodic potential is described by an eigenfunction of the stationary Schrödinger equation:

$$-\frac{\hbar^2}{2m}\Delta\Psi - qV_L(x)\Psi = \varepsilon\Psi, \quad x \in \mathbb{R}^3, \tag{2.1}$$

where $\Psi : \mathbb{R}^3 \to \mathbb{C}$ is called *wave function.* In this equation, the physical parameters are the reduced Planck constant $\hbar = h/2\pi$, the electron mass (at rest) m, the elementary charge q, and the total energy ε. In fact, (2.1) is an eigenvalue problem, and we have to find eigenfunction-eigenvalue pairs (Ψ, ε). We consider two examples.

Example 2.1 (Free electron motion)
First, consider a free electron moving in a one-dimensional vacuum, i.e., $V_L(x) = 0$ for all $x \in \mathbb{R}$. Then the solutions of (2.1) are given by

$$\Psi_k(x) = Ae^{ikx} + Be^{-ikx}, \quad x \in \mathbb{R},$$

and

$$\varepsilon = \frac{\hbar^2 k^2}{2m},$$

for any $k \in \mathbb{R}$. We disregard purely imaginary values of $k = i\gamma$ ($\gamma \in \mathbb{R}$) which also yield real values of the energy, since this leads to unbounded solutions of the type $\exp(\pm ikx) = \exp(\mp\gamma x)$ for $x \in \mathbb{R}$. (Notice that the integral of $|\Psi|^2$ over $\mathbb{R}$ can be interpreted as the particle mass which should be finite.) Thus the eigenvalue problem (2.1) has infinitely many solutions corresponding to different energies. The functions $\exp(\pm ikx)$ are called *plane waves.*

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

$$\Psi(0) = \Psi(L) = 0$$

and potential $V_L(x) = 0$ for $x \in (0, L)$ (see Figure 2.1). The eigenfunctions of (2.1) are given by

$$\Psi_k(x) = A_k \sin(kx), \quad x \in (0, L),$$

with discrete eigenvalues $k = \pi n/L$, $n \in \mathbb{N}$, and energies

$$\varepsilon(k) = \frac{\hbar^2 k^2}{2m}, \quad k = \frac{\pi n}{L}.$$

The system only allows *discrete* energy states.
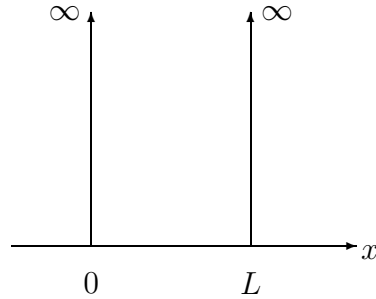
Figure 2.1: Infinite square-well potential.

Although the lattice potential V_L is periodic, the wave function is defined on $\mathbb{R}^3$ in order to account for electron motion from one lattice cell to the next. In fact, the so-called Bloch theorem states that the Schrödinger equation (2.1) can be reduced to a Schrödinger equation on a so-called primitive cell of the lattice. Before we state the result, we need some definitions.

Definition 2.3 *(1) The* reciprocal lattice *L^* corresponding to L is defined by*

$$L^* = \{\ell\vec{a_1^*} + m\vec{a_2^*} + n\vec{a_3^*} : \ell, m, n \in \mathbb{Z}\},$$

where the basis vectors (or primitive vectors) $\vec{a_1^*}, \vec{a_2^*}, \vec{a_3^*} \in \mathbb{R}^3$ satisfy the relation

$$\vec{a_m} \cdot \vec{a_n^*} = 2\pi \delta_{mn}.$$

(2) *The connected set $D \subset \mathbb{R}^3$ is called* primitive cell *of L (or of L^*) if and only if*

- *the volume of D equals the volume of the parallelepiped spanned by the basis vectors of L (or of L^*);*
- *the whole space $\mathbb{R}^3$ is covered by the union of translates of D by the primitive vectors.*

(3) *The (first)* Brillouin zone *$B \subset \mathbb{R}^3$ is that primitive cell of the reciprocal lattice L^* which consists of those points, which are closer to the origin than to any other point of L^* (see Figure 2.2).*
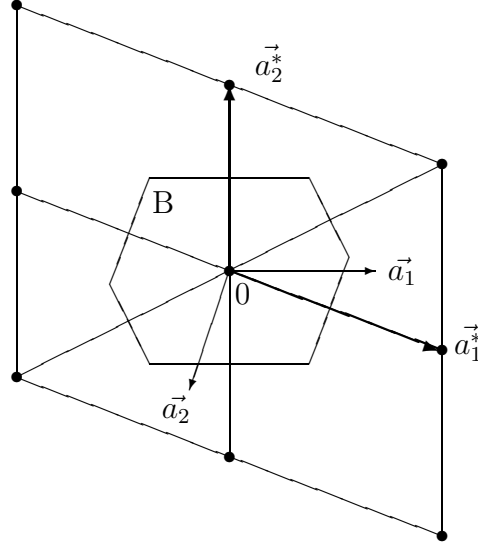
Figure 2.2: The primitive vectors of a two-dimensional lattice L and its reciprocal lattice L^* and the Brillouin zone B.

We give some explainations of the above definition. As the wave planes $\exp(\pm ik \cdot x)$ suggest, k can be considered as a reciprocal variable which leads to the definitions of the reciprocal lattice and the Brillouin zone. In particular, for any $x \in L$ and $k \in L^*$ with

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

we obtain

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

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 *pseudo-wave vector*; see below.)

The Brillouin zone can be constructed as follows. Draw arrows from a lattice point to its nearest neighbors and cut them in half. The planes through these intermediate points perpendicular to the arrows form the surface of the (bounded) Brillouin zone. In two space dimensions, the Brillouin zone is always a hexagon (or a square; see Figure 2.2).

Now we are able to apply the *Bloch theorem*. It says that the Schrödinger equation (2.1) is equivalent to the system of Schrödinger equations

$$-\frac{\hbar^2}{2m}\Delta\Psi_k - qV_L\Psi_k = \varepsilon(k)\Psi_k, \quad x \in D, \tag{2.2}$$

indexed by $k \in B$, with pseudo-periodic boundary conditions

$$\Psi_k(x+y) = \Psi_k(x)e^{ik\cdot y} \quad \text{for } x, x+y \in \partial D, \tag{2.3}$$

where D is a primitive cell of L. (This follows from the fact that the Hamilton operator $H = -(\hbar^2/2m)\Delta + qV_L(x)$ commutes with all the translation operators T_ℓ associated with the lattice, where $(T_\ell\Psi)(x) = \Psi(x+\ell)$ for $\ell \in L$, $x \in \mathbb{R}^3$.) By functional analysis, the eigenvalue problem (2.2)-(2.3), for any $k \in B$, possesses a sequence of eigenfunctions Ψ_k^n with associated eigenvalues $\varepsilon_n(k)$, $n \in \mathbb{N}_0$. The relation between Ψ_k and Ψ is given by

$$\Psi_k(x) = \sum_{\ell \in L} e^{-ik\cdot\ell}\Psi(x+\ell).$$

We can also write Ψ_k^n as distorted waves

$$\Psi_k^n(x) = u_k^n(x)e^{ik\cdot x}, \tag{2.4}$$

where

$$u_k^n(x) = \sum_{\ell \in L} e^{-ik\cdot(x+\ell)}\Psi(x+\ell)$$

is clearly periodic in L. In some sense, Ψ_k^n are plane waves which are modulated by a periodic function u_k^n taking into account the influence of the lattice of atoms. The functions (2.4) are also called *Bloch functions*. Since the vector k appears in distorted plane waves, it is not termed wave vector but pseudo-wave vector.

The function $k \mapsto \varepsilon_n(k)$ is called the *disperson relation* or the n-th *energy band*. It shows how the energy of the n-th band depends on the wave vector

k. The union of the ranges of ε_n over $n \in \mathbb{N}$ is not necessarily the whole set $\mathbb{R}$, i.e., there may exist energies ε^* for which there is no $n \in \mathbb{N}$ and no $k \in B$ such that $\varepsilon_n(k) = \varepsilon^*$. The connected components of the set of energies with this non-existence property are called *energy gaps*. We illustrate this property by an example.

Example 2.4 (Kronig-Penney model)
We study a simple one-dimensional model for the periodic potential generated by the lattice atoms. We use the square-well potential (see Figure 2.3)

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

and V_L is extended to $\mathbb{R}$ with period $a + b$, i.e.

$$V_L(x) = V_L(x + a + b) \quad \forall x \in \mathbb{R},$$

where $a, b > 0$ and $V_0 \in \mathbb{R}$.
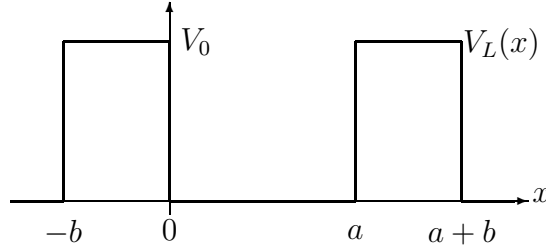
Figure 2.3: Periodic square-well potential V_L.

In order to solve the Schrödinger equation (2.1), we make the Bloch ansatz

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

with an $(a + b)$-periodic function u. Set

$$\alpha = \sqrt{\frac{2m\varepsilon}{\hbar^2}}, \quad \beta = \sqrt{\frac{2m(\varepsilon - qV_0)}{\hbar^2}}.$$

If $\varepsilon < qV_0$ then β can be written as $\beta = i\gamma$ with $\gamma > 0$. A computation shows that the general solution of (2.1) in the interval $(-b, a)$ is given by

$$u(x) = \begin{cases} Ae^{i(\alpha-k)x} + Be^{-i(\alpha+k)x} & : & 0 < x \leq a \\ Ce^{i(\beta-k)x} + De^{-i(\beta+k)x} & : & -b < x \leq 0. \end{cases}$$

15

The constants A, B, C and D can be determined from the boundary conditions. More precisely, we assume that u is continuously differentiable and periodic in $\mathbb{R}$:

$$u(a + 0) = u(a - 0), \quad u(a + 0) = u(a - 0),$$
$$u(-b) = u(a), \qquad u'(-b) = u'(a),$$

where $u(a + 0) = \lim_{x \searrow a} u(x)$ etc. This gives four linear equations for the four unknowns A, B, C and D. They can be casted into a linear system

$$M(A, B, C, D)^\top = 0$$

with matrix $M \in \mathbb{C}^{4 \times 4}$. (The sign "$\top$" means the transposed of a vector.) The necessary conditions for getting non-trivial solutions is that $\det M = 0$. An elementary, but lenghty calculation shows that this conditon is equivalent to

$$\cos(k(a + b)) = \sqrt{1 + \left(\frac{\alpha^2 - \beta^2}{2\alpha\beta} \right)^2 \sin^2(\beta b)} \cdot \cos(\alpha a - \delta), \qquad (2.5)$$

where

$$\tan \delta = -\frac{\alpha^2 + \beta^2}{2\alpha\beta} \tan(\beta b).$$

The equation (2.5) is implicit in k and real-valued for all values of the energy. Indeed, if $\varepsilon > qV_0$ we have $\beta = i\gamma$ and $\sin(\beta b)/\beta = \sinh(\gamma b)/\gamma$ since $\sin(ix) = i \sinh(x)$.

The right-hand side of (2.5) may be larger than $+1$ or smaller than -1 for certain values of ε. Then (2.5) is solvable only for imaginary wave vectors $k = i\kappa$, $\kappa \in \mathbb{R}$. But then $\Psi(x) = u(x) \exp(ikx) = u(x) \exp(-\kappa x)$ and this gives non-integrable solutions in $\mathbb{R}$. We conclude that there exist values of ε for which (2.5) has *no* real solution k. Every connected subset of $[0, \infty) \backslash R(\varepsilon)$, where $R(\varepsilon) = \{\varepsilon_0 \in [0, \infty) : \exists k \in \mathbb{R} : \varepsilon(k) = \varepsilon_0\}$, is an energy gap.

In Table 2.1 some values of energy gaps for some semiconductor materials are collected. The energy gap separates two energy bands. The nearest energy band

Material	Energy gap ε_g (eV)
Silicon	1.12
Germanium	0.664
Gallium-Arsenide	1.424

Table 2.1: Energy gaps of selected semiconductors.

below the energy gap is called *valence band*; the nearest energy band above the energy gap is termed *conductor band* (see Figure 2.4).

Now we are able to state the definition of a *semiconductor:* It is a solid with an energy gap whose value is positive and smaller than about $4\,\mathrm{eV}$.
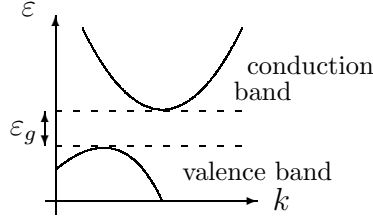
Figure 2.4: Schematic band structure with energy gap ε_g.

2.2 Mean electron velocity and effective mass approximation

In this section we will motivate the following two expressions:

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

$$v_n(k) = \frac{1}{\hbar}\nabla_k \varepsilon_n. \tag{2.6}$$

- The *effective mass* tensor m^* is defined by

$$(m^*)^{-1} = \frac{1}{\hbar^2}\frac{d^2\varepsilon_n}{dk^2}. \tag{2.7}$$

Motivation of (2.6). The result (2.6) can be obtained as follows. We omit in the following the index n and define the group velocity by

$$v(k) = \left(\int_D |\Psi|^2 \, dx\right)^{-1} \int_D \hat{v}_k |\Psi_k|^2 \, dx,$$

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

$$\hat{v}_k = -\frac{J_k}{qn_k} = \frac{\hbar}{m}\frac{\mathrm{Im}(\overline{\Psi}_k\nabla_x\Psi_k)}{|\Psi_k|^2}. \tag{2.8}$$

In order to show (2.6) we have to compute the integral

$$\int_D \mathrm{Im}(\overline{\Psi}_k\nabla_x\Psi_k) \, dx.$$

We take the derivative of (2.2) with respect to k, use the distorted wave plane

$$\Psi_k(x) = u_k(x)e^{ik\cdot x}$$

and

$$
\begin{aligned}
\Delta_x(\nabla_k\Psi_k) &= \Delta_x\left(e^{ik\cdot x}\nabla_k u_k + ix\Psi_k\right) \\
&= \Delta_x\left(e^{ik\cdot x}\nabla_k u_k\right) + 2i\nabla_x\Psi_k + ix\Delta_x\Psi_k
\end{aligned}
$$

17

to obtain

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

Multiplying this equation by $\overline{\Psi}_k$, integrating over D and integrating by parts yields

$$
\begin{aligned}
&\nabla_k\varepsilon\int_D |\Psi_k|^2\, dx + i\frac{\hbar^2}{m}\int_D(\nabla_x\Psi_k)\overline{\Psi}_k\, dx \\
&= \int_D e^{ikx}\nabla_k u_k\left(-\frac{\hbar^2}{2m}\Delta_x - qV_L - \varepsilon\right)\overline{\Psi}_k\, dx = 0,
\end{aligned}
$$

using (2.2) again. Thus

$$
\nabla_k\varepsilon = -\frac{i\hbar^2\int_D(\nabla_x\Psi_k)\overline{\Psi}_k\, dx}{m\int_D |\Psi_k|^2\, dx}.
$$

Taking the real part of both sides of this equation gives

$$
\nabla_k\varepsilon = \frac{\hbar^2}{m}\frac{\mathrm{Im}\int_D(\nabla_x\Psi_k)\overline{\Psi}_k\, dx}{\int_D |\Psi_k|^2\, dx} = \hbar v(k),
$$

which shows (2.6).

The expression (2.6) has some consequences. The change of energy with respect to time equals the product of force F and velocity v_n:

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

By the chain rule,

$$
\partial_t\varepsilon_n(k) = \nabla_k\varepsilon_n(k)\partial_t k
$$

and hence

$$
F = \partial_t(\hbar k). \tag{2.9}
$$

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

$$
p = \hbar k.
$$

Motivation of (2.7). Another consequence follows from (2.6) and (2.9): Differentiating (2.6) leads to

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

The momentum equals $m^* v_n$, where m^* is the (effective) mass. Using again Newton's law $F = \partial_t p = m^*\partial_t v_n$ we obtain

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

We consider this equation as a definition of the *effective mass* m^*. The second derivative of ε_n with respect to k is a 3×3 matrix, so the *symbol* $(m^*)^{-1}$ is also a matrix. If we evaluate the Hessian of ε_n near a local minimum, i.e. $\nabla_k\varepsilon_n(k_0) = 0$, then $d^2\varepsilon_n(k_0)/dk^2$ is a symmetric positive matrix which can be diagonalized and the diagonal elements are positive:

$$\begin{pmatrix} 1/m_x^* & 0 & 0 \\ 0 & 1/m_y^* & 0 \\ 0 & 0 & 1/m_z^* \end{pmatrix} = \frac{1}{\hbar^2}\frac{d^2\varepsilon_n}{dk^2}(k_0). \tag{2.10}$$

Assume that the energy values are shifted in such a way that the energy vanishes at the local minimum k_0. For wave vectors k "close" to k_0, we have from Taylor's formula and (2.10)

$$\begin{aligned} \varepsilon_n(k) &= \varepsilon_n(k_0) + \nabla_k\varepsilon_n(k_0)\cdot k + \frac{1}{2}k^\top\frac{d^2\varepsilon_n}{dk^2}(k_0)k + O(|k-k_0|^3) \\ &= \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_x^*} + \frac{k_y^2}{m_y^*} + \frac{k_z^2}{m_z^*}\right) + O(|k-k_0|^3), \end{aligned}$$

where $k = (k_x, k_y, k_z)^\top$. If the effective masses are equal in all directions, i.e. $m^* = m_x^* = m_y^* = m_z^*$, we can write, neglecting higher-order terms,

$$\varepsilon_n(k) = \frac{\hbar^2}{2m^*}|k|^2 \tag{2.11}$$

for wave vectors k "close" to a local band minimum. In this situation, m^* is called the *isotropic effective mass.* We conclude that the energy of an electron near a band minimum equals the energy of a free electron in vacuum (see Example 2.1) where the (rest) mass of the electron is replaced by the effective mass. The effects of the crystal potential are represented by the effective mass.

The expression (2.11) is referred to as the *parabolic band approximation* and usually, the range of wave vectors k is extended to the whole space, i.e., $k = \mathbb{R}^3$ is assumed. In order to account for non-parabolic effects, the following *non-parabolic band approximation* in the sense of Kane is used [41, Ch. 2.1]:

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

where $\alpha > 0$ is a non-parabolicity parameter. In silicon, for instance, $\alpha = 0.5\,(\mathrm{eV})^{-1}$.

Definition of holes. When we consider the effective mass near a band maximum, we find that the Hessian of ε_n is negative definite which would lead to a negative effective mass. However, in the derivation of the mean velocity and consequently, of the effective mass definition, we have used in (2.8) that the charge of the electrons is negative. Using a positive charge leads to a positive effective mass. The corresponding particles are calles *holes.* 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 rather the motion of the valence band vacancies that the motion of the electrons moving from one vacant orbital to the next (Figure 2.5).
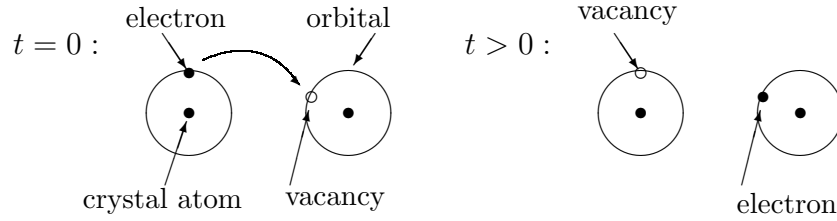
Figure 2.5: Motion of a valence band electron to a neighboring vacant orbital or, equivalently, motion of a hole in the inverse direction.

Semi-classical picture. In the so-called *semi-classical picture,* the motion of an electron in the n-th band can be approximately described by a point particle moving with the velocity $v_n(k)$. Denoting by $(x(t), v_n(k, t))$ the trajectory of an electron in the position-velocity phase space, we obtain from Newton's law (see (2.9))

$$\partial_t x = v_n(k) = \frac{1}{\hbar}\nabla_k \varepsilon_n, \qquad \hbar\partial_t k = F, \tag{2.12}$$

where F represents a driving force, for instance, $F = -q\nabla_x V_L(x, t)$. Notice that band transitions are excluded since the band index n is fixed in the equations.

Non-periodic potentials. If a non-periodic potential (or driving force) is superimposed to V_L, the Schrödinger equation (2.1) cannot be decomposed into the decoupled Schrödinger equations (2.2), and the above computations are no longer valid. In fact, 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.

2.3 Semiconductor statistics

Number densities. We want to determine the number of electrons in the conduction band and the number of holes in the valence band per unit volume. In view of the very high number of particles (typically $> 10^9\,\mathrm{cm}^{-3}$) it seems appropriate to use statistical methods. We make the following assumptions:

- Electrons cannot be distinguished from one another.

- The Pauli exclusion principle holds, i.e., each level of a band can be occupied by not more than two electrons with opposite spin.

First we compute the number of possible states within all energy bands per unit volume $(2\pi)^3$. In the continuum limit, this number equals:

$$g(\varepsilon) = \frac{2}{(2\pi)^3} \sum_{\nu} \int_B \delta(\varepsilon - \varepsilon_\nu(k))\, dk. \qquad (2.13)$$

The function $g(\varepsilon)$ is called *density of states*. The factor 2 comes from the two possible states of the spin of an electron. The set B is the Brillouin zone (see Section 2.1), and the function δ is the delta distribution defined by

$$\int_{-\infty}^{\infty} \delta(\varepsilon_0 - \varepsilon)\phi(\varepsilon)\, d\varepsilon = \phi(\varepsilon_0) \qquad (2.14)$$

for all appropriate functions ϕ. Mathematically, δ is a functional operating in some function space and the above integral has to be interpreted as the value of $\delta(\varepsilon_0 - \varepsilon)$ acting on $\phi(\varepsilon)$. Formally, it holds

$$\int_{-\infty}^{\infty} \delta(\varepsilon_0 - \varepsilon)\phi(\varepsilon)\, d\varepsilon = \int_{-\infty}^{\infty} \delta(\varepsilon - \varepsilon_0)\phi(\varepsilon)\, d\varepsilon = \int_{\varepsilon_1}^{\infty} \delta(\varepsilon - \varepsilon_0)\phi(\varepsilon)\, d\varepsilon, \qquad (2.15)$$

for any $\varepsilon_1 < \varepsilon_0$.

Example 2.5 (Density of states in the parabolic band approximation)
In the case of the parabolic band approximation we can compute the integral (2.13). Consider a single band near the bottom of the conduction band with isotropic effective mass. The energy is given by

$$\varepsilon(k) = \varepsilon_c + \frac{\hbar^2}{2m_e^*}|k|^2, \quad k \in \mathbb{R}^3.$$

Transforming k to the spherical coordinates (ρ, θ, ϕ) and then substituting $z = \hbar^2\rho^2/2m_e^*$, we obtain

$$
\begin{aligned}
g_c(\varepsilon) &= \frac{1}{4\pi^3} \int_{\mathbb{R}^3} \delta(\varepsilon - \varepsilon_c - \hbar^2|k|^2/2m_e^*)\, dk \\
&= \frac{1}{4\pi^3} \int_0^{2\pi} \int_0^{\pi} \int_0^{\infty} \delta(\varepsilon - \varepsilon_c - \hbar^2\rho^2/2m_e^*)\, \rho^2 \sin\theta\, d\rho\, d\theta\, d\phi \\
&= \frac{4\pi}{4\pi^3} \frac{1}{2} \left(\frac{2m_e^*}{\hbar^2}\right)^{3/2} \int_0^{\infty} \delta(\varepsilon - \varepsilon_c - z)\, \sqrt{z}\, dz.
\end{aligned}
$$

Define the Heaviside function H by

$$H(x) = \begin{cases} 0 & : & x < 0 \\ 1 & : & x > 0. \end{cases}$$

Then, by (2.14),

$$\begin{aligned} g_c(\varepsilon) &= \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2} \right)^{3/2} \int_{-\infty}^{\infty} \delta(\varepsilon - \varepsilon_c - z) \sqrt{z} H(z) \, dz \\ &= \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2} \right)^{3/2} \sqrt{\varepsilon - \varepsilon_c} \, H(\varepsilon - \varepsilon_c). \end{aligned}$$

The electron density n is the integral of the probability density of occupancy of an energy state per unit volume:

$$n = \frac{2}{(2\pi)^3} \sum_\nu \int_B f(\varepsilon_\nu(k)) \, dk,$$

where $f(\varepsilon)$ is a distribution function. Using (2.14), (2.15) and the definition (2.13) of g_c, we have

$$\begin{aligned} n &= \frac{1}{4\pi^3} \sum_\nu \int_B \int_{-\infty}^{\infty} \delta(\varepsilon - \varepsilon_\nu(k)) f(\varepsilon) \, d\varepsilon \, dk \\ &= \int_{-\infty}^{\infty} g_c(\varepsilon) f(\varepsilon) \, d\varepsilon. \end{aligned} \tag{2.16}$$

Fermi-Dirac distribution. We will motivate that the electron distribution follows the *Fermi-Dirac distribution*

$$f(\varepsilon) = \frac{1}{1 + e^{(\varepsilon - \varepsilon_F)/k_B T}}, \tag{2.17}$$

where the number ε_F which is called *Fermi energy* or Fermi level will be interpreted later. Consider a system of non-interacting particles with discrete quantum states. The system is put into contact with a reservoir at temperature T. Let ε_i and ε_j be the total energies of a particle in the state i and j, respectively. Let W_{ij} be the probability that an electron of state i will go to state j in a unit time. The average number of particles that make a transition from state i to state j is

$$A(T) e^{-\varepsilon_i/k_B T} W_{ij},$$

where $A(T)$ is some function. The temperature T enters because the transition will absorb energy from, or omit energy to, the reservoir. Conversely, the average number of particles that go from state j to state i is

$$A(T) e^{-\varepsilon_j/k_B T} W_{ji}.$$

22

We use the *principle of detailed balance.* It says that in thermal equilibrium, the two number must coincide. Therefore

$$\frac{W_{ji}}{W_{ij}} = e^{(\varepsilon_j - \varepsilon_i)/k_B T}. \tag{2.18}$$

Now let f_i be the average number of electrons in the state i. By Pauli's exclusion principle, each state can only have 0 or 1 electron in it at any time. Hence, $0 \leq f_i \leq 1$, and f_i can be seen as the fraction of configurations in which the state i is occupied. Similarly, $1 - f_i$ is the fraction of configurations in which the state i is empty. The number of transitions from state i to state j of a large ensemble M of configurations is

$$M f_i (1 - j_j) W_{ij},$$

since the transition only takes place if the state i is occupied and the state j is empty. By the principle of detailed balance,

$$M f_i (1 - f_j) W_{ij} = M f_j (1 - f_i) W_{ji}.$$

This equation is equivalent to

$$\frac{f_i}{1 - f_i} = \frac{f_j}{1 - f_j} \frac{W_{ji}}{W_{ij}}$$

or, by (2.18),

$$\frac{f_i}{1 - f_i} e^{\varepsilon_i / k_B T} = \frac{f_j}{1 - f_j} e^{\varepsilon_j / k_B T}.$$

This relation holds for any state i and any state j. Therefore, both sides are independent of i and j :

$$\frac{f_i}{1 - f_i} e^{\varepsilon_i / k_B T} = \text{const.} = e^{\varepsilon_F / k_B T}.$$

This defines the number ε_F. Solving the above equation for f_i yields

$$f_i = \frac{1}{1 + e^{(\varepsilon_i - \varepsilon_F)/k_B T}}.$$

Hence, (2.17) is motivated.

At zero temperature, we expect that all states below a certain energy are occupied and all states above that energy are empty. Indeed, as $T \to 0$, we obtain from (2.17)

$$f(\varepsilon) \to \begin{cases} 1 & : \quad \varepsilon < \varepsilon_F \\ 0 & : \quad \varepsilon > \varepsilon_F \end{cases}$$

and $f(\varepsilon_F) = 1/2$. This provides a physical interpretation of the Fermi energy ε_F for any temperature: For energies below the Fermi level, the states are rather

occupied ($f(\varepsilon) > 1/2$ for $\varepsilon < \varepsilon_F$), and for energies above the Fermi level, the states are rather empty ($f(\varepsilon) < 1/2$ for $\varepsilon > \varepsilon_F$).

For energies which are much larger than the Fermi energy in the scale of $k_B T$, i.e. $\varepsilon - \varepsilon_F \gg k_B T$, we can approximate the Fermi-Dirac distribution by the *Maxwell-Boltzmann distribution*

$$f(\varepsilon) = e^{-(\varepsilon - \varepsilon_F)/k_B T}$$

since $1 + \exp((\varepsilon - \varepsilon_F)/k_B T) \sim \exp((\varepsilon - \varepsilon_F)/k_B T)$ for $\varepsilon - \varepsilon_F \gg k_B T$. We call this the *non-degenerate case.* A semiconductor in which Fermi-Dirac statistics have to be used is called *degenerate.*

Formulas for the number densities. We use the Fermi-Dirac distribution (2.17) in the formula (2.16) for the electron density

$$n = \int_{-\infty}^{\infty} \frac{g_c(\varepsilon)\, d\varepsilon}{1 + e^{(\varepsilon - \varepsilon_F)/k_B T}}.$$

Similar expressions hold for the hole density p for energies below the valence band minimum ε_v:

$$
\begin{aligned}
p &= \int_{-\infty}^{\infty} g_v(\varepsilon) \left(1 - \frac{1}{1 + e^{(\varepsilon - \varepsilon_F)/k_B T}} \right) d\varepsilon \\
&= \int_{-\infty}^{\infty} \frac{g_v(\varepsilon)\, d\varepsilon}{1 + e^{(\varepsilon_F - \varepsilon)/k_B T}}.
\end{aligned}
$$

The electron and hole densities can be computed more explicitly in the parabolic band approximation with isotropic effective mass:

$$
\begin{aligned}
n &= \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2} \right)^{3/2} \int_{\varepsilon_c}^{\infty} \frac{\sqrt{\varepsilon - \varepsilon_c}\, d\varepsilon}{1 + e^{(\varepsilon - \varepsilon_F)/k_B T}} = N_c F_{1/2} \left(\frac{\varepsilon_F - \varepsilon_c}{k_B T} \right), \\
p &= \frac{1}{2\pi^2} \left(\frac{2m_h^*}{\hbar^2} \right)^{3/2} \int_{-\infty}^{\varepsilon_v} \frac{\sqrt{\varepsilon_v - \varepsilon}\, d\varepsilon}{1 + e^{(\varepsilon_F - \varepsilon)/k_B T}} = N_v F_{1/2} \left(\frac{\varepsilon_v - \varepsilon_F}{k_B T} \right),
\end{aligned}
$$

where we have introduced the *effective densities of states*

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

and the Fermi integral

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

In the non-degenerate case $\varepsilon_c - \varepsilon_F \gg k_B T$ and $\varepsilon_F - \varepsilon_v \gg k_B T$ we can replace the Fermi integral by a simpler function:

$$F_{1/2}(y) \sim \exp(y) \quad \text{as } y \to -\infty.$$

Then the particle densities become

$$n = N_c \exp\left(\frac{\varepsilon_F - \varepsilon_c}{k_B T}\right), \quad p = N_v \exp\left(\frac{\varepsilon_v - \varepsilon_F}{k_B T}\right). \tag{2.19}$$

Intrinsic semiconductors. A pure semiconductor crystal with no impurities is called an *intrinsic semiconductor.* In this case, the conduction band electrons can only have come from formerly occupied valence band levels leaving holes behind them. The number of conduction band electrons is therefore equal to the number of valence band holes:

$$n = p =: n_i.$$

The *intrinsic density* n_i can be computed in the non-degenerate case from (2.19):

$$n_i = \sqrt{np} = \sqrt{N_c N_v} \exp\left(\frac{\varepsilon_v - \varepsilon_c}{k_B T}\right) = \sqrt{N_c N_v} \exp\left(-\frac{\varepsilon_g}{2 k_B T}\right) \tag{2.20}$$

since the energy gap is $\varepsilon_g = \varepsilon_c - \varepsilon_v$. The Fermi energy of an intrinsic semiconductor can be computed using (2.19) and (2.20):

$$\begin{aligned}
\varepsilon_F &= \varepsilon_c + k_B T \ln(n/N_c) \\
&= \varepsilon_c + k_B T \ln(n_i/N_c) \\
&= \varepsilon_c - \frac{1}{2}\varepsilon_g + \frac{1}{2}k_B T \ln\frac{N_v}{N_c} \\
&= \frac{1}{2}(\varepsilon_c + \varepsilon_v) + \frac{3}{4}k_B T \ln\frac{m_h^*}{m_e^*}.
\end{aligned}$$

This asserts that as $T \to 0$, 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 $k_B T$. In most semiconductors and at room temperature, the energy gap ε_g is much larger than $k_B T$ (silicon: $\varepsilon_g = 1.12\,\text{eV}$, $k_B T = 0.0259\,\text{eV}$ at room temperature). This shows that the non-degeneracy assumptions

$$\varepsilon - \varepsilon_F \geq \varepsilon_c - \varepsilon_F = \frac{1}{2}\varepsilon_g + \frac{3}{4}k_B T \ln\frac{m_h^*}{m_e^*} \gg k_B T,$$

$$\varepsilon_F - \varepsilon \geq \varepsilon_F - \varepsilon_v = \frac{1}{2}\varepsilon_g + \frac{3}{4}k_B T \ln\frac{m_h^*}{m_e^*} \gg k_B T$$

are satisfied and that the result is consistent with the assumptions (see Figure 2.6).

Doping. The intrinsic density is too small to result in a significant conductivity (for instance, silicon: $n_i = 6.93 \cdot 10^9\,\text{cm}^{-3}$). However, it is possible to replace some atoms in a semiconductor crystal by atoms which provide free electrons in the conduction band or free holes in the valence band. This process is called *doping.* Impurities that contribute to the carrier density are called
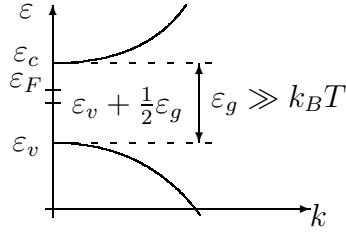
Figure 2.6: Illustration of the energy gap ε_g in relation to ε_c, ε_v and ε_F.

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 germanium crystal, whose atoms have each 4 valence electrons, with arsenic, whose atoms have each 5 valence electrons, each arsenic impurity provides one additional electron. These additional electrons are only weakly binded to the arsenic atom. Indeed, the binding energy is about $0.013\,\text{eV}$ and thermal excitation provides enough energy to excite the additional electrons to the conduction band. Generally speaking, let ε_d and ε_a be the energy level of a donor electron and an acceptor hole, respectively. Then $\varepsilon_c - \varepsilon_d$ and $\varepsilon_a - \varepsilon_v$ are small compared to $k_B T$ (Figure 2.7). This means that the additional particles contribute at room temperature to the electron and hole density.

Figure 2.7: Illustration of the donor and acceptor energy levels ε_d and ε_a.

3 Classical Kinetic Transport Models
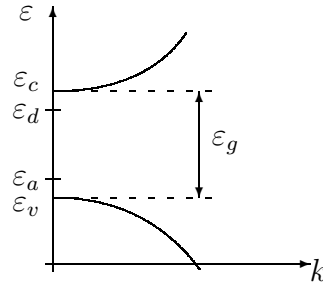
3.1 The Liouville equation

We first analyze the motion of an ensemble of M electrons in a vacuum under the action of an electric field E. The electrons will be discribed 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 space dimension is usually three, but also one- or two-dimensional models can be considered. Since the electrons have equal mass m, the trajectories $(x_i(t), v_i(t))$ of the ensemble satisfy the system of ordinary differential equations in the $(2d \cdot M)$-dimensional ensemble position-velocity space

$$\partial_t x_i = v_i, \quad \partial_t v_i = \frac{F_i}{m}, \qquad t > 0, \ i = 1, \dots, M, \tag{3.1}$$

where $F_i = F_i(x, v, t)$ are forces and $x = (x_1, \dots, x_M)$, $v = (v_1, \dots, v_M)$. We set $F = (F_1, \dots, F_M)$. The initial conditions are given by

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

The system (3.1)-(3.2) constitutes an initial-value problem for the trajectory $w(t; x_0, v_0) = (x(t), v(t))$. For instance, the force can be given by an electric field acting on the electron ensemble:

$$F_i = -qE(x, t).$$

In this case the forces are independent of the velocity.

In semiconductor applications, M is a very large number (typically, $M \sim 10^5$) and therefore, the (numerical) solution of (3.1) is very expensive or even not feasible. It seems reasonable to use a statistical description. We assume that instead of the precise initial position x_0 and initial velocity v_0 we are given the joint probability density $f_I(x, v)$ of the initial position and velocity of the electrons. This density has the properties

$$f_I(x, v) \geq 0 \quad \text{for } x, v \in \mathbb{R}^{dM}, \quad \int_{\mathbb{R}^{dM}} \int_{\mathbb{R}^{dM}} f_I(x, v) \, dx \, dv = 1. \tag{3.3}$$

Then

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

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

Let $f(x, v, t)$ be the probability density of the electron ensemble at time t. We wish to derive a differential equation for f. For this we use the *Lionville*

theorem [16]. It states that the function f does not change along the trajectories $w(t; x_0, v_0) = (x(t), v(t))$, i.e.

$$f(x(t), v(t), t) = f(w(t; x_0, v_0), t) = f_I(x_0, v_0) \tag{3.4}$$

for all $x_0, v_0 \in \mathbb{R}^{dM}$ and $t \geq 0$. Differentiating this equation with respect to time gives

$$
\begin{aligned}
0 &= \frac{d}{dt} f(x(t), v(t), t) \\
&= \partial_t f + \partial_t x \cdot \nabla_x f + \partial_t v \cdot \nabla_v f \\
&= \partial_t f + v \cdot \nabla_x f + \frac{1}{m} F \cdot \nabla_v f.
\end{aligned}
\tag{3.5}
$$

This equation is referred to as the *classical Liouville equation* for an electron ensemble. When the force is given by the electric field,

$$F = -qE = q\nabla_x V(x, t),$$

where V is the electrostatic potential, the Liouville equation reads

$$\partial_t f + v \cdot \nabla_x f + \frac{q}{m} \nabla_x V \cdot \nabla_v f = 0, \quad x, v \in \mathbb{R}^{dM}, \ t > 0, \tag{3.6}$$

with initial condition

$$f(x, v, 0) = f_I(x, v), \quad x, v \in \mathbb{R}^{dM}. \tag{3.7}$$

This equation governs the evolution of the position-velocity probability density $f(x, v, t)$ of an electron ensemble in the electric field $-q\nabla_x V$ under the assumptions that

- the electrons move according to the laws of classical mechanics and

- the electrons move in a vacuum.

In the following we describe some properities of the solution of (3.6)-(3.7). We assume that f_I satisfies (3.3).

- *Non-negativity:* From (3.4) we conclude that

$$f(x, v, t) \geq 0 \quad \forall x, v \in \mathbb{R}^{dM}$$

  and for all $t \geq 0$ for which a solution exists.

- *Conservation property:* We integrate (3.6) over $\mathbb{R}^{dM} \times \mathbb{R}^{dM}$ and use the divergence theorem:

$$\partial_t \int_{\mathbb{R}^{dM}} \int_{\mathbb{R}^{dM}} f(x, v, t) \, dx \, dv$$

$$= -\int_{\mathbb{R}^{dM}} \int_{\mathbb{R}^{dM}} \left(\text{div}_x(vf) + \frac{q}{m} \text{div}_v(f \nabla_x V) \right) dx \, dv$$

$$= 0.$$

By (3.3), we conclude for all $t \geq 0$

$$\int_{\mathbb{R}^{dM}} \int_{\mathbb{R}^{dM}} f(x, v, t) \, dx \, dv = \int_{\mathbb{R}^{dM}} \int_{\mathbb{R}^{dM}} f_I(x, v) \, dx \, dv = 1. \qquad (3.8)$$

- *Moments:* We define the zeroth- and first-order moments

$$n(x, t) = \int_{\mathbb{R}^{dM}} f(x, v, t) \, dv,$$

$$J(x, t) = -q \int_{\mathbb{R}^{dM}} v f(x, v, t) \, dv,$$

where q is the elementary charge. The functions n and J are interpreted as the position probability density (or particle density) and as the particle current density, respectively. The conservation property (3.8) can be restated as

$$\int_{\mathbb{R}^{dM}} n(x, t) \, dx = \int_{\mathbb{R}^{dM}} n_I(x) \, dx, \quad t \geq 0, \qquad (3.9)$$

with the initial electron density

$$n_I(x) = \int_{\mathbb{R}^{dM}} f_I(x, v) \, dv.$$

This means that the total number of electrons is conserved in time. By formally integrating (3.6) over $v \in \mathbb{R}^{dM}$ we obtain the so-called particle continuity equation

$$\partial_t n - \frac{1}{q} \text{div}_x J = 0, \quad x \in \mathbb{R}^{dM}, \ t > 0. \qquad (3.10)$$

We consider an electron ensemble moving in a semiconductor crystal. As explained in Chapter 2, the ions in the crystal lattice induce a lattice-periodic potential which influences the motion of the charged particles. Let $k \mapsto \varepsilon_n(k)$ be the n-th energy band of the crystal for wave vectors k of the Brillouin zone B. The corresponding mean velocity is given by (see Section 2.2)

$$v_n(k) = \frac{1}{\hbar} \nabla_k \varepsilon_n(k).$$

Let $k_i \in \mathbb{R}^d$ be the wave vector and $x_i \in \mathbb{R}^d$ the position vector of the i-th electron and set $k = (k_1, \ldots, k_M) \in B^M$, $x = (x_1, \ldots, x_M) \in \mathbb{R}^{dM}$. We fix the energy band and omit therefore the index n. The motion of the electrons in the (x, k)-space can be described semi-classically by the equations

$$\partial_t x = v(k) = \frac{1}{\hbar}\nabla_k \varepsilon(k), \quad \hbar \partial_t k = F,$$

where $v(k) = (v(k_1), \ldots, v(k_M))$, and $F = (F_1, \ldots, F_M)$ are some driving forces. Notice that band transitions are excluded in this formulation since the band index is fixed. Notice also that the equation $\hbar \partial_t k = F$ is an expression of Newton's law since $p = \hbar k$ is the crystal momentum vector of the electrons. An analogous computation as in (3.5) yields the *semi-classical Liouville equation* for the distribution function $f(x, k, t)$ of an electron ensemble:

$$\partial_t f + v(k) \cdot \nabla_x f + \frac{1}{\hbar} F \cdot \nabla_k f = 0, \quad x \in \mathbb{R}^{dM}, \; k \in B^M, \; t > 0, \qquad (3.11)$$

where

$$v(k) \cdot \nabla_x f = \sum_{j=1}^{M} v(k_j) \cdot \nabla_{x_j} f.$$

As the Brillouin zone is a bounded subset, we have to impose boundary conditions for k. We choose the periodic boundary conditions

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

for all $j = 1, \ldots, M$. This formulation makes sense since B is point symmetric to the origin, i.e. $k \in B$ if and only if $-k \in B$.

The macroscopic particle and current densities are now defined by

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

The periodicity of f in k_i and the point-symmetry of B imply the conservation property (3.9) and the conservation law (3.10). Here we also need the assumption $\mathrm{div}\,_k F = 0$ such that terms arising in the integration by parts cancel. This assumption is satisfied if F is given by an electric field

$$F = -qE(x, t)$$

or by an electro-magnetic field

$$F = -q(E(x, t) + v \times B_{\mathrm{ind}}(x, t)),$$

where B_{ind} is the induction vector.

When the parabolic band approximation (see Section 2.2)

$$\varepsilon(k) = \frac{\hbar^2}{2m^*}|k|^2, \quad k \in \mathbb{R}^{dM},$$

for electrons is used, then $v(k) = \hbar k/m$ and $\nabla_k f = (\hbar/m^*)\nabla_v f$, and the semi-classical Liouville (3.11) reduces to its classical counterpart (3.6).

3.2 The Vlasov equation

The main disadvantage of the (semi-)classical Liouville equation is that it has to be solved in a very high-dimensional phase space: typically, $M \sim 10^5$, $d = 3$, and the dimension of the (x, k)-space is $3 \cdot 10^5$! In the following we will formally derive a low-dimensional equation, the so-called Vlasov equation. The idea of the derivation is first to assume a certain structure of the interaction (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 with equal mass and denote by $x = (x_1, \ldots, x_M) \in \mathbb{R}^{dM}$, $v = (v_1, \ldots, v_M) \in \mathbb{R}^{dM}$ the position and velocity coordinates of the electrons, respectively. We impose the following assumptions:

- **Assumption 1:** The electrons move in a vacuum.

- **Assumption 2:** The force field F only depends on position and time.

- **Assumption 3:** The motion is governed by an external electric field and by two-particle interaction forces.

The first assumption will be discarded later on. The second assumption means that magnetic fields are ignored (see [34, Ch. 1.3] for an inclusion of magnetic effects). The last assumption is crucial for the derivation of the Vlasov equation. It means that the force field F_i exerted on the i-th electron is given by the sum of an electric field acting on the i-th electron and of the sum of $M - 1$ two-particle interaction forces exerted on the i-th electron by the other electrons:

$$F_i(x, t) = -qE_{\text{ext}}(x_i, t) - q\sum_{j=1,\ j\neq i}^{M} E_{\text{int}}(x_i, x_j), \quad i = 1, \ldots, M. \qquad (3.12)$$

The interaction force E_{int} is independent of the electron indices, which interprets the fact that the electrons are indistinguishable. We assume that the force exerted by the i-th electron on the j-th electron is equal to the negative force exerted by the j-th electron on the i-th electron:

$$E_{\text{int}}(x_i, x_j) = -E_{\text{int}}(x_j, x_i) \quad \forall x_i, x_j \in \mathbb{R}^d, \qquad (3.13)$$

which implies $E_{\text{int}}(x, x) = 0$.

The classical Liouville equation for the density $f(x, v, t)$ of the ensemble reads:

$$\partial_t f + \sum_{i=1}^{M} v_i \cdot \nabla_{x_i} f - \frac{q}{m} \sum_{i=1}^{M} E_{\text{ext}}(x_i, t) \cdot \nabla_{v_i} f$$

$$- \frac{q}{m} \sum_{i,j=1}^{M} E_{\text{int}}(x_i, x_j) \cdot \nabla_{v_i} f = 0. \qquad (3.14)$$

We assume that the initial density is independent of the numbering of the particles (**Assumption 4**):

$$f_I(x_1, \ldots, x_M, v_1, \ldots, v_M) = f_I(x_{\pi(1)}, \ldots, x_{\pi(M)}, v_{\pi(1)}, \ldots, v_{\pi(M)}) \qquad (3.15)$$

for all $x_i, v_i \in \mathbb{R}^d$, $i = 1, \ldots, M$, and for all permutations π of $\{1, \ldots, M\}$. Then the property (3.13) implies that also $f(x, v, t)$ is independent of the numbering of the particles for all $t > 0$. We introduce the density $f^{(a)}$ of a subensemble consisting of $a < M$ electrons:

$$f^{(a)}(x_1, \ldots, x_a, v_1, \ldots, v_a, t) = \int_{\mathbb{R}^{2d(M-a)}} f(x, v, t) \, dx_{a+1} \cdots dx_M \, dv_{a+1} \cdots dv_M.$$

We integrate the above Liouville equation with respect to $x_{a+1}, \ldots, x_M, v_{a+1}, \ldots, v_M$ in order to obtain an equation for $f^{(a)}$. By the divergence theorem,

$$\sum_{i=1}^{M} \int_{\mathbb{R}^{2d(M-a)}} v_i \cdot \nabla_{x_i} f \, dx_{a+1} \cdots dx_M \, dv_{a+1} \cdots dv_M$$

$$= \sum_{i=1}^{a} v_i \cdot \nabla_{x_i} \int_{\mathbb{R}^{2d(M-a)}} f \, dx_{a+1} \cdots dx_M \, dv_{a+1} \cdots dv_M$$

$$+ \sum_{i=a+1}^{M} \int_{\mathbb{R}^{2d(M-a)}} \text{div}_{x_i}(v_i f) \, dx_{a+1} \cdots dx_M \, dv_{a+1} \cdots dv_M$$

$$= \sum_{i=1}^{a} v_i \cdot \nabla_{x_i} f^{(a)}.$$

Similarly,

$$- \frac{q}{m} \sum_{i=1}^{M} \int_{\mathbb{R}^{2d(M-a)}} \text{div}_{v_i}(E_{\text{ext}}(x_i, t) f) \, dx_{a+1} \cdots dx_M \, dv_{a+1} \cdots dv_M$$

$$= - \frac{q}{m} \sum_{i=1}^{a} \text{div}_{v_i}(E_{\text{ext}}(x_i, t) f^{(a)}).$$

The last integral becomes

$$-\frac{q}{m}\sum_{i=1}^{M}\int_{\mathbb{R}^{2d(M-a)}} \operatorname{div}_{v_i}(E_{\text{int}}(x_i,x_j)f)\,dx_{a+1}\cdots dx_M\,dv_{a+1}\cdots dv_M$$

$$= \quad -\frac{q}{m}\sum_{i,j=1}^{a} E_{\text{int}}(x_i,x_j)\cdot\nabla_{v_i}f^{(a)}$$

$$-\frac{q}{m}\sum_{i=a+1}^{M}\sum_{j=1}^{M}\int_{\mathbb{R}^{2d(M-a)}} \operatorname{div}_{v_i}(E_{\text{int}}(x_i,x_j)f)\,dx_{a+1}\cdots dx_M\,dv_{a+1}\cdots dv_M$$

$$-\frac{q}{m}\sum_{i=1}^{a}\sum_{j=a+1}^{M}\int_{\mathbb{R}^{2d(M-a)}} \operatorname{div}_{v_i}(E_{\text{int}}(x_i,x_j)f)\,dx_{a+1}\cdots dx_M\,dv_{a+1}\cdots dv_M.$$

The second integral vanishes by the divergence theorem. The last integral is equal to

$$-\frac{q}{m}\sum_{i=1}^{a}(M-a)\operatorname{div}_{v_i}\int_{\mathbb{R}^{2d}} E_{\text{int}}(x_i,x_{a+1})$$

$$\left(\int_{\mathbb{R}^{2d(M-a)}} f\,dx_{a+2}\cdots dx_M\,dv_{a+2}\cdots dv_M\right) dx_{a+1}\,dv_{a+1}$$

$$= \quad -\frac{q}{m}\sum_{i=1}^{a}(M-a)\operatorname{div}_{v_i}\int_{\mathbb{R}^{2d}} E_{\text{int}}(x_i,x_*)$$

$$\times f^{(a+1)}(x_1,\ldots,x_a,x_*,v_1,\ldots,v_d,v_*,t)\,dx_*\,dv_*,$$

using the indepency of f on the numbering of the particles.

We illustrate the last argument by the example $M=3$, $a=1$:

$$\sum_{j=2}^{3}\int_{\mathbb{R}^{4d}} \operatorname{div}_{v_1}(E_{\text{int}}(x_1,x_j)f(x,v,t))\,dx_2\,dx_3\,dv_2\,dv_3$$

$$= \quad \int_{\mathbb{R}^{4d}} \operatorname{div}_{v_1}(E_{\text{int}}(x_1,x_2)f(x_1,x_2,x_3,v_1,v_2,v_3,t))\,dx_2\,dx_3\,dv_2\,dv_3$$

$$+\int_{\mathbb{R}^{4d}} \operatorname{div}_{v_1}(E_{\text{int}}(x_1,x_3)f(x_1,x_2,x_3,v_1,v_3,v_2,t))\,dx_2\,dx_3\,dv_2\,dv_3$$

$$= \quad 2\operatorname{div}_{v_1}\int_{\mathbb{R}^{2d}} E_{\text{int}}(x_1,x_2)$$

$$\times\left(\int_{\mathbb{R}^{2}} f(x_1,x_2,x_3,v_1,v_2,v_3,t)\,dx_3\,dv_3\right) dx_2\,dv_2$$

$$= \quad 2\operatorname{div}_{v_1}\int_{\mathbb{R}^{2}} E_{\text{int}}(x_1,x_*)f^{(2)}(x_1,x_*,v_1,v_*,t)\,dx_*dv_*.$$

We summarize the above computations and obtain from (3.14)

$$\partial_t f^{(a)} + \sum_{i=1}^{a} v_i \cdot \nabla_{x_i} f^{(a)} - \frac{q}{m} \sum_{i=1}^{a} E_{\text{ext}}(x_i, t) \cdot \nabla_{v_i} f^{(a)}$$

$$- \frac{q}{m} \sum_{i,j=1}^{a} E_{\text{int}}(x_i, x_j) \cdot \nabla_{v_i} f^{(a)} \tag{3.16}$$

$$- \frac{q}{m}(M-a) \sum_{i=1}^{a} \operatorname{div}_{v_i} \int_{\mathbb{R}^{2d}} E_{\text{int}}(x_i, x_*) f_*^{(a+1)} \, dx_* \, dv_* = 0,$$

where

$$f_*^{(a+1)} = f^{(a+1)}(x_1, \ldots, x_a, x_*, v_1, \ldots, v_a, v_*, t).$$

These equations for $1 \leq a \leq M-1$ are called the Bogoliubov-Born-Green-Kirkwood-Yvon (BBGKY) hierarchy. We wish to perform the formal limit "$M \to \infty$". For this, we assume that $|E_{\text{int}}|$ is of order $1/M$ (**Assumption 5**) such that $|F_i|$ is of order one as $M \to \infty$. Then the third term in (3.16) vanishes in the formal limit, whereas the expression $(M-a)E_{\text{int}}$ in the last term on the left-hand side of (3.16) remains finite. For $M \gg 1$, we can substitute this expression by ME_{int}. Thus, for $M \gg 1$, we obtain from (3.16), neglecting terms of order $1/M$,

$$\partial_t f^{(a)} + \sum_{i=1}^{a} v_i \cdot \nabla_{x_i} f^{(a)} - \frac{q}{m} \sum_{i=1}^{a} E_{\text{ext}}(x_i, t) \cdot \nabla_{v_i} f^{(a)}$$

$$- \frac{q}{m} \sum_{i=1}^{a} \operatorname{div}_{v_i} \int_{\mathbb{R}^{2d}} M f_*^{(a+1)} E_{\text{int}}(x_i, x_*) \, dx_* \, dv_* = 0. \tag{3.17}$$

In order to solve this equation we make the ansatz

$$f^{(a)}(x_1, \ldots, x_a, v_1, \ldots, v_a, t) = \prod_{i=1}^{a} P(x_i, v_i, t). \tag{3.18}$$

The interpretation of this factorization is that we assume that the electrons of the subensemble move independently of each other. Using this ansatz in (3.17) for $a = 1$ gives

$$\partial_t P + v_1 \cdot \nabla_{x_1} P - \frac{q}{m} E_{\text{eff}}(x_1, t) \cdot \nabla_{v_1} P = 0,$$

where

$$E_{\text{eff}}(x, t) = E_{ext}(x, t) + \int_{\mathbb{R}^{2d}} M P(x_2, v_2, t) E_{\text{int}}(x, x_2) \, dx_2 \, dv_2.$$

It can be shown that (3.18) is indeed a particular solution of (3.16) if P satisfies the above equation. Clearly, we need to assume that the initial density $f^{(a)}(\cdot, \cdot, 0)$

admits such a factorization for all $a \in \mathbb{N}$ (**Assumption 6**). The whole information of the evolution of the electron subensemble is thus contained in the function P. We define

$$\begin{aligned}
F(x, v, t) &= MP(x, v, t), \\
n(x, t) &= \int_{\mathbb{R}^d} F(x, v, t) \, dv, \quad x, v \in \mathbb{R}^d, \ t \geq 0.
\end{aligned}$$

The quantity $n(x, t)$ represents the density of the electrons per unit volume. The number density F satisfies the so-called *classical Vlasov equation*

$$\partial_t F + v \cdot \nabla_x F - \frac{q}{m} E_{\text{eff}} \cdot \nabla_v F = 0, \quad x, v \in \mathbb{R}^d, \ t > 0, \tag{3.19}$$

$$E_{\text{eff}}(x, t) = E_{\text{ext}}(x, t) + \int_{\mathbb{R}^d} n(x_*, t) \, E_{\text{int}}(x, x_*) \, dx_*. \tag{3.20}$$

This equation has the form of a single-particle Liouville equation with the force E_{eff}. Many-particle physics enters only through the effective field E_{eff}, which in turn depends on the density n and hence on F. This means that (3.19) is a nonlinear equation with a nonlocal nonlinearity of quadratic type. It provides a macroscopic description of the motion of many-particle systems with weak long-range forces. However, it does not account for strong short-range forces such as scattering of particles (see Section 3.3).

The Vlasov equation (3.19)-(3.20) is supplemented by the initial condition

$$F(x, v, 0) = F_I(x, v), \quad x, v \in \mathbb{R}^d.$$

Since the solution F of (3.19)-(3.20) can be interpreted as the probability of a particle to be in the state (x, v) at time t, we expect that

$$0 \leq F(x, k, t) \leq 1, \quad x, v \in \mathbb{R}^d, \ t > 0.$$

Indeed, assuming that $0 \leq F_I(x, k) \leq 1$, we obtain from the trajectory equations

$$\partial_t x = v, \quad \partial_t v = -\frac{q}{m} E_{\text{eff}}, \quad x(0) = x_0, \quad v(0) = v_0,$$

the expression

$$\begin{aligned}
0 &= \partial_t F + v \cdot \nabla_x F - \frac{q}{m} E_{\text{eff}} \cdot \nabla_v F \\
&= \frac{d}{dt} F(x(t), v(t), t)
\end{aligned}$$

and therefore

$$F(x(t), v(t), t) = F_I(x_0, v_0) \in [0, 1].$$

Example 3.1 (Coulomb force in $\mathbb{R}^3$)

The most important long-range force acting between two electrons is the Coulomb force:

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

where the *permittivity* ε_s is a material constant. (The following considerations are also valid in non-vacuum.) Then the effective field reads (see (3.20))

$$E_{\text{eff}}(x, t) = E_{\text{ext}}(x, t) - \frac{q}{4\pi\varepsilon_s} \int_{\mathbb{R}^3} n(z, t) \frac{x - z}{|x - z|^3} \, dz.$$

It is well known that the function

$$\phi(x) = \frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{f(z)}{|x - z|} \, dz, \quad x \in \mathbb{R}^3,$$

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

$$f(x) = \frac{1}{4\pi} \int_{\mathbb{R}^3} f(z) \Delta_x \frac{1}{|x - z|} \, dz = \frac{1}{4\pi} \int_{\mathbb{R}^3} f(z) \operatorname{div}_x \frac{x - z}{|x - z|^3} \, dz$$

and

$$0 = \operatorname{curl}_x \nabla_x \phi(x) = \frac{1}{4\pi} \operatorname{curl}_x \int_{\mathbb{R}^3} f(z) \frac{x - z}{|x - z|^3} \, dz.$$

We conclude that

$$\begin{aligned} \operatorname{div} E_{\text{eff}}(x, t) &= \operatorname{div} E_{\text{ext}}(x, t) - \frac{q}{\varepsilon_s} n(x, t), \\ \operatorname{curl} E_{\text{eff}}(x, t) &= \operatorname{curl} E_{\text{ext}}(x, t), \quad x \in \mathbb{R}^3, \; t > 0. \end{aligned}$$

We assume that the external field is generated by ions of charge $+q$ which are present in the material (for instance, doping atoms in the semiconductor crystal). Then, by Coulomb's law

$$E_{ext}(x, t) = \frac{+q}{4\pi\varepsilon_s} \int_{\mathbb{R}^3} C(z, t) \frac{x - z}{|x - z|^3} \, dz,$$

where C is the number density of the background ions. Since

$$\operatorname{div} E_{\text{ext}} = \frac{q}{\varepsilon_s} C, \quad \operatorname{curl} E_{\text{ext}} = 0,$$

we obtain

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

Since E_{eff} is vortex-free, there is a potential V_{eff} such that $E_{\text{eff}} = -\nabla V_{\text{eff}}$. Then we can rewrite the above equation as

$$\varepsilon_s \Delta V_{\text{eff}} = q(n - C), \quad x \in \mathbb{R}^3. \tag{3.21}$$

This equation which also holds in $\mathbb{R}^d$, $d \geq 1$, is called *Poisson equation*. The Vlasov equation (3.19) with the Coulomb interaction field, i.e. with $E_{\text{eff}} = -\nabla V_{\text{eff}}$ and (3.21), is referred to as *Vlasov-Poisson system*.

Instead of taking the classical Liouville equation (3.6) as basis for the derivation of the Vlasov equation, we can also start from the semi-classical formulation (3.11). Under the assumptions (3.12) and (3.13) on the force fields, we obtain by proceeding as above the *semi-classical Vlasov equation*

$$\partial_t F + v(k) \cdot \nabla_x F - \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k F = 0, \quad x \in \mathbb{R}^d, \ k \in B, \ t > 0, \quad (3.22)$$

$$E_{\text{eff}}(x,t) = E_{\text{ext}}(x,t) + \int_{\mathbb{R}^d} n(x_*, t) \, E_{\text{int}}(x, x_*) \, dx_* \quad (3.23)$$

with the particle position density

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

and the initial condition

$$F(x,k,0) = F_I(x,k), \quad x \in \mathbb{R}^d, \ k \in B.$$

The number density is assumed to satisfy periodic boundary equations in k :

$$F(x,k,t) = F(x,-k,t), \quad x \in \mathbb{R}^d, \ k \in \partial B, \ t > 0. \quad (3.24)$$

We notice that the precise meaning of the number density F is the following: $F(x,k,t)$ is the ratio of the number of occupied states in the "volume" $dx \, dk$ in the conduction band and the total number of quantum states in this volume in the conduction band. Then (see also Section 2.3)

$$0 \leq F(x,k,t) \leq 1, \quad x \in \mathbb{R}^d, \ k \in B, \ t > 0,$$

if $0 \leq F_I(x,k) \leq 1$, and $F(x,k,t)$ is also called the *occupation number* of state k in x at time t.

3.3 The Boltzmann equation

The Vlasov equation accounts for long-range particle interactions, like the Coulomb force (see Example 3.1). Short-range interactions, like collisions of the particles with other particles or with the crystal lattice, are not included. We wish to extend the Vlasov equation to include collision mechanisms, which will lead to the Boltzmann equation. We present only a phenomenological derivation, formulated first by Boltzmann in 1872 for the description of non-equilibrium

phenomena in dilute gases. For details on a more rigorous derivation, we refer to the literature in [9, Sec. 1.5.3] and [16].

The starting point of the derivation is to postulate that the rate of change of the number density $F(x, v, t)$ of the particle ensemble due to convection and the effective field, dF/dt, and the rate of change of F due to collisions, $Q(F)$, balance:

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

This equation has to be understood along the particle trajectories. Explicitely, it reads (in the classical case, see (3.19))

$$\partial_t F + v \cdot \nabla_x F - \frac{q}{m} E_{\text{eff}} \cdot \nabla_v F = Q(F), \quad x, v \in \mathbb{R}^d, \ t > 0, \tag{3.25}$$

where the effective field E_{eff} is given by (3.20). In order to derive an expression for $Q(F)$ we assume that the rate $P(x, v' \to v, t)$ of a particle at (x, t) to change its velocity v' into v due to a scattering event is proportional to

- the occupation probability $F(x, v', t)$ and

- the probability $1 - F(x, v, t)$ that the state (x, v) is not occupied at time t.

Thus

$$P(x, v' \to v, t) = s(x, v', v)F(x, v', t)(1 - F(x, v, t)),$$

where $s(x, v', v)$ is the so-called *scattering rate*. (We made a similiar consideration in Section 2.3.) Then the rate of change of F due to collisions is the sum of

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

for all velocities v' in the volume element dv'. In the continuum limit, the sum is in fact an integral and we get

$$
\begin{aligned}
(Q(F))(x, v, t) &= \int_{\mathbb{R}^d} [P(x, v' \to v, t) - P(x, v \to v', t)] \, dv' \tag{3.26} \\
&= \int_{\mathbb{R}^d} [s(x, v', v)F'(1 - F) - s(x, v, v')F(1 - F')] \, dv',
\end{aligned}
$$

where $F = F(x, v, t)$, $F' = F(x, v', t)$. The equation (3.25), together with the effective field equation (3.20) and the definition of the collision operator (3.26), is called the *classical Boltzmann equation*. When the Coulomb force is used to model the long-range interactions, then (3.25), (3.20), (3.26), and (3.21) are termed *Boltzmann-Poisson system*.

The Boltzmann equation has two nonlinearities: a quadratic nonlocal nonlinearity caused by the self-consistent field E_{eff} and another quadratic nonlocal nonlinearity caused by the collision integral (3.26). A rigorous mathematical

analysis of the initial-value problem (3.25)-(3.26) (existence and uniqueness of solutions) is very difficult. We refer to the research papers [23, 36] and to the review paper [4] for details on the mathematical analysis of this problem.

Starting from the semi-classical Vlasov equation (3.22) we can include the collision effects similiarly as above and obtain the *semi-classical Boltzmann equation*

$$\partial_t F + v(k) \cdot \nabla_x F - \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k F = Q(F), \quad x \in \mathbb{R}^d, \ k \in B, \ t > 0, \qquad (3.27)$$

where the effective field E_{eff} is defined in (3.23). This equation models the semi-classical motion of electrons incorporating the quantum effects of the semiconductor crystal lattice as explained in Section 2.2. The Boltzmann equation (3.27) is supplemented by the initial condition

$$F(x, k, 0) = F_I(x, k), \quad x \in \mathbb{R}^d, \ k \in B, \qquad (3.28)$$

and by the periodic boundary conditions (3.24).

In semiconductor crystals, there are three main classes of scattering events:

- electron-phonon scattering,

- ionized impurity scattering, and

- electron-electron scattering.

We explain these collision mechanisms now in detail.

At finite temperature, the atoms in a crystal lattice undergo variations about their fixed equilibrium positions. These lattice vibrations are quantized and the quantum of lattice vibrations are called *phonons.* We can distinguish so-called *acoustic phonons* and *optical phonons.* These names refer to the vibration models of the lattice (see [14, Ch. 9.2] or [28, Ch. 7]). The collision operator for electron-phonon interactions is given by

$$(Q_\alpha(F))(x, k, t) = \int_B [s_\alpha(x, k', k)F'(1 - F) - s_\alpha(x, k, k')F(1 - F')] \, dk',$$

where $F = F(x, k, t)$, $F'(F(x, k', t)$, and the scattering rate is

$$\begin{aligned} s_\alpha(x, k, k') &= \sigma_\alpha(x, k, k')[(1 + N_\alpha)\delta(\varepsilon(k') - \varepsilon(k) + \hbar\omega_\alpha) \\ &\quad + N_\alpha \delta(\varepsilon(k') - \varepsilon(k) - \hbar\omega_\alpha)], \end{aligned}$$

the function $\sigma_\alpha(x, k, k')$ is symmetric in k and k', $\hbar\omega_\alpha$ is the (constant) phonon energy, N_α is the phonon occupation number given by Bose-Einstein statistics

$$N_\alpha = \left(\exp\left(\frac{\hbar\omega_\alpha}{k_B T} \right) - 1 \right)^{-1},$$

the index α refers to either "opt" for optical phonons or "ac" to acoustic phonons, and δ is the delta distribution used in Section 2.3. The above transition rate is non-zero only if $\varepsilon(k') - \varepsilon(k) = \pm\hbar\omega_\alpha$, i.e., the rate accounts for phonon emission and absorption of energy $\hbar\omega_\alpha$. The above expression for s_α shows that the scattering rates can be highly non-smooth; in fact, Q_α is a distribution.

When an atom different from the semiconductor material is introduced into the semiconductor it may donate either an electron or a hole, leaving behind an ionized charged impurity center. These ionized impurities can act as scattering agents for free carriers propagating through the crystal. The interaction of carriers with neutral impurities is also possible, but we do not consider it here. The collision operator reads

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

where $F = F(x,k,t)$, $F' = F(x,k',t)$ and σ_{imp} is symmetric in k and k' [11].

The electron-electron interaction arises from the long-range nature of the Coulomb force. The corresponding collision operator

$$
\begin{aligned}
(Q_{\mathrm{ee}}(F))(x,k,t) &= \int_{B^3} [s_{\mathrm{ee}}(x,k,k_1,k',k_1')F'F_1'(1-F)(1-F_1) \\
&\quad - s_{\mathrm{ee}}(x,k',k_1',k,k_1)FF_1(1-F')(1-F_1')]\,dk'\,dk_1\,dk_1',
\end{aligned}
$$

where $F = F(x,k,t)$, $F' = F(x,k',t)$, $F_1 = F(x,k_1,t)$, and $F_1' = F(x,k_1',t)$, is a nonlocal nonlinearity of fourth order. The influence of electron-electron interaction 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 2.3). Also electron-hole interactions are possible but we do not consider these collision events here. For more details of scattering processes, we refer to [41, Ch. 6] or [14, Ch. 9].

The collision operator $Q(F)$ in (3.27) can generally be written as

$$Q(F) = Q_{\mathrm{opt}}(F) + Q_{\mathrm{ac}}(F) + Q_{\mathrm{imp}}(F) + Q_{\mathrm{ee}}(F).$$

In the following we assume that the collision operator is given by

$$(Q(F))(x,k,t) = \int_B [s(x,k',k)F'(1-F) - s(x,k,k')F(1-F')]\,dk', \quad (3.29)$$

where we have again set $F = F(x,k,t)$ and $F' = F(x,k',t)$.

We study now some properties of the collision integral (3.29). Since $s(x,k,k')$ is the transition rate from state k to state k', the principle of detailed balance implies that

$$\frac{s(x,k,k')}{s(x,k',k)} = e^{(\varepsilon(k) - \varepsilon(k'))/k_B T} \quad \forall x \in \mathbb{R}^d,\ k,k' \in B, \quad (3.30)$$

holds (compare with (2.18)).

Lemma 3.2 *(see [36]) Let (3.30) hold for some function $\varepsilon(k)$ and let $s(x, k, k') > 0$ for all $x \in \mathbb{R}^d$, $k, k' \in B$.*

(1) It holds

$$\int_B (Q(F))(x, k, t) \, dk = 0 \quad \forall x \in \mathbb{R}^d, \ t > 0,$$

for all (regular) functions $F : \mathbb{R}^3 \times B \times [0, \infty) \to [0, 1]$.

(2) It holds for all functions $F : \mathbb{R}^3 \times B \times [0, \infty) \to [0, 1]$ and all non-decreasing functions $\chi : \mathbb{R} \to \mathbb{R}$

$$\int_B (Q(F))(x, k, t) \, \chi \left(\frac{F(x, k, t)}{1 - F(x, k, t)} e^{\varepsilon(k)/k_B T} \right) \, dk \ \leq \ 0,$$

$$\int_B (Q(F))(x, k, t) \, \chi \left(\frac{1 - F(x, k, t)}{F(x, k, t)} e^{-\varepsilon(k)/k_B T} \right) \, dk \ \geq \ 0.$$

(3) The property $Q(F) = 0$ is equivalent to the existence of $-\infty \leq \varepsilon_F \leq +\infty$ such that

$$F(k) = \frac{1}{1 + \exp((\varepsilon(k) - \varepsilon_F)/k_B T)}, \quad k \in B.$$

Let F be a solution to the Boltzmann equation (3.27), (3.28). Then Lemma 3.2(1) implies that the total number of electrons is conserved in time:

$$n(x, t) = \int_B F(x, k, t) \, dk = \int_B F_I(x, k) \, dk = n_I(x, t), \quad x \in \mathbb{R}^d, \ t > 0.$$

Physically, this is reasonable: collisions neither destroy nor generate particles. The statements of Lemma 3.2(2) are also called *H-theorems.* Finally, Lemma 3.2(3) means that precisely the Fermi-Dirac distribution functions are in the kernel of the collision operator.

Proof of Lemma 3.2. (1) Changing k and k' in the second integral gives

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

where $F = F(x, k, t)$ and $F' = F(x, k', t)$.

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

$$M(k) = e^{-\varepsilon(k)/k_B T}, \quad h(k) = \frac{1 - F(k)}{F(k)} M(k).$$

Then (3.30) is equivalent to

$$\frac{s(x, k', k)}{M(k)} = \frac{s(x, k, k')}{M(k')},$$ (3.31)

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

$$\int_B Q(F)\chi(h) \, dk$$
$$= \int_{B^2} s(x, k, k') \left[\frac{M}{M'} F'(1 - F)\chi(h) - F(1 - F')\chi(h) \right] \, dk' \, dk$$
$$= \int_{B^2} \frac{s(x, k, k')}{M'} FF' \left[\frac{1 - F}{F} M\chi(h) - \frac{1 - F'}{F'} M'\chi(h) \right] \, dk' \, dk$$
$$= \int_{B^2} \frac{s(x, k, k')}{M'} FF'(h - h')\chi(h) \, dk' \, dk,$$ (3.32)

and, by changing k and k' and using (3.31),

$$\int_B Q(F)\chi(h) \, dk$$
$$= \int_{B^2} \frac{s(x, k', k)}{M} F'F \left[\frac{1 - F'}{F'} M'\chi(h') - \frac{1 - F}{F} M\chi(h') \right] \, dk \, dk'$$
$$= \int_{B^2} \frac{s(x, k, k')}{M'} F'F \left[\frac{1 - F'}{F'} M'\chi(h') - \frac{1 - F}{F} M\chi(h') \right] \, dk' \, dk$$
$$= \int_{B^2} \frac{s(x, k, k')}{M'} F'F(h' - h)\chi(h') \, dk' \, dk.$$ (3.33)

Adding (3.32) and (3.33) yields

$$\int_B Q(F)\chi(h) \, dk = \frac{1}{2} \int_B \frac{s(x, k, k')}{M'} FF'(h - h')(\chi(h) - \chi(h')) \, dk \, dk'$$
$$\geq 0,$$

since $s(x, k, k')$, M', F, and F' are non-negative and χ is non-decreasing.

(3) By (2), with $\chi(x) = x$, the property $Q(F) = 0$ is equivalent to

$$FF'(h - h')^2 = 0 \quad \text{for almost all } k, k' \in B.$$

This implies $F = 0$ or $h = h'$ almost everywhere. The equation $h = h'$ is equivalent to

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

Thus, both sides are constant and we call this constant $\exp(-\varepsilon_F/k_B T)$ with $\varepsilon_F \in \mathbb{R}$. Notice that the constant must be positive since $0 \le F \le 1$ expect if $F \equiv 1$. But then

$$\frac{1 - F(k)}{F(k)} = \frac{e^{-\varepsilon_F/k_B T}}{M(k)} = e^{(\varepsilon(k) - \varepsilon_F)/k_B T}$$

and $F(k) = 1 + \exp((\varepsilon(k) - \varepsilon_F)/k_B T))^{-1}$. Choosing $\varepsilon_F = \pm\infty$ yields the other two possibilities $F \equiv 0$ or $F \equiv 1$. $\qquad\square$

In the literature, two approximations of the collision operator are frequently used:

- the *low-density approximation*:

$$(Q(F))(x, k, t) = \int_B \sigma(x, k', k)(MF' - M'F) \, dk', \qquad (3.34)$$

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

- the *relaxation-time approximation*:

$$(Q(F))(x, k, t) = -\frac{F(x, k, t) - M(k)n(x, t)}{\tau(x, k)}, \qquad (3.35)$$

  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).

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

$$0 \le F(x, k, t) \ll 1.$$

Then $1 - F(k)$ is approximated by one in (3.28) and, employing (3.30),

$$
\begin{aligned}
(Q(F))(x, k, t) &= \int_B [s(x, k', k)F' - s(x, k, k')F] \, dk' \\
&= \int_B \frac{s(x, k', k)}{M(k)}[M(k)F' - M(k')F] \, dk' \\
&= \int_B \sigma(s, k', k)(MF' - M'F) \, dk',
\end{aligned}
$$

where $M(k) = \exp(-\varepsilon(k)/k_B T)$ and $M' = M(k')$. Notice that by (3.30), the collision cross-section $\sigma(x, k', k)$ is symmetric in k and k'.

43

When the initial datum F_I is close to a multiple of the so-called *Maxwellian* $M(k)$ it is reasonable to approximate F' in (3.34) by $n(x, t)M(k')$. Then we obtain from (3.34)

$$
\begin{aligned}
(Q(F))(x, k, t) &= \int_B \sigma(x, k', k)(nMM' - M'F) \, dk' \\
&= \int_B s(x, k, k')(nM - F) \, dk' \\
&= \int_B s(x, k, k') \, dk' \cdot (n(x, t)M(k) - F(x, k, t)) \\
&= -\frac{F(x, k, t) - n(x, t)M(k)}{\tau(x, k)}.
\end{aligned}
$$

Along the trajectories $(x(t), k(t))$, where

$$
\partial_t x = v(k), \quad \hbar \partial_t k = -qE_{\text{eff}}, \quad x(0) = x_0, \quad k(0) = k_0,
$$

the Boltzmann equation with the relaxation-time approximation for *constant* τ reads

$$
\frac{dF}{dt} = -\frac{1}{\tau}(F - nM), \quad t > 0.
$$

The solution of this linear ordinary differential equation (for given $n(x, t)$) is

$$
F(x(t), k(t), t) = e^{-t/\tau} F_I(x_0, k_0) + \frac{1}{\tau} \int_0^t n(x(s), s)M(k(s))e^{(s-t)/\tau} \, ds.
$$

It is possible to show that

$$
e^{t/\tau} \left[F(x(t), k(t), t) - n(x(t), t)M(k(t)) \right] \to 1 \text{ as } t \to \infty,
$$

i.e., the function relaxes to the equilibrium density nM from the perturbed state F_I along the trajectories after a time of order τ. This explains the name relaxation time.

A summary of the main models derived in Section 3.1-3.3 and its relations are presented in Figure 3.1.

3.4 Extensions

We discuss two extensions of the Boltzmann equation:

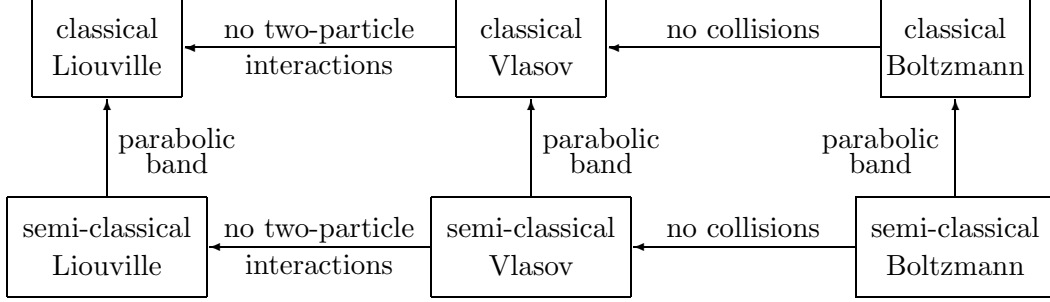- multi-valley models and

- bipolar models.

Figure 3.1: Relations between the (semi-)classical Liouville, Vlasov and Boltzmann equations.
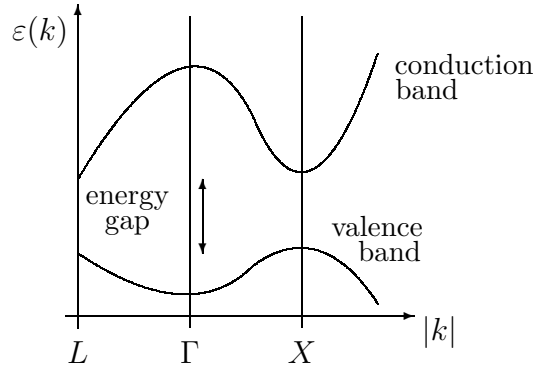
Figure 3.2: Schematic band structure of GaAs.

In the semiconductor material GaAs, the (conduction band) energy $\varepsilon(k)$ has one maximum and several minima which are called energy valleys. More precisely, GaAs has three valleys: the low-energy Γ-valley and the higher energetic L- and X-valleys; see Figure 3.2 [28, Ch. 4]. As the highest energy X-valley can be occupied only at very high electric field strengths, we will consider only the Γ- and L-valleys. We assume furthermore the parabolic band approximation. Then, after a suitable translation and rotation of the three-dimensional k-space,

$$
\begin{aligned}
\varepsilon_\Gamma(k) &= \Delta_\Gamma - \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_{\Gamma x}^*} + \frac{k_y^2}{m_{\Gamma y}^*} + \frac{k_z^2}{m_{\Gamma z}^*}\right), \\
\varepsilon_L(k) &= \Delta_L + \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_{Lx}^*} + \frac{k_y^2}{m_{Ly}^*} + \frac{k_z^2}{m_{Lz}^*}\right), \quad k = (k_x, k_y, k_z)^\top,
\end{aligned}
$$

where $m_{\Gamma_\alpha}^*$, $m_{L_\alpha}^*$ are the effective masses of an electron in the Γ-valley, L-valley, respectively, in the different directions. Since we assume low field strengths, we split the distribution function F into a part F_Γ corresponding to the Γ-valley and

a part F_L corresponding to the L-valley. Indeed, the electrons move only within each valley since the transfer to a higher-energetic valley requires the presence of high electric fields. Thus, for a single L-valley,

$$F(x, k, t) = F_\Gamma(x, k, t) + F_L(x, k, t),$$

and the functions F_Γ and F_L are satisfying the semi-classical Boltzmann equations

$$\partial_t F_\Gamma + v_\Gamma(k) \cdot \nabla_x F_\Gamma - \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k F_\Gamma = Q_\Gamma(F_\Gamma) + Q_{\Gamma \to L}(F_\Gamma, F_L),$$

$$\partial_t F_L + v_L(k) \cdot \nabla_x F_L - \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k F_L = Q_L(F_L) + Q_{L \to \Gamma}(F_\Gamma, F_L),$$

where

$$v_\Gamma(k) = \frac{1}{\hbar} \nabla_k \varepsilon_\Gamma(k), \quad v_L(k) = \frac{1}{\hbar} \nabla_k \varepsilon_L(k)$$

denote the mean velocities of the electrons in the Γ- and L-valleys, respectively (see Section 2.2), $Q_\Gamma(F_\Gamma)$ and $Q_L(F_L)$ are the intravalley low-density collision operators

$$Q_\alpha(F_\alpha) = \int_{\mathbb{R}^3} \sigma_\alpha(x, k', k)(M_\alpha(k)F'_\alpha - M_\alpha(k')F_\alpha) \, dk', \quad \alpha = \Gamma, L,$$

$Q_{\Gamma \to L}$ and $Q_{L \to \Gamma}$ are the intervalley low-density collision operators

$$Q_{\Gamma \to L}(F_\Gamma, F_L) = \int_{\mathbb{R}^3} (\sigma_{L \to \Gamma}(x, k', k)M_\Gamma(k)F'_L - \sigma_{\Gamma \to L}(x, k, k')M_L(k')F_\Gamma) \, dk'$$

$$Q_{L \to \Gamma}(F_\Gamma, F_L) = \int_{\mathbb{R}^3} (\sigma_{\Gamma \to L}(x, k', k)M_L(k)F'_\Gamma - \sigma_{L \to \Gamma}(x, k, k')M_\Gamma(k')F_L) \, dk'$$

$s_{\Gamma \to L}(x, k, k')$ is the cross-section from the state (x, k) of the Γ-valley into a state (x, k') of the L-valley, similarly for $s_{L \to \Gamma}(x, k, k')$, and the Maxwellians M_Γ and M_L are given by

$$M_\alpha(k) = N^*_\alpha \exp\left(-\frac{\varepsilon_\alpha(k)}{k_B T}\right), \quad N^*_\alpha = \left(\int_{\mathbb{R}^3} \exp\left(-\frac{\varepsilon_\alpha(k)}{k_B T}\right) dk\right)^{-1}, \quad \alpha = \Gamma, L.$$

The effective field E_{eff} is computed from (3.23), i.e.

$$E_{\text{eff}}(x, t) = E_{\text{ext}}(x, t) + \int_{\mathbb{R}^3} (n_\Gamma(x_*, t) + n_L(x_*, t)) \, E_{\text{int}}(x, x_*) \, dx_*,$$

$$n_\alpha(x, t) = \int_{\mathbb{R}^3} F_\alpha(x, k, t) \, dk, \quad \alpha = \Gamma, L.$$

It can be shown similarly as in Lemma 3.2 that under the assumption

$$\sigma_{\Gamma \to L}(x, k, k') = \sigma_{L \to \Gamma}(x, k', k) \quad \forall x, k, k' \in \mathbb{R}^d,$$

46

the property $Q_{\Gamma \to L}(F_\Gamma, F_L) = Q_{L \to \Gamma}(F_\Gamma, F_L) = 0$ is equivalent to

$$(F_\Gamma, F_L) = \text{const.}(M_\Gamma, M_L).$$

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. For such a process, which is termed *generation* of an electron-hole pair, energy absorption is necessary. The inverse process that an electron of the conduction band moves to the lower energetic valence band, occupying an empty state, is called *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, and

- 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 then excited to a higher energy state. The inverse Auger process, in which an electron-hole pair is generated, is called impact ionization. The energy comes from the collision of a high-energy free carrier with the lattice.

In a radiative recombination event, an electron from the conduction band recombines with a hole from the valence band emitting a photon. The energy lost by the electron is equal to the energy gap of the material, and a photon of this amount of energy is produced. Radiative generation events occur from the absorption of a photon of energy greater than or equal to the bandgap energy.

Finally, thermal recombination or generation events arise from phonon emission or absorption, respectively (see Figure 3.3).
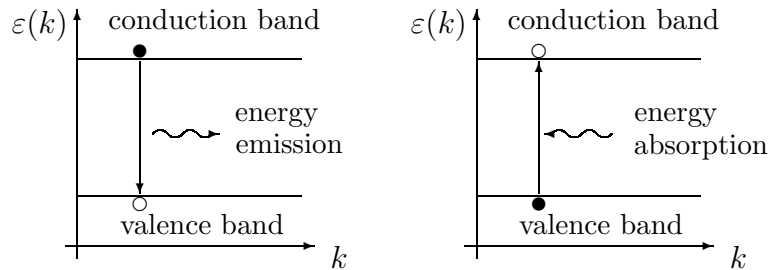

Figure 3.3: Recombination (left) and generation (right) of an electron-hole pair.

The evolution of the distribution functions F_n of the electrons and F_p of the holes is given by the Boltzmann equation including an operator accounting for the recombination-generation processes. More precisely, in the semi-classical framework,

$$\partial_t F_n + v_n(k) \cdot \nabla_x F_n - \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k F_n = Q_n(F_n) + I_n(F_n, F_p), \quad (3.36)$$

$$\partial_t F_p + v_p(k) \cdot \nabla_x F_p + \frac{q}{\hbar} E_{\text{eff}} \cdot \nabla_k F_p = Q_p(F_n) + I_p(F_n, F_p), \quad (3.37)$$

where

$$v_n = \frac{1}{\hbar} \nabla_k \varepsilon_n, \quad v_p = \frac{1}{\hbar} \nabla_k \varepsilon_p$$

are the mean velocities related to the electron conduction band ε_n and the hole valence band ε_p, respectively, Q_n and Q_p are the collision operators

$$(Q_\alpha(F_\alpha))(x, k, t) = \int_B [s_\alpha(x, k', k)F'(1 - F) - s_\alpha(x, k, k')F(1 - F)] \ dk',$$

$\alpha = n, p$, and the recombination-generation operators are given by

$$(I_n(F_n, F_p))(x, k, t) = \int_B \left[g(x, k', k)(1 - F_n)(1 - F_p') - r(x, k, k')F_n F_p' \right] \ dk',$$

$$(I_p(F_n, F_p))(x, k, t) = \int_B \left[g(x, k, k')(1 - F_n')(1 - F_p) - r(x, k', k)F_n' F_p \right] \ dk'.$$

Here $g(x, k, k') \geq 0$ is the rate of generation of an electron at the state (x, k) and of a hole at the state (x, k'), and $r(x, k', k) \geq 0$ is the analogous recombination rate. Similarly as for the transition rates s_n and s_p (see (3.30)) the relation

$$r(x, k, k') = \exp\left(\frac{\varepsilon_n(k) - \varepsilon_p(k')}{k_B T} \right) g(x, k', k)$$

is assumed to hold. Then the operators I_n and I_p can be written, in the low-density approximation, as follows:

$$(I_n(F_n, F_p))(x, k, t) = \int_B g(x, k', k) \left[1 - \exp\left(\frac{\varepsilon_n(k) - \varepsilon_p(k')}{k_B T} \right) F_n F_p' \right] \ dk',$$

$$(I_n(F_n, F_p))(x, k, t) = \int_B g(x, k, k') \left[1 - \exp\left(\frac{\varepsilon_n(k') - \varepsilon_p(k)}{k_B T} \right) F_n' F_p \right] \ dk'.$$

The positive sign for the term involving E_{eff} in the Botzmann equation (3.37) for the holes comes from the opposite flow direction of the positively charged holes in the electric field E_{eff}.

In the case of Coulomb forces in $\mathbb{R}^3$, the effective field is given by

$$E_{\text{eff}}(x, t) = \frac{1}{4\pi\varepsilon_s} \int_{\mathbb{R}^3} \rho(x_*, t) \frac{x - x_*}{|x - x_*|^3} \ dx_*,$$

48

where ρ is the total space charge density. The charge density 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 crystal is doped (see Section 2.3):

$$\rho = q(-n + p - N_A + N_D),$$

with corresponding charges $+q$ or $-q$. Here, we have set

$$n(x,t) = \int_B F_n(x,k,t)\,dk, \quad p(x,t) = \int_B F_p(x,k,t)\,dk.$$

Thus the Poisson equation (3.21) for the electrostatic potential defined by $E_{\text{eff}} = -\nabla V_{\text{eff}}$ writes

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

where $C = N_D - N_A$ is the doping profile.

The total number of each type of particles is not conserved anymore, due to recombination-generation effects. However, taking the difference of the Boltzmann equations (3.36) and (3.37) and integrating over $x \in \mathbb{R}^d$ and $k \in B$ yields

$$\partial_t \int_{\mathbb{R}^d} (n-p)(x,t)\,dx = \int_{\mathbb{R}^d} \int_B (I_n(F_n, F_p) - I_p(F_n, F_p))\,dk\,dx = 0,$$

by Lemma 3.2(1) and by definition of I_n and I_p. If the doping profile C does not depend on t, we obtain the conservation of the total charge density

$$\partial_t \int_{\mathbb{R}^d} \rho(x,t)\,dx = \partial_t \int_{\mathbb{R}^d} (n(x,t) - p(x,t) - C(x))\,dx = 0.$$

4 Quantum Kinetic Transport Equations

4.1 The Wigner equation

As the dimensions of a semiconductor device decrease, quantum mechanical transport phenomena play an important rule in the function of the device. Therefore, it is of great importance to devise transport models which, on one hand, are capable of describing quantum effects and, on the other hand, are sufficiently simple to allow for efficient numerical simulations. In order to derive such models, we start with the Schrödinger equation.

The motion of an electron ensemble consisting of M particles in a vacuum under the action of an electric field $E = -\nabla_x V$ is described by the Schrödinger equation

$$i\hbar \partial_t \psi = -\frac{\hbar^2}{2m} \sum_{j=1}^{M} \Delta_{x_j} \psi - qV(x,t)\psi, \quad x \in \mathbb{R}^{dM}, \ t > 0, \qquad (4.1)$$

$$\psi(x,0) = \psi_I(x), \quad x \in \mathbb{R}^{dM},$$

where $x = (x_1, \ldots, x_M)$, $i^2 = -1$, $\hbar = h/2\pi$ is the reduced Planck constant, m the electron mass, q the elementary charge, V the real-valued electrostatic potential, and $\psi = \psi(x,t)$ is called the *wave function* of the electron ensemble. Macroscopic variables are defined by

$$\text{electron density:} \qquad n = |\psi|^2,$$
$$\text{electron current density:} \quad J = -\frac{\hbar q}{m} \text{Im}(\overline{\psi} \nabla_x \psi),$$

where $\overline{\psi}$ is the complex conjugate of ψ.

Another formulation of the motion of the electron ensemble can be obtained using the so-called *density matrix*

$$\rho(r,s,t) = \overline{\psi(r,t)}\psi(s,t), \quad r,s \in \mathbb{R}^{dM}, \ t > 0. \qquad (4.2)$$

The electron density and current density, respectively, are given in terms of the density matrix by

$$n(x,t) = \rho(x,x,t), \qquad (4.3)$$

$$J(x,t) = \frac{i\hbar q}{2m} \left(\overline{\psi} \nabla_x \psi - \psi \nabla_x \overline{\psi} \right)(x,t)$$

$$= \frac{i\hbar q}{2m} (\nabla_s - \nabla_r)\rho(x,x,t). \qquad (4.4)$$

Differentiating (4.2) with respect to t and using the Schrödinger equation (4.1)

leads to the *Heisenberg equation*

$$
\begin{aligned}
i\hbar\partial_t\rho(r,s,t) &= i\hbar\left(\overline{\partial_t\psi(r,t)}\psi(s,t) + \overline{\psi(r,t)}\partial_t\psi(s,t)\right)\\
&= \left(\frac{\hbar^2}{2m}\Delta_r\overline{\psi} + qV\overline{\psi}\right)(r,t)\psi(s,t)\\
&\quad + \overline{\psi(r,t)}\left(-\frac{\hbar^2}{2m}\Delta_s\psi - qV\psi\right)(s,t) \qquad (4.5)\\
&= -\frac{\hbar^2}{2m}(\Delta_s\rho - \Delta_r\rho)(r,s,t) - q(V(s,t) - V(r,t))\rho(r,s,t),
\end{aligned}
$$

with $r, s \in \mathbb{R}^{dM}$, $t > 0$.

The kinetic form of the quantum equations is obtained by a change of unknowns and a Fourier transformation. We recall that the Fourier transform F is defined by

$$
(Fg)(\eta) = \hat{g}(\eta) = \frac{1}{(2\pi)^{dM/2}}\int_{\mathbb{R}^{dM}} g(v)e^{-i\eta\cdot v}\,dv
$$

for (sufficiently smooth) functions $g : \mathbb{R}^{dM} \to \mathbb{C}$ with inverse

$$
(F^{-1}h)(v) = \check{h}(v) = \frac{1}{(2\pi)^{dM/2}}\int_{\mathbb{R}^{dM}} h(\eta)e^{i\eta\cdot v}\,d\eta
$$

for functions $h : \mathbb{R}^{dM} \to \mathbb{C}$. Introduce the change of coordinates

$$
r = x + \frac{\hbar}{2m}\eta, \quad s = x - \frac{\hbar}{2m}\eta
$$

and set

$$
u(x,\eta,t) = \rho\left(x + \frac{\hbar}{2m}\eta, x - \frac{\hbar}{2m}\eta, t\right). \qquad (4.6)
$$

Since $(\hbar/2m)\eta$ has the dimension of length and $\hbar/2m$ has the dimension of cm^2/s, η has the dimension of inverse velocity s/cm. Thus the variable v in the Fourier transform of a function $g(v)$ has the dimension of velocity cm/s. We compute

$$
\begin{aligned}
\operatorname{div}_\eta(\nabla_x u)(x,\eta,t) &= \operatorname{div}_\eta(\nabla_r\rho + \nabla_s\rho)\left(x + \frac{\hbar}{2m}\eta, x - \frac{\hbar}{2m}\eta, t\right)\\
&= \frac{\hbar}{2m}(\Delta_r\rho - \Delta_s\rho)(r,s,t).
\end{aligned}
$$

Hence the transformed Heisenberg equation for u reads

$$
\partial_t u + i\operatorname{div}_\eta(\nabla_x u) + \frac{q}{m}(\delta V)u = 0, \quad x, \eta \in \mathbb{R}^{dM},\ t > 0, \qquad (4.7)
$$

where

$$
(\delta V)(x,\eta,t) = \frac{im}{\hbar}\left[V\left(x + \frac{\hbar}{2m}\eta, t\right) - V\left(x - \frac{\hbar}{2m}\eta, t\right)\right]. \qquad (4.8)
$$

The initial condition is

$$u(x, \eta, 0) = \overline{\psi_I \left(x + \frac{\hbar}{2m}\eta \right)} \psi_I \left(x - \frac{\hbar}{2m}\eta \right), \quad x, \eta \in \mathbb{R}^{dM}.$$

We prefer to work with the variables velocity v and space x instead of inverse velocity η and space x and apply therefore the Fourier transformation to u. The function

$$w = (2\pi)^{-dM/2} F^{-1} u = (2\pi)^{-dM/2} \breve{u}$$

or, in terms of the wave function,

$$w(x, v, t) = \frac{1}{(2\pi)^{dM}} \int_{\mathbb{R}^{dM}} \overline{\psi \left(x + \frac{\hbar}{2m}\eta \right)} \psi \left(x - \frac{\hbar}{2m}\eta \right) e^{i\eta \cdot v} \, d\eta$$

is called *Wigner function.* It was introduced by Wigner in 1932 [43].

The macroscopic electron density n and the current density J can now be written as

$$n(x, t) \;=\; \int_{\mathbb{R}^{dM}} w(x, v, t) \, dv, \tag{4.9}$$

$$J(x, t) \;=\; -q \int_{\mathbb{R}^{dM}} v w(x, v, t) \, dv, \tag{4.10}$$

since the first identity follows from (4.3) and

$$n(x, t) = \rho(x, x, t) = u(x, 0, t) = (2\pi)^{dM/2} \hat{w}(x, 0, t) = \int_{\mathbb{R}^{dM}} w(x, v, t) \, dv,$$

and for the second identity we use the fact that the Fourier transform translates differential operators into a multiplication:

$$
\begin{aligned}
i(\nabla_\eta \hat{w})(x, \eta, t) \;&=\; \frac{i}{(2\pi)^{dM/2}} \int_{\mathbb{R}^{dM}} w(x, v, t) \nabla_\eta e^{-i\eta \cdot v} \, dv \\
&=\; \frac{1}{(2\pi)^{dM/2}} \int_{\mathbb{R}^{dM}} w(x, v, t) v e^{-i\eta \cdot v} \, dv \\
&=\; \widehat{vw}(x, \eta, t),
\end{aligned}
$$

and therefore, using (4.4),

$$
\begin{aligned}
J(x, t) \;&=\; \frac{i\hbar q}{2m}(\nabla_s - \nabla_r)\rho(x, x, t) = -iq(\nabla_\eta u)(x, 0, t) \\
&=\; -iq(2\pi)^{dM/2}(\nabla_\eta \hat{w})(x, 0, t) = -q(2\pi)^{dM/2}\widehat{vw}(x, 0, t) \\
&=\; -q \int_{\mathbb{R}^{dM}} v w(x, v, t) \, dv.
\end{aligned}
$$

The integrals (4.9) and (4.10) are called the zeroth and first moments of the Wigner function, respectively, in analogy to the classical situation (see Section 3.1).

The transport equation for w is obtained by taking the inverse Fourier transform of the equation (4.7) for u :

$$\partial_t \check{u} + i(\operatorname{div}_\eta \nabla_x u)^\vee + \frac{q}{m} \left[(\delta V) u \right]^\vee = 0. \tag{4.11}$$

Using

$$\begin{aligned}
i(\operatorname{div}_\eta \nabla_x u)^\vee (x, v, t) &= \frac{i}{(2\pi)^{dM/2}} \int_{\mathbb{R}^{dM}} (\operatorname{div}_\eta \nabla_x u(x, \eta, t)) e^{iv \cdot \eta} \, d\eta \\
&= \frac{v}{(2\pi)^{dM/2}} \cdot \int_{\mathbb{R}^{dM}} \nabla_x u(x, \eta, t) e^{iv \cdot \eta} \, d\eta \\
&= v \cdot \nabla_x \check{u}(x, v, t) \\
&= (2\pi)^{dM/2} v \cdot \nabla_x w(x, v, t),
\end{aligned}$$

where we have employed integration by parts, and

$$\begin{aligned}
(\theta[V]w)(x, v, t) &:= \frac{1}{(2\pi)^{dM/2}} \left[(\delta V) u \right]^\vee (x, v, t) \tag{4.12} \\
&= \frac{1}{(2\pi)^{dM}} \int_{\mathbb{R}^{dM}} (\delta V)(x, \eta, t) \, u(x, \eta, t) \, e^{i\eta \cdot v} \, d\eta \\
&= \frac{1}{(2\pi)^{dM}} \int_{\mathbb{R}^{dM}} \int_{\mathbb{R}^{dM}} (\delta V)(x, \eta, t) \, w(x, v', t) \, e^{i\eta \cdot (v - v')} \, dv' \, d\eta,
\end{aligned}$$

it follows from (4.11)

$$\partial_t w + v \cdot \nabla_x w + \frac{q}{m} \theta[V] w = 0, \quad x, v \in \mathbb{R}^{dM}, \ t > 0. \tag{4.13}$$

This equation is called *Wigner equation* or *quantum Liouville equation*. The initial condition for w reads

$$w(x, v, 0) = w_I(x, v), \quad x, v \in \mathbb{R}^{dM}, \tag{4.14}$$

where

$$w_I(x, v) = \frac{1}{(2\pi)^{dM}} \int_{\mathbb{R}^{dM}} \overline{\psi_I \left(x + \frac{\hbar}{2m} \eta \right)} \psi_I \left(x - \frac{\hbar}{2m} \eta \right) e^{iv \cdot \eta} \, d\eta, \quad x, v \in \mathbb{R}^{dM}. \tag{4.15}$$

The operator (4.12) is a pseudo-differential operator. Generally, an operator, whose Fourier transform acts as a multiplication operator on the Fourier transform of the function, is called a linear pseudo-differential operator [42]. Indeed, it holds

$$\widehat{\theta[V]w}(x, \eta, t) = (\delta V)(x, \eta, t) \hat{w}(x, \eta, t),$$

and so, $\theta[V]$ is a pseudo-differential operator and the Wigner equation (4.13) is a linear pseudo-differential equation.

We discuss now some properties of the Wigner equation, namely

- the semi-classical limit,

- the relation between the solution of the Wigner equation and the solution of the Schrödinger equation, and

- the question of positivity of the solution of the Wigner equation.

It is interesting to perform the formal *semi-classical limit* "$\hbar \to 0$". We compute, as "$\hbar \to 0$",

$$(\delta V)(x, \eta, t) \to i\nabla_x V(x, t) \cdot \eta$$

and, using $i(\eta u)^\vee = \nabla_v \check{u}$,

$$\theta[V]w \to i(2\pi)^{-dM/2} \left[(\nabla_x V \cdot \eta)\, u \right]^\vee = (2\pi)^{-dM/2} \nabla_x V \cdot \nabla_v \check{u} = \nabla_x V \cdot \nabla_v w.$$

Therefore, (4.13) becomes, as "$\hbar \to 0$",

$$\partial_t w + v \cdot \nabla_x w + \frac{q}{m} \nabla_x V \cdot \nabla_v w = 0,$$

and we recover the classical Liouville equation (3.6). Clearly, the above limits have to be understood in a formal way.

The solution of the initial-value problem (4.13)-(4.14) is given by

$$w(x, v, t) = \frac{1}{(2\pi)^{dM}} \int_{\mathbb{R}^{dM}} \overline{\psi\left(x + \frac{\hbar}{2m}\eta, t\right)} \psi\left(x - \frac{\hbar}{2m}\eta, t\right) e^{iv \cdot \eta}\, d\eta, \quad t > 0,$$

$$(4.16)$$

where ψ solves (4.1) if and only if the initial datum w_I satisfies (4.15). If (4.16) holds the quantum state of the electron is fully described by the single wave function ψ. In quantum physics this is referred to as a *pure quantum state.* For more general initial data, we cannot expect that (4.16) holds true. However, we can derive a similar expression. For this we choose the initial data

$$w_I \in L^2(\mathbb{R}^{dM} \times \mathbb{R}^{dM}) = \left\{ w : \mathbb{R}^{dM} \times \mathbb{R}^{dM} \to \mathbb{C} : \int_{\mathbb{R}^{2dM}} |w(x, v)|^2\, dx\, dv < \infty \right\}.$$

From functional analysis, there exists a complete orthonormal system $(\phi^{(k)})_{k \in \mathbb{N}}$ of functions $\phi^{(k)}$ in $L^2(\mathbb{R}^{dM})$ (defined similarly as above). It can be shown that $(\overline{\phi^{(j)}}\phi^{(k)})_{j,k \in \mathbb{N}}$ is a complete orthonormal system in $L^2(\mathbb{R}^{dM} \times \mathbb{R}^{dM})$ and that

$$\rho_I(r, s) = \hat{w}_I(x, \eta)$$

can be expanded in the series

$$\rho_I(r,s) = \sum_{k \in \mathbb{N}} \lambda_k \overline{\phi^{(k)}(r)} \phi^{(k)}(s)$$

with

$$\lambda_k = \int_{\mathbb{R}^{dM}} \int_{\mathbb{R}^{dM}} \rho_I(r,s)\, \phi^{(k)}(r) \overline{\phi^{(k)}(s)}\, dr\, ds.$$

Let $\psi^{(k)}$ be the solution of the Schrödinger equation

$$\begin{aligned} i\hbar \partial_t \psi^{(k)} &= -\frac{\hbar^2}{2m} \Delta_x \psi^{(k)} - qV\psi^{(k)}, \quad x \in \mathbb{R}^{dM},\ t > 0, \\ \psi^{(k)}(x,0) &= \phi^{(k)}(x), \quad x \in \mathbb{R}^{dM}. \end{aligned}$$

Then the solution of the Wigner equation (4.13)-(4.14) is given by

$$w(x,v,t) = \frac{1}{(2\pi)^{dM}} \sum_{k \in \mathbb{N}} \lambda_k \int_{\mathbb{R}^{dM}} \overline{\psi^{(k)}\left(x + \frac{\hbar}{2m}\eta, t\right)} \psi^{(k)}\left(x - \frac{\hbar}{2m}\eta, t\right) e^{iv \cdot \eta}\, d\eta,$$

for $t > 0$. This means that the Wigner equation is capable of describing so-called *mixed quantum states*.

The electron density is

$$\begin{aligned} n(x,t) &= \int_{\mathbb{R}^{dM}} w(x,v,t)\, dv \\ &= (2\pi)^{dM/2} \hat{w}(x,0,t) \\ &= \frac{1}{(2\pi)^{dM/2}} \sum_{k \in \mathbb{N}} \lambda_k \left[\overline{\psi^{(k)}\left(x + \frac{\hbar}{2m}\eta, t\right)} \psi^{(k)}\left(x - \frac{\hbar}{2m}\eta, t\right) \right]_{\eta=0} \\ &= \sum_{k \in \mathbb{N}} \frac{\lambda_k}{(2\pi)^{dM/2}} |\psi^{(k)}(x,t)|^2. \end{aligned}$$

If the initial electron density is non-negative or, more precisely, if $\lambda_k \geq 0$ for all $k \in \mathbb{N}$, the electron density is non-negative for all $t > 0$. One may ask if also the Wigner function is non-negative for all $t > 0$ if this is true initially. This would allow for a probabilistic interpretation of the Wigner function. However, generally this is not true. It is shown in [30] that for a pure quantum state,

$$w(x,v,t) = \frac{1}{(2\pi)^{dM}} \int_{\mathbb{R}^{dM}} \overline{\psi\left(x + \frac{\hbar}{2m}\eta, t\right)} \psi\left(x - \frac{\hbar}{2m}\eta, t\right) e^{iv \cdot \eta}\, d\eta$$

is non-negative if and only if either $\psi \equiv 0$ or

$$\psi(x,t) = \exp\left(-x^\top A(t)x - a(t) \cdot x + b(t)\right), \quad x \in \mathbb{R}^{dM},\ t > 0,$$

where $A(t)$ is a $\mathbb{C}^{dM \times dM}$ matrix with symmetric positive definite real part and $a(t) \in \mathbb{C}^{dM}$, $b(t) \in \mathbb{C}$. A necessary condition for the non-negativity of w for mixed quantum states is not known.

The Wigner equation (4.13) does not account for the effect of the crystal lattice on the motion of the electron ensemble. The motion of *one* electron in the crystal is described by the Schrödinger equation

$$i\hbar\partial_t\psi = -\frac{\hbar^2}{2m}\Delta_x\psi - q(V_L + V)\psi, \quad x \in \mathbb{R}^3, \ t > 0,$$

where V_L is the lattice potential satisfying

$$V_L(x + \gamma\vec{a_j}) = V_L(x) \quad x \in \mathbb{R}^3, \ j = 1, 2, 3,$$

with the primitive vectors $\vec{a_j}$ of the lattice and the lattice length scale γ (see Section 2.1). By using a Bloch decomposition and performing the limit $\gamma \to 0$, it is possible to derive the Wigner equation in a crystal:

$$\partial_t w + v(k) \cdot \nabla_x w + \frac{q}{\hbar}\theta[V]w = 0, \tag{4.17}$$

where $v(k) = (1/\hbar)\nabla_k\varepsilon(k)$ and $\varepsilon(k)$ is the energy band in which the electrons are moving. The pseudo-differential operator $\theta[V]$ is defined as in (4.12), and the function w is the inverse of the Fourier transform of the density matrix transformed from (r, s) to (x, η) as in (4.6). We refer to [6, 38] for details on the derivation.

4.2 The quantum Vlasov and quantum Boltzmann equation

The quantum Liouville equation has the same disadvantages as its classical analogue:

- The equation has to be solved in a very high-dimensional phase space since $M \gg 1$, and its numerical solution is almost unfeasible.

- Short-range or long-range interactions are not included.

In this section we derive the quantum analogue of the classical Vlasov equation, the quantum Vlasov equation which acts on a low-dimensional (x, v)-space of dimension $2d$. We proceed similarly as in Section 3.2. In particular, we impose the following assumptions.

Consider an ensemble of M electrons with mass m in a vacuum under the action of the real-valued electrostatic potential $V(x, t)$, $x \in \mathbb{R}^{dM}$, $t > 0$. The motion of the particle ensemble is described by the wave function ψ which is a solution of the many-particle Schrödinger equation (4.1). We assume:

- The initial wave function is anti-symmetric:

$$\psi(x_1, \ldots, x_M, 0) = \text{sign}(\pi)\, \psi(x_{\pi(1)}, \ldots, x_{\pi(M)}, 0) \qquad (4.18)$$

for all permutations π of the set $\{1, \ldots, M\}$ and for all $x = (x_1, \ldots, x_M) \in \mathbb{R}^{dM}$.

- The potential can be decomposed as

$$V(x_1, \ldots, x_M, t) = \sum_{j=1}^{M} V_{\text{ext}}(x_j, t) + \frac{1}{2} \sum_{i,j=1}^{M} V_{\text{int}}(x_i, x_j), \qquad (4.19)$$

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 of the order of magnitude $1/M$ as $M \to \infty$.

We discuss the assumptions. The factor $\frac{1}{2}$ in (4.19) is necessary since each electron-electron pair in the sum of two-particle interactions is counted twice. Notice that the symmetry of V_{int} implies

$$V(x_1, \ldots, x_M, t) = V(x_{\pi(1)}, \ldots, x_{\pi(M)}, t)$$

for all permutations. It is possible to show that then, the anti-symmetry of ψ is conserved for all time if ψ is anti-symmetric initially.

The first hypothesis implies that the ensemble density matrix

$$\rho(r, s, t) = \overline{\psi(r, t)}\psi(s, t), \quad r = (r_1, \ldots, r_M), \ s = (s_1, \ldots, s_M) \in \mathbb{R}^{dM}, \ t \geq 0,$$

remains invariant under any permutations of the r- and s-arguments:

$$\rho(r_1, \ldots, r_M, s_1, \ldots, s_M, t) = \rho(r_{\pi(1)}, \ldots, r_{\pi(M)}, s_{\pi(1)}, \ldots, s_{\pi(M)}, t) \qquad (4.20)$$

for all permutations π of $\{1, \ldots, M\}$ and all $r_i, s_j \in \mathbb{R}^d$, $t \geq 0$. This expresses the fact that the electrons are indistinguishable.

It is possible to use, instead of (4.18), the condition (4.20) as a hypothesis which has the advantage that the condition can be easily interpreted physically. However, (4.20) does not imply (4.18). In fact, (4.20) is satisfied if either the wave function is anti-symmetric or symmetric. The anti-symmetry property represents the Pauli exclusion principle since it implies that

$$\psi(x_1, \ldots, x_M, t) = 0 \quad \text{if } x_i = x_j \text{ for } i \neq j,$$

meaning that double occupancy of states is prohibited. We would need to assume (4.20) *and* the Pauli exclusion principle as the hypotheses. We prefer to use the slightly stronger condition (4.18) as a hypothesis.

We wish to model the evolution of subensembles. The density matrix of a subensemble of particles is defined by

$$\rho^{(a)}(r^{(a)}, s^{(a)}, t) = \int_{\mathbb{R}^{d(M-a)}} \rho(r^{(a)}, u_{a+1}, \ldots, u_M, s^{(a)}, u_{a+1}, \ldots, u_M) \, du_{a+1} \ldots du_M,$$

where

$$r^{(a)} = (r_1, \ldots, r_a), \quad s^{(a)} = (s_1, \ldots, s_a) \in \mathbb{R}^{da}.$$

Due to (4.20), the subensemble density matrices satisfy

$$\rho^{(a)}(r_1, \ldots, r_a, s_1, \ldots, s_a, t) = \rho^{(a)}(r_{\pi(1)}, \ldots, r_{\pi(a)}, s_{\pi(1)}, \ldots, s_{\pi(a)}, t) \qquad (4.21)$$

for all permutations π of $\{1, \ldots, a\}$ and all $r_i, s_j \in \mathbb{R}^d$, $t \geq 0$. The evolution of the complete electron ensemble is governed by the Heisenberg equation (see (4.5))

$$
\begin{aligned}
i\hbar \partial_t \rho \;=\; & -\frac{\hbar^2}{2m} \sum_{j=1}^{M} \left(\Delta_{s_j} - \Delta_{r_j} \right) \rho - q \sum_{j=1}^{M} (V_{\text{ext}}(s_j, t) - V_{\text{ext}}(r_j, t)) \rho \\
& -\frac{q}{2} \sum_{j, \ell=1}^{M} \left(V_{\text{int}}(s_j, s_\ell) - V_{\text{int}}(r_j, r_\ell) \right) \rho.
\end{aligned}
$$

We set $u_j = s_j = r_j$ for $j = a+1, \ldots, M$ in the above equation, integrate over $(u_{a+1}, \ldots, u_M) \in \mathbb{R}^{d(M-a)}$ and use the indistinguishability property (4.21) to obtain, after an analogous calculation as in Section 3.2, a quantum equivalent of the BBGKY hierarchy:

$$
\begin{aligned}
i\hbar \partial_t \rho^{(a)} \;=\; & -\frac{\hbar^2}{2m} \sum_{j=1}^{M} \left(\Delta_{s_j} - \Delta_{r_j} \right) \rho^{(a)} - q \sum_{j=1}^{M} (V_{\text{ext}}(s_j, t) - V_{\text{ext}}(r_j, t)) \rho^{(a)} \\
& -q(M-a) \sum_{j=1}^{a} \int_{\mathbb{R}^d} \left(V_{\text{int}}(s_j, u_*) - V_{\text{int}}(r_j, u_*) \right) \rho_*^{(a+1)} \, du_*,
\end{aligned}
$$

for $1 \leq a \leq M - 1$, where

$$\rho_*^{(a+1)} = \rho^{(a+1)}(r^{(a)}, u_*, s^{(a)}, u_*, t).$$

Since we assume that V_{int} is of the order of magnitude $1/M$ as $M \to \infty$, we get for fixed $a \geq 1$ and $M \gg 1$, neglecting terms of order $1/M$,

$$
\begin{aligned}
i\hbar \partial_t \rho^{(a)} \;=\; & -\frac{\hbar^2}{2m} \sum_{j=1}^{M} \left(\Delta_{s_j} - \Delta_{r_j} \right) \rho^{(a)} - q \sum_{j=1}^{M} \left(V_{\text{ext}}(s_j, t) - V_{\text{ext}}(r_j, t) \right) \rho^{(a)} \\
& -q \sum_{j=1}^{a} \int_{\mathbb{R}^d} \left(V_{\text{int}}(s_j, u_*) - V_{\text{int}}(r_j, u_*) \right) M \rho_*^{(a+1)} \, du_*.
\end{aligned}
$$

Similarly to the classical case, a particular solution of this equation is given by the so-called Hartree ansatz

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

where R solves the equation

$$i\hbar \partial_t R = -\frac{\hbar^2}{2m}(\Delta_s - \Delta_r)R - q(V_{\text{eff}}(s, t) - V_{\text{eff}}(r, t))R, \quad r, s \in \mathbb{R}^d, \ t > 0, \quad (4.22)$$

with the effective potential

$$V_{\text{eff}}(x, t) = V_{\text{ext}}(s, t) + \int_{\mathbb{R}^d} MR(z, z, t) \, V_{\text{int}}(x, z) \, dz.$$

The kinetic formulation of (4.22) is derived as in Section 4.1. We multiply (4.22) by M, introduce the change of coordinates

$$r = x + \frac{\hbar}{2m}\eta, \quad s = x - \frac{\hbar}{2m}\eta$$

and set $U(x, \eta, t) = MR(r, s, t)$. Then U solves the equation

$$\partial_t U + i\operatorname{div}_\eta \nabla_x U + i\frac{q}{\hbar}\left(V_{\text{eff}}\left(x + \frac{\hbar}{2m}\eta, t\right) - V_{\text{eff}}\left(x - \frac{\hbar}{2m}\eta\right)\right) U = 0.$$

Finally, the inverse Fourier transform $W = (2\pi)^{-d/2}\check{U}$ is a solution of the *quantum Vlasov equation*

$$\partial_t W + v \cdot \nabla_x W + \frac{q}{m}\theta[V_{\text{eff}}]W = 0, \quad x, v \in \mathbb{R}^d, \ t > 0. \quad (4.23)$$

The pseudo-differential operator $\theta[V_{\text{eff}}]$ is defined as in (4.12), and the effective potential is

$$V_{\text{eff}}(x, t) = V_{\text{ext}}(x, t) + \int_{\mathbb{R}^d} n(z, t)V_{\text{int}}(x, z) \, dz, \quad (4.24)$$

where

$$n(x, t) = \int_{\mathbb{R}^d} W(x, v, t) \, dv = U(x, 0, t) = MR(x, x, t)$$

is the quantum electron number density. As the effective potential depends on the Wigner function W, it is a *nonlinear* pseudo-differential equation.

Contrary to the classical Vlasov equation, the quantum Vlasov equation does not preserve the non-negativity of the solution W (cf. the discussion in Section 4.1). However, the number density n remains non-negative for all times, if the initial single-particle density matrix $R(r, s, 0)$ is positive semi-definite.

In the semi-classical limit "$\hbar \to 0$" the quantum Vlasov equation formally converges to the classical Vlasov equation

$$\partial_t W + v \cdot \nabla_x W + \frac{q}{m} \nabla_x V_{\text{eff}} \cdot \nabla_v W = 0.$$

In semiconductor modelling, a usual choice for V_{int} is the Coulomb potential

$$V_{\text{int}}(x,y) = -\frac{q}{4\pi\varepsilon_s} \frac{1}{|x-y|}, \quad x, y \in \mathbb{R}^3, \ x \neq y,$$

where ε_s denotes the permittivity of the semiconductor crystal (see Example 3.1). In Section 3.2 has been shown that the effective potential

$$V_{\text{eff}}(x,t) = V_{\text{ext}}(x,t) - \frac{q}{4\pi\varepsilon_s} \int_{\mathbb{R}^3} \frac{n(z,t)}{|z-x|} \, dz$$

solves the Poisson equation

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

where

$$C(x) = -\frac{\varepsilon_s}{q} V_{\text{ext}}(x)$$

is the doping concentration if V_{ext} is generated by ions of charge $+q$ in the semiconductor material.

The presented quantum Vlasov equation models the motion of an ensemble of many particles in a vacuum taking into account long-range particle interactions. For more realistic models, the following issues should be modeled too:

- The electrons are moving in a crystal and not in a vacuum.

- As the semiconductor device is bounded, the equations have to be solved in a bounded domain with appropriate boundary conditions at the device contacts and insulating surfaces.

- Short-range interactions modelled by scattering events of particles have to be included in the model.

For the first two generalizations, we refer to [34, Ch. 1] (also see (4.17)). The quantum mechanical modelling of collisions of electrons (with phonons, for instance) is a very difficult task. One approach is to formulate the so-called *quantum Boltzmann equation*

$$\partial_t W + v \cdot \nabla_x W + \frac{q}{m} \theta[V_{\text{eff}}]W = Q(W), \quad x, v \in \mathbb{R}^d, \ t > 0,$$

by adding a heuristic collision term to the right-hand side of the quantum Vlasov equation (4.23). In numerical studies, the *relaxation-time model*

$$Q(W) = \frac{1}{\tau}\left(\frac{n}{n_0}W_0 - W\right),$$

or the *Fokker-Planck model*

$$Q(W) = \frac{1}{\tau}\mathrm{div}_v\left(\frac{k_B T_0}{m}\nabla_v W + vW\right),$$

is often used, where τ is the relaxation time,

$$n(x,t) = \int_{\mathbb{R}^d} W(x,v,t)\,dv, \quad n_0(x) = \int_{\mathbb{R}^d} W_0(x,v)\,dv,$$

W_0 is the density of the quantum mechanical thermal equilibrium (see [39, Sec. 2] for its definition), and T_0 is the lattice temperature. In the case of the energy-band quantum Boltzmann equation

$$\partial_t W + v(k) \cdot \nabla_x W + \frac{q}{m}\theta[V_{\mathrm{eff}}]W = Q(W), \tag{4.25}$$

the above collision models read as follows: The relaxation-time collision term is

$$Q(W) = \frac{1}{\tau}\left(\frac{n}{n_0}W_0 - W\right),$$

where now

$$n(x,t) = \int_{\mathbb{R}^d} W(x,k,t)\,dk, \quad n_0(x) = \int_{\mathbb{R}^d} W_0(x,k)\,dk,$$

and the Fokker-Planck terms is given by

$$Q(W) = \frac{1}{\tau}\mathrm{div}_k\left(\frac{mk_B T_0}{\hbar^2}\nabla_k W + kW\right). \tag{4.26}$$

A summary of the kinetic models derived in this and the previous section is presented in Figure 4.1. 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.
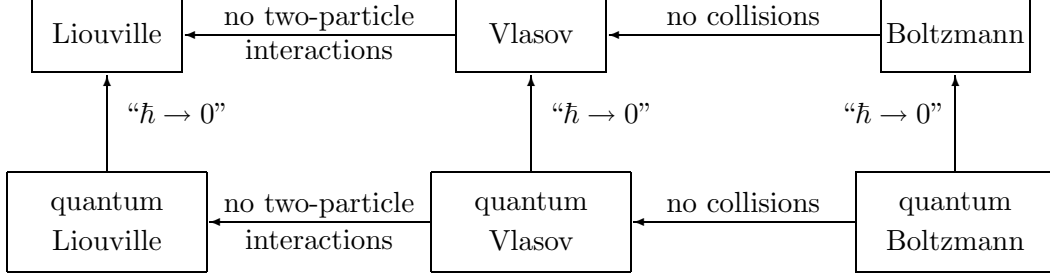
Figure 4.1: Relations between the classical and quantum kinetic equations.

5 From Kinetic to Fluiddynamical Models

5.1 The drift-diffusion equations: first derivation

We derive the drift-diffusion equations (formally) from the bipolar Boltzmann equations

$$\partial_t f_n + v_n(k) \cdot \nabla_x f_n - \frac{q}{\hbar} E \cdot \nabla_k f_n = Q_n(f_n) + I_n(f_n, f_p), \qquad (5.1)$$

$$\partial_t f_p + v_p(k) \cdot \nabla_x f_p + \frac{q}{\hbar} E \cdot \nabla_k f_p = Q_p(f_p) + I_p(f_n, f_p) \qquad (5.2)$$

with the low-density collision operators

$$
\begin{aligned}
(Q_n(f_n))(x, k, t) &= \int_B \phi_n(x, k, k') \\
&\quad \times \left[\exp\left(-\frac{\varepsilon_n(k)}{k_B T} \right) f_n' - \exp\left(-\frac{\varepsilon_n(k')}{k_B T} \right) f_n \right] \, dk', \quad (5.3) \\
(Q_p(f_p))(x, k, t) &= \int_B \phi_p(x, k, k') \\
&\quad \times \left[\exp\left(-\frac{\varepsilon_p(k)}{k_B T} \right) f_p' - \exp\left(-\frac{\varepsilon_p(k')}{k_B T} \right) f_p \right] \, dk', \quad (5.4)
\end{aligned}
$$

and the recombination-generation rates

$$
\begin{aligned}
(I_n(f_n, f_p))(x, k, t) &= -\int_B g(x, k, k') \\
&\quad \times \left[\exp\left(\frac{\varepsilon_n(k) - \varepsilon_p(k')}{k_B T} \right) f_n f_p' - 1 \right] \, dk', \quad (5.5) \\
(I_p(f_n, f_p))(x, k, t) &= -\int_B g(x, k, k') \\
&\quad \times \left[\exp\left(\frac{\varepsilon_n(k') - \varepsilon_p(k)}{k_B T} \right) f_n' f_p - 1 \right] \, dk'. \quad (5.6)
\end{aligned}
$$

62

Here, f'_j means evaluation at k', i.e. $f'_j = f_j(x, k', t)$, $j = n, p$, and B denotes the Brillouin zone. We assume the parabolic band approximation

$$\varepsilon_n(k) = \varepsilon_c + \frac{\hbar^2}{2m_n}|k|^2, \quad \varepsilon_p(k) = \varepsilon_v - \frac{\hbar^2}{2m_p}|k|^2,$$

where ε_c denotes the conduction band minimum, ε_v the valence band maximum, and m_n and m_p are the effective masses of the electrons and holes, respectively. Then the velocities are given by

$$
\begin{aligned}
v_n(k) &= \frac{1}{\hbar}\nabla_k \varepsilon_n(k) = \frac{\hbar k}{m_n}, \\
v_p(k) &= -\frac{1}{\hbar}\nabla_k \varepsilon_p(k) = \frac{\hbar k}{m_p}.
\end{aligned}
$$

In particular, we can replace B by $\mathbb{R}^d$ in the above integrals.

The drift-diffusion equations are derived under the assumption that collisions occur on a much shorter time scale than recombination-generation events. In order to make this statement more precise, we scale the Boltzmann equations (5.1) and (5.2). For this, we make the following assumptions:

(1) The effective masses of the electrons and holes are of the same order such that we can define the reference velocity $\bar{v} = \sqrt{k_B T/m_n}$. This means that we assume that the thermal energy $k_B T$ is of the same order as the kinetic energy $m_n \bar{v}^2/2$.

(2) Collisions occur on a much shorter time scale than recombination-generation events. Thus, scaling

$$Q_n = \frac{1}{\tau_c}Q_{ns}, \qquad I_n = \frac{1}{\tau_R}I_{ns}, \tag{5.7}$$

$$Q_p = \frac{1}{\tau_c}Q_{ps}, \qquad I_p = \frac{1}{\tau_R}I_{ps} \tag{5.8}$$

with the collision relaxation time τ_c and the recombination-generation relaxation time τ_R, we assume that

$$\tau_c \ll \tau_R.$$

This is our main assumption.

(3) The reference length ι_0 is given by the geometric average of the mean free paths $\iota_c := \tau_c \bar{v}$ and $\iota_R := \tau_R \bar{v}$:

$$\iota_0 = \sqrt{\iota_R \iota_c}.$$

The *mean free path* is the average time between two successive collisions.

63

(4) We use the reference time τ_R, the reference wave vector $m_n\bar{v}/\hbar$, and the reference field strength $k_BT/q\iota_0$:

$$t = \tau_R t_s, \quad k = \frac{m_n\bar{v}}{\hbar}k_s, \quad E = \frac{k_BT}{\iota_0 q}E_s. \tag{5.9}$$

By assumption (2), the parameter $\alpha^2 := \iota_c/\iota_R = \tau_c/\tau_R$ satisfies $\alpha^2 \ll 1$. With the scaling (5.7)-(5.9) we can rewrite (5.1) (omitting the index s):

$$\frac{1}{\tau_R}\partial_t f_n + \frac{\bar{v}}{\iota_0}k \cdot \nabla_x f_n - \frac{k_BT}{\iota_0 m_n \bar{v}}E \cdot \nabla_k f_n = \frac{1}{\tau_c}Q_n(f_n) + \frac{1}{\tau_R}I_n(f_n, f_p).$$

Multiplying this equation by $\tau_c = \iota_c/\bar{v}$ and using $\alpha = \iota_c/\iota_0$ and $k_BT = m_n\bar{v}^2$, we obtain

$$\alpha^2\partial_t f_n + \alpha(k \cdot \nabla_x f_n - E \cdot \nabla_k f_n) = Q_n(f_n) + \alpha^2 I_n(f_n, f_p). \tag{5.10}$$

In a similar way, we have

$$\alpha^2\partial_t f_p + \alpha\left(\frac{m_n}{m_p}k \cdot \nabla_x f_p - E \cdot \nabla_k f_p\right) = Q_p(f_p) + \alpha^2 I_p(f_n, f_p). \tag{5.11}$$

The scaled collision and recombination-generation terms have the same form as (5.3)-(5.6) but the exponential terms

$$\exp\left(-\frac{\varepsilon_n(k)}{k_BT}\right) \quad \text{and} \quad \exp\left(-\frac{\varepsilon_p(k)}{k_BT}\right)$$

are replaced by

$$\exp\left(-\frac{\varepsilon_c}{k_BT} - \frac{|k|^2}{2}\right) \quad \text{and} \quad \exp\left(-\frac{\varepsilon_v}{k_BT} + \frac{m_n}{m_p}\frac{|k|^2}{2}\right)$$

and the rates ϕ_n, ϕ_p and g are multiplied by $(m_n\bar{v}/\hbar)^d$.

We want to study the scaled Boltzmann equations for "small" α. First we analyze the collision operators Q_n and Q_p. We rewrite them in the form

$$Q_j(f_j) = \int_{\mathbb{R}^d} \Phi_j(x, k, k') \left[M_j(k)f_j' - M_j(k')f_j\right] dk', \quad j = n, p, \tag{5.12}$$

where

$$\Phi_j(x, k, k') = N_j\phi_n(x, k, k'), \quad j = n, p,$$

and

$$M_n(k) = \frac{1}{N_n}\exp\left(-\frac{|k|^2}{2}\right), \quad M_p(k) = \frac{1}{N_p}\exp\left(-\frac{m_n}{m_p}\frac{|k|^2}{2}\right)$$

64

are the scaled Maxwellians. The constants

$$N_n = (2\pi)^{d/2}, \quad N_p = (2\pi m_p/m_n)^{d/2}$$

are chosen such that the integrals of the Maxwellians over $k \in \mathbb{R}^d$ are equal to one.

For the following analysis define the functions

$$\lambda_j(k) = \int_{\mathbb{R}^d} \Phi_j(x, k, k') M_j(k') \, dk', \quad k \in \mathbb{R}^d, \ j = n, p,$$

for fixed $x \in \mathbb{R}^d$, and the Banach spaces

$$
\begin{aligned}
X_j &= \{f : \mathbb{R}^d \to \mathbb{R} \text{ measurable: } \|f\|_{X_j} < \infty\}, \\
Y_j &= \{f : \mathbb{R}^d \to \mathbb{R} \text{ measurable: } \|f\|_{Y_j} < \infty\}
\end{aligned}
$$

with associated norms

$$
\begin{aligned}
\|f\|_{X_j} &= \left(\int_{\mathbb{R}^d} f(k)^2 \lambda_j(k) M_j(k)^{-1} \, dk \right)^{1/2}, \\
\|f\|_{Y_j} &= \left(\int_{\mathbb{R}^d} f(k)^2 (\lambda_j(k) M_j(k))^{-1} \, dk \right)^{1/2}.
\end{aligned}
$$

Lemma 5.1 *Let $j = n$ or $j = p$, let Φ_n, $\Phi_p > 0$ be symmetric in k and k' and assume that for all isometric matrices A it holds*

$$\Phi(x, Ak, Ak') = \Phi(x, k, k') \quad \forall x, k, k' \in \mathbb{R}^d.$$

Then

(1) *The kernel $N(Q_j) = \{f \in X : Q_j(f) = 0\}$ is given by*

$$N(Q_j) = \{\sigma M : \sigma = \sigma(x) \in \mathbb{R}\}.$$

(2) *The equation $Q_j(f) = g$ with $g \in Y_j$ has a solution $f \in X_j$ if and only if*

$$\int_{\mathbb{R}^d} g(k) \, dk = 0.$$

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.

(3) *The solution $f \in X_j$ of $Q_j(f_j) = g$ is unique if the orthogonality relation*

$$\int_{\mathbb{R}^d} f(k) \lambda_j(k) dk = 0 \tag{5.13}$$

holds.

(4) The equations

$$Q_n(h_{ni}) = k_i M_n(k), \quad Q_p(h_{pi}) = \frac{m_n}{m_p} k_i M_p(k), \quad i = 1, \dots, d,$$

have solutions $h_n(x, k) = (h_{ni}(x, k))_{i=1}^d$ and $h_p(x, k) = (h_{pi}(x, k))_{i=1}^d$ with the property that there exist $\mu_n(x)$, $\mu_p(x) \geq 0$ such that

$$\int_{\mathbb{R}^d} k \otimes h_n \, dk = -\mu_n(x) \, \text{Id}, \tag{5.14}$$

$$\int_{\mathbb{R}^d} \frac{m_n}{m_p} k \otimes h_p \, dk = -\mu_p(x) \, \text{Id}, \tag{5.15}$$

where Id is the unit matrix in $\mathbb{R}^{d \times d}$ and $a \otimes b = a^\top b$ for $a, b \in \mathbb{R}^d$.

In the statement of the lemma we have omitted some technical assumptions on Φ_n and Φ_p (regularity conditions) which are particularly needed in the proof of Lemma 5.1(4). We only sketch the proof of part (3) since properties of so-called Hilbert-Schmidt operators are needed, and refer to [37] for a complete proof.

Proof. In the following we omit the index j. We proceed as in [37, Prop. 1].

(1) 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

$$
\begin{aligned}
Q_s(f_s) &= (\lambda M)^{-1/2} \left(M \int_{\mathbb{R}^d} \Phi(x, k, k') f' \, dk' - \lambda f \right) \\
&= \int_{\mathbb{R}^d} \Phi(x, k, k') \left(\frac{MM'}{\lambda \lambda'} \right)^{1/2} f_s' \, dk' - f_s,
\end{aligned}
$$

where $M' = M(k')$ and $\lambda' = \lambda(k')$. Since Φ is symmetric in k and k' by assumption, the operator $Q_s : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is self-adjoint.

Next we analyze the kernel of Q_s. By definition of λ we have

$$\int_{\mathbb{R}^d} \Phi(x, k, k') M(k') / \lambda(k) \, dk' = 1$$

and therefore

$$
\begin{aligned}
0 \;\leq\; & \frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Phi(x,k,k') M M' \left(\frac{f_s}{\sqrt{\lambda M}} - \frac{f'_s}{\sqrt{\lambda' M'}} \right)^2 dk\, dk' \\
= \;& \frac{1}{2} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \Phi(x,k,k') \frac{M'}{\lambda} \, dk' \right) f_s^2 \, dk \\
& + \frac{1}{2} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \Phi(x,k,k') \frac{M}{\lambda'} \, dk \right) (f'_s)^2 \, dk' \\
& - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Phi(x,k,k') \left(\frac{M M'}{\lambda \lambda'} \right)^{1/2} f'_s f_s \, dk\, dk' \\
= \;& \int_{\mathbb{R}^d} \left(f_s^2 - \int_{\mathbb{R}^d} \Phi(x,k,k') \left(\frac{M M'}{\lambda \lambda'} \right)^{1/2} f'_s f_s \, dk' \right) dk \\
= \;& - \int_{\mathbb{R}^d} Q_s(f_s) f_s \, dk.
\end{aligned}
$$

Hence $Q_s(f_s) = 0$ implies

$$
\frac{f_s}{\sqrt{\lambda M}} - \frac{f'_s}{\sqrt{\lambda' M'}} = 0 \quad \text{for } k, k' \in \mathbb{R}^d
$$

and $f_s / \sqrt{\lambda M} = \sigma = \text{const.}$, where $\sigma = \sigma(x)$ is a parameter. In original variables, $Q(f) = 0$ implies $f = \sigma M$ for all $\sigma \in \mathbb{R}$. Conversely, if $f = \sigma M$ for some $\sigma \in \mathbb{R}$ then the formulation (5.12) of $Q(f)$ immediately yields $Q(f) = 0$. Thus $N(Q) = \{f \in X : \; Q(f) = 0\} = \{\sigma M : \; \sigma \in \mathbb{R}\}$.

(2) It is not difficult to see that the operator $Q_s : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is linear, continuous and has closed range $R(Q_s)$ where

$$
R(Q_s) = \{g_s \in L^2(\mathbb{R}^d) : \; \exists f_s \in L^2(\mathbb{R}^d) : \; Q_s(f_s) = g_s\}.
$$

From functional analysis follows that $R(Q_s) = N(Q_s^*)^\perp$, where Q_s^* is the adjoint of Q_s and $N(Q_s^*)$ is the kernel of Q_s^*. In fact, since Q_s is selfadjoint, $Q_s^* = Q_s$. We obtain

$$
\begin{aligned}
Q_s(f_s) = g_s \text{ has a solution} \quad &\Leftrightarrow \quad g_s \in R(Q_s) \\
&\Leftrightarrow \quad g_s \in N(Q_s^*)^\perp = N(Q_s)^\perp \\
&\Leftrightarrow \quad \forall h \in L^2(\mathbb{R}^d), \; Q_s(h) = 0 : \int_{\mathbb{R}^d} g_s h \, dk = 0.
\end{aligned}
$$

The equivalence between the solvability of $Q_s(f_s) = g$ and $g \in N(Q_s^*)^\perp$ is also known as the *Fredholm alternative* [44, Appendix, (39)]. As the kernel of Q_s is spanned by $\sqrt{\lambda M}$, $Q_s(f_s) = g_s$ has a solution if and only if

$$
\int_{\mathbb{R}^d} g_s \sqrt{\lambda M} \, dk = 0
$$

or, in the original variables,

$$0 = \int_{\mathbb{R}^d} g_s \sqrt{\lambda M} \, dk = \int_{\mathbb{R}^d} g \, dk.$$

(3) We only give a sketch of the proof and refer to [37] for details. It is possible to show that the operator $-Q$ is coercive in the following sense:

$$\int_{\mathbb{R}^d} (-Q(f))(k)f(k)M(k)^{-1} \, dk \geq c\|f\|_X^2 \tag{5.16}$$

for all $f \in X$ satisfying (5.13). Clearly, this implies that Q is one-to-one on the subset of functions satisfying (5.13), and the uniqueness is proved. In order to show the coerciveness property prove that $\mathrm{Id} + Q$ is a Hilbert-Schmidt operator and use general properties of those operators (see [37]).

(4) The existence of solutions h_i of $(Q(h_i))(k) = k_i M(k)$ follows from part (2) since

$$\int_{\mathbb{R}^d} k_i M_j(k) \, dk = 0 \quad \text{for } i = 1, \ldots, d, \ j = n, p.$$

We only show (5.14) since the proof of (5.15) is similar.

Let A be the matrix of a rotation with axis k_1. Then $(Ak)_1 = k_1$ for all $k \in \mathbb{R}^d$ and

$$(Ak)_1 M(Ak) = k_1 N^{-1} \exp(-|Ak|^2/2) = k_1 N^{-1} \exp(-|k|^2/2) = k_1 M(k), \tag{5.17}$$

since A is isometric. We have for all $k \in \mathbb{R}^d$, using the assumption $\Phi(x, Ak, Ak') = \Phi(x, k, k')$ and (5.17),

$$
\begin{aligned}
(Q(h_1 \circ A))(k) &= \int_{\mathbb{R}^d} \Phi(x, k, k')[M(k)h_1(Ak') - M(k')h_1(Ak)] \, dk' \\
&= \int_{\mathbb{R}^d} \Phi(x, Ak, Ak')[M(Ak)h_1(Ak') - M(Ak')h_1(Ak)] \, dk' \\
&= \int_{\mathbb{R}^d} \Phi(x, Ak, w)[M(Ak)h_1(w) - M(w)h_1(Ak)] \, dw \\
&= (Q(h_1))(Ak) \\
&= (Ak)_1 M(Ak) \\
&= k_1 M(k) \\
&= (Q(h_1))(k)
\end{aligned}
$$

and thus $Q(h_1 \circ A - h_1) = 0$. Another computation leads to

$$
\begin{aligned}
\int_{\mathbb{R}^d} h_1(Ak)\lambda(k) \ dk &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Phi(x, k, k') h_1(Ak) M(k') \ dk' \ dk \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Phi(x, Ak, Ak') h_1(Ak) M(Ak') \ dk' \ dk \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Phi(x, v, w) h_1(v) M(w) \ dw \ dv \\
&= \int_{\mathbb{R}^d} h_1(v)\lambda(v) \ dv
\end{aligned}
$$

or

$$
\int_{\mathbb{R}^d} (h_1 \circ A - h_1)(k)\lambda(k) \ dk = 0.
$$

This is the orthogonality condition (5.13) which ensures the uniqueness of the solution of the equation $Q(h_1 \circ A - h_1) = 0$. Therefore, $h_1 \circ A - h_1 = 0$. We conclude that h_1 remains invariant under a rotation with axis k_1, and we can write h_1 as

$$
h_1(k) = H(k_1, |k|^2 - k_1^2).
$$

Now let A be the isometric matrix of the mapping $k \mapsto (-k_1, k_2, \ldots, k_d)$. Since $k \mapsto k_1 M(k)$ is an odd function, a similar computation as above yields $Q(h_1 \circ A) = -Q(h_1)$ and

$$
\int_{\mathbb{R}^d} (h_1 \circ A + h_1)\lambda(k) \ dk = 0.
$$

This implies as above that $h_1 \circ A + h_1 = 0$ and thus, h_1 is an odd function with respect to k_1.

In a similar way, we can show that

$$
h_i(k) = H_i(k_i, |k|^2 - k_i^2), \quad i = 2, \ldots, d,
$$

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

$$
Q(H(k_1, k_2^2 + \cdots + k_d^2)) = k_1 M(k_1^2 + \cdots + k_d^2)
$$

(with slight abuse of notation) leads to

$$
Q(H(k_2, k_1^2 + k_3^2 + \cdots + k_d^2)) = k_2 M(k_1^2 + \cdots + k_d^2) = Q(H_2(k_2, |k|^2 - k_2^2))
$$

and hence, by a similar argument as above, $H = H_2$.

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 $i \neq j$

$$
\int_{\mathbb{R}^d} k_i h_j(k) dk = \int_{\mathbb{R}^d} k_i H(k_j, |k|^2 - k_j^2) dk = 0.
$$

Furthermore

$$\int_{\mathbb{R}^d} k_i h_i(k) dk = \int_{\mathbb{R}^d} k_i H(k_i, |k|^2 - k_i^2) dk$$

$$= \int_{\mathbb{R}^d} k_j H(k_j, |k|^2 - k_j^2) dk$$

$$= \int_{\mathbb{R}^d} k_j h_j(k) dk,$$

for all i and j. Thus, the above integral is independent of i and we can set

$$\mu := - \int_{\mathbb{R}^d} k_1 h_1(k) dk.$$

The constant μ depends on the parameter x since h_1 depends on x through Q. Moreover, by (5.16),

$$\mu(x) = - \int_{\mathbb{R}^d} (Q(h_1))(k) h_1(k) M(k)^{-1} \, dk \geq 0,$$

and this implies

$$\int_{\mathbb{R}^d} k \otimes h(k) \, dk = -\mu(x) \, \mathrm{Id}.$$

The theorem is proved. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Setting $\alpha = 0$ in the scaled Boltzmann equations (5.10)-(5.11) gives

$$Q_n(f_n) = 0, \quad Q_p(f_p) = 0.$$

By Lemma 5.1(1) these equations possess the solutions

$$f_{n0} = n(x, t) M_n, \quad f_{p0} = p(x, t) M_p, \qquad\qquad (5.18)$$

respectively, where $n(x, t)$ and $p(x, t)$ are some parameters. Since

$$\int_{R^d} f_{n0} \, dk = n(x, t), \quad \int_{\mathbb{R}^d} f_{p0} \, dk = p(x, t),$$

we interpret n and p as the scaled number densities of electrons and holes, respectively. In order to obtain more informations from (5.10)-(5.11) we use the *Hilbert expansion method.* The idea is to expand f_n and f_p in terms of powers of α,

$$f_n = f_{n0} + \alpha f_{n1} + \alpha^2 f_{n2} + \cdots,$$
$$f_p = f_{p0} + \alpha f_{p1} + \alpha^2 f_{p2} + \cdots,$$

and to derive equations for f_{nj} and f_{pj}. Substituting this ansatz into (5.10)-(5.11) and equating coefficients of equal powers of α yields (notice that Q_n and Q_p are linear)

- for the terms of order 1:

$$Q_n(f_{n0}) = 0, \quad Q_p(f_{p0}) = 0; \tag{5.19}$$

- for the terms of order α:

$$k \cdot \nabla_x f_{n0} - E \cdot \nabla_k f_{n0} = Q_n(f_{n1}), \tag{5.20}$$
$$k \cdot \nabla_x f_{p0} + E \cdot \nabla_k f_{p0} = Q_n(f_{p1}); \tag{5.21}$$

- and for the terms of order α^2:

$$\partial_t f_{n0} + k \cdot \nabla_x f_{n1} - E \cdot \nabla_k f_{n1} = Q_n(f_{n2}) + I_n(f_{n0}, f_{p0}), \tag{5.22}$$
$$\partial_t f_{p0} + k \cdot \nabla_x f_{p1} + E \cdot \nabla_k f_{p1} = Q_p(f_{p2}) + I_p(f_{n0}, f_{p0}). \tag{5.23}$$

We already solved (5.19). By (5.18) and $\nabla_k M_j(k) = -k M_j(k)$ $(j = n, p)$ we can write (5.20)-(5.21) as

$$Q_n(f_{n1}) = M_n k \cdot (\nabla_x n + nE),$$
$$Q_p(f_{p1}) = \frac{m_n}{m_p} M_p k \cdot (\nabla_x p - nE).$$

Lemma 5.1(2) shows that these equations have solutions and that any solution can be expressed as

$$f_{n1} = (\nabla_x n + nE) \cdot h_n + \sigma_n M_n,$$
$$f_{p1} = (\nabla_x p - pE) \cdot h_p + \sigma_p M_p$$

for some unspecified parameters $\sigma_n(x, t)$, $\sigma_p(x, t)$. It is convenient to define

$$J_n(x, t) = \mu_n(\nabla_x n + nE), \quad J_p(x, t) = -\mu_p(\nabla_x p - pE), \tag{5.24}$$

where μ_n μ_p are introduced in Lemma 5.1(4). We will see below that J_n and J_p can be interpreted as scaled current densities.

The equation (5.22) is solvable, by Lemma 5.1(2), if and only if

$$0 = \int_{\mathbb{R}^d} (\partial_t f_{n0} + k \cdot \nabla_x f_{n1} - \operatorname{div}_k(E f_{n1}) - I_n(f_{n0}, f_{p0})) \, dk$$
$$= \partial_t n + \int_{\mathbb{R}^d} k \cdot \nabla_x f_{n1} \, dk - \int_{\mathbb{R}^d} I_n(f_{n0}, f_{p0}) \, dk.$$

From (5.14) follows that

$$\int_{\mathbb{R}^d} k \cdot \nabla_x f_{n1} \, dk = \int_{\mathbb{R}^d} k \cdot \nabla_x(\mu_n^{-1} J_n \cdot h_n + \sigma_n M_n) \, dk$$
$$= \sum_{i,j=1}^d \frac{\partial}{\partial x_i} \left(\mu_n^{-1} J_{nj} \int_{\mathbb{R}^d} k_i h_{nj} \, dk \right) + \nabla_x \sigma_n \cdot \int_{\mathbb{R}^d} k M_n(k) \, dk$$
$$= -\sum_{i,j=1}^d \frac{\partial}{\partial x_i} J_{nj} \delta_{ij}$$
$$= -\operatorname{div}_x J_n.$$

Furthermore

$$
\begin{aligned}
R & := -\int_{\mathbb{R}^d} I_n(f_{n0}, f_{p0}) \\
& = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} g(x, k, k') \Big[\exp\Big(\frac{\varepsilon_c - \varepsilon_v}{k_B T}\Big) \exp\Big(\frac{|k|^2}{2} + \frac{m_n}{m_p}\frac{|k'|^2}{2}\Big) \\
& \qquad \times np M_n(k) M_p(k') - 1 \Big] \, dk' \, dk \\
& = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} g(x, k, k') \Big[\exp\Big(\frac{\varepsilon_c - \varepsilon_v}{k_B T}\Big) \frac{np}{N_n N_p} - 1 \Big] \, dk' \, dk \\
& = A(x)(np - n_i^2),
\end{aligned}
$$

where

$$
A(x) = \frac{1}{n_i^2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} g(x, k, k') \, dk \, dk'
$$

and

$$
n_i = \sqrt{N_n N_p} \exp\Big(\frac{\varepsilon_v - \varepsilon_c}{2 k_B T}\Big)
$$

is the scaled *intrinsic density.* We conclude that (5.22) and (by a similar computation) (5.23) are solvable if and only if

$$
\partial_t n - \operatorname{div}_x J_n = -R, \quad \partial_t p + \operatorname{div}_x J_p = -R. \tag{5.25}
$$

Here we can see that the quantities J_n and J_p can be indeed interpreted as current densities. In order to scale back to the physical variables we notice that the scaled number densities, now called n_s and p_s, are obtained from

$$
\begin{aligned}
n_s & = \int_{\mathbb{R}^d} f_{n0} \, dk_s = \Big(\frac{\hbar}{m_n \bar{v}}\Big)^d \int_{\mathbb{R}^d} f_{n0} \, dk = \frac{\hbar^d n}{(k_B T m_n)^{d/2}}, \\
p_s & = \int_{\mathbb{R}^d} f_{p0} \, dk_s = \frac{\hbar^d p}{(k_B T m_n)^{d/2}},
\end{aligned}
$$

where n and p are now the unscaled variables. Thus, using

$$
t_s = \frac{t}{\tau_R}, \quad E_s = \frac{\iota_0}{U_T} E, \quad x_s = \frac{x}{\iota_0}, \quad \mu_{ns} = \frac{m_n}{\tau_c q} \mu_n,
$$

we get from (5.24)-(5.25) after some computations:

$$
\partial_t n - \operatorname{div}_x(\mu_n U_T \nabla_x n + \mu_n n E) = -A(np - n_i^2),
$$

where $U_T = k_B T / q$,

$$
\frac{\hbar^d}{(k_B T m_n)^{d/2}} A_s = A \quad \text{and} \quad \frac{(k_B T m_n)^{d/2}}{\hbar^d} n_{is}^2 = n_i^2.
$$

In particular, the unscaled intrinsic density reads (cf. (2.20))

$$n_i = \left(\frac{2\pi k_B T \sqrt{m_n m_p}}{\hbar^2} \right)^d \exp \left(\frac{\varepsilon_v - \varepsilon_c}{2 k_B T} \right).$$

The unscaled current density and the evolution equation for n are given by

$$\partial_t n - \frac{1}{q} \operatorname{div}_x J_n = -R, \quad J_n = q\mu_n(U_T \nabla_x n + nE). \qquad (5.26)$$

Similarly, we can compute,

$$\partial_t p + \frac{1}{q} \operatorname{div}_x J_p = -R, \quad J_p = -q\mu_p(U_T \nabla_x p - pE). \qquad (5.27)$$

The equations (5.26)-(5.27) are referred to as the *drift-diffusion equations* for given electric field. For a selfconsistent treatment of the electric field, (5.26)-(5.27) have to be supplemented by the Poisson equation

$$-\operatorname{div}(\varepsilon_s E) = q(n - p - C),$$

where ε_s is the semiconductor permittivity and C the doping concentration.

Mathematically, (5.26)-(5.27) are parabolic convection-diffusion equations with diffusion coefficients

$$D_n := \mu_n U_T, \quad D_p := \mu_p U_T$$

and *mobilities* μ_n, μ_p. The quotient of diffusivity and mobility is constant, and the equations

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = U_T$$

are called *Einstein relations*.

Usually, the drift-diffusion equations are considered in a bounded domain such that appropriate boundary conditions have to be prescribed. We refer to Chapter 6 for the choice of the boundary and initial conditions which complete (5.26)-(5.27).

The derivation of the drift-diffusion model is mainly based on the following hypotheses:

- The free mean path ι_c between two consecutive scattering events is much smaller than the free mean path ι_R between two recombination-generation events (typically, $\iota_c \sim 10^{-7}$ m, $\iota_R \sim 10^{-4}$ m).

- The device diameter is of the order of $\iota_0 = \sqrt{\iota_R \iota_c}$ (typically, $\iota_0 \sim 10^{-5}$ m).

- The electrostatic potential is of the order of $U_T = 0.026$ V (at $T = 300$ K).

Thus, the drift-diffusion model is appropriate for semiconductor devices with characteristic lengths not smaller than $1 \ldots 10\,\mu$m and applied voltages much smaller than 1 V. However, in application this model is used also for higher applied voltages. It gives reasonable results as long as the characteristic length is not much smaller than $1\,\mu$m.

5.2 The drift-diffusion equations: second derivation

Using a relaxation approximation of the collision operator in the Boltzmann equationm we can give a much shorter derivation of the drift-diffusion model than presented in Section 5.1. The derivation combines the Hilbert expansion method and the moment method. For simplicity, we start with the unipolar classical Boltzmann equation in the diffusion scale (5.10) neglecting recombination-generation effects:

$$\alpha^2 \partial_t f + \alpha \left(v \cdot \nabla_x f - \frac{q}{m} E \cdot \nabla_v f \right) = Q(f), \quad x, v \in \mathbb{R}^d, \ t > 0, \qquad (5.28)$$

where $E = E(x,t)$ is the electric field and $\alpha > 0$ is a parameter which is small compared to one. We assume a low-density collision operator $Q(f)$ with a scattering rate independent of v and v' (see (5.12)):

$$(Q(f))(x,v,t) = \int_{\mathbb{R}^d} \phi(x) \left[M(v) f(x,v',t) - M(v') f(x,v,t) \right] \, dv', \qquad (5.29)$$

where

$$M(v) = \left(\frac{m}{2\pi k_B T} \right)^{d/2} \exp\left(-\frac{m|v|^2}{2k_B T} \right)$$

is the Maxwellian. We introduce the average of f in the velocity space by

$$[f](x,t) = \int_{\mathbb{R}^d} f(x,v,t) \, dv.$$

Then the collision operator (5.29) can be written as

$$\begin{aligned}
(Q(f))(x,v,t) &= \phi(x)(M(v)[f] - f(v)) \\
&= \frac{M(v)[f] - f(v)}{\tau(x)}, \qquad (5.30)
\end{aligned}$$

where $\tau(x) := 1/\phi(x)$ is called *relaxation time,* and the operator (5.30) is termed *relaxation-time operator* (also see (3.35)).

We use the Hilbert expansion

$$f = f_0 + \alpha f_1 + \alpha^2 f_2 + \cdots$$

in the Boltzmann equation (5.28) and identify terms of the same order of α:

- terms of order α^0:
$$M(v)[f_0] - f_0 = 0, \qquad (5.31)$$

- terms of order α^1:
$$v \cdot \nabla_x f_0 - \frac{q}{m} E \cdot \nabla_v f_0 = \frac{M(v)[f_1] - f_1}{\tau(x)}, \qquad (5.32)$$

- terms of order α^2:

$$\partial_t f_0 + v \cdot \nabla_x f_1 - \frac{q}{m} E \cdot \nabla_v f_1 = \frac{M(v)[f_2] - f_2}{\tau(x)}. \qquad (5.33)$$

Equation (5.31) implies

$$f_0(x, v, t) = M(v)[f_0](x, t) = M(v)n(x, t),$$

where $n := [f_0]$ is the particle density. Multiplying (5.32) by $-qv$ and integrating over $v \in \mathbb{R}^d$ implies

$$
\begin{aligned}
J &:= -q \int_{\mathbb{R}^d} v f_1 \, dv \\
&= q\tau(x) \int_{\mathbb{R}^d} \left(v \cdot \nabla_x f_0 - \frac{q}{m} E \cdot \nabla_v f_0 \right) v \, dv + q[f_1] \int_{\mathbb{R}^d} M(v) v \, dv \\
&= q\tau(x) \left(\int_{\mathbb{R}^d} v \otimes v M(v) \, dv \cdot \nabla_x n - \frac{q}{m} \int_{\mathbb{R}^d} v \otimes \nabla_v M(v) \, dv \cdot En \right).
\end{aligned}
$$

Since

$$
\begin{aligned}
\int_{\mathbb{R}^d} v_i v_j M(v) \, dv &= 0 \quad \text{for } i \neq j, \\
\int_{\mathbb{R}^d} v_i^2 M(v) \, dv &= \frac{k_B T}{(2\pi)^{1/2} m} \int_{\mathbb{R}} z^2 e^{-z^2/2} \, dz = \frac{k_B T}{m}, \\
\int_{\mathbb{R}^d} v \otimes \nabla_v M(v) \, dv &= -\frac{m}{k_B T} \int_{\mathbb{R}^d} v \otimes v M(v) \, dv = -\mathrm{Id},
\end{aligned}
$$

we have

$$J = \frac{q^2 \tau(x)}{m} \left(\frac{k_B T}{q} \nabla_x n + nE \right) = q\mu_n (U_T \nabla_x n + nE),$$

where $\mu_n = q\tau/m$ is the electron mobility and $U_T = k_B T/q$ the thermal voltage. This shows that $J(x, t)$ can be interpreted as the electron current density. Finally, we integrate (5.33) over $v \in \mathbb{R}^d$:

$$\partial_t n + \int_{\mathbb{R}^d} v \cdot \nabla_x f_1 \, dv - \frac{q}{m} E \cdot \int_{\mathbb{R}^d} \nabla_v f_1 \, dv = \frac{1}{\tau(x)} \int_{\mathbb{R}^d} (M(v)[f_2] - f_2) \, dv.$$

The third term of the left-hand side vanishes, by the divergence theorem. The integral of the right-hand side also vanishes. Finally, the second term on the left-hand side equals $-(1/q)\mathrm{div}\, J$. Thus we have derived the *drift-diffusion equations*

$$
\begin{aligned}
\partial_t n - \frac{1}{q} \mathrm{div}\, J &= 0 \\
J &= q\mu_n (U_T \nabla_x n + nE).
\end{aligned}
$$

5.3 The hydrodynamic equations

We derive the hydrodynamic model from the Boltzmann equation by employing the so-called *moment method*. The idea of this method is to multiply the Boltzmann equation by powers of the velocity components, to integrate over the velocity space and to derive evolution equations for the integrals. We consider the classical Boltzmann equation for one type of charge carriers (say, electrons):

$$\partial_t f + v \cdot \nabla_x f - \frac{q}{m_n} E \cdot \nabla_v f = Q(f) \tag{5.34}$$

with the low-density collision operator

$$Q(f) = \int_{\mathbb{R}^d} \phi(x, v, v')(Mf' - M'f) \, dv',$$

where the Maxwellian is given by

$$M(v) = \left(\frac{m_n}{2\pi k_B T_0} \right)^{d/2} \exp \left(-\frac{m_n |v|^2}{2 k_B T_0} \right) \tag{5.35}$$

and f' and M' means evaluation at v'. We assume that the scattering rate ϕ is symmetric in v and v'. The first moments of the distribution function f are defined by

$$
\begin{aligned}
\langle 1 \rangle(x, t) &= \int_{\mathbb{R}^d} f(x, v, t) \, dv, \\
\langle v_j \rangle(x, t) &= \int_{\mathbb{R}^d} v_j \, f(x, v, t) \, dv, \\
\langle v_i v_j \rangle(x, t) &= \int_{\mathbb{R}^d} v_i v_j \, f(x, v, t) \, dv, \quad i, j = 1, \dots, d,
\end{aligned}
$$

and so on. We write

$$\langle v \rangle = (\langle v_j \rangle)_j, \quad \langle v^2 \rangle = \langle v \otimes v \rangle = (\langle v_i v_j \rangle)_{ij},$$

$$\langle v^3 \rangle = (\langle v_i v_j v_k \rangle)_{ijk}.$$

Multiply (5.34) by powers of v_j and integrate over $v \in \mathbb{R}^d$. This leads to the moment equations

$$
\begin{aligned}
\partial_t \langle 1 \rangle + \operatorname{div}_x \langle v \rangle &= \int_{\mathbb{R}^d} Q(f) \, dv, \\
\partial_t \langle v \rangle + \operatorname{div}_x \langle v^2 \rangle + \frac{q}{m_n} \langle 1 \rangle E &= \int_{\mathbb{R}^d} v \, Q(f) \, dv, \\
\partial_t \langle v^2 \rangle + \operatorname{div}_x \langle v^3 \rangle + \frac{2q}{m_n} \langle v \rangle \otimes E &= \int_{\mathbb{R}^d} v \otimes v \, Q(f) \, dv,
\end{aligned}
$$

and so on. The symmetry of ϕ implies

$$
\begin{aligned}
\int_{\mathbb{R}^d} Q(f)\, dv &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \phi(x,v,v')M(v)\, dv \right) f'\, dv' \\
&\quad - \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \phi(x,v,v')M(v')\, dv' \right) f\, dv \\
&= 0.
\end{aligned}
$$

The moments are related to physical quantities. In fact, we *define* the number density n, the current density J, the energy tensor $\mathcal{E}$, and the energy e by

$$
n = \langle 1 \rangle, \quad J = -q\langle v \rangle,
$$

$$
\mathcal{E} = \frac{m_n}{2}\frac{1}{n}\langle v \otimes v \rangle, \quad e = \frac{m_n}{2}\frac{1}{n}\sum_{j=1}^d \langle v_j v_j \rangle = \frac{m_n}{2}\frac{1}{n}\langle |v|^2 \rangle.
$$

With these definitions we obtain the conservation laws of mass, momentum and energy:

$$
\partial_t n - \frac{1}{q}\operatorname{div}_x J = 0, \tag{5.36}
$$

$$
\partial_t J - \frac{2q}{m_n}\operatorname{div}_x(n\mathcal{E}) - \frac{q^2}{m_n}nE = -q\int_{\mathbb{R}^d} v\, Q(f)\, dv, \tag{5.37}
$$

$$
\partial_t(ne) + \frac{m_n}{2}\operatorname{div}_x\langle v|v|^2 \rangle - J\cdot E = \frac{m_n}{2}\int_{\mathbb{R}^d} |v|^2\, Q(f)\, dv. \tag{5.38}
$$

The moment method has two difficulties. First, a truncation of the hierarchy of moment equations does not give a closed system, i.e., the equation for the j-th moment contains always a moment of order $j+1$. Second, the terms originating from the collision operator generally do not depend on the moments in a simple way. The first difficulty can be overcome by making an ansatz for the distribution function. In order to deal with the integrals involving the collision operator, we choose a particular scattering rate ϕ.

Lemma 5.1 shows that the Maxwellian (5.35) lies in the kernel of $Q(f)$: The ansatz $f = nM$ would yield $J = -q\langle v \rangle = 0$. More interesting equations can be obtained from the so-called *shifted Maxwellian*

$$
f_e(x,v,t) = n\left(\frac{m_n}{2\pi k_B T}\right)^{d/2} \exp\left(-\frac{m_n|v-\bar{v}|^2}{2k_B T}\right),
$$

where $n = n(x,t)$, $T = T(x,t)$ and $\bar{v} = \bar{v}(x,t)$ are interpreted as the electron number density, electron temperature and mean velocity, respectively. Substitut-

ing $z = \sqrt{m_n/k_B T}(v - \bar{v})$ and using the identities

$$\int_{\mathbb{R}} e^{-x^2/2} \, dx = \sqrt{2\pi},$$

$$\int_{\mathbb{R}} x e^{-x^2/2} \, dx = 0,$$

$$\int_{\mathbb{R}} x^2 e^{-x^2/2} \, dx = \int_{\mathbb{R}} 1 \cdot e^{-x^2/2} \, dx = \sqrt{2\pi},$$

$$\int_{\mathbb{R}} x^3 e^{-x^2/2} \, dx = 2 \int_{\mathbb{R}} x e^{-x^2/2} \, dx = 0,$$

the moments corresponding to the above ansatz are

$$\langle 1 \rangle = \int_{\mathbb{R}^d} f_e \, dv = n,$$

$$\langle v \rangle = \int_{\mathbb{R}^d} v f_e \, dv = \int_{\mathbb{R}^d} \left(\bar{v} + \sqrt{\frac{k_B T}{m_n}} z \right) \frac{n}{(2\pi)^{d/2}} e^{-|z|^2/2} \, dz = n\bar{v},$$

$$\langle v \otimes v \rangle = \int_{\mathbb{R}^d} \left(\bar{v} + \sqrt{\frac{k_B T}{m_n}} z \right) \otimes \left(\bar{v} + \sqrt{\frac{k_B T}{m_n}} z \right) \frac{n}{(2\pi)^{d/2}} e^{-|z|^2/2} \, dz$$

$$= (\bar{v} \otimes \bar{v})n + \frac{k_B T}{m_n} \frac{n}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} z \otimes z \, e^{-|z|^2/2} \, dz$$

$$= (\bar{v} \otimes \bar{v})n + \frac{k_B T}{m_n} n \mathrm{Id}, \qquad \mathrm{Id} = \text{identity matrix in } \mathbb{R}^{d\times d},$$

$$\langle v_j |v|^2 \rangle = \sum_{i=1}^{d} \int_{\mathbb{R}^d} \left(\bar{v}_j + \sqrt{\frac{k_B T}{m_n}} z_j \right) \left(\bar{v}_i + \sqrt{\frac{k_B T}{m_n}} z_i \right)^2 \frac{n}{(2\pi)^{d/2}} e^{-|z|^2/2} \, dz$$

$$= \bar{v}_j |\bar{v}|^2 n + \frac{k_B T}{m_n} \frac{n}{(2\pi)^{d/2}} \sum_{i=1}^{d} \bar{v}_j \int_{\mathbb{R}^d} |z_i|^2 e^{-|z|^2/2} \, dz$$

$$+ 2\frac{k_B T}{m_n} \frac{n}{(2\pi)^{d/2}} \sum_{i=1}^{d} \bar{v}_i \int_{\mathbb{R}^d} z_i z_j \, e^{-|z|^2/2} \, dz$$

$$+ \left(\frac{k_B T}{m_n} \right)^{3/2} \frac{n}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} z_j |z|^2 \, e^{-|z|^2/2} \, dz$$

$$= \bar{v}_j |\bar{v}|^2 n + \frac{k_B T}{m_n} n \sum_{i=1}^{d} (\bar{v}_j + 2\bar{v}_i \delta_{ij})$$

$$= \frac{2}{m_n} \bar{v}_j n \left(\frac{m_n}{2} |\bar{v}|^2 + \frac{d+2}{2} k_B T \right).$$

Thus, the energy is the sum of kinetic and thermal energy:

$$e = \frac{m_n}{2}|\bar{v}|^2 + \frac{d}{2}k_B T = \frac{m_n}{2q^2}\frac{|J|^2}{n^2} + \frac{d}{2}k_B T,$$

and the third-order term can be written as

$$\langle v|v|^2 \rangle = \frac{2}{m_n}\bar{v}n(e + k_B T).$$

In order to compute the terms coming from the collision operator, we assume that the scattering rate is independent of v and v', i.e. $\phi(x, v, v') = \phi_0(x)$. Then

$$
\begin{aligned}
\int_{\mathbb{R}^d} v\, Q(f_e)\, dv &= \phi_0 \int_{\mathbb{R}^d} Mv\, dv \int_{\mathbb{R}^d} f'_e\, dv' - \phi_0 \int_{\mathbb{R}^d} M'\, dv' \int_{\mathbb{R}^d} f_e v\, dv \\
&= -\phi_0 \langle v \rangle \\
&= \frac{\phi_0}{q} J, \\
\int_{\mathbb{R}^d} |v|^2\, Q(f_e)\, dv &= \phi_0 \int_{\mathbb{R}^d} M\, |v|^2\, dv \cdot n - \phi_0 \int_{\mathbb{R}^d} M'\, dv' \cdot \langle |v|^2 \rangle \\
&= \phi_0 n \left(d\frac{k_B T_0}{m_n} - \frac{2}{m_n}e \right).
\end{aligned}
$$

We introduce the relaxation time by $\tau = 1/\phi_0$. Then

$$-q \int_{\mathbb{R}^d} v\, Q(f_e)\, dv = -\frac{J}{\tau}, \quad \frac{m_n}{2} \int_{\mathbb{R}^d} |v|^2\, Q(f_e)\, dv = -\frac{n}{\tau}\left(e - \frac{d}{2}k_B T_0 \right).$$

Notice that these terms vanish in the thermal equilibrium state $J = 0$, $T = T_0$.

We conclude that the moment equations (5.36)-(5.38) can be written as

$$\partial_t n - \frac{1}{q}\text{div}_x J = 0, \tag{5.39}$$

$$\partial_t J - \frac{1}{q}\text{div}_x \left(\frac{J \otimes J}{n} \right) + \frac{qk_B}{m_n}\nabla(Tn) - \frac{q^2}{m_n}nE = -\frac{J}{\tau}, \tag{5.40}$$

$$\partial_t(ne) - \frac{1}{q}\text{div}_x[J(e + k_B T)] - J \cdot E = -\frac{n}{\tau}\left(e - \frac{d}{2}k_B T_0 \right). \tag{5.41}$$

For vanishing right-hand sides, these equations are the Euler equations of gas dynamics for a gas of charged particles in an electric field. Sometimes an additional term $-\text{div}(\kappa(n, t)\nabla T)$ is added to the left-hand side of (5.41). This term whose presence is purely heuristic is a diffusion term and makes the equation (5.41) of parabolic type. This is desirable since we expect the temperature to satisfy an equation related to the parabolic heat equation. The heat conductivity $\kappa(n, T)$ is usually modeled (in $\mathbb{R}^3$) by [10]

$$\kappa(n, T) = \frac{3}{2}\frac{k_B^2 \tau}{m_n}nT.$$

The equations (5.39)-(5.41) including the aboce heat flow (but with different relaxation times in the momentum and energy equations) has been first derived by Bløtekjær [13] and Baccarani and Wordeman [10].

The equations (5.39) and (5.40) are of hyperbolic type. The system of equations (5.39)-(5.41), with or without the additional heat flux term, are referred to as the *hydrodynamic equations*. For constant temperature $T = T_0$, (5.39)-(5.40) are called the *isothermal hydrodynamic equations.*

In a similar way, equations for the hole density, the hole current density and the hole energy density can be derived from the corresponding Boltzmann equation for holes (see (5.2)). The equations are as above with $-q$ replaced by q. The corresponding equations for the electrons and holes are referred to as the *bipolar hydrodynamic model.*

5.4   The Spherical Harmonic Expansion (SHE) model

We start with the semi-classical Boltzmann equation for the distribution function $f(x, k, t)$ (see Section 3.3)

$$\partial_t f + \frac{1}{\hbar}\nabla_k \varepsilon \cdot \nabla_x f + \frac{q}{\hbar}\nabla_x V \cdot \nabla_k f = Q(f), \quad x \in \mathbb{R}^d, \ k \in B, \ t > 0. \qquad (5.42)$$

In the following we only consider the evolution of the electrons in a single conduction band $\varepsilon(k)$. We proceed similarly as in [11, 18]. The collision operator is assumed to be the sum of lattice-defect collision terms (due to ionized impurities, acoustic and optical phonons) and electron-electron collision terms. These terms are separated in an elastic collision part and an inelastic collision part:

$$Q(f) = Q_{\text{el}}(f) + \alpha^2 Q_{\text{inel}}(f).$$

We have assumed that the electric field is so large that the typical energy qU of the electrons, where U is the voltage bias between the contacts, is much larger than the energy $\hbar\omega$ of the phonons:

$$\alpha^2 = \frac{\hbar\omega}{qU} \ll 1,$$

and we have expanded the collision operator in terms of α^2. By Section 3.3, we can write the elastic collision operator as

$$(Q_{\text{el}}(f))(x, k) = \int_B \phi(x, k, k')\delta(\varepsilon(k') - \varepsilon(k))(f(k') - f(k)) \, dk' \qquad (5.43)$$

for functions $f : \mathbb{R}^d \times B \to \mathbb{R}$, where

$$\phi(x, k, k') = \sigma_{\text{imp}}(x, k, k') + (2N_{\text{op}} + 1)\sigma_{\text{op}}(x, k, k') + (2N_{\text{ac}} + 1)\sigma_{\text{ac}}(x, k, k'),$$

is the scattering rate, N_{op}, N_{ac} are the occupation numbers of optical or acoustic phonons, respectively, and δ is the delta distribution. Clearly, the integral (5.43) has to be understood symbolically. The operator $Q_{\mathrm{inel}}(f)$ contains the inelastic correction to phonon and electron-electron collisions.

We also impose the following assumptions:

- The inelastic collision operator Q_{inel} is linear.

- The scattering rate ϕ is positive and symmetric in k, k':

$$\phi(x, k, k') > 0, \quad \phi(x, k, k') = \phi(x, k', k) \quad \forall x, k, k'. \tag{5.44}$$

The Spherical Harmonic Expansion (SHE) model is derived in the so-called *diffusion scaling*, i.e. by introducing the space and time scale

$$x_s = \alpha x, \quad t_s = \alpha^2 t.$$

Then (x_s, t_s) are called *macroscopic variables.* In these variables, the Boltzmann equation (5.42) reads

$$\alpha^2 \partial_t f + \alpha \left(\frac{1}{\hbar} \nabla_k \varepsilon \cdot \nabla_x f + \frac{q}{\hbar} \nabla_x V \cdot \nabla_k f \right) = Q_{\mathrm{el}}(f) + \alpha^2 Q_{\mathrm{inel}}(f), \tag{5.45}$$

where we have written again (x, t) instead of (x_s, t_s). Notice that this is exactly the same scaling used in the derivation of the drift-diffusion equations (see Section 5.1).

We use now the Hilbert expansion method to "solve" (5.45). Inserting the expansion

$$f = f_0 + \alpha f_1 + \alpha^2 f_2 + \cdots$$

into (5.45), using the linearity of the operators Q_{el} and Q_{inel} and identifying terms of the same order in α, we find:

- terms of order α^0:

$$Q_{\mathrm{el}}(f_0) = 0, \tag{5.46}$$

- terms of order α^1:

$$Q_{\mathrm{el}}(f_1) = \frac{1}{\hbar} \nabla_k \varepsilon \cdot \nabla_x f_0 + \frac{q}{\hbar} \nabla_x V \cdot \nabla_k f_0, \tag{5.47}$$

- terms of order α^2:

$$Q_{\mathrm{el}}(f_2) = \partial_t f_0 + \frac{1}{\hbar} \nabla_k \varepsilon \cdot \nabla_x f_1 + \frac{q}{\hbar} \nabla V \cdot \nabla_k f_1 - Q_{\mathrm{inel}}(f_0). \tag{5.48}$$

First we wish to solve (5.46). For this, we need to prove some properties of the operator Q_{el} defined on the space

$$L^2(B) = \left\{ f : B \to B \text{ measurable:} \int_B |f(k)|^2 \, dk < \infty \right\}.$$

Lemma 5.2 *Assume that (5.44) hold. Then*

(1) $-Q_{\text{el}}$ *is a self-adjoint and non-negative operator on* $L^2(B)$.

(2) The kernel of Q_{el} *and its orthogonal complement are given by*

$$
\begin{aligned}
N(Q_{\text{el}}) &= \{ f \in L^2(B) : \exists g : \mathbb{R} \to \mathbb{R} : \forall k \in B : f(k) = g(\varepsilon(k)) \}, \\
N(Q_{\text{el}})^\perp &= \left\{ f \in L^2(B) : \forall e \in R(\varepsilon) : \int_B f(k)\delta(e - \varepsilon(k)) \, dk = 0 \right\}.
\end{aligned}
$$

(3) For all functions $g = g(\varepsilon(k))$ *and* $f \in L^2(B)$ *it holds*

$$Q_{\text{el}}(gf) = g \, Q_{\text{el}}(f).$$

In the statement of the lemma we have omitted some technical assumptions on the regularity of ϕ and $g(\varepsilon(k))$. The orthogonal complement $N(Q_{\text{el}})^\perp$ is the set of all functions $f \in L^2(B)$ such that

$$\langle f, F \rangle = \int_B f(k) \, F(k) \, dk = 0 \quad \forall F \in N(Q_{\text{el}}).$$

The range $R(\varepsilon)$ of ε is defined by $R(\varepsilon) = \{ \varepsilon \in \mathbb{R} : \exists k \in B : \varepsilon(k) = e \}$. The operator $-Q_{\text{el}}$ is called non-negative if

$$- \int_B Q_{\text{el}}(f) f \, dk \geq 0 \quad \forall f \in L^2(B).$$

Proof of Lemma 5.2. (1) Let $x \in \mathbb{R}^d$, $f, g \in L^2(B)$ and set $f = f(k)$, $f' = f(k')$, $g = g(k)$, $g' = g(k')$, $\varepsilon = \varepsilon(k)$ and $\varepsilon' = \varepsilon(k')$. Then, using the property

$$\int_B \delta(\varepsilon' - \varepsilon)\psi(k, k') \, dk' = \int_B \delta(\varepsilon - \varepsilon')\psi(k, k') \, dk'$$

for any function ψ and the symmetry of ϕ, we obtain

$$
\begin{aligned}
\int_B Q_{\text{el}}(f) g \, dk &= \frac{1}{2} \int_B \int_B \phi(x, k, k')\delta(\varepsilon' - \varepsilon)(f' - f)g \, dk' \, dk \\
&\quad + \frac{1}{2} \int_B \int_B \phi(x, k', k)\delta(\varepsilon - \varepsilon')(f - f')g' \, dk \, dk' \\
&= -\frac{1}{2} \int_B \int_B \phi(x, k, k')\delta(\varepsilon' - \varepsilon)(f' - f)(g' - g) \, dk' \, dk \\
&= \int_B Q_{\text{el}}(g) f \, dk.
\end{aligned}
$$

82

This shows that Q_{el} is symmetric and hence self-adjoint. Taking $f = g$ gives

$$\int_B Q_{el}(f)f \ dk = -\frac{1}{2}\int_B\int_B \phi(x,k,k')\delta(\varepsilon' - \varepsilon)(f' - f)^2 \ dk' \ dk \leq 0 \qquad (5.49)$$

and thus, $-Q_{el}$ is non-negative.

(2) Let $f \in N(Q_{el})$, i.e. $Q_{el}(f) = 0$. By (5.49), this implies

$$\delta(\varepsilon' - \varepsilon)(f' - f)^2 = 0 \quad \text{for almost all } k, k' \in B.$$

The delta distribution has the "property" $\delta(z) = 0$ for all $z \neq 0$. (We treat the delta distribution as a function; this is mathematically not correct but avoids technical calculations.) Therefore

$$\varepsilon(k') = \varepsilon(k) \quad \text{and} \quad f(k') = f(k) \quad \text{for almost all } k, k' \in B.$$

Hence f must be constant on each energy surface $\{k : \varepsilon(k) = e\}$. This implies that f is a function of $\varepsilon(k)$ and proves the first assertion.

Let $f \in N(Q_{el})^\perp$ and $F \in N(Q_{el})$. Then $F(k) = g(\varepsilon(k))$ for some function g and

$$\begin{aligned}
0 &= \int_B f(k)F(k) \ dk \\
&= \int_B f(k)g(\varepsilon(k)) \ dk \\
&= \int_B f(k)\left(\int_{\mathbb{R}} g(e)\delta(\varepsilon(k) - e) \ de\right) \ dk \\
&= \int_{\mathbb{R}}\left(\int_B f(k)\delta(\varepsilon(k) - e) \ dk\right)g(e) \ de. \qquad (5.50)
\end{aligned}$$

Here we have used the definition of the delta distribution

$$\int_{\mathbb{R}} \delta(z - e)\psi(e) \ de = \psi(z) \quad \text{for any (regular) function } \psi. \qquad (5.51)$$

Equation (5.50) holds for all $F \in N(Q_{el})$ and thus for any function g. We infer

$$\int_B f(k)\delta(\varepsilon(k) - e) \ dk = 0 \quad \text{for almost all } e \in R(\varepsilon).$$

This proves (2).

(3) The first integral in

$$\begin{aligned}
(Q_{el}(gf))(k) &= \int_B \phi(x,k,k')\delta(\varepsilon(k') - \varepsilon(k))f(k')g(\varepsilon(k')) \ dk' \\
&\quad - \int_B \phi(x,k,k')\delta(\varepsilon(k') - \varepsilon(k))f(k)g(\varepsilon(k)) \ dk'
\end{aligned}$$

equals, after the change of unknown $e = \varepsilon(k')$ and the definition (5.51),

$$\int_{R(\varepsilon)} \phi(x,k,k')\delta(e - \varepsilon(k))f(k')g(e) \left| \det \frac{d\varepsilon}{dk}(k') \right|^{-1} de$$

$$= \phi(x,k,k')f(k')g(\varepsilon(k)) \left| \det \frac{d\varepsilon}{dk}(k') \right|^{-1} \Bigg|_{e=e(k)=\varepsilon(k')}$$

$$= g(\varepsilon(k)) \int_{R(\varepsilon)} \phi(x,k,k')\delta(e - \varepsilon(k))f(k') \left| \det \frac{d\varepsilon}{dk}(k') \right|^{-1} de$$

$$= g(\varepsilon(k)) \int_{B} \phi(x,k,k')\delta(\varepsilon(k') - \varepsilon(k))f(k')\, dk'.$$

Notice that the use of the transformation formula is rather formal since ε may not satisfy the required technical assumptions. However, the above computation can be made mathematically rigorous by employing the coarea formula (see [11, 18] for details). We infer

$$(Q_{\mathrm{el}}(gf))(k) = g(\varepsilon(k)) \int_{B} \phi(x,k,k')\delta(\varepsilon(k') - \varepsilon(k))f(k')\, dk'$$

$$-g(\varepsilon(k)) \int_{B} \phi(x,k,k')\delta(\varepsilon(k') - \varepsilon(k))f(k)\, dk'$$

$$= g(\varepsilon(k))(Q_{\mathrm{el}}(f))(k).$$

This proves (3). $\qquad\qquad\square$

Remark 5.3 From the proof of Lemma 5.2(3) we conclude the following general result: For all $\psi(k,k')$ and $g(\varepsilon(k))$ it holds

$$\int_{B} \psi(k,k')\delta(\varepsilon(k') - \varepsilon(k))g(\varepsilon(k'))\, dk' = g(\varepsilon(k)) \int_{B} \psi(k,k')\delta(\varepsilon(k') - \varepsilon(k))\, dk'.$$

$$(5.52)$$

We can now solve the equations (5.46)-(5.48). By Lemma 5.2(2), any solution of (5.46) can be written as

$$f_0(x,k,t) = F(x,\varepsilon(k),t)$$

for some function F. Thus we can reformulate (5.47) as

$$Q_{\mathrm{el}}(f_1) = \frac{1}{\hbar}\nabla_k\varepsilon \cdot \left(\nabla_x F + q\nabla_x V \frac{\partial F}{\partial \varepsilon} \right).$$

As $\nabla_x F + q\nabla_x V(\partial F/\partial\varepsilon)$ only depends on $\varepsilon(k)$ (and on the parameters x,t), any solution of this equation can be written, by Lemma 5.2(3), as

$$f_1(x,k,t) = -\lambda(x,k) \cdot \left(\nabla_x F + q\nabla_x V \frac{\partial F}{\partial \varepsilon} \right) + F_1(x,\varepsilon(k),t)$$

84

where $F_1 \in N(Q_{\text{el}})$ and $\lambda(x, k)$ is a solution of

$$Q_{\text{el}}(\lambda) = -\frac{1}{\hbar}\nabla_k\varepsilon. \tag{5.53}$$

More precisely, this equation has to be considered componentwise, i.e.

$$Q_{\text{el}}(\lambda_i) = -\hbar^{-1}\frac{\partial\varepsilon}{\partial k_i}, \quad i = 1, \ldots, d.$$

Notice that λ depends on the parameter x since ϕ depends on x. The above equation is solvable if and only if $\nabla_k\varepsilon \in R(Q_{\text{el}})$ (the range of Q_{el}). It is not difficult to see that $R(Q_{\text{el}})$ is closed. From functional analysis (and the self-adjointness of Q_{el}) follows that $R(Q_{\text{el}}) = N(Q_{\text{el}})^\perp$. Thus (5.53) is solvable if and only if $\nabla_k\varepsilon \in N(Q_{\text{el}})^\perp$. By Lemma 5.2(2) this is equivalent to

$$\int_B \nabla_k\varepsilon(k)\delta(e - \varepsilon(k))\, dk = 0 \quad \forall e \in R(\varepsilon).$$

Now, using the relation

$$\frac{dH}{dz}(z) = \delta(z),$$

where H is the Heaviside function, we have

$$\int_B \nabla_k\varepsilon(k)\delta(e - \varepsilon(k))\, dk = -\int_B \nabla_k H(e - \varepsilon(k))\, dk = 0, \tag{5.54}$$

since ε is periodic on B. Thus, (5.53) is solvable. We choose the particular solution of (5.47) as

$$f_1(x, k, t) = -\lambda(x, k) \cdot \left(\nabla_x F + q\nabla_x V \frac{\partial F}{\partial\varepsilon}\right).$$

It remains to solve (5.48). Again, (5.48) is solvable if and only if the right-hand side of (5.48) is an element of $N(Q_{\text{el}})^\perp$ or, equivalently,

$$\int_B \left(\partial_t f_0 + \frac{1}{\hbar}\nabla_k\varepsilon \cdot \nabla_x f_1 + \frac{q}{\hbar}\nabla_x V \cdot \nabla_k f_1 - Q_{\text{el}}(f_0)\right)\delta(e - \varepsilon(k))\, dk = 0 \tag{5.55}$$

for all $e \in R(\varepsilon)$. We define the *density of states* of energy e by

$$N(e) = \int_B \delta(e - \varepsilon(k))\, dk \tag{5.56}$$

(compare with (2.13) in Section 2.3). Then, by (5.52),

$$\begin{aligned}
\int_B \partial_t f_0 \delta(e - \varepsilon(k))\, dk &= \partial_t \int_B F(x, \varepsilon(k), t)\delta(e - \varepsilon(k))\, dk \\
&= \partial_t F(x, e, t) \int_B \delta(e - \varepsilon(k))\, dk \\
&= N(e)\partial_t F(x, e, t).
\end{aligned}$$

85

Furhermore, we introduce the *electron current density* by

$$J(x, e, t) = -q \int_B \frac{1}{\hbar} \nabla_k \varepsilon(k) f_1(x, k, t) \delta(e - \varepsilon(k)) \, dk.$$

Then we can write the second term in (5.55) as

$$\int_B \frac{1}{\hbar} \nabla_k \varepsilon \cdot \nabla_x f_1 \delta(e - \varepsilon(k)) \, dk = -\frac{1}{q} \mathrm{div}_x J(x, e, t).$$

Using the expression for f_1 and the property (5.52), we have

$$
\begin{aligned}
J(x, e, t) &= q \int_B \frac{1}{\hbar} \nabla_k \varepsilon(k) \lambda(x, k) \cdot \left(\nabla_x F + q \nabla_x V \frac{\partial F}{\partial \varepsilon} \right)(x, \varepsilon(k), t) \\
&\quad \times \delta(e - \varepsilon(k)) \, dk \\
&= D(x, e) \left(\nabla_x F + q \nabla_x V \frac{\partial F}{\partial \varepsilon} \right)(x, e, t),
\end{aligned}
$$

where

$$D(x, e) = \frac{q}{\hbar} \int_B \nabla_k \varepsilon(k) \otimes \lambda(x, k) \delta(e - \varepsilon(k)) \, dk \in \mathbb{R}^{d \times d} \qquad (5.57)$$

is called diffusion matrix.

The last term in (5.55) is a collision term, averaged over the energy surface $e = \varepsilon(k)$:

$$(S(F))(x, e, t) = \int_B (Q_{\mathrm{inel}}(F))(x, k, t) \delta(e - \varepsilon(k)) \, dk. \qquad (5.58)$$

It remains to compute the third term in (5.55). For any (smooth) function ψ we obtain, using the definition of δ and integrating by parts,

$$
\begin{aligned}
&\int_{R(\varepsilon)} \psi(e) \int_B \nabla_k f_1(x, k, t) \delta(e - \varepsilon(k)) \, dk \, de \\
&= \int_B \psi(\varepsilon(k)) \nabla_k f_1(x, k, t) \, dk \\
&= -\int_B f_1(x, k, t) \frac{d\psi}{d\varepsilon}(\varepsilon(k)) \nabla_k \varepsilon(k) \, dk \\
&= -\int_{R(\varepsilon)} \frac{d\psi}{d\varepsilon}(e) \int_B \nabla_k \varepsilon(k) f_1(x, k, t) \delta(e - \varepsilon(k)) \, dk \, de \\
&= \frac{\hbar}{q} \int_{R(\varepsilon)} \frac{d\psi}{d\varepsilon}(e) J(x, e, t) \, de \\
&= -\frac{\hbar}{q} \int_{R(\varepsilon)} \psi(e) \frac{\partial J}{\partial \varepsilon}(x, e, t) \, de.
\end{aligned}
$$

Since ψ is arbitrary, it follows

$$\int_B \nabla_k f_1(x, k, t) \delta(e - \varepsilon(k)) \, dk = -\frac{\hbar}{q} \frac{\partial J}{\partial \varepsilon}(x, e, t)$$

and

$$\frac{q}{\hbar} \int_B \nabla_x V \cdot \nabla_k f_1 \delta(e - \varepsilon(k)) \, dk = -\nabla_x V \cdot \frac{\partial J}{\partial \varepsilon}(x, e, t).$$

We have shown that the Hilbert expansion (5.46)-(5.48) is solvable if and only if $f_0(x, k, t) = F(x, \varepsilon(k), t)$, where F is a solution of

$$N(\varepsilon)\partial_t F - \frac{1}{q}\mathrm{div}_x J - \nabla_x V \cdot \frac{\partial J}{\partial \varepsilon} = S(F), \qquad (5.59)$$

$$J(x, \varepsilon, t) = D(x, \varepsilon)\left(\nabla_x F + q\nabla_x V \frac{\partial F}{\partial \varepsilon}\right), \quad x \in \mathbb{R}^d, \ \varepsilon \in R(\varepsilon), \ t > 0, (5.60)$$

where $D(x, \varepsilon)$ is given by (5.57). The equations (5.59)-(5.60) are referred to as *Spherically Harmonic Expansion (SHE) model.* It has been first derived in the physical literature for spherically symmetric band diagrams and rotationally invariant collision operators from the Boltzmann equation [26, 27]. In fact, the above derivation of the SHE model does not require any assumption of spherical symmetry. The equations can be written more clearly in terms of the total energy $u = \varepsilon - qV(x, t)$. Introducing the change of unknowns

$$
\begin{aligned}
F(x, \varepsilon, t) &= f(x, u, t) = f(x, \varepsilon - qV(x, t), t), \\
D(x, \varepsilon) &= d(x, u, t) = d(x, \varepsilon - qV(x, t), t), \\
N(\varepsilon) &= n(x, u, t) = N(u + qV(x, t)),
\end{aligned}
$$

we obtain

$$\nabla_x F = \nabla_x f - q\nabla_x V \frac{\partial f}{\partial u}, \qquad \partial_t F = \partial_t f - q\partial_t V \frac{\partial f}{\partial u}$$

and therefore

$$n(x, u, t)\partial_t f - \mathrm{div}_x(d(x, u, t)\nabla_x f) = S(f) + n(x, u, t)q\partial_t V \frac{\partial f}{\partial u}.$$

This shows, together with the following proposition, that the SHE model is of parabolic type.

Proposition 5.4 *The diffusion matrix D is a symmetric non-negative $d \times d$ matrix. Moreover, there exists a constant $K > 0$ such that for all $z \in \mathbb{R}^d$ and all $x \in \mathbb{R}^d$ and $e \in R(\varepsilon)$,*

$$z^\top D(x, e)z \geq \frac{K}{N(\varepsilon)} \int_B |\nabla_k \varepsilon(k) \cdot z|^2 \delta(e - \varepsilon(k)) \, dk.$$

The fact that the matrix $D(x, \varepsilon)$ is non-negative is a direct consequence of the non-negativity of the operator $-Q_{\mathrm{el}}$ (see Lemma 5.2(1)). The symmetry of $D(x, e)$ is related to the so-called Onsager reciprocity relation of non-equilibrium thermodynamics [22].

Proof. We only prove that $D(x, \varepsilon) = (D_{ij})_{ij}$ is symmetric. The proof of the second assertion is more technical and can be found in [11, Sec. III.4]. We compute, for (smooth) functions ψ,

$$
\begin{aligned}
\int_{\mathbb{R}} D_{ij} \psi(e) \, de &= -\frac{q}{\hbar} \int_{\mathbb{R}} \int_B \frac{\partial \varepsilon}{\partial k_i} \lambda_j \delta(e - \varepsilon(k)) \, dk \, \psi(e) \, de \\
&= -q \int_B Q_{\text{el}}(\lambda_i) \lambda_j \psi(\varepsilon(k)) \, dk \\
&\quad \text{(using the definition of } \delta) \\
&= -q \int_B \lambda_i Q_{\text{el}}(\lambda_j \psi(\varepsilon)) \, dk \\
&\quad \text{(since } Q_{\text{el}} \text{ is self-adjoint by Lemma 5.2(1))} \\
&= -q \int_B \lambda_i Q_{\text{el}}(\lambda_j) \psi(\varepsilon) \, dk \\
&\quad \text{(by Lemma 5.2(3))} \\
&= \frac{q}{\hbar} \int_{\mathbb{R}} \int_B \lambda_i \frac{\partial \varepsilon}{\partial k_j} \delta(e - \varepsilon(k)) \, dk \, \psi(e) \, de \\
&= \int_{\mathbb{R}} D_{ji} \psi(e) \, de.
\end{aligned}
$$

Since ψ is arbitrary, $D_{ij}(x, e) = D_{ji}(x, e)$ for $x \in \mathbb{R}^d$, $e \in R(\varepsilon)$. $\square$

For more explicit expressions for $N(e)$ and $D(x, e)$ we consider the following example.

Example 5.5 (Spherically symmetric energy bands)
Assume that the scattering rate only depends on $\varepsilon(k)$, i.e.

$$
\phi(x, k, k') = \phi(x, \varepsilon(k)) \quad \text{for all } k, k' \text{ with } \varepsilon(k') = \varepsilon(k),
$$

that the energy band is spherically symmetric (thus, $B = \mathbb{R}^d$) and strictly monotone in $|k|$, i.e.

$$
\varepsilon = \varepsilon(|k|),
$$

and that $d = 3$. Notice that the first assumption makes sense since due to the term $\delta(\varepsilon(k') - \varepsilon(k))$ in the definition of Q_{el}, the scattering rate only needs to be defined on the energy surface $\{k' : \varepsilon(k') = \varepsilon(k)\}$ of energy $\varepsilon(k)$. The second assumption implies that $\varepsilon(|k|)$ is invertible.

From the first assumption and the definition (5.56) of the density of states $N(e)$ follows, with $\varepsilon = \varepsilon(k)$, $\varepsilon' = \varepsilon(k')$,

$$
\begin{aligned}
(Q_{\text{el}}(f))(x, k, t) &= \phi(x, \varepsilon) \int_{\mathbb{R}^d} \delta(\varepsilon' - \varepsilon) f(k') \, dk' - \phi(x, \varepsilon) N(\varepsilon) f(k) \\
&= \frac{1}{\tau(x, \varepsilon)} \left([f] - f\right)(x, k, t), \quad (5.61)
\end{aligned}
$$

where

$$\tau(x, \varepsilon) = \frac{1}{\phi(x, \varepsilon)N(\varepsilon)}$$

is called relaxation time and

$$[f](k) = \frac{1}{N(\varepsilon)} \int_{\mathbb{R}^d} \delta(\varepsilon' - \varepsilon) f(k') \, dk'$$

is the average of f on the energy surface $\{k' \in B : \varepsilon(k') = \varepsilon(k)\}$. The expression (5.61) is called *relaxation-time operator*.

It follows that a solution $\lambda(x, k)$ of $Q_{\mathrm{el}}(\lambda) = -\hbar^{-1}\nabla_k\varepsilon$ (see (5.53)) is given by

$$\lambda(x, k) = \tau(x, \varepsilon(k))\hbar^{-1}\nabla_k\varepsilon(k).$$

Indeed, since, by (5.52) and (5.54),

$$
\begin{aligned}
[\lambda] &= \frac{1}{N(\varepsilon)} \int_{\mathbb{R}^d} \delta(\varepsilon' - \varepsilon)\tau(x, \varepsilon')\hbar^{-1}\nabla_k\varepsilon(k') \, dk' \\
&= \frac{1}{N(\varepsilon)}\tau(x, \varepsilon)\hbar^{-1} \int_{\mathbb{R}^d} \delta(\varepsilon' - \varepsilon)\nabla_k\varepsilon(k') \, dk' \\
&= 0,
\end{aligned}
$$

we obtain

$$Q_{\mathrm{el}}(\lambda) = \frac{1}{\tau(x, \varepsilon)}([\lambda] - \lambda) = -\frac{\lambda}{\tau(x, \varepsilon)} = -\frac{1}{\hbar}\nabla_k\varepsilon.$$

Thus, again using (5.52), the diffusion matrix becomes

$$
\begin{aligned}
D(x, e) &= q\hbar^{-2} \int_{\mathbb{R}^d} \nabla_k\varepsilon(k) \otimes \nabla_k\varepsilon(k)\tau(x, \varepsilon)\delta(e - \varepsilon) \, dk \\
&= q\hbar^{-2}\tau(x, e) \int_{\mathbb{R}^d} \nabla_k\varepsilon(k) \otimes \nabla_k\varepsilon(k)\delta(e - \varepsilon(k)) \, dk.
\end{aligned}
$$

This expression can be further simplified under the second assumption. As ε only depends on $|k|$, it is convenient to introduce spherical coordinates $k = \rho\omega$, where $\rho = |k| \in [0, \infty)$ is the modulus of k and $\omega \in S^{d-1}$ are the angle variables on the unit sphere $S^{d-1} \subset \mathbb{R}^d$. Then

$$\nabla_k\varepsilon(|k|) = \varepsilon'(|k|)\frac{k}{|k|} = \varepsilon'(\rho)\omega$$

and the volume element dk transforms symbolically to $\rho^{d-1}F(\omega) \, d\rho \, d\omega$, where $F(\omega)$ is the modulus of the determinant of the Jacobian of the angle transform.

We compute

$$
\begin{aligned}
D(x,e) &= q\tau(x,e)\hbar^{-2}\int_0^\infty\int_{S^{d-1}}\varepsilon'(\rho)^2\omega\otimes\omega\delta(e-\varepsilon(\rho))\rho^{d-1}F(\omega)\,d\rho\,d\omega \\
&= q\tau(x,e)\hbar^{-2}\int_0^\infty\varepsilon'(\rho)^2\delta(e-\varepsilon(\rho))\rho^{d-1}\,d\rho\cdot\int_{S^{d-1}}\omega\otimes\omega F(\omega)\,d\omega \\
&= q\tau(x,e)\hbar^{-2}\int_{R(\varepsilon)}\varepsilon'(\rho)\delta(e-\eta)\rho^{d-1}\,d\eta\cdot\int_{S^{d-1}}\omega\otimes\omega F(\omega)\,d\omega \\
&\quad\text{(transforming }\eta=\varepsilon(\rho)\text{ and }d\eta=\varepsilon'(\rho)\,d\rho\text{)} \\
&= q\tau(x,e)\hbar^{-2}\varepsilon'(\rho)\rho^{d-1}\Big|_{e=\varepsilon(\rho)}\cdot\int_{S^{d-1}}\omega\otimes\omega F(\omega)\,d\omega.
\end{aligned}
$$

In the three dimensional case $d=3$ (third assumption) this becomes

$$
D(x,e)=\frac{4\pi q}{3\hbar^2}\tau(x,e)\varepsilon'(|k|)|k|^2\,\mathrm{Id}, \tag{5.62}
$$

where $|k|$ is such that $e=\varepsilon(|k|)$ and $\mathrm{Id}\in\mathbb{R}^{3\times3}$ is the identity matrix. Indeed, from

$$
\omega=\begin{pmatrix}\sin\theta\,\cos\phi\\\sin\theta\,\sin\phi\\\cos\theta\end{pmatrix},\quad 0\le\phi<2\pi,\ 0\le\theta<\pi,
$$

and $F(\omega)\,d\omega=\sin\theta\,d\theta\,d\phi$ we obtain

$$
\int_{S^2}(\omega\otimes\omega)_{ij}F(\omega)\,d\omega=\int_0^{2\pi}\int_0^\pi\omega_i\omega_j\sin\theta\,d\theta\,d\phi.
$$

For $i\neq j$ the integral vanishes. We compute for

$$
\begin{aligned}
i=j=1:\quad &\int_0^{2\pi}\cos^2\phi\,d\phi\cdot\int_0^\pi\sin^3\theta\,d\theta=\pi\cdot\frac{4}{3}=\frac{4\pi}{3}, \\
i=j=2:\quad &\int_0^{2\pi}\sin^2\phi\,d\phi\cdot\int_0^\pi\sin^3\theta\,d\theta=\pi\cdot\frac{4}{3}=\frac{4\pi}{3}, \\
i=j=3:\quad &2\pi\cdot\int_0^\pi\cos^2\theta\sin\theta\,d\theta=2\pi\cdot\frac{2}{3}=\frac{4\pi}{3},
\end{aligned}
$$

and (5.62) follows.

Furthermore,

$$
\begin{aligned}
N(e) &= \int_{\mathbb{R}^d}\delta(e-\varepsilon(|k|))\,dk=\int_0^\infty\int_{S^{d-1}}\delta(e-\varepsilon(\rho))\rho^{d-1}F(\Omega)\,d\rho\,d\omega \\
&= \int_{\mathbb{R}}\delta(e-\eta)\frac{\rho^{d-1}}{\varepsilon'(\rho)}d\eta\cdot\int_{S^{d-1}}F(\omega)\,d\omega \\
&= \frac{S^{d-1}}{\varepsilon'(\rho)}\Big|_{e=\varepsilon(\rho)}\int_{S^{d-1}}F(\omega)\,d\omega,
\end{aligned}
$$

90

and for $d = 3$ we get

$$\int_{S^{d-1}} F(\omega) \, d\omega = \int_0^{2\pi} \int_0^\pi \sin\theta \, d\theta \, d\phi = 4\pi$$

and therefore

$$N(e) = 4\pi \frac{|k|^2}{\varepsilon'(|k|)} \quad \text{with } e = \varepsilon(|k|). \tag{5.63}$$

It is convenient to express $D(x, e)$ and $N(e)$ in terms of the function γ, where

$$|k|^2 = \gamma(\varepsilon(|k|)).$$

Differentiating this equation with respect to $|k|$ yields

$$2|k| = \gamma'(\varepsilon(|k|))\varepsilon'(|k|)$$

and thus

$$\varepsilon'(|k|) = \frac{2\sqrt{\gamma(e)}}{\gamma'(e)} \quad \text{with } e = \varepsilon(|k|).$$

It follows from (5.62) and (5.63) for $d = 3$

$$N(e) = 4\pi \frac{\gamma(e)\gamma'(e)}{2\sqrt{\gamma'(e)}} = 2\pi \sqrt{\gamma(e)}\gamma'(e),$$

$$D(x, e) = \frac{4\pi q}{3\hbar^2} \frac{\gamma(e)}{2\pi\phi(x, e)\sqrt{\gamma(e)}\gamma'(e)} \frac{2\sqrt{\gamma(e)}}{\gamma'(e)} \text{Id} = \frac{4q}{3\hbar^2} \frac{\gamma(e)}{\phi(x, e)\gamma'(e)^2} \text{Id}.$$

Example 5.6 (Parabolic band approximation)
We assume that $d = 3$, $\phi = \phi(x, \varepsilon(k))$ and

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

This implies $\gamma(e) = 2m^*e/\hbar^2$ and (see Example 5.5)

$$N(e) = 2\pi \left(\frac{2m^*}{\hbar^2}\right)^{3/2} \sqrt{e},$$

$$D(x, e) = \frac{2q}{3m^*} \frac{e}{\phi(x, e)} \text{Id}, \quad e > 0.$$

It remains to determine the average collision operator $S(F)$ (see (5.58). The precise structure of this term depends on the assumptions on the inelastic collision operator. Here, we only notice that a simplified expression is given by the *Fokker-Planck approximation*

$$S(F) = \frac{\partial}{\partial\varepsilon}\left[A(\varepsilon)\left(F + k_B T_L \frac{\partial F}{\partial\varepsilon}\right)\right], \tag{5.64}$$

where T_L is the lattice temperature and $A(\varepsilon)$ depends on the scattering rate of the inelastic collisions, averaged on the energy surface (see [40]).

91

5.5 The energy-transport equations

The SHE model is much simpler than the Boltzmann equation since the SHE model has a parabolic structure and it has to be solved in a $(d + 2)$-dimensional (x, e, t)-space instead of the $(2d+1)$-dimensional (x, k, t)-space for the Boltzmann equation. 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 [11].

In [12] an energy-transport model has been derived directly from the Boltzmann equation. In this approach, the electron-electron collision term is assumed to be dominant compared to the phonon collisions which can be criticized. Here we assume that the electron-electron collision term is dominant compared to the *inelastic part* of the phonon collision term. More precisely, we start with the SHE model

$$N(\varepsilon)\partial_t F - \frac{1}{q}\text{div}_x J - \nabla_x V \cdot \frac{\partial J}{\partial \varepsilon} = S(F), \tag{5.65}$$

$$J = D(x,\varepsilon)\left(\nabla_x F + q\nabla_x V \cdot \frac{\partial F}{\partial \varepsilon}\right), \quad x \in \mathbb{R}^d, \ \varepsilon \in \mathbb{R}, \ t > 0, \tag{5.66}$$

for the distribution function $F(x, \varepsilon, t)$. The density of states $N(\varepsilon)$, the diffusion matrix $D(x, \varepsilon)$, and the averaged inelastic collision operator $S(F)$ are defined in (5.56), (5.57), and (5.58), respectively. We recall that the inelastic collision operator contains the inelastic correction to phonon collisions and electron-electron collisions. Let β^2 be the ratio of the typical electron-electron collision time and the phonon-collision time. We assume that $\beta^2 \ll 1$ which corresponds to so-called hot-electron regimes. Moreover, we suppose that

- $S(F) = \beta^2 S_{\text{ph}}(F) + S_e(F)$,

- $x_s = \beta x, \ t_s = \beta^2 t, \ J_s = J/\beta$,

- S_e is a linear operator.

In the first assumption, S_{ph} and S_e are the averaged inelastic phonon and electron-electron operators, respectively. The second assumption means that we are considering a macroscopic space and time scale and "small" current densities. Finally, the third assumption is only for convenience (see below). In the macroscopic scale, the SHE equation (5.65) reads, writing again (J, x, t) instead of (J_s, x_s, t_s),

$$N(\varepsilon)\partial_t F - \frac{1}{q}\text{div}_x J - \nabla_x V \cdot \frac{\partial J}{\partial \varepsilon} = S_{\text{ph}}(F) + \frac{1}{\beta^2}S_e(F), \tag{5.67}$$

$$J = D(x,\varepsilon)\left(\nabla_x F + q\nabla_x V \cdot \frac{\partial F}{\partial \varepsilon}\right). \tag{5.68}$$

We use the Hilbert expansion

$$F = F_0 + \beta F_1 + \cdots, \qquad J = J_0 + \beta J_1 + \cdots$$

in (5.67) and identify equal powers of β. This leads to

- terms of order β^{-2}:

$$S_e(F_0) = 0, \tag{5.69}$$

- terms of order β^0:

$$N(\varepsilon)\partial_t F_0 - \frac{1}{q}\mathrm{div}\,_x J_0 - \nabla_x V \cdot \frac{\partial J_0}{\partial \varepsilon} - S_{\mathrm{ph}}(F_0) = S_e(F_1). \tag{5.70}$$

If S_e is not a linear operator, we would need to replace $S_e(F_1)$ by $(D_{F_0} S_e)(F_1)$, the derivative of S_e at F_0 applied to F_1.

We need the following properties to the collision operator S_e:

Lemma 5.7 *Assume that the operators S_e, S_{ph} are given by (5.58), where*

$$(S_e(F))(x, e, t) = \int_B (Q_e(F))(x, k, t)\delta(e - \varepsilon(k))\ dk,$$

$$(S_{\mathrm{ph}}(F))(x, e, t) = \int_B (Q_{\mathrm{ac}}(F) + Q_{\mathrm{op}}(F))(x, k, t)\delta(e - \varepsilon(k))\ dk,$$

and Q_e, Q_{ac} and Q_{op} are the operators of electron-electron, acoustic phonon and optical phonon collisions, respectively (see [11] for precise definitions of these operators).

(1) It holds for any (smooth) function $F : \mathbb{R} \to \mathbb{R}$

$$\int_{\mathbb{R}} (S_e(F))(\varepsilon)\ d\varepsilon = 0, \quad \int_{\mathbb{R}} (S_e(F))(\varepsilon)\varepsilon\ d\varepsilon = 0, \quad \int_{\mathbb{R}} (S_{\mathrm{ph}}(F))(\varepsilon)\ d\varepsilon = 0.$$

(2) The kernel of S_e is given by

$$N(S_e) = \{F : \mathbb{R} \to \mathbb{R} : \exists \mu \in \mathbb{R},\ T > 0 :\ F = F_{\mu,T}\},$$

where

$$F_{\mu,T}(\varepsilon) = \frac{1}{1 + \exp((\varepsilon - q\mu)/k_B T)},$$

and k_B is the Boltzmann constant.

(3) The equation $S_e(F) = G$ is solvable if and only if

$$\int_{\mathbb{R}} G(\varepsilon)\ d\varepsilon = 0 \quad and \quad \int_{\mathbb{R}} G(\varepsilon)\varepsilon\ d\varepsilon = 0.$$

Proof. See [12] and [11]. The proof is similar to that of Lemma 5.2. $\qquad\square$

The variable μ is called *chemical potential* and T is the *electron temperature*. The quantity $\varepsilon_F = q\mu$ is termed *Fermi energy* (see Section 2.3).

From Lemma 5.7(2) follows that the solutions of (5.69) are given by

$$F_0(x, \varepsilon, t) = F_{\mu(x,t),\tau(x,t)}(\varepsilon),$$

for some parameters $\mu(x,t) \in \mathbb{R}$, $T(x,t) > 0$. Lemma 5.7(3) implies that (5.70) is solvable if and only if

$$\int_{\mathbb{R}} \left(N(\varepsilon)\partial_t F_{\mu,T} - \frac{1}{q}\mathrm{div}\,_x J_0 - \nabla_x V \cdot \frac{\partial J_0}{\partial \varepsilon} - S_{\mathrm{ph}}(F_0) \right) \begin{pmatrix} 1 \\ \varepsilon \end{pmatrix}\, d\varepsilon = 0. \quad (5.71)$$

We introduce the particle density and the particle energy density, respectively, by

$$n(\mu, T) = \int_{\mathbb{R}} F_{\mu,T}(\varepsilon)N(\varepsilon)\, d\varepsilon, \quad \mathcal{E}(\mu, T) = \int_{\mathbb{R}} F_{\mu,T}(\varepsilon)\varepsilon N(\varepsilon)\, d\varepsilon, \quad (5.72)$$

and the macroscopic particle and energy current densities, respectively, by

$$J(x,t) = \int_{\mathbb{R}} J_0(x, \varepsilon, t)\, d\varepsilon, \quad S(x,t) = \frac{1}{q}\int_{\mathbb{R}} J_0(x, \varepsilon, t)\varepsilon\, d\varepsilon.$$

Then, using Lemma 5.7(1), we can write (5.71) as

$$\partial_t n - \frac{1}{q}\mathrm{div}\,_x J = \nabla_x V \cdot \int_{\mathbb{R}} \frac{\partial J_0}{\partial \varepsilon}\, d\varepsilon + \int_{\mathbb{R}} S_{\mathrm{ph}}(F_{\mu,T})\, d\varepsilon = 0,$$

$$\partial_t \mathcal{E} - \mathrm{div}\,_x S = \nabla_x V \cdot \int_{\mathbb{R}} \frac{\partial J_0}{\partial \varepsilon}\varepsilon\, d\varepsilon + \int_{\mathbb{R}} S_{\mathrm{ph}}(F_{\mu,T})\varepsilon\, d\varepsilon$$

$$= -\nabla_x V \cdot \int_{\mathbb{R}} J_0\, d\varepsilon + \int_{\mathbb{R}} S_{\mathrm{ph}}(F_{\mu,T})\varepsilon\, d\varepsilon$$

$$= -\nabla_x V \cdot J + W(\mu, T),$$

where

$$W(\mu, T) = \int_{\mathbb{R}} \left(S_{\mathrm{ph}}(F_{\mu,T}) \right)(\varepsilon)\varepsilon\, d\varepsilon \quad (5.73)$$

is called *relaxation term*. It can be seen that W also depends on the lattice temperature T_L, i.e. $W(\mu, T) = W(\mu, T, T_L)$ (this requires knowledge about the precise structure of S_{ph}; we refer to [12]).

It remains to compute the fluxes J and S. We can interpret $F_{\mu,T}(\varepsilon)$ as a function of the variables $u_1 := q\mu/k_B T$, $u_2 := -1/k_B T$ and ε:

$$F(u_1, u_2, \varepsilon) = F\left(\frac{qu}{k_B T}, \frac{-1}{k_B T}, \varepsilon \right) := F_{\mu,T}(\varepsilon).$$

Then

$$F(u_1, u_2, \varepsilon) = \frac{1}{1 + e^{-(\varepsilon u_2 + u_1)}}$$

and

$$\frac{\partial F}{\partial u_1} = \frac{e^{-(\varepsilon u_2 + u_1)}}{(1 + e^{-(\varepsilon u_2 + u_1)})^2} = F_{\mu,T}(1 - F_{\mu,T}),$$

$$\frac{\partial F}{\partial u_2} = \frac{\varepsilon e^{-(\varepsilon u_2 + u_1)}}{(1 + e^{-(\varepsilon u_2 + u_1)})^2} = \varepsilon F_{\mu,T}(1 - F_{\mu,T}),$$

$$\frac{\partial F}{\partial \varepsilon} = \frac{u_2\, e^{-(\varepsilon u_2 + u_1)}}{(1 + e^{-(\varepsilon u_2 + u_1)})^2} = -\frac{1}{k_B T} F_{\mu,T}(1 - F_{\mu,T}).$$

This yields

$$J_0 = D(x, \varepsilon)\left[\frac{\partial F}{\partial u_1}\nabla\left(\frac{qu}{k_B T}\right) + \frac{\partial F}{\partial u_2}\nabla\left(\frac{-1}{k_B T}\right) + q\nabla_x V \cdot \frac{\partial F}{\partial \varepsilon}\right]$$

and

$$J = \mathcal{D}_{11}\nabla\left(\frac{qu}{k_B T}\right) + \mathcal{D}_{12}\nabla\left(\frac{-1}{k_B T}\right) - \mathcal{D}_{11}\frac{q\nabla_x V}{k_B T}, \tag{5.74}$$

$$qS = \mathcal{D}_{21}\nabla\left(\frac{qu}{k_B T}\right) + \mathcal{D}_{22}\nabla\left(\frac{-1}{k_B T}\right) - \mathcal{D}_{21}\frac{q\nabla_x V}{k_B T}, \tag{5.75}$$

where the coefficients

$$\mathcal{D}_{ij}(x, \mu, T) = \int_{\mathbb{R}} D(x, \varepsilon)F_{\mu,T}(1 - F_{\mu,T})\varepsilon^{i+j-2}\, d\varepsilon, \quad i, j = 1, 2, \tag{5.76}$$

are $d \times d$ matrices.

Summarizing, we can write the *energy transport equations* as

$$\partial_t n - \frac{1}{q}\operatorname{div}_x J = 0, \tag{5.77}$$

$$\partial_t \mathcal{E} - \operatorname{div}_x S + \nabla_x V \cdot J = W(\mu, T), \tag{5.78}$$

where $n(\mu, T)$, $\mathcal{E}(\mu, T)$ are defined in (5.72), $W(\mu, T)$ is given by (5.73), and finally, $J(\mu, T)$ and $S(\mu, T)$ are given by (5.74) and (5.75), respectively.

The $(2d \times 2d)$-matrix $\mathcal{D} = (\mathcal{D}_{ij}(x, \mu, T))_{ij}$ and the relaxation term have the following properties.

Proposition 5.8 *It holds*

(1) *The $(2d \times 2d)$-matrix $\mathcal{D} = (\mathcal{D}_{ij})_{ij}$ is symmetric, i.e. $\mathcal{D}_{12} = \mathcal{D}_{21}$.*

(2) *The $(d \times d)$-matrices $\mathcal{D}_{ij}$ are symmetric, i.e. $\mathcal{D}_{ij}^\top = \mathcal{D}_{ij}$ for $i, j = 1, 2$.*

95

(3) Assume that the 2d functions

$$\frac{\partial \varepsilon}{\partial k_1}, \dots, \frac{\partial \varepsilon}{\partial k_d}, \ \varepsilon\frac{\partial \varepsilon}{\partial k_1}, \dots, \varepsilon\frac{\partial \varepsilon}{\partial k_d}$$

are linearly independent. Then $\mathcal{D}(x, \mu, T)$ is symmetric positive definitive for any $\mu \in \mathbb{R}$, $T > 0$.

(4) The relaxation term $W(\mu, T, T_L)$ is monotone in the sense of operators, i.e.

$$W(\mu, T, T_L) \cdot (T - T_L) \leq 0.$$

The hypothesis of part (3) of the above proposition is a geometric assumption on the band structure. It expresses for the case $d = 3$ that the energy band needs to have a real three-dimensional structure, excluding bands depending only on one or two variables, for instance. Notice that the statement of part (4) justifies the name "relaxation term": The temperature of the system is expected to relax to the (constant) lattice temperature if there are no forces influencing the temperature.

Proof. We only prove parts (1)-(3) since for the proof of (4) we need the precise structure of S_{ph} (see [12, Lemma 4.11] for a detailed proof). Part (1) follows from the symmetry of the matrix, and part (2) is a consequence of the symmetry of D. It remains to prove part (3).

Let $\xi = (\xi^{(1)}, \xi^{(2)})^\top \in \mathbb{R}^{2d}$ with $\xi^{(i)} \in \mathbb{R}^d$, $i = 1, 2$, such that $\xi \neq 0$. Then

$$
\begin{aligned}
\xi^\top \mathcal{D}\xi &= \int_{\mathbb{R}} \xi^\top \begin{pmatrix} D(x,\varepsilon) & \varepsilon D(x,\varepsilon) \\ \varepsilon D(x,\varepsilon) & \varepsilon^2 D(x,\varepsilon) \end{pmatrix} \xi F_{\mu,T}(1 - F_{\mu,T})\, d\varepsilon \\
&\quad \text{(by the definition of } \mathcal{D}) \\
&= \int_{\mathbb{R}} \left[(\xi^{(1)})^\top D\xi^{(1)} + 2\varepsilon (\xi^{(1)})^\top D\xi^{(2)} + \varepsilon^2 (\xi^{(2)})^\top D\xi^{(2)} \right] \\
&\quad \times F_{\mu,T}(1 - F_{\mu,T})\, d\varepsilon \qquad \text{(since } D = D(x,\varepsilon) \text{ is symmetric)} \\
&= \int_R (\xi^{(1)} + \varepsilon\xi^{(2)})^\top D (\xi^{(1)} + \varepsilon\xi^{(2)}) F_{\mu,T}(1 - F_{\mu,T})\, d\varepsilon \\
&\geq \frac{K}{N(\varepsilon)} \int_{\mathbb{R}} \int_B |\nabla_k\varepsilon(k) \cdot (\xi^{(1)} + \varepsilon\xi^{(2)})|^2 \delta(\varepsilon - \varepsilon(k)) \\
&\quad \times F_{\mu,T}(1 - F_{\mu,T})\, dk\, d\varepsilon \qquad \text{(by Proposition 5.4)} \\
&= \frac{K}{N(\varepsilon)} \int_{\mathbb{R}} \int_B \left| \begin{pmatrix} \nabla_k\varepsilon(k) \\ \varepsilon(k)\nabla_k\varepsilon(k) \end{pmatrix} \cdot \xi \right|^2 \delta(\varepsilon - \varepsilon(k)) F_{\mu,T}(1 - F_{\mu,T})\, dk\, d\varepsilon \\
&> 0;
\end{aligned}
$$

since otherwise would

$$\begin{pmatrix} \nabla_k\varepsilon(k) \\ \varepsilon(k)\nabla_k\varepsilon(k) \end{pmatrix} \cdot \xi = 0, \quad \xi \neq 0,$$

imply that $(\nabla_k \varepsilon, \varepsilon(k)\nabla_k \varepsilon)$ is linearly dependent, which is a contradiction. $\qquad\square$

More explicit expressions for n, $\mathcal{E}$ and $\mathcal{D}$ can be derived under the assumptions of Example 5.5 and using Maxwell-Boltzmann statistics.

Example 5.9 (Spherically symmetric energy bands)
We impose the following hypotheses:

- The space dimension is $d = 3$ (to simplify the computations).

- The distribution function $F_{\mu,T}$ is approximated by $\exp(-(\varepsilon - q\mu)/k_B T)$. This is possible if $\varepsilon - q\mu \gg k_B T$. In particular,

$$F_{\mu,T}(1 - F_{\mu,T}) \approx e^{-(\varepsilon - q\mu)/k_B T} \cdot 1.$$

- The scattering rate ϕ depends only on x and $\varepsilon(k)$.

- The energy band ε is spherically symmetric and strictly monotone in $|k|$.

From Example 5.5 and (5.72) follows (if $R(\varepsilon) = \mathbb{R}$)

$$n(\mu, T) \;=\; \int_{\mathbb{R}} e^{-(\varepsilon - q\mu)/k_B T} N(\varepsilon)\, d\varepsilon = 2\pi e^{q\mu/k_B T} \int_{\mathbb{R}} \sqrt{\gamma(\varepsilon)}\gamma'(\varepsilon)e^{-\varepsilon/k_B T}\, d\varepsilon,$$

$$\mathcal{E}(\mu, T) \;=\; \int_{\mathbb{R}} e^{-(\varepsilon - q\mu)/k_B T} \varepsilon N(\varepsilon)\, d\varepsilon = 2\pi e^{q\mu/k_B T} \int_{\mathbb{R}} \sqrt{\gamma(\varepsilon)}\gamma'(\varepsilon)\varepsilon e^{-\varepsilon/k_B T}\, d\varepsilon,$$

where $|k|^2 = \gamma(\varepsilon(k))$. The diffusion matrices (5.76) become

$$\mathcal{D}_{ij}(\mu, T) = \frac{4q}{3\hbar^2} e^{q\mu/k_B T} \int_{\mathbb{R}} \frac{\gamma(\varepsilon)\varepsilon^{i+j-2}}{\phi(x,\varepsilon)\gamma'(\varepsilon)^2} e^{-\varepsilon/k_B T}\, d\varepsilon \cdot \mathrm{Id}. \qquad (5.79)$$

Example 5.10 (Parabolic band approximation)
We impose the same assumptions as in Example 5.9 and additionally,

- The energy band is given by the parabolic approximation

$$\varepsilon(|k|) = \frac{\hbar^2}{2m^*}|k|^2, \quad k \in \mathbb{R}^3.$$

- The scattering rate is given by the so-called *Chen model* [17, 21]

$$\phi(x, \varepsilon) = \phi_0(x)\sqrt{\varepsilon}.$$

By Example 5.6,

$$N(\varepsilon) = 2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \sqrt{\varepsilon},$$

and therefore

$$
\begin{aligned}
n(\mu, T) &= 2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2} e^{q\mu/k_B T} \int_0^\infty e^{-\varepsilon/k_B T} \sqrt{\varepsilon} \, d\varepsilon \\
&= 2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2} e^{q\mu/k_B T} \frac{1}{\sqrt{2}} (k_B T)^{3/2} \int_0^\infty e^{-z^2/2} z^2 \, dz,
\end{aligned}
$$

after substituting $\varepsilon/k_B T = z^2/2$. From

$$\int_0^\infty e^{-z^2/2} z^2 \, dz = \frac{1}{2} \int_{\mathbb{R}} e^{-z^2/2} z^2 \, dz = \frac{1}{2} \int_{\mathbb{R}} e^{-z^2/2} \cdot 1 \, dz = \frac{1}{2} \sqrt{2\pi} = \sqrt{\frac{\pi}{2}}$$

we conclude

$$n(\mu, T) = \pi^{3/2} \left(\frac{2m^* k_B T}{\hbar^2} \right)^{3/2} e^{q\mu/k_B T} = (4\pi^2)^{3/2} N_c(T) e^{q\mu/k_B T},$$

where

$$N_c(T) = 2 \left(\frac{m^* k_B T}{2\pi \hbar^2} \right)^{3/2}$$

is the effective density of states defined in Section 2.3. For the energy density we compute

$$
\begin{aligned}
\mathcal{E}(\mu, T) &= 2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2} e^{q\mu/k_B T} \int_0^\infty e^{-\varepsilon/k_B T} \varepsilon^{3/2} \, d\varepsilon \\
&= 2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2} e^{q\mu/k_B T} \frac{1}{\sqrt{8}} (k_B T)^{5/2} \frac{1}{2} \int_{\mathbb{R}} e^{-z^2/2} z^4 \, dz \\
&= \frac{\pi}{\sqrt{8}} \left(\frac{2m^* k_B T}{\hbar^2} \right)^{3/2} e^{q\mu/k_B T} (k_B T) \cdot 3 \int_{\mathbb{R}} e^{-z^2/2} z^2 \, dz \\
&= \frac{3}{2} (4\pi^2)^{3/2} N_c(T) e^{q\mu/k_B T} (k_B T) \\
&= \frac{3}{2} n(\mu, T) \cdot k_B T.
\end{aligned}
$$

The diffusion matrices become, using $\gamma(\varepsilon) = 2m^*\varepsilon/\hbar^2$ and $\phi(x, \varepsilon) = \phi_0(x)\sqrt{\varepsilon}$,

$$
\begin{aligned}
\mathcal{D}_{ij}(\mu, T) &= \frac{2q}{3m^*} e^{q\mu/k_B T} \int_0^\infty \frac{\varepsilon^{i+j-1}}{\phi(x, \varepsilon)} e^{-\varepsilon/k_B T} \, d\varepsilon \cdot \mathrm{Id} \\
&= \frac{2q}{3m^* \phi_0(x)} e^{q\mu/k_B T} \int_0^\infty \varepsilon^{i+j-3/2} e^{-\varepsilon/k_B T} \, d\varepsilon \cdot \mathrm{Id},
\end{aligned}
$$

and from the relations

$$\int_0^\infty \varepsilon^{1/2} e^{-\varepsilon/k_B T}\, d\varepsilon = \frac{1}{\sqrt{2}} (k_B T)^{3/2} \int_0^\infty z^2 e^{-z^2/2}\, dz = \frac{\sqrt{\pi}}{2}(k_B T)^{3/2},$$

$$\int_0^\infty \varepsilon^{3/2} e^{-\varepsilon/k_B T}\, d\varepsilon = \frac{1}{\sqrt{2}^3} (k_B T)^{5/2} \int_0^\infty z^4 e^{-z^2/2}\, dz = \frac{3}{2} \cdot \frac{\sqrt{\pi}}{2}(k_B T)^{5/2},$$

$$\int_0^\infty \varepsilon^{5/2} e^{-\varepsilon/k_B T}\, d\varepsilon = \frac{1}{\sqrt{2}^5} (k_B T)^{7/2} \int_0^\infty z^6 e^{-z^2/2}\, dz = \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{\sqrt{\pi}}{2}(k_B T)^{7/2},$$

we conclude

$$
\begin{aligned}
\mathcal{D}(\mu, T) &= \frac{\sqrt{\pi} q}{3 m^* \phi_0(x)} (k_B T)^{3/2} e^{q\mu/k_B T}
\begin{pmatrix}
\mathrm{Id} & \frac{3}{2}(k_B T)\mathrm{Id} \\
\frac{3}{2}(k_B T)\mathrm{Id} & \frac{15}{4}(k_B T)^2\mathrm{Id}
\end{pmatrix} \\
&= \frac{q\hbar^3 n(\mu, T)}{3\sqrt{8}\pi (m^*)^2 \phi_0(x)}
\begin{pmatrix}
\mathrm{Id} & \frac{3}{2}(k_B T)\mathrm{Id} \\
\frac{3}{2}(k_B T)\mathrm{Id} & \frac{15}{4}(k_B T)^2\mathrm{Id}
\end{pmatrix} \in \mathbb{R}^{6\times 6}.
\end{aligned}
$$

Notice that $\det \mathcal{D} > 0$ as long as $n > 0$ and $T > 0$, and the energy-transport equations are of parabolic type. Moreover, the matrix $\mathcal{D}(\mu, T)$ can be identified by a $\mathbb{R}^2 \times \mathbb{R}^2$ matrix.

Example 5.11 (Fokker-Planck relaxation term)
The relaxation term (5.73) can be written explicitly in terms of μ and T using the Fokker-Planck approximation (5.64) with $A(\varepsilon) = \phi_0 \sqrt{\varepsilon} N(\varepsilon)^2$ and assuming the hypotheses of Example 5.10. Then

$$
\begin{aligned}
W(\mu, T, T_L) &= \int_0^\infty \frac{\partial}{\partial \varepsilon}\left[A(\varepsilon)\left(F_{\mu, T} + k_B T_L \frac{\partial F_{\mu, T}}{\partial \varepsilon} \right)\right] \varepsilon\, d\varepsilon \\
&= \phi_0 e^{q\mu/k_B T} \int_0^\infty \frac{\partial}{\partial \varepsilon}\left[\sqrt{\varepsilon} N(\varepsilon)^2 e^{-\varepsilon/k_B T}\left(1 - \frac{T_L}{T} \right)\right] \varepsilon\, d\varepsilon \\
&= -\phi_0 e^{q\mu/k_B T}\left(1 - \frac{T_L}{T} \right) \int_0^\infty \sqrt{\varepsilon} N(\varepsilon)^2 e^{-\varepsilon/k_B T}\, d\varepsilon \\
&= -4\pi^2 \phi_0 \left(\frac{2m^*}{\hbar^2} \right)^3 e^{q\mu/k_B T}\left(1 - \frac{T_L}{T} \right) \int_0^\infty \varepsilon^{3/2} e^{-\varepsilon/k_B T}\, d\varepsilon \\
&= -3\pi^{5/2} \phi_0 \left(\frac{2m^*}{\hbar^2} \right)^3 e^{q\mu/k_B T}\left(1 - \frac{T_L}{T} \right)(k_B T)^{5/2} \\
&= -3\pi \phi_0 \left(\frac{2m^*}{\hbar^2} \right)^{3/2} n(\mu, T) \cdot k_B(T - T_L) \\
&= \frac{3}{2} \frac{n(\mu, T) k_B(T - T_L)}{\tau_0},
\end{aligned}
$$

where

$$\tau_0 = \frac{1}{2\pi \phi_0}\left(\frac{\hbar^2}{2m^*} \right)^{3/2}$$

is called *relaxation time*.

In this chapter we have derived four semiconductor models: the hydrodynamic equations, the SHE model, the energy-transport equations, and the drift-diffusion model. In Figure 5.1 we summarize the derivations of the models and their relations.
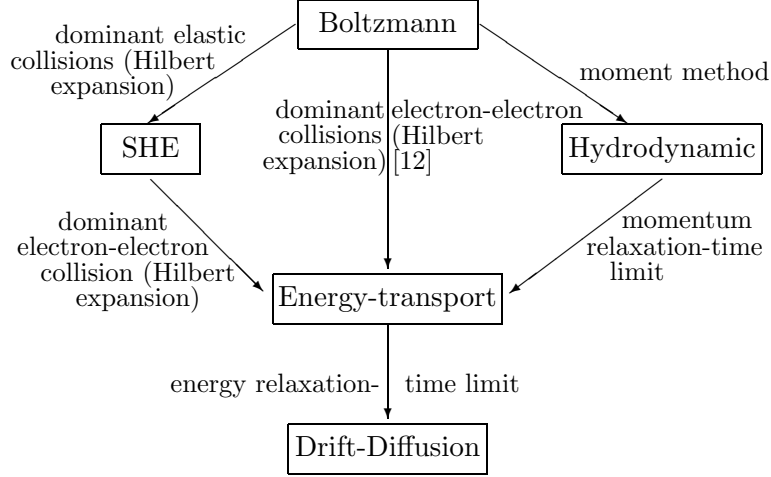

Figure 5.1: Hierarchy of classical semiconductor models. The drift-diffusion model can be directly derived from the Boltzmann equation via a moment method or from the hydrodynamic model in the combined momentum and energy relaxation-time limit.

5.6 Relaxation-time limits

In this section we will derive the drift-diffusion and energy-transport equations from the (full) hydrodynamic model by performing so-called relaxation-time limits. We recall the full hydrodynamic model:

$$\partial_t n - \frac{1}{q} \operatorname{div} J = 0, \tag{5.80}$$

$$\partial_t J - \frac{1}{q} \operatorname{div} \left(\frac{J \otimes J}{n} \right) - \frac{q k_B}{m} \nabla (nT) + \frac{q^2}{m} n \nabla V = -\frac{J}{\tau_p}, \tag{5.81}$$

$$\partial_t (ne) - \frac{1}{q} \operatorname{div} \left[J(e + k_B T) \right] + J \cdot \nabla V - \operatorname{div} \left(\frac{\tau_p k_B^2}{m} \kappa_0 nT \nabla T \right) \tag{5.82}$$

$$= -\frac{n}{\tau_w} \left(e - \frac{d}{2} k_B T_0 \right) \quad \text{in } \Omega \subset \mathbb{R}^d,$$

where τ_p and τ_w are the momentum energy relaxation times, respectively, and

$$e = \frac{m}{q^2} \frac{|J|^2}{2n^2} + \frac{d}{2} k_B T$$

100

is the total energy. The coupling of the electrostatic potential V to the electron density through the Poisson equation needs not to be considered in the subsequent considerations. Therefore, V will be treated as a given function.

We scale the equations (5.80)-(5.82) by introducing reference values for the physical variables (see Table 5.1). Here, $L > 0$ is a reference length and $\overline{C}$ is a reference value for the particle density. When the Poisson equation is added to the above system of equations, $\overline{C}$ can be chosen, for instance, as the supremum of the doping profile $C = C(x)$. 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 needed to cross the domain in time τ_0, τ_p, respectively:

$$k_B T_0 = \sqrt{m\left(\frac{L}{\tau_0}\right)^2}\sqrt{m\left(\frac{L}{\tau_p}\right)^2}.$$

The reference velocity u_0 is equal to the velocity needed to cross the domain in time τ_0.

variable	reference value	definition
x	L	e.g. $\mathrm{diam}(\Omega)$
t	τ_0	$\tau_0 = L^2 m / k_B T \tau_p$
$n(x,t)$	$\overline{C}$	e.g. $\sup_\Omega |C|$
$T(x,t)$	T_0	lattice temperature
$V(x,t)$	U_T	$U_T = k_B T_0 / q$
$J(x,T)$	$q\overline{C}u_0$	$u_0 = L/\tau_0$
$e(x,t)$	$k_B T_0$	thermal energy

Table 5.1: Reference values for the physical variables.

With these reference values we can define the non-dimensional variables

$$\begin{aligned}
x &= L x_s, & T &= T_0 T_s, \\
t &= \tau_0 t_s, & V &= U_T V_s, \\
n &= \overline{C} n_s, & J &= q\overline{C}u_0 \cdot J_s.
\end{aligned}$$

Replacing the dimensional variables in (5.80)-(5.82) by the scaled ones we obtain the scaled equations (writing x, t, n etc. instead of x_s, t_s, n_s etc.)

$$\partial_t n - \operatorname{div} J = 0, \tag{5.83}$$

$$\frac{\tau_p}{\tau_0}\partial_t J - \frac{\tau_p}{\tau_0}\operatorname{div}\left(\frac{J \otimes J}{n}\right) - \nabla(nT) + n\nabla V = -J, \tag{5.84}$$

$$\partial_t(ne) - \operatorname{div}\left[J(e+T)\right] + J \cdot \nabla V - \operatorname{div}\left(\kappa_0 nT\nabla T\right) = -\frac{\tau_0}{\tau_w}\left(e - \frac{d}{2}\right), \tag{5.85}$$

and the scaled energy becomes

$$e = \frac{\tau_p}{\tau_0}\frac{|J|^2}{2n^2} + \frac{d}{2}T. \tag{5.86}$$

We consider the following physical situations:

Drift-diffusion limit in the hydrodynamic equations. The kinetic energy needed to cross the domain in time τ_0 is assumed to be much smaller than the thermal energy, and the momentum and energy relaxation times are assumed to be of the same order. This means

$$1 \gg \frac{mu_0^2}{k_B T_0} = \frac{L^2 m}{k_B T_0 \tau_0^2} = \frac{\tau_p}{\tau_0} \quad \text{and} \quad O(1) = \frac{\tau_p}{\tau_w} = \frac{\tau_p}{\tau_0} \cdot \frac{\tau_0}{\tau_w}.$$

The second condition implies

$$\frac{\tau_0}{\tau_w} \gg 1.$$

The drift-diffusion limit means that we perform (formally) the limits

$$\frac{\tau_p}{\tau_0} \to 0 \quad \text{and} \quad \frac{\tau_w}{\tau_0} \to 0.$$

Then (5.85)-(5.86) become in the limit

$$e = \frac{d}{2}T = \frac{d}{2}.$$

This means that the limit temperature equals one. Scaling back to the unscaled variables the physical limit temperature equals the lattice temperature. From (5.83)-(5.84) we conclude

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

In the original variables this reads

$$\partial_t n - \frac{1}{q}\operatorname{div} J = 0, \quad J = q\mu_n(U_T\nabla n - n\nabla V), \tag{5.87}$$

where

$$\mu_n = \frac{q\tau_p}{m}$$

is called the *electron mobility.* The equations (5.87) together with the Poisson equation are the drift-diffusion equations derived in Sections 5.1 and 5.2.

Energy-transport limit in the hydrodynamic equations. The kinetic energy needed to cross the domain in time τ_0 is assumed to be much smaller than

the thermal energy, and the reference time is of the same order as the energy relaxation time, i.e.

$$\frac{\tau_p}{\tau_0} \ll 1 \quad \text{and} \quad O(1) = \frac{\tau_0}{\tau_w} = \frac{\tau_0}{\tau_p} \cdot \frac{\tau_p}{\tau_w},$$

which implies

$$\frac{\tau_p}{\tau_w} \ll 1.$$

In the energy-transport limit we only perform (formally) the limit

$$\frac{\tau_p}{\tau_0} \to 0.$$

Equations (5.83)-(5.86) become in the limit

$$\partial_t n - \operatorname{div} J = 0, \quad J = \nabla(nT) - n\nabla V, \tag{5.88}$$

$$\partial_t \left(\frac{d}{2} nT \right) - \operatorname{div} \left(\frac{d+2}{2} JT + \kappa_0 nT\nabla T \right) + J \cdot \nabla V = -\frac{d}{2}\frac{\tau_0}{\tau_w} n(T-1), \tag{5.89}$$

or, in the original variables,

$$\partial_t n - \frac{1}{q}\operatorname{div} J = 0, \quad J = q\mu_n \left(U_T \nabla(nT) - n\nabla V \right),$$

$$\partial_t \left(\frac{d}{2} k_B Tn \right) - \operatorname{div} S + J \cdot \nabla V = -\frac{d}{2}\frac{nk_B(T-T_0)}{\tau_w},$$

$$S = \frac{J}{q} k_B T + \frac{\tau_p k_B^2}{m} \kappa_0 nT\nabla T.$$

In order to see that this system of equations can be written as an energy-transport model, we introduce the chemical potential μ by

$$n = N(k_B T)^{d/2} \, e^{q\mu/k_B T},$$

where N is the effective density of states (see Example 5.10). A lengthy computation shows that the above system of equations can be written equivalently in the variables $q\mu/k_B T$ and $-1/k_B T$ as

$$\partial_t n - \frac{1}{q}\operatorname{div} J = 0, \quad \partial_t \mathcal{E} - \operatorname{div} S + J \cdot \nabla V = W(\mu, T),$$

where

$$\mathcal{E} = \frac{d}{2} k_B T \cdot n, \quad W(\mu, T) = -\frac{d}{2}\frac{nk_B(T-T_0)}{\tau_w}$$

and

$$J = D_{11}\nabla \left(\frac{q\mu}{k_B T} \right) + D_{12}\nabla \left(\frac{-1}{k_B T} \right) - D_{11}\frac{q\nabla V}{k_B T},$$

$$S = D_{21}\nabla \left(\frac{q\mu}{k_B T} \right) + D_{22}\nabla \left(\frac{-1}{k_B T} \right) - D_{21}\frac{q\nabla V}{k_B T},$$

103

with the diffusion matrix

$$(D_{ij})_{ij} = \frac{q\tau_p}{m} n k_B T \begin{pmatrix} 1 & \frac{d+2}{2} k_B T \\ \frac{d+2}{2} k_B T & \left(\left(\frac{d+2}{2} \right)^2 + \kappa_0 \right) (k_B T)^2 \end{pmatrix}.$$

The above equations form an energy-transport model as discussed in Section 5.5.

Notice that the determinant of this matrix is positive for positive particle densities and temperatures if and only if $\kappa_0 > 0$. Hence, the heat conductivity term $\mathrm{div}\,(\kappa_0 n T \nabla T)$ is necessary to obtain a uniformly parabolic set of equations. We recall, however, that this term is introduced heuristically, and no justification is available.

Drift-diffusion limit in the energy-transport equations. It is possible to derive the drift-diffusion model from the energy-transport equations. For this we consider the energy-transport model in the scaled form (5.88)-(5.89) and assume that the reference time τ_0 is much larger than the energy relaxation time:

$$\frac{\tau_0}{\tau_w} \gg 1.$$

This is exactly the second condition needed in the drift-diffusion limit. The formal limit $\tau_w/\tau_0 \to 0$ in (5.88)-(5.89) yields $T = 1$ and, after scaling back to the original variables, the drift-diffusion equations (5.87).

The relaxation-time limits studied in this section are summarized in Figure 5.2.
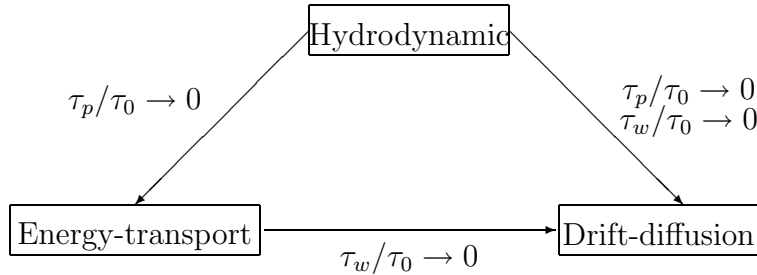

Figure 5.2: Relaxation-time limits in the hydrodynamic and energy-transport models

5.7 The extended hydrodynamic model

The heuristic addition of the heat flux $-\mathrm{div}\,(\kappa \nabla T)$ in the energy equation of the hydrodynamic model derived in Section 5.4 is not very satisfactory. In this section

we present an alternative way of deriving hydrodynamic models based on the so-called *maximum entropy principle*. This method provides a set of closed moment equations consisting of the conservation laws of mass, momentum, energy and energy flux, which we refer to as the *extended hydrodynamic model*. In this model, no heuristic terms need to be included. The maximum entropy method has been first used by Anile and Pennisi [2] for the derivation of charge transport equations for semiconductors. It is based on principles of extended thermodynamics [35].

We start with the semiclassical Boltzmann equation in $\mathbb{R}^3$

$$\partial_t f + v(k) \cdot \nabla_x f - \frac{q}{\hbar} E \cdot \nabla_k f = Q(f), \quad x \in \mathbb{R}^3, \ k \in B, \ t > 0, \qquad (5.90)$$

where $v(k) = \hbar^{-1} \nabla_k \varepsilon(k)$. The collision operator $Q(f)$ is assumed to satisfy

$$\int_B Q(f) \, dk = 0.$$

In order to simplify the presentation, we assume parabolic bands

$$\varepsilon(k) = \frac{\hbar^2 |k|^2}{2m^*}, \quad k \in B = \mathbb{R}^3,$$

such that $v(k) = \hbar k/m^*$. The maximum entropy principle also works for general band diagrams (see [3] for details). We define the first moments

$$\langle M_i \rangle = \int_{\mathbb{R}^3} f(x, k, t) \, M_i(k) \, dk,$$

where

$$M_0(k) = 1, \quad M_1(k) = k, \quad M_2(k) = |k|^2, \quad M_3(k) = k|k|^2.$$

As in Section 5.4 we obtain the moment equations by multiplying the Boltzmann equation (5.90) by M_i and integrating over $k \in \mathbb{R}^3$:

$$\partial_t \langle 1 \rangle + \frac{\hbar}{m^*} \mathrm{div}\,_x \langle k \rangle = 0, \qquad (5.91)$$

$$\partial_t \langle k \rangle + \frac{\hbar}{m^*} \mathrm{div}\,_x \langle k \otimes k \rangle + \frac{q}{\hbar} E \langle 1 \rangle = \int_{\mathbb{R}^3} Q(f) k \, dk, \qquad (5.92)$$

$$\partial_t \langle |k|^2 \rangle + \frac{\hbar}{m^*} \mathrm{div}\,_x \langle k|k|^2 \rangle + \frac{2q}{\hbar} E \cdot \langle k \rangle = \int_{\mathbb{R}^3} Q(f) |k|^2 \, dk, \qquad (5.93)$$

$$\partial_t \langle k|k|^2 \rangle + \frac{\hbar}{m^*} \mathrm{div}\,_x \langle k \otimes k|k|^2 \rangle + \frac{q}{\hbar} E \cdot \langle \nabla_k \otimes (k|k|^2) \rangle \qquad (5.94)$$

$$= \int_{\mathbb{R}^3} Q(f) k|k|^2 \, dk,$$

since, using integrating by parts

$$
\begin{aligned}
-\frac{q}{\hbar}E \cdot \int_{\mathbb{R}^3} \nabla_k f |k|^2 \, dk &= \frac{q}{\hbar}E \cdot \int_{\mathbb{R}^3} f \nabla_k |k|^2 \, dk \\
&= \frac{2q}{\hbar}E \cdot \int_{\mathbb{R}^3} f k \, dk \\
&= \frac{2q}{\hbar}E \cdot \langle k \rangle.
\end{aligned}
$$

We can write these moment equations in a more compact form as

$$
\partial_t \langle M_i \rangle + \operatorname{div}_x F_i + E \cdot G_i = P_i, \quad i = 0, 1, 2, 3,
$$

where F_i, G_i and P_i are defined in an obvious way. The moments $\langle M_i \rangle$ can be considered as the fundamental variables since they have a direct physical interpretation. Indeed, we have:

$$
\begin{aligned}
n &= \langle M_0 \rangle && \text{electron density,} \\
J &= -\frac{q\hbar}{m^*}\langle M_1 \rangle && \text{electron current density,} \\
ne &= \frac{\hbar^2}{2m^*}\langle M_2 \rangle && \text{electron energy density,} \\
S &= -\frac{\hbar^3}{2(m^*)^2}\langle M_3 \rangle && \text{energy flux.}
\end{aligned}
$$

The goal now is to express F_i, G_i and P_i in terms of the moments $\langle M_i \rangle$. Since

$$
F_0 = \langle M_1 \rangle, \quad F_2 = \langle M_3 \rangle, \quad G_0 = 0, \quad G_1 = \langle M_0 \rangle, \quad G_2 = \langle M_1 \rangle,
$$

we only have to compute F_1, F_3, G_3 and the production terms $P_0, \dots, P_3$. For this we choose a distribution function f_0 which can be used to evaluate the unknown moments, i.e., we set

$$
\begin{aligned}
F_1 &= \langle k \otimes k \rangle = \int_{\mathbb{R}^3} f_0 (k \otimes k) \, dk, \\
F_3 &= \langle k \otimes k |k|^2 \rangle = \int_{\mathbb{R}^3} f_0 (k \otimes k) |k|^2 \, dk,
\end{aligned}
$$

and so on. The function f_0 is defined as the extremal of the entropy functional

$$
s(f) = \int_{\mathbb{R}^3} f (\log f - 1) \, dk
$$

under the constraints that it yields exactly the known moments $\langle M_i \rangle$. This means that we have to minimize s (or, equivalently, to *maximize* the physical entropy $-s$) under the constraints

$$
\int_{\mathbb{R}^3} M_i(k) f_0(k) \, dk = \langle M_i \rangle, \quad i = 0, \dots, 3. \tag{5.95}
$$

Using Lagrange multipliers, we have to analyze the functional

$$H(f) = \int_{\mathbb{R}^3} f(\log f - 1) \, dk + \sum_{i=0}^{3} \lambda_i \left(\int_{\mathbb{R}^3} M_i \, f \, dk - \langle M_i \rangle \right).$$

A simple computation shows that

$$\frac{dH(f)}{df}(g) = \int_{\mathbb{R}^3} g \log f \, dk + \sum_{i=0}^{3} \lambda_i \int_{\mathbb{R}^3} M_i \, g \, dk.$$

The necessary condition

$$\frac{dH(f_0)}{df}(g) = 0 \quad \forall g$$

for the extremal f_0 leads to

$$\int_{\mathbb{R}^3} g \left(\log f_0 + \sum_{i=0}^{3} \lambda_i M_i \right) dk = 0 \quad \forall g$$

and hence

$$f_0(k) = \exp \left(- \sum_{i=0}^{3} \lambda_i M_i(k) \right), \quad k \in \mathbb{R}^3. \tag{5.96}$$

The Lagrange multipliers are obtained by inserting this expression into (5.95):

$$\langle M_i \rangle = \int_{\mathbb{R}^3} M_i(k) \exp \left(- \sum_{i=0}^{3} \lambda_i M_i(k) \right) dk. \tag{5.97}$$

This is an implicit equation for $\lambda_0, \ldots, \lambda_3$. Once we have found the values of $\lambda_0, \ldots, \lambda_3$ (which are functions of M_i and $\langle M_i \rangle$), we can define the maximum entropy distribution function f_0 from (5.96), and inserting the expression for f_0 into the definitions of F_i, G_i and P_i gives finally a closed set of equations.

Unfortunately, the implicit equation (5.97) usually cannot be inverted directly to give explicit expressions for the Lagrange multipliers. If we assume that the distribution function f_0 is in some sense not far from the thermal equilibrium distribution function

$$f_{\text{eq}}(k) = \exp \left(\frac{\varepsilon_F}{k_B T} - \frac{\varepsilon(k)}{k_B T} \right) \tag{5.98}$$

(see Lemma 3.2(3) for Boltzmann statistics), we can develop the Lagrange multipliers in terms of a small parameter $\delta > 0$. Comparing (5.98) with (5.96) suggests that the moments $\langle M_1 \rangle$ and $\langle M_3 \rangle$ (which appear in (5.98)) are of order 1 whereas the moments $\langle M_2 \rangle$ and $\langle M_4 \rangle$ (which do not appear in (5.98)) are of order $\delta \ll 1$.

Our mean assumption is now that we can develop λ_i in terms of the moments $\langle M_i \rangle$ in the following way:

$$
\begin{aligned}
\lambda_i &= a_{i1} + a_{i2}\,\delta\langle M_1 \rangle \cdot \delta\langle M_3 \rangle + a_{i3}\,\delta\langle M_1 \rangle \cdot \delta\langle M_1 \rangle \\
&\quad + a_{i4}\,\delta\langle M_3 \rangle \cdot \delta\langle M_3 \rangle + O(\delta^3), \quad i = 0, 2, \\
\lambda_j &= a_{j1} \cdot \delta\langle M_1 \rangle + a_{j2} \cdot \delta\langle M_3 \rangle + O(\delta^3), \quad j = 1, 3,
\end{aligned}
$$

where we replaced $\langle M_1 \rangle$ and $\langle M_3 \rangle$ by the rescaled moments $\delta\langle M_1 \rangle$ and $\delta\langle M_3 \rangle$, respectively, and where the functions a_{ij} depend on $\langle M_0 \rangle$ and $\langle M_2 \rangle$. The facts that λ_0 and λ_2 do not contain first-order terms in δ and that λ_1 and λ_3 do not contain zero- and second-order terms in δ can be regarded here as an assumption. The argument can be made more rigorous by assuming that f_0 is anisotrop and δ is a small anisotropy parameter (see [3]). The above equations can be written, up to second order in δ, as

$$
\begin{aligned}
\lambda_0 &= \lambda_0^{(0)} + \delta^2 \lambda_0^{(2)}, \\
\lambda_1 &= \delta \lambda_1^{(1)}, \\
\lambda_2 &= \lambda_2^{(0)} + \delta^2 \lambda_2^{(2)}, \\
\lambda_3 &= \delta \lambda_3^{(1)}.
\end{aligned}
$$

We insert this expansion into (5.96) and use the approximation

$$
e^{-x} = 1 - x + \frac{x^2}{2} + O(x^3) \quad (x \to 0)
$$

to obtain, up to second order in δ,

$$
\begin{aligned}
f_0 &= \exp\left[-(\lambda_0^{(0)} + \delta^2 \lambda_0^{(2)}) - \delta\lambda_1^{(1)} \cdot k - (\lambda_2^{(0)} + \delta^2 \lambda_2^{(2)})|k|^2 - \delta\lambda_3^{(1)} \cdot k|k|^2 \right] \\
&= \exp\left[-(\lambda_0^{(0)} + \lambda_2^{(0)}|k|^2) \right] \exp\left[-\delta(\lambda_1^{(1)} + \lambda_3^{(1)}|k|^2) \cdot k - \delta^2(\lambda_0^{(2)} + \lambda_2^{(2)}|k|^2) \right] \\
&= \exp\left[-(\lambda_0^{(0)} + \lambda_2^{(0)}|k|^2) \right] \left\{ 1 - \delta(\lambda_1^{(1)} + \lambda_3^{(1)}|k|^2) \cdot k - \delta^2(\lambda_0^{(2)} + \lambda_2^{(2)}|k|^2) \right. \\
&\quad \left. + \frac{\delta^2}{2}((\lambda_1^{(1)} + \lambda_3^{(1)}|k|^2) \cdot k)^2 \right\}.
\end{aligned}
\tag{5.99}
$$

Using this expression in the constraints (5.95) allows to compute the Lagrange multipliers.

First we compute $\lambda_0^{(0)}$ and $\lambda_2^{(0)}$. Setting $A = \lambda_1^{(1)} + \lambda_3^{(1)}|k|^2$ and $B = \lambda_0^{(2)} + \lambda_2^{(2)}|k|^2$, we infer

$$
n = \langle M_0 \rangle = e^{-\lambda_0^{(0)}} \int_{\mathbb{R}^3} e^{-\lambda_2^{(0)}|k|^2} \left[1 - \delta A \cdot k - \delta^2 B + \frac{\delta^2}{2}(A \cdot k)^2 \right] \, dk.
$$

Equating equal powers of δ gives

$$
\begin{aligned}
n &= e^{-\lambda_0^{(0)}} \int_{\mathbb{R}^3} e^{-\lambda_2^{(0)}|k|^2}\, dk \\
&= (2\lambda_2^{(0)})^{-3/2} e^{-\lambda_0^{(0)}} \int_{\mathbb{R}^3} e^{-|u|^2/2}\, du \\
&= \pi^{3/2}(\lambda_2^{(0)})^{-3/2} e^{-\lambda_0^{(0)}}. \tag{5.100}
\end{aligned}
$$

Similarly, the expression

$$
\begin{aligned}
ne &= \frac{\hbar^2}{2m^*}\langle M_2\rangle \\
&= \frac{\hbar^2}{2m^*} e^{-\lambda_0^{(0)}} \int_{\mathbb{R}^3} e^{-\lambda_2^{(0)}|k|^2}|k|^2 \left[1 - \delta A\cdot k - \delta^2 B + \frac{\delta^2}{2}(A\cdot k)^2\right]\, dk
\end{aligned}
$$

yields

$$
\begin{aligned}
ne &= \frac{\hbar^2}{2m^*} e^{-\lambda_0^{(0)}} \int_{\mathbb{R}^3} e^{-\lambda_2^{(0)}|k|^2}|k|^2\, dk \\
&= \frac{\hbar^2}{2m^*}(2\lambda_2^{(0)})^{-5/2} e^{-\lambda_0^{(0)}} \int_{\mathbb{R}^3} e^{-|u|^2/2}|u|^2\, du \\
&= \frac{\hbar^2}{2m^*}(2\lambda_2^{(0)})^{-5/2} e^{-\lambda_0^{(0)}} \cdot 3(2\pi)^{3/2} \\
&= \frac{3\hbar^2\pi^{3/2}}{4m^*}(\lambda_2^{(0)})^{-5/2} e^{-\lambda_0^{(0)}}. \tag{5.101}
\end{aligned}
$$

The equations (5.100) and (5.101) can be used to compute $\lambda_0^{(0)}$ and $\lambda_2^{(0)}$. From

$$
e = \frac{ne}{n} = \frac{3}{4}\frac{\hbar^2}{m^*}(\lambda_3^{(0)})^{-1}
$$

follows

$$
\lambda_2^{(0)} = \frac{3\hbar^2}{4m^*e}, \quad \lambda_0^{(0)} = -\ln n + \frac{3}{2}\ln\left(\frac{4\pi m^*}{3\hbar^2}e\right).
$$

Now we compute the remaining Lagrangian multipliers $\lambda_1^{(1)}$, $\lambda_3^{(1)}$, $\lambda_0^{(2)}$, and $\lambda_2^{(2)}$. Recalling that we have replaced $\langle M_1\rangle$ and $\langle M_3\rangle$ by the rescaled expressions $\delta\langle M_1\rangle$ and $\delta\langle M_3\rangle$ we obtain

$$
\begin{aligned}
\delta J &= -\frac{q\hbar}{m^*}\delta\langle M_1\rangle \\
&= -\frac{q\hbar}{m^*} e^{-\lambda_0^{(0)}} \int_{\mathbb{R}^3} e^{-\lambda_2^{(0)}|k|^2} k\left[1 - \delta A\cdot k - \delta^2 B + \frac{\delta^2}{2}(A\cdot k)^2\right]\, dk
\end{aligned}
$$

and

$$\delta S = -\frac{\hbar^3}{2(m^*)^2}\delta\langle M_3\rangle$$

$$= -\frac{\hbar^3}{2(m^*)^2}e^{-\lambda_0^{(0)}}\int_{\mathbb{R}^3}e^{-\lambda_2^{(0)}|k|^2}k|k|^2\left[1-\delta A\cdot k-\delta^2 B+\frac{\delta^2}{2}(A\cdot k)^2\right]dk.$$

Equating equal powers of δ leads to

$$J = -\frac{q\hbar}{m^*}e^{-\lambda_0^{(0)}}\int_{\mathbb{R}^3}e^{-\lambda_2^{(0)}|k|^2}k\left(\lambda_1^{(1)}+\lambda_3^{(1)}|k|^2\right)\cdot k\,dk, \qquad (5.102)$$

$$0 = \int_{\mathbb{R}^3}e^{-\lambda_2^{(0)}|k|^2}k\left[\frac{1}{2}\left((\lambda_1^{(1)}+\lambda_3^{(1)}|k|^2)\cdot k\right)^2-(\lambda_0^{(2)}+\lambda_2^{(2)}|k|^2)\right]dk, \quad (5.103)$$

and

$$S = -\frac{\hbar^3}{2(m^*)^2}e^{-\lambda_0^{(0)}}\int_{\mathbb{R}^3}e^{-\lambda_2^{(0)}|k|^2}k|k|^2\left(\lambda_1^{(1)}+\lambda_3^{(1)}|k|^2\right)\cdot k\,dk, \qquad (5.104)$$

$$0 = \int_{\mathbb{R}^3}e^{-\lambda_2^{(0)}|k|^2}k|k|^2\left[\frac{1}{2}\left((\lambda_1^{(1)}+\lambda_3^{(1)}|k|^2)\cdot k\right)^2-(\lambda_0^{(2)}+\lambda_2^{(2)}|k|^2)\right]dk \quad (5.105)$$

The multipliers $\lambda_1^{(1)}$ and $\lambda_3^{(1)}$ can be computed by solving the linear system given by (5.102) and (5.104). Then the remaining parameters $\lambda_0^{(2)}$ and $\lambda_2^{(2)}$ are solutions of the linear system (5.103) and (5.105). The general structure of the multipliers is as follows:

$$\lambda_1^{(1)} = a_{11}(n,e)J+a_{12}(n,e)S,$$

$$\lambda_3^{(1)} = a_{31}(n,e)J+a_{32}(n,e)S,$$

$$\lambda_0^{(2)} = a_{01}(n,e)J\cdot J+a_{02}(n,e)J\cdot S+a_{03}(n,e)S\cdot S,$$

$$\lambda_2^{(2)} = a_{21}(n,e)J\cdot J+a_{22}(n,e)J\cdot S+a_{23}(n,e)S\cdot S.$$

Explicit expressions for the coefficients are given in [3, Section 3.4].

With the above relations for the Lagrange multipliers, the maximum entropy distribution function f_0 in (5.99) becomes

$$f_0 = n\left(\frac{3\hbar^2}{4\pi m^*e}\right)^{3/2}\exp\left(-\frac{3\hbar^2|k|^2}{4m^*e}\right) \qquad (5.106)$$

$$\times\left\{1-(a_{21}J+a_{22}S)\cdot k-(a_{11}+a_{31}|k|^2)J\cdot J-(a_{12}+a_{32}|k|^2)J\cdot S\right.$$

$$\left.-(a_{13}+a_{33}|k|^2)S\cdot S+\frac{\delta^2}{2}\left[(a_{21}+a_{41}|k|^2)J\cdot k+(a_{22}+a_{42}|k|^2)S\cdot k\right]^2\right\}.$$

Hence, the unknown moments are given by

$$F_1 = \langle k\otimes k\rangle=\int_{\mathbb{R}^3}f_0(k\otimes k)\,dk=F_1(n,e,J,S),$$

$$F_3 = \langle k\otimes k|k|^2\rangle=F_3(n,e,J,S),$$

$$G_3 = \langle\nabla_k\otimes(k|k|^2)\rangle=\langle|k|^2\rangle\mathrm{Id}+2\langle k\otimes k\rangle=\frac{m^*}{\hbar^2}ne\cdot\mathrm{Id}+F_1(n,e,J,S).$$

From (5.106) we see that the dependence of F_1 and F_3 on J, S is at most quadratic, i.e., F_1 and F_3 are in fact functions of $J \cdot J$, $J \otimes J$, $J \cdot S$, $S \otimes S$ and so on. The precise formulas for F_1 and F_3 are given in [3, (124), (125)].

The production terms P_i are approximated by

$$P_1 = \int_{\mathbb{R}^3} Q(f_0) k \, dk,$$

$$P_2 = \int_{\mathbb{R}^3} Q(f_0) |k|^2 \, dk,$$

$$P_3 = \int_{\mathbb{R}^3} Q(f_0) k |k|^2 \, dk.$$

In order to get explicit formulas for P_i as functions of n, J, e and S, the collision operator $Q(f)$ has to be specified. In [3] the production terms are computed for (acoustic or optical) phonon scattering or for scattering with impurities; we refer to this work for the precise expressions.

Finally, we want to compare the extended hydrodynamic model (5.91)-(5.94) with the hydrodynamic equations of Section 5.4. We need to rewrite the equations (5.91)-(5.94) by introducing the electron temperature. For this, introduce the deviations of the microscopic velocity $v(k)$ from the averaged macroscopic velocity $u = -J/qn$:

$$\bar{u}(k) = v(k) + u = \frac{\hbar k}{m^*} + u. \tag{5.107}$$

The electron temperature T and the heat flux σ are defined as moments of the deviations:

$$n k_B T = \frac{m^*}{3} \int_{\mathbb{R}^3} f \, |\bar{u}|^2 \, dk,$$

$$n \sigma = \frac{m^*}{2} \int_{\mathbb{R}^3} f \, |\bar{u}|^2 \bar{u} \, dk.$$

Then we can reformulate the energy density, using (5.107), as

$$
\begin{aligned}
ne &= \frac{\hbar}{2m^*} \int_{\mathbb{R}^3} f \, |k|^2 \, dk \\
&= \frac{m^*}{2} \int_{\mathbb{R}^3} f \left(|u|^2 - 2u \cdot \bar{u} + |\bar{u}|^2 \right) dk \qquad \text{(using (5.107))} \\
&= \frac{m^*}{2} n |u|^2 + \frac{3}{2} n k_B T,
\end{aligned}
$$

since

$$\int_{\mathbb{R}^3} f \, \bar{u} \, dk = \int_{\mathbb{R}^3} f \left(\frac{\hbar k}{m^*} - u \right) dk = -\frac{J}{q} - nu = 0. \tag{5.108}$$

The energy flux can be written as

$$
\begin{aligned}
S &= -\frac{\hbar^3}{2(m^*)^2} \int_{\mathbb{R}^3} f k |k|^2 \, dk \\
&= -\frac{m^*}{2} \int_{\mathbb{R}^3} f \, (\bar{u} - u)(|\bar{u}|^2 - 2u \cdot \bar{u} + |u|^2) \, dk \quad \text{(using (5.107)}\\
&= -\frac{m^*}{2} \int_{\mathbb{R}^3} f \, (\bar{u}|\bar{u}|^2 - u|\bar{u}|^2 + 2(u \cdot \bar{u})\bar{u} - u|u|^2) \, dk \quad \text{(using (5.108)}\\
&= -n\sigma + \frac{3}{2} n k_B T u - m^* \int_{\mathbb{R}^3} f \, (u \cdot \bar{u})\bar{u} + \frac{m^*}{2} n u |u|^2.
\end{aligned}
$$

Computing

$$
\begin{aligned}
\int_{\mathbb{R}^3} f \, (u \cdot \bar{u})\bar{u}_j \, dk &= \sum_i u_i \int_{\mathbb{R}^3} f \left(\frac{\hbar k_i}{m^*} + u_i \right) \left(\frac{\hbar k_j}{m^*} + u_j \right) \, dk \\
&= \sum_i u_i \left[\left(\frac{\hbar^2}{m^*} \right)^2 \int_{\mathbb{R}^3} f k_i k_j \, dk + n u_i u_j + \frac{\hbar}{m^*} \int_{\mathbb{R}^3} f(k_i u_j + k_j u_i) \, dk \right] \\
&= \left[\left(\frac{\hbar^2}{m^*} \right)^2 \langle k \otimes k \rangle \cdot u + 3 n u |u|^2 \right]_j,
\end{aligned}
$$

we obtain

$$
S = ne - n\sigma + \left(\frac{\hbar}{m^*} \right)^2 \langle k \otimes k \rangle \cdot u + 3 n u |u|^2.
$$

Introducing

$$
\bar{F}_1 = q \left(\frac{\hbar}{m^*} \right)^2 F_1, \quad \bar{F}_3 = \frac{\hbar^4}{2(m^*)^3} F_3
$$

and

$$
\bar{P}_1 = -\frac{q\hbar}{m^*} P_1, \quad \bar{P}_2 = \frac{\hbar^2}{2m^*} P_2, \quad \bar{P}_3 = -\frac{\hbar^3}{2(m^*)^3} P_3,
$$

we can write the system (5.91)-(5.94) as follows:

$$
\partial_t n - \frac{1}{q} \operatorname{div} J = 0,
$$

$$
\partial_t J - \operatorname{div} \bar{F}_1(n, e, S, J) - \frac{q^2}{m^*} E n = \bar{P}_1,
$$

$$
\partial_t(ne) - \operatorname{div} S - E \cdot J = \bar{P}_2,
$$

$$
\partial_t S - \operatorname{div} \bar{F}_3(n, e, S, J) + \frac{1}{2} \left(\frac{q}{m^*} ne \cdot \operatorname{Id} + \bar{F}_1 \right) \cdot E = \bar{P}_3,
$$

where

$$
\begin{aligned}
ne &= \frac{m^* |J|^2}{2q^2 n} + \frac{3}{2} n k_B T, \\
S &= ne - n\sigma + \frac{1}{q} \bar{F}_1(n, e, J, S) - \frac{3J |J|^2}{q^3 n^2},
\end{aligned}
$$

and $\bar{F}_1$, $\bar{F}_3$, $\bar{P}_1$, $\bar{P}_2$ and $\bar{F}_1$ are given explicitly in [3]. The above system of equations is referred to as the *extended hydrodynamic model* for parabolic band diagrams. For non-parabolic band diagrams, similar equations can be derived; we refer to [3] for details. Whereas the independent variables in the system (5.91)-(5.94) are n, J, *ne* and S, the independent variables of the above equations are n, J, T and σ. Notice that in the above derivation, there was no need to introduce heuristic terms.

6 The Drift-Diffusion Equations

In this chapter the drift-diffusion model is studied in more detail. We recall the equations for electrons and holes:

$$\partial_t n - \frac{1}{q}\operatorname{div} J_n = -R(n,p), \tag{6.1}$$

$$\partial_t p + \frac{1}{q}\operatorname{div} J_p = -R(n,p), \tag{6.2}$$

$$J_n = q\mu_n(U_T\nabla n - n\nabla V), \tag{6.3}$$

$$J_p = -q\mu_p(U_T\nabla p + p\nabla V), \tag{6.4}$$

$$\varepsilon_s\Delta V = q(n - p - C), \quad x \in \Omega,\ t > 0, \tag{6.5}$$

with the recombination-generation term

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

Here, $\Omega \subset \mathbb{R}^d$ ($d \geq 1$) is a bounded domain.

6.1 Thermal equilibrium state and boundary conditions

The thermal equilibrium state is a steady state with no current flow, i.e.

$$\partial_t n = \partial_t p = 0 \quad \text{and} \quad J_n = J_p = 0 \quad \text{in } \Omega.$$

This implies $R(n,p) = 0$ or $np = n_i^2$ in Ω and

$$\begin{aligned}
0 &= U_T\nabla n - n\nabla V = n\nabla(U_T\ln n - V), \\
0 &= U_T\nabla p + p\nabla V = p\nabla(U_T\ln p + V).
\end{aligned}$$

It is physically reasonable to assume that the particle densities n and p are positive in Ω. (In fact, this can be proved using maximum principle arguments if the boundary data is positive; see [33]). This yields

$$\alpha := U_T\ln n - V = \text{const.}, \quad \beta := U_T\ln p + V = \text{const.} \tag{6.6}$$

or

$$n = e^{\alpha/U_T}e^{V/U_T}, \quad p = e^{\beta/U_T}e^{-V/U_T} \quad \text{in } \Omega.$$

We wish to determine the constants α and β. For this we use the equation $np = n_i^2$ and the fact that V is determined only up to an additive constant. Then the sum of the two equations in (6.6) yields $\alpha + \beta = U_T\ln(np) = 2U_T\ln n_i$ and, substituting V by $V + \gamma$, where $\gamma := U_T\ln n_i - \alpha$, gives

$$n = e^{(\alpha+\gamma)/U_T}e^{V/U_T} = n_i e^{V/U_T} \tag{6.7}$$

and
$$p = e^{(\beta - \gamma)/U_T} e^{-V/U_T} = e^{(\beta + \alpha - U_T \ln n_i)/U_T} e^{-V/U_T} = n_i e^{-V/U_T}. \qquad (6.8)$$

The electrostatic potential satisfies the semilinear elliptic equation

$$\begin{aligned} \varepsilon_s \Delta V &= q(n - p - C) = q(n_i e^{V/U_T} - n_i e^{-V/U_T} - C) \\ &= q\left(2n_i \sinh \frac{V}{U_T} - C\right) \quad \text{in } \Omega. \end{aligned}$$

In order to get a unique solution of this differential equation, we have to impose boundary conditions.

We assume that the boundary of the semiconductor domain consists of two parts: one part, called the Dirichlet part, on which the particle densities and the potential are prescribed,

$$n = n_D, \ p = p_D, \ V = V_D \quad \text{on } \Gamma_D, \qquad (6.9)$$

and another part, the Neumann part, which models the insulating boundary segments on which the normal components of the current densities and the electric field vanish,

$$J_n \cdot \nu = J_p \cdot \nu = \nabla V \cdot \nu = 0 \quad \text{on } \Gamma_N. \qquad (6.10)$$

In view of the expressions (6.3)-(6.4) for J_n and J_p this is equivalent to

$$\nabla n \cdot \nu = \nabla p \cdot \nu = \nabla V \cdot \nu = 0 \quad \text{on } \Gamma_N. \qquad (6.11)$$

It remains to determine the boundary functions n_D, p_D and V_D. We assume:

- The total space charge vanishes on Γ_D :

$$n_D - p_D - C = 0. \qquad (6.12)$$

- The densities are in equilibrium on Γ_D :

$$n_D p_D = n_i^2 \qquad (6.13)$$

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

$$V_D = V_{\text{bi}} + U. \qquad (6.14)$$

Clearly, in thermal equilibrium we have $U = 0$. The built-in potential is the potential corresponding to the equilibrium densities given by (6.7) and (6.8):

$$n_D = n_i e^{V_{\text{bi}}/U_T}, \quad p_D = n_i e^{-V_{\text{bi}}/U_T}. \qquad (6.15)$$

Substituting (6.12) in (6.13) gives $n_D^2 - Cn_D = n_i^2$ such that

$$n_D = \frac{1}{2}\left(C + \sqrt{C^2 + 4n_i^2}\right), \quad p_D = \frac{1}{2}\left(-C + \sqrt{C^2 + 4n_i^2}\right) \quad \text{on } \Gamma_D.$$

Thus we get from (6.15)

$$V_{\mathrm{bi}} = U_T \ln \frac{n_D}{n_i} = U_T \ln \left(\frac{C}{2n_i} + \sqrt{\frac{C^2}{4n_i^2} + 1} \right) = U_T \mathrm{arsinh} \frac{C}{2n_i}. \qquad (6.16)$$

Therefore, the thermal equilibrium state (n_e, p_e, V_e) is the (unique) solution of

$$\varepsilon_s \Delta V_e = q \left(2n_i \sinh \frac{V_e}{U_T} - C \right) \quad \text{in } \Omega,$$

$$V_e = V_{\mathrm{bi}} \quad \text{on } \Gamma_D, \quad \nabla V_e \cdot \nu = 0 \quad \text{on } \Gamma_N,$$

where V_{bi} is given by (6.16), and

$$n_e = n_i e^{V_e/U_T}, \quad p_e = n_i e^{-V_e/U_T} \quad \text{in } \Omega.$$

Moreover, the drift-diffusion equations (6.1)-(6.5) are complemented with the boundary conditions (6.9)-(6.11), where (n_D, p_D, V_D) are defined by (6.14)-(6.16), and the initial conditions

$$n(\cdot, 0) = n_I, \quad p(\cdot, 0) = p_I \quad \text{in } \Omega.$$

6.2 Scaling of the equations

The drift-diffusion model (6.1)-(6.5), (6.9)-(6.11) contains several physical parameters. By bringing the equations in a non-dimensional form, we can identify the relevant scaled parameters. In fact, we show that there are only two relevant parameters. Define

$$\overline{C} = \sup_{x \in \Omega} |C(x)|, \quad L = \mathrm{diam}(\Omega).$$

We expect that the particle densities are of the order of $\overline{C}$ and the potential is of the order of U_T. Then the scaled variables n_s, p_s and V_s, defined by

$$n = \overline{C} n_s, \quad p = \overline{C} p_s, \quad V = U_T V_s,$$

are of the order 1. We also set

$$x = L x_s, \quad C = \overline{C} C_s, \quad t = \tau t_s$$

with

$$\tau = \frac{L^2 \overline{\mu}}{U_T}, \quad \overline{\mu} = \max(\mu_n, \mu_p).$$

This fixes the time scale to the order τ. A simple computation shows that the scaled quantities satisfy the equations (where we omitted the index "s"):

$$\partial_t n - \mu_n' \mathrm{div} \left(\nabla n - n \nabla V \right) = -R(n, p), \qquad (6.17)$$

$$\partial_t p - \mu_p' \mathrm{div} \left(\nabla p + p \nabla V \right) = -R(n, p), \qquad (6.18)$$

$$\lambda^2 \Delta V = n - p - C, \qquad (6.19)$$

116

where $\mu'_\alpha = \mu_\alpha/\overline{\mu}$ ($\alpha = n, p$), and

$$\lambda = \sqrt{\frac{\varepsilon_s U_T}{q L^2 \overline{C}}}$$

is called the *scaled Debye length*. (The Debye length is the quantity $\sqrt{\varepsilon_s U_T/q\overline{C}} = L\lambda$.) The scaled recombination-generation term reads

$$R(n, p) = \frac{np - \delta^4}{\tau'_p(n + \delta^2) + \tau'_n(p + \delta^2)},$$

where $\tau'_\alpha = \tau_\alpha/\tau$ ($\alpha = n, p$) and

$$\delta^2 = \frac{n_i}{\overline{C}}$$

is the ratio of the intrinsic density and the maximal doping concentration. The relevant parameters are λ and δ. Typically, they are small compared to one. For instance, in silicon we have

$$\varepsilon_s = 1.05 \cdot 10^{-10}\,\text{As/Vm}, \qquad n_i = 10^{16}\,\text{m}^{-3},$$
$$\mu_n = 0.15\,\text{m}^2/\text{Vs}, \qquad\qquad \mu_p = 0.045\,\text{m}^2/\text{Vs}.$$

Choosing

$$L = 10^{-6}\,\text{m}, \quad \overline{C} = 10^{23}\,\text{m}^{-3}, \quad U_T = 0.026\,\text{V} \quad (\text{i.e. } T = 300\,\text{K}),$$

we obtain

$$\lambda^2 = 1.7 \cdot 10^{-4}, \quad \delta^2 = 10^{-7}.$$

The scaled current densities can now be defined by

$$J_n = \nabla n - n\nabla V, \quad J_p = -(\nabla p + p\nabla V).$$

The scaled Dirichlet boundary conditions read

$$n_D = \frac{1}{2}(C + \sqrt{C^2 + 4\delta^4}), \quad p_D = \frac{1}{2}(-C + \sqrt{C^2 + 4\delta^4}), \qquad (6.20)$$

$$V_D = V_{\text{bi}} + U = \text{arsinh}\frac{C}{2\delta^2} + U \quad \text{on } \Gamma_D, \qquad\qquad (6.21)$$

and the built-in potential can be written as

$$V_{\text{bi}} = \ln\left(\frac{C}{2\delta^2} + \sqrt{\frac{C^2}{4\delta^4} + 1}\right) = \ln\frac{n_D}{\delta^2} = \ln\frac{\delta^2}{p_D}. \qquad (6.22)$$

117

6.3 Static current-voltage characteristic of a diode

In this subsection we derive the current-voltage characteristic of a *pn*-junction diode in the case of "small" current densities. The diode is defined by

$$C(x) = \begin{cases} C_n & : & x \in \Omega_n \\ -C_p & : & x \in \Omega_p, \end{cases}$$

where $C_n > 0$ is the doping concentration in the *n*-region Ω_n, $-C_p < 0$ is the doping density in the *p*-region Ω_p, and $\Omega = \Omega_n \cup \Omega_p$, $\Omega_n \cap \Omega_p = \emptyset$. The Dirichlet boundary Γ_D is supposed to consist of the two boundary parts $\Gamma_1 \subset \overline{\Omega}_n$ and $\Gamma_2 \subset \overline{\Omega}_p$ (see Figure 6.1). We set $\Gamma = \overline{\Omega}_n \cap \overline{\Omega}_p$. The applied voltage U is given by

$$U(x) = \begin{cases} 0 & : & x \in \Gamma_1 \\ U_0 & : & x \in \Gamma_2, \end{cases}$$
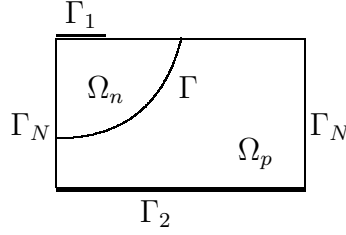
where U_0 is a constant.

Figure 6.1: Geometry of a *pn*-junction diode.

The total current is defined by

$$I = \int_{\Gamma_1} (J_n + J_p) \cdot \nu \, ds,$$

where ν is the exterior unit normal to $\partial \Omega_n$. We assume (for simplicity) that $\mu'_n = \mu'_p = 1$ (see (6.17)-(6.18)). Then, taking the difference of the stationary versions of (6.17)-(6.18), we have div $(J_n + J_p) = 0$ and, by the divergence theorem and $J_n \cdot \nu = J_p \cdot \nu = 0$ on Γ_N (see (6.10)),

$$0 = \int_{\Omega_n} \mathrm{div}\,(J_n + J_p) \, dx = \int_{\Gamma_1} (J_n + J_p) \cdot \nu \, dx + \int_{\Gamma} (J_n + J_p) \cdot \nu \, dx.$$

Thus

$$I = \int_{\Gamma_1} (J_n + J_p) \cdot \nu \, ds = - \int_{\Gamma} (J_n + J_p) \cdot \nu \, ds. \tag{6.23}$$

Our main goal is to describe the function $I = I(U_0)$ in more detail.

Our first main assumption is to consider "small" current densities. More precisely, we assume that the current densities are of the order of δ^4. Then we can rescale J_n and J_p by

$$J_n = \delta^4 j_n, \quad J_p = \delta^4 j_p. \tag{6.24}$$

We see later in this section why we use the exponent 4 in the rescaling. The new variables j_n and j_p are of the order $O(1)$. It is convenient to introduce new variables u, w by

$$n = \delta^2 e^V u, \quad p = \delta^2 e^{-V} w.$$

The functions u and w are called *Slotboom variables.* The advantage is that they symmetrize the current relations in the following sense:

$$
\begin{aligned}
\delta^2 e^V \nabla u = e^V \nabla(e^{-V} n) = \nabla n - n \nabla V &= J_n, \\
-\delta^2 e^{-V} \nabla w = -e^{-V} \nabla(e^V p) = -(\nabla p + p \nabla V) &= J_p.
\end{aligned}
$$

We consider the stationary equations for u, w and V (see (6.17)-(6.19)):

$$
\begin{aligned}
\delta^2 \mathrm{div}\,(e^V \nabla u) &= R(\delta^2 e^V u, \delta^2 e^{-V} w), & (6.25) \\
\delta^2 \mathrm{div}\,(e^{-V} \nabla w) &= R(\delta^2 e^V u, \delta^2 e^{-V} w), & (6.26) \\
\lambda^2 \Delta V &= \delta^2 e^V u - \delta^2 e^{-V} w - C \quad \text{in } \Omega. & (6.27)
\end{aligned}
$$

The boundary conditions for u and w follow from (6.20)-(6.22):

$$
\begin{aligned}
u &= \delta^{-2} e^{-V} n = \delta^{-2} e^{-V_{\mathrm{bi}}} n_D = \delta^{-2} e^{-\ln(n_D/\delta^2)} n_D = 1, \\
w &= \delta^{-2} e^V p = \delta^{-2} e^{V_{\mathrm{bi}}} p_D = \delta^{-2} e^{\ln(\delta^2/p_D)} p_D = 1, \\
V &= V_{\mathrm{bi}} + U = V_{\mathrm{bi}} \qquad \text{on } \Gamma_1,
\end{aligned}
$$

and

$$
\begin{aligned}
u &= \delta^{-2} e^{-V_{\mathrm{bi}} - U} n_D = e^{-U_0}, \\
w &= \delta^{-2} e^{V_{\mathrm{bi}} + U} p_D = e^{U_0}, \\
V &= V_{\mathrm{bi}} + U_0 \qquad \text{on } \Gamma_2.
\end{aligned}
$$

On the Neumann boundary we get

$$\nabla u \cdot \nu = \nabla w \cdot \nu = \nabla V \cdot \nu = 0 \quad \text{on } \Gamma_N.$$

The equations (6.25)-(6.27) together with the above boundary conditions are still equivalent to the full problem.

In order to get some informations on the current-voltage characteristic $I = I(U_0)$, we have to simplify the full problem. In Section 6.2 we have seen that the parameters λ and δ are small compared to one. Hence the solution of the *reduced problem* $\lambda = 0$ and $\delta = 0$ may be close to the solution of the full problem. In

119

fact, this is our second main assumption. (It is possible to prove this rigorously by means of asymptotic analysis; see [33]).

First, we study the reduced problem $\lambda = 0$. From (6.27) we obtain

$$0 = \delta^2 e^V u - \delta^2 e^{-V} w - C \quad \text{in } \Omega$$

and therefore

$$\delta^2 e^V = \frac{1}{2u}\left(C + \sqrt{C^2 + 4\delta^4 uw}\right), \quad \delta^2 e^{-V} = \frac{1}{2w}\left(-C + \sqrt{C^2 + 4\delta^4 uw}\right).$$

We can discard one solution of the above quadratic equation, since we require $u, w \geq 0$. Thus (6.25) and (6.26) become

$$\operatorname{div}\left[\frac{1}{2u}\left(C + \sqrt{C^2 + 4\delta^4 uw}\right)\nabla u\right] = R, \tag{6.28}$$

$$\operatorname{div}\left[\frac{1}{2w}\left(-C + \sqrt{C^2 + 4\delta^4 uw}\right)\nabla w\right] = R, \tag{6.29}$$

with the boundary conditions

$$u = w = 1 \quad \text{on } \Gamma_1, \qquad u = e^{-U_0}, \ w = e^{U_0} \quad \text{on } \Gamma_2, \tag{6.30}$$

$$\nabla u \cdot \nu = \nabla w \cdot \nu = 0 \quad \text{on } \Gamma_N. \tag{6.31}$$

Transforming back to the original variables, the particle densities n and p are functions of u and w :

$$n = \delta^2 e^V u = \frac{1}{2}(C + \sqrt{C^2 + 4\delta^4 uw}),$$

$$p = \delta^2 e^{-V} w = \frac{1}{2}(-C + \sqrt{C^2 + 4\delta^4 uw}).$$

Notice that the doping profil C is discontinuous across Γ. This means that also the diffusion coefficients of the elliptic equations (6.28) and (6.29) are discontinuous, and we cannot expect to obtain classical (i.e. twice continuously differentiable) solutions. In fact, the elliptic problem (6.28)-(6.31) has to be solved in a generalized (weak) sense. The particle densities are also discontinuous across Γ. Thus, we expect that the particle densities solving the full problem have an internal layer (i.e. sharp gradients) near Γ.

Next we consider the reduced problem $\delta = 0$. Recall that

$$\delta^4 j_n = J_n = \delta^2 e^V \nabla u = \frac{1}{2u}(C + \sqrt{C^2 + 4\delta^4 uw})\nabla u, \tag{6.32}$$

$$\delta^4 j_p = J_p = -\delta^2 e^{-V} \nabla w = -\frac{1}{2w}(-C + \sqrt{C^2 + 4\delta^4 uw})\nabla w, \tag{6.33}$$

and

$$R = \delta^4 \frac{uw - 1}{\tau_p'(n + \delta^2) + \tau_n'(p + \delta^2)}.$$

120

Hence we can write (6.28)-(6.29) as

$$\text{div } j_n = \frac{uw - 1}{\tau_p'(n + \delta^2) + \tau_n'(p + \delta^2)}, \quad \text{div } j_p = -\frac{uw - 1}{\tau_p'(n + \delta^2) + \tau_n'(p + \delta^2)}. \quad (6.34)$$

The Taylor expansion $\sqrt{1 + x} = 1 + x/2 + O(x^2)$ around $x = 0$ yields in Ω_p

$$\begin{aligned}
C + \sqrt{C^2 + 4\delta^4 uw} &= -C_p + \sqrt{C_p^2 + 4\delta^4 uw} \\
&= -C_p + C_p \left(1 + \frac{2\delta^4 uw}{C_p^2} + O(\delta^8)\right) \\
&= \frac{2uw}{C_p}\delta^4 + O(\delta^8), \\
-C + \sqrt{C^2 + 4\delta^4 uw} &= 2C_p + \frac{2uw}{C_p}\delta^4 + O(\delta^8) \quad \text{as } \delta \to 0
\end{aligned}$$

and in Ω_n

$$\begin{aligned}
C + \sqrt{C^2 + 4\delta^4 uw} &= 2C_n + \frac{2uw}{C_n}\delta^4 + O(\delta^8) \\
-C + \sqrt{C^2 + 4\delta^4 uw} &= -C_n + \sqrt{C^2 + 4\delta^4 uw} \\
&= \frac{2uw}{C_n}\delta^4 + O(\delta^8) \quad \text{as } \delta \to 0.
\end{aligned}$$

Therefore, (6.32) and (6.33) give

$$\begin{aligned}
j_n &= \frac{w\nabla u}{C_p} + O(\delta^4), \quad & \delta^4 j_p &= \frac{C_p \nabla w}{w} + O(\delta^4) \quad & \text{in } \Omega_p, \\
\delta^4 j_n &= \frac{C_n \nabla u}{u} + O(\delta^4), \quad & j_p &= -\frac{u\nabla w}{C_n} + O(\delta^4) \quad & \text{in } \Omega_n.
\end{aligned}$$

Here we see why we have used the scaling factor δ^4 in (6.24). Exponents larger or smaller than 4 lead to useless relations. Moreover,

$$\begin{aligned}
n = C_n, \quad p = 0 \quad & \text{in } \Omega_n, \\
n = 0, \quad p = C_p \quad & \text{in } \Omega_p.
\end{aligned}$$

Now we can perform the limit $\delta \to 0$ in (6.34) and in the above equations:

$$\begin{aligned}
\nabla u = 0, \quad j_p = -\frac{u\nabla w}{C_n}, \quad \text{div } j_p = -\frac{uw - 1}{\tau_p' C_n} \quad & \text{in } \Omega_n, \\
\nabla w = 0, \quad j_n = \frac{w\nabla u}{C_p}, \quad \text{div } j_n = \frac{uw - 1}{\tau_n' C_p} \quad & \text{in } \Omega_p.
\end{aligned}$$

From the boundary conditions (6.30) and (6.31) we conclude

$$u = 1 \quad \text{in } \Omega_n, \qquad w = e^{U_0} \quad \text{in } \Omega_p.$$

121

Thus

$$j_p = -\frac{\nabla w}{C_n} \quad \text{in } \Omega_n, \quad j_n = \frac{e^{U_0}}{C_p}\nabla u \quad \text{in } \Omega_p \tag{6.35}$$

and

$$\Delta w = -C_n \text{div} j_p = \frac{w-1}{\tau'_p} \quad \text{in } \Omega_n,$$

$$\Delta u = C_p e^{-U_0} \text{div} j_n = \frac{u - e^{-U_0}}{\tau'_n} \quad \text{in } \Omega_p.$$

The boundary conditions for w and u read

$$w = 1 \quad \text{on } \Gamma_1, \quad w = e^{U_0} \quad \text{on } \Gamma, \quad \nabla w \cdot \nu = 0 \quad \text{on } \partial\Omega_n \cap \Gamma_N$$

and

$$u = e^{-U_0} \quad \text{on } \Gamma_2, \quad u = 1 \quad \text{on } \Gamma, \quad \nabla u \cdot \nu = 0 \quad \text{on } \partial\Omega_p \cap \Gamma_N.$$

We wish to separate the influence of the applied potential. We claim that the solutions w and u of the above boundary-value problems can be written as

$$w = 1 + \left(e^{U_0} - 1\right) g \quad \text{in } \Omega_n, \quad u = e^{-U_0} + \left(1 - e^{-U_0}\right) f \quad \text{in } \Omega_p,$$

where f and g are the (unique) solutions of

$$\Delta f = \frac{f}{\tau'_n} \quad \text{in } \Omega_p, \tag{6.36}$$

$$f = 0 \quad \text{on } \Gamma_2, \quad f = 1 \quad \text{on } \Gamma, \quad \nabla f \cdot \nu = 0 \quad \text{on } \partial\Omega_p \cap \Gamma_N$$

and

$$\Delta g = \frac{g}{\tau'_p} \quad \text{in } \Omega_n, \tag{6.37}$$

$$g = 0 \quad \text{on } \Gamma_1, \quad g = 1 \quad \text{on } \Gamma, \quad \nabla g \cdot \nu = 0 \quad \text{on } \partial\Omega_n \cap \Gamma_N.$$

Indeed, it holds

$$\Delta w = \left(e^{U_0} - 1\right) \Delta g = \frac{\left(e^{U_0} - 1\right) g}{\tau'_p} = \frac{w-1}{\tau'_p} \quad \text{in } \Omega_n,$$

$$\Delta u = \left(1 - e^{-U_0}\right) \Delta f = \frac{\left(1 - e^{-U_0}\right) f}{\tau'_n} = \frac{u - e^{-U_0}}{\tau'_n} \quad \text{in } \Omega_p,$$

and it is easy to verify the boundary conditions for u and w. The advantage of the introduction of the functions f and g is that they depend on the geometry of the diode but not on the applied potential.

We can now compute the total current (6.23). The current densities (6.35) are only known in Ω_n or Ω_p; therefore we calculate I on Γ. Assuming continuity of the normal components of j_n in Ω_p and j_p in Ω_n, respectively, we get from

$$
\begin{aligned}
j_n &= \frac{e^{U_0}}{C_p}\nabla u = \frac{e^{U_0}-1}{C_p}\nabla f \quad \text{in } \Omega_p, \\
j_p &= -\frac{\nabla w}{C_n} = -\frac{e^{U_0}-1}{C_n}\nabla g \quad \text{in } \Omega_n
\end{aligned}
$$

and from (6.23)

$$
I = -\int_\Gamma (j_n + j_p)\cdot\nu\,ds = -\left(e^{U_0}-1\right)\int_\Gamma\left(\frac{\nabla f}{C_p} - \frac{\nabla g}{C_n}\right)\cdot\nu\,ds.
$$

Notice that also the total current is scaled by the factor δ^4. Introducing the *leakage current*,

$$
I_s := \int_\Gamma\left(\frac{\nabla g}{C_n} - \frac{\nabla f}{C_p}\right)\cdot\nu\,ds,
$$

we obtain the scaled *Shockley equation* (see Figure 6.2)

$$
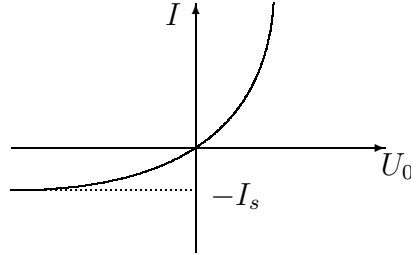I = I_s\left(e^{U_0}-1\right). \tag{6.38}
$$

Figure 6.2: The Shockley equation for a *pn*-junction diode.

The leakage current is positive. Indeed, we use integration by parts and the differential equations for f to infer

$$
\begin{aligned}
-\int_\Gamma\frac{1}{C_p}\nabla f\cdot\nu ds &= \int_\Gamma\frac{f}{C_p}\nabla f\cdot(-\nu)ds + \int_{\Omega_p}\frac{1}{C_p}\left(-f\Delta f + \frac{f^2}{\tau'_n}\right)\,dx \\
&\quad \text{(using } f = 1 \text{ on } \Gamma \text{ and the differential equation for } f) \\
&= \int_{\Omega_p}\frac{1}{C_p}\left(|\nabla f|^2 + \frac{f^2}{\tau'_n}\right)\,dx \\
&\quad \text{(using integration by parts)} \\
&> 0.
\end{aligned}
$$

123

Notice that $-\nu$ is the *exterior* unit normal to $\partial\Omega_p$. Similarly,

$$\int_{\Omega_n} \frac{1}{C_n}\left(|\nabla g|^2 + \frac{1}{\tau_p'}g^2\right)\,dx = \int_\Gamma \frac{1}{C_n}\nabla g \cdot \nu\,ds.$$

Therefore

$$I_s = \int_{\Omega_p} \frac{1}{C_p}\left(|\nabla f|^2 + \frac{1}{\tau_n'}f^2\right)\,dx + \int_{\Omega_n} \frac{1}{C_n}\left(|\nabla g|^2 + \frac{1}{\tau_p'}g^2\right)\,dx > 0.$$

Notice that the unscaled Shockley equation reads

$$I = I_s\left(e^{U_0/U_T} - 1\right).$$

Shockley's equation is valid for "small" current densities or, equivalently, for "small" applied voltages U_0. Experiments show that (6.38) indeed holds for sufficiently small applied potentials but not for "large" values of $|U_0|$. There exists a threshold voltage $U_{\text{th}} < 0$ such that $I(U_0) \to -\infty$ as $U_0 \searrow U_{\text{th}}$. This break-down effect is used in so-called Zener diodes.

We study the high-injection case $U_0 \gg 1$ for a simple one-dimensional symmetric diode modeled by the reduced problem $\lambda = 0$. For this we define $\Omega = (-1, 1)$,

$$C(x) = \begin{cases} -1 &:& x \in (-1, 0) \\ 1 &:& x \in (0, 1) \end{cases}, \qquad U(x) = -\frac{U_0}{2}x, \quad x \in \Omega.$$

We assume that recombination-generation effects are neglegible. Then $J_n = J_p =: I$ is constant. We are looking for solutions that satisfy

$$n(-x) = p(x), \quad V(-x) = -V(x), \quad x \in \Omega.$$

Then $V(0) = 0$, and it is sufficient to solve

$$I = n_x - nV_x, \quad I = -p_x - pV_x, \quad 0 = n - p - 1 \quad \text{in } (0,1) \qquad (6.39)$$

subject to the boundary conditions (see (6.20)-(6.21))

$$V(0) = 0, \quad V(1) = V_{\text{bi}}(1) + U(1) = \operatorname{arsinh}\frac{1}{2\delta^2} - \frac{U_0}{2},$$

$$p(1) = \frac{1}{2}(-1 + \sqrt{1 + 4\delta^4}), \quad n(1) = \frac{1}{2}(1 + \sqrt{1 + 4\delta^4}).$$

Taking the difference of the first two equations in (6.39) gives

$$0 = (n + p)_x - (n - p)V_x = (n + p)_x - V_x$$

and

$$V_x = (n + p)_x \quad \text{in } (0, 1). \qquad (6.40)$$

Thus, adding the first two equations in (6.39) yields

$$2I = -(n+p)V_x = -\frac{1}{2}\frac{d}{dx}[(n+p)^2].$$

This equation can be integrated from x to 1:

$$(n+p)^2(x) = (n+p)^2(1) + 4I(1-x),$$

and the boundary conditions at $x = 1$ and the last equation in (6.39) give

$$\sqrt{1 + 4\delta^4 + 4I(1-x)} = (n+p)(x) = 2n(x) - 1.$$

We conclude

$$\begin{aligned}
n(x) &= \frac{1}{2}(1 + \sqrt{1 + 4\delta^4 + 4I(1-x)}), \\
p(x) &= \frac{1}{2}(-1 + \sqrt{1 + 4\delta^4 + 4I(1-x)}).
\end{aligned}$$

The electrostatic potential is computed from (6.40) by integrating from x to 1:

$$\begin{aligned}
V(x) &= V(1) + (n+p)(x) - (n+p)(1) \\
&= \operatorname{arsinh}\frac{1}{2\delta^2} - \frac{U_0}{2} + \sqrt{1 + 4\delta^4 + 4I(1-x)} - \sqrt{1 + 4\delta^4}.
\end{aligned}$$

At $x = 0$ we obtain

$$\frac{U_0}{2} = \operatorname{arsinh}\frac{1}{2\delta^2} + \sqrt{1 + 4\delta^4 + 4I(1-x)} - \sqrt{1 + 4\delta^4}.$$

If $U_0 \to \infty$ or, equivalently, $I \to \infty$,

$$I \sim U_0^2. \tag{6.41}$$

This result means that the growth of the voltage-current characteristic slows down in high injection. The relation (6.39) is known as *Mott's law*.

For sufficiently large voltages we have *Ohm's law*

$$\frac{I}{U_0} \to \text{const.} \quad (U_0 \to \infty).$$

Whereas the Zener effect is not modeled in our drift-diffusion system (see [34] for a more general system), we can derive the Ohm law for $U_0 \gg 1$. To see this, we rescale the current densities and the potential,

$$J_n = U_0^\alpha j_n, \quad J_p = U_0^\beta j_p, \quad V = U_0^\gamma \phi,$$

where $\alpha, \beta, \gamma > 0$, and we assume that the particle densities are of order $O(1)$ when $U_0 \to \infty$. Also j_n, j_p and ϕ are supposed to be of order $O(1)$ as $U_0 \to \infty$. The Poisson equation then becomes

$$\lambda^2 \Delta\phi = U_0^{-\gamma}(n - p - C) \to 0 \quad \text{in } \Omega, \quad \text{as } U_0 \to \infty,$$
$$\phi = 0 \quad \text{on } \Gamma_1, \quad \phi = U_0^{1-\gamma} \quad \text{on } \Gamma_2, \quad \nabla\phi \cdot \nu = 0 \quad \text{on } \Gamma_N.$$

When $\gamma > 1$, ϕ solves the Laplace equation with homogenous boundary conditions in the limit $U_0 \to \infty$ and therefore, $\phi = 0$ in Ω. This solution is useless. On the other hand, when $\gamma < 1$, $\phi \to \infty$ on Γ_2 ($U_0 \to \infty$), which is not the right scaling. Hence $\gamma = 1$. The scaled current densities are

$$j_n = U_0^{-\alpha}\nabla n - U_0^{1-\alpha} n\nabla\phi, \quad j_p = -U_0^{-\beta}\nabla p - U_0^{1-\beta} p\nabla\phi.$$

When $\alpha > 1$, we get $j_n = 0$ in the limit which is useless. When $\alpha < 1$, we have $n\nabla\phi = 0$ which does not allow to compute j_n. Therefore, $\alpha = 1$. Similarly, we conclude $\beta = 1$. This yields, as $U_0 \to \infty$,

$$j_n = -n\nabla\phi, \quad j_p = -p\nabla\phi.$$

Notice that ϕ solves the problem

$$\Delta\phi = 0 \quad \text{in } \Omega,$$
$$\phi = 0 \quad \text{on } \Gamma_1, \quad \phi = 1 \quad \text{on } \Gamma_2, \quad \nabla\phi \cdot \nu = 0 \quad \text{on } \Gamma_N,$$

and does not depend on U_0. We obtain (see (6.20))

$$
\begin{aligned}
\frac{I}{U_0} &= \frac{1}{U_0}\int_{\Gamma_1}(U_0 j_n + U_0 j_p) \cdot \nu \, ds \\
&= -\int_{\Gamma_1}(n + p)\nabla\phi \cdot \nu \, ds \\
&= -\sqrt{C_n^2 + 4\delta^4}\int_{\Gamma_1}\nabla\phi \cdot \nu \, ds \\
&= \text{const.}
\end{aligned}
$$

We remark that the assumption that n and p are of order $O(1)$ as $U_0 \to \infty$ is not very realistic. However, since we evaluate n and p only on Γ_1 and they are fixed by the boundary conditions, we expect to obtain the same formula as above without this assumption.

Whereas the Mott law $I \sim U_0^2$ holds for "large" U_0, we needed to perform the limit $U_0 \to \infty$ to derive the Ohm law $I \sim U_0$. Therefore, we expect that the Mott law holds for "large" applied voltages and the Ohm law for "very large" applied voltages. This can be confirmed by physical experiments. Conditions on the size of the applied voltage which allow to distinguish the different regimes for "small", "large" and "very large" applied bias, unfortunately, do not follow from the above asymptotic analysis.

6.4 Numerical discretization of the stationary equations

The discretization of the drift-diffusion equations using standard methods may lead to unsatisfactory results. Indeed, for "large" electric fields $-\nabla V$, the corresponding elliptic problem is convection-dominant, and it is well known that the numerical discretization of such problems requires special care. In a simplified situation, we can see this directly from the equations.

Consider the one-dimensional stationary equations for the electron density in $\Omega = (0,1)$ with $\mu'_n = 1$ and $R = 0$ (see (6.17)):

$$J_n = n_x - nV_x, \quad J_{n,x} = 0 \quad \text{in } (0,1),$$
$$n(0) = n_{D0}, \quad n(1) = n_{D1},$$

for given electrostatic potential. We introduce the grid $x_i = ih$, $i = 0, \ldots, N$, with $h > 0$ and $N = 1/h \in \mathbb{N}$. With the approximations n_i, V_i of $n(x_i)$, $V(x_i)$ respectively, a standard finite-difference scheme is given by

$$J_{n,i+1/2} = \frac{n_{i+1} - n_i}{h} - \frac{n_{i+1} + n_i}{2} \frac{V_{i+1} - V_i}{h}, \tag{6.42}$$

$$\frac{1}{h}(J_{n,i+1/2} - J_{n,i-1/2}) = 0, \quad i = 1, \ldots, N-1. \tag{6.43}$$

Then $J_{n,i+1/2}$ can be regarded as an approximation of $J(x_{i+1/2}) = J(x_i + h/2)$. Setting $c_i := V_i - V_{i-1}$, we can rewrite (6.42)-(6.43) as a difference equation

$$(2 - c_{i+1})n_{i+1} + (c_i - c_{i+1} - 4)n_i + (2 + c_i)n_{i-1} = 0, \quad i = 1, \ldots, N-1,$$
$$n_0 = n_{D0}, \quad n_N = n_{D1}.$$

Suppose that the electric field is almost constant in Ω such that $c := c_i = \text{const.} > 0$ for all $i = 1, \ldots, N$ and that the grid is such that $c \neq 2$. Then the above difference equation has the explicit solution

$$n_i = (n_{D1} - n_{D0})\frac{r^i - 1}{r^N - 1} + n_{D0}, \quad i = 0, \ldots, N, \quad \text{with } r = \frac{2+c}{2-c}.$$

The discrete electron density n_i should be positive for all i. However, with the discretization (6.42)-(6.43) this may be not true. Indeed, let (for simplicity) $n_{D0} = 2$, $n_{D1} = 1$, N even and let

$$c > 2\frac{2^{1/N} + 1}{2^{1/N} - 1}.$$

Then $c > 2$, $-2^{1/N} < r < -1$ and $1 < r^N < 2$. If i is odd, then $r^i < 0$, $r^i + r^N < 2$ and

$$n_i = \frac{r^i - 1}{r^N - 1} + 1 = \frac{r^i + r^N - 2}{r^N - 1} < 0.$$

In fact, $n_i < 0$ holds for all odd indices i and $n_i > 0$ for all even indices i. In order to avoid the oscillatory behaviour, one has to choose the grid in such a way that $c = V_i - V_{i-1} < 2$. In Section 6.3, we have found that the potential can be written as

$$V = \ln\left[\frac{1}{2\delta^2 u}\left(C + \sqrt{C^2 + 4\delta^4 uw}\right)\right] \quad \text{in } \Omega$$

for the reduced problem $\lambda = 0$, and u, w solve the problem (6.28)-(6.31). This means that if C is discontinuous in Ω, we expect large gradients for V. To get a non-oscillating numerical scheme, we need to choose a very fine grid at least near the points where the field is "large" (i.e. $V_i - V_{i-1} < 2$). Particularly for multi-dimensional problems the restriction that the potential only varies weakly on each of the subintervals (x_i, x_{i+1}) may require extremely many grid points and is very expensive and sometimes even impossible to guarantee. The scheme (6.42)-(6.43) is therefore not practical.

One possibility to solve the drift-diffusion equations numerically accurately is to use the *Scharfetter-Gummel discretization*. The idea is to symmetrize first the current densities and then to discretize the equations. We illustrate the idea for the one-dimensional stationary equations

$$\begin{aligned}
J_{n,x} &= 0, \quad J_n = \mu_n(n_x - nV_x), \\
J_{p,x} &= 0, \quad J_p = -\mu_p(p_x + pV_x), \\
\lambda^2 V_{xx} &= n - p - C \quad \text{in } \Omega = (0,1)
\end{aligned}$$

subject to the boundary conditions

$$n(0) = n_{D0}, \quad n(1) = n_{D1}, \quad p(0) = p_{D0}, \quad p(1) = p_{D1},$$

$$V(0) = 0, \quad V(1) = U_0.$$

In the following we consider only the equation for n since the treatment of the equation for p is analogous. Define the *Slotboom variable* $z = \mu_n ne^{-V}$ (see Section 6.3). Then

$$e^V z_x = \mu_n(n_x - nV_x) = J_n$$

and we have to discretize

$$J_{n,x} = 0, \quad z_x = J_n e^{-V} \quad \text{in } (0,1). \tag{6.44}$$

Introduce grid points $0 = x_0 < x_1 < \ldots < x_N = 1$ with $h_i := x_{i+1} - x_i$, $i = 0, \ldots, N-1$. We use central finite differences to discretize the Poisson equation:

$$\begin{aligned}
n_i - p_i - C(x_i) &= \frac{\lambda^2}{h_{i-1}}\left(\frac{V_{i+1} - V_i}{h_i} - \frac{V_i - V_{i-1}}{h_{i-1}}\right) \\
&= \frac{\lambda^2}{h_{i-1}h_i}\left[V_{i+1} - \left(1 + \frac{h_i}{h_{i-1}}\right)V_i + \frac{h_i}{h_{i-1}}V_{i-1}\right],
\end{aligned}$$

where n_i, p_i, and V_i are approximations of $n(x_i)$, $p(x_i)$, and $V(x_i)$, respectively. We define an approximate potential V_h as the linear interpolation of the V_i:

$$V_h(x) = V_i + (x - x_i)\frac{V_{i+1} - V_i}{h_i}, \quad x \in [x_i, x_{i+1}].$$

We use the approximate potential in the last equation of (6.44), integrate over (x_i, x_{i+1}) and use the fact that J_n is constant on Ω:

$$z_{i+1} - z_i = \int_{x_i}^{x_{i+1}} z_{h,x}(x)\, dx = J_n \int_{x_i}^{x_{i+1}} e^{-V_h(x)}\, dx.$$

Here, $z_h(x_i) = z_i$ are approximations of $z(x_i)$. The last integral can be computed explicitely since V_h is piecewise linear:

$$z_{i+1} - z_i = J_n e^{-V_i} \frac{h_i}{V_{i+1} - V_i}\left(1 - e^{-(V_{i+1} - V_i)}\right).$$

This gives a definition for J_n (after transforming back to $n_i = z_i e^{V_i}/\mu_n$):

$$
\begin{aligned}
J_n &= \mu_n \frac{V_{i+1} - V_i}{h_i} \frac{e^{V_i}}{1 - e^{V_i - V_{i+1}}}\left(e^{-V_{i+1}} n_{i+1} - e^{-V_i} n_i\right) \\
&= \frac{\mu_n}{h_i}\left(B(V_{i+1} - V_i)n_{i+1} - B(V_i - V_{i+1})n_i\right), \quad\quad (6.45)
\end{aligned}
$$

with the *Bernoulli function*

$$B(x) = \frac{x}{e^x - 1}, \quad x \in \mathbb{R}.$$

In order to understand the difference of the Scharfetter-Gummel discretization (6.45) and the standard finite-difference scheme (6.42), we rewrite (6.45) as

$$
\begin{aligned}
J_n &= \mu_n \frac{V_{i+1} - V_i}{h_i}\left[\left(\frac{1}{2} + \frac{e^{V_i - V_{i+1}}}{1 - e^{V_i - V_{i+1}}}\right)n_{i+1} - \left(\frac{1}{1 - e^{V_i - V_{i+1}}} - \frac{1}{2}\right)n_i\right] \\
&\quad - \mu_n \frac{n_i + n_{i+1}}{2}\frac{V_{i+1} - V_i}{h_i} \\
&= \mu_n\left((V_{i+1} - V_i)\coth\frac{V_{i+1} - V_i}{2}\right)\frac{n_{i+1} - n_i}{h_i} - \mu_n \frac{n_i + n_{i+1}}{2}\frac{V_{i+1} - V_i}{h_i}.
\end{aligned}
$$

The function

$$f(V_{i+1} - V_i) = (V_{i+1} - V_i)\coth\frac{V_{i+1} - V_i}{2}$$

plays the role of an upwind factor. Indeed, for slowly varying potentials $|V_{i+1} - V_i| \ll 1$ we have

$$f(x) = 1 \quad \text{as } x \to 0,$$

whereas for large gradients $|V_{i+1} - V_i| \gg 1$ it holds

$$f(x) \sim x \quad \text{as } x \to \pm\infty.$$

Thus, for diffusion-dominated problems $|V_{i+1} - V_i| \ll 1$, we obtain the standard discretization (6.42), and for convection-dominated problems $|V_{i+1} - V_i| \gg 1$, the upwind factor provides a correction such that the diffusion term

$$\left((V_{i+1} - V_i) \coth \frac{V_{i+1} - V_i}{2} \right) \frac{n_{i+1} - n_i}{h_i} \sim (V_{i+1} - V_i) \frac{n_{i+1} - n_i}{h_i}$$

is of the same order as the convection term

$$\frac{n_{i+1} + n_i}{2} \frac{V_{i+1} - V_i}{h_i}.$$

The current density for the holes is discretized analogously:

$$J_p = -\mu_p \left((V_{i+1} - V_i) \coth \frac{V_{i+1} - V_i}{2} \right) \frac{p_{i+1} - p_i}{h_i} - \mu_p \frac{p_{i+1} + p_i}{2} \frac{V_{i+1} - V_i}{h_i}.$$

We obtain three discrete nonlinear systems

$$
\begin{aligned}
a_{n,i} n_{i+1} + b_{n,i} n_i + c_{n,i} n_{i-1} &= 0, \\
a_{p,i} p_{i+1} + b_{p,i} p_i + c_{p,i} p_{i-1} &= 0, \\
\alpha_i V_{i+1} + \beta_i V_i + \gamma_i V_{i-1} &= \lambda^{-2} h_{i-1} h_i (n_i - p_i - C(x_i)),
\end{aligned}
$$

where we have defined

$$
\begin{aligned}
a_{n/p,i} &= \frac{1}{h_i} \left(f(V_{i+1} - V_i) \mp \frac{1}{2}(V_{i+1} - V_i) \right), \\
b_{n/p,i} &= -\frac{1}{h_i} \left(f(V_{i+1} - V_i) \mp \frac{1}{2}(V_{i+1} - V_i) \right) \\
&\quad - \frac{1}{h_{i-1}} \left(f(V_i - V_{i-1}) \mp \frac{1}{2}(V_i - V_{i-1}) \right), \\
c_{n/p,i} &= \frac{1}{h_{i-1}} \left(f(V_i - V_{i-1}) \pm \frac{1}{2}(V_i - V_{i-1}) \right),
\end{aligned}
$$

and

$$\alpha_i = 1, \quad \beta_i = -1 - \frac{h_i}{h_{i-1}}, \quad \gamma_i = -\frac{h_i}{h_{i-1}}.$$

7 The Energy-Transport Equations

In this chapter, we analyze the energy-transport equations for spherically symmetric and strictly monotone energy bands $\varepsilon(|k|)$ in $\mathbb{R}^3$, as derived in Section 5.5 (see Example 5.9):

$$\partial_t n(\mu, T) - \frac{1}{q}\operatorname{div} J_1 = 0, \tag{7.1}$$

$$\partial_t \mathcal{E}(\mu, T) - \operatorname{div} J_2 + \nabla V \cdot J_1 = W(\mu, T), \tag{7.2}$$

$$\varepsilon_s \Delta V = q(n(\mu, T) - C(x)), \quad x \in \Omega, \ t > 0, \tag{7.3}$$

where the electron and energy current densities are given by (see (5.74)-(5.75))

$$J_1 = \mathcal{D}_{11}\nabla\left(\frac{q\mu}{k_B T}\right) + \mathcal{D}_{12}\nabla\left(\frac{-1}{k_B T}\right) - \mathcal{D}_{11}\frac{q\nabla V}{k_B T}, \tag{7.4}$$

$$J_2 = \mathcal{D}_{21}\nabla\left(\frac{q\mu}{k_B T}\right) + \mathcal{D}_{22}\nabla\left(\frac{-1}{k_B T}\right) - \mathcal{D}_{21}\frac{q\nabla V}{k_B T}, \tag{7.5}$$

the electron and energy density are defined, respectively, by

$$n(\mu, T) = 2\pi e^{q\mu/k_B T}\int_0^\infty \sqrt{\gamma(\varepsilon)}\gamma'(\varepsilon)e^{-\varepsilon/k_B T}\, d\varepsilon, \tag{7.6}$$

$$\mathcal{E}(\mu, T) = 2\pi e^{q\mu/k_B T}\int_0^\infty \sqrt{\gamma(\varepsilon)}\gamma'(\varepsilon)e^{-\varepsilon/k_B T}\varepsilon\, d\varepsilon, \tag{7.7}$$

V is the electrostatic potential, $W(\mu, T)$ a relaxation term satisfying

$$W(\mu, T)(T - T_L) \le 0 \quad \text{for all } \mu \in \mathbb{R}, \ T > 0, \tag{7.8}$$

μ is the chemical potential, T the electron temperature, T_L the lattice temperature, and $C(x)$ the doping profile. We assume Maxwell-Boltzmann statistics and space-independent scattering rates, so that the diffusion matrices $\mathcal{D}_{ij}$ can be written as (see (5.76))

$$\mathcal{D}_{ij}(\mu, T) = e^{q\mu/k_B T}\int_0^\infty D(\varepsilon)e^{-\varepsilon/k_B T}\varepsilon^{i+j-2}\, d\varepsilon, \quad i, j = 1, 2. \tag{7.9}$$

The function γ is defined by $|k|^2 = \gamma(\varepsilon(|k|))$. The equations (7.1)-(7.3) are considered in the bounded domain $\Omega \subset \mathbb{R}^d$ ($d \ge 1$), and we impose the following initial and boundary conditions

$$\mu(\cdot, 0) = \mu_I, \quad T(\cdot, 0) = T_I \quad \text{in } \Omega,$$

$$\mu = \mu_D, \quad T = T_D, \quad V = V_D \quad \text{on } \Gamma_D \times (0, \infty),$$

$$J_1 \cdot \nu = J_2 \cdot \nu = \nabla V \cdot \nu = 0 \quad \text{on } \Gamma_N \times (0, \infty).$$

We have assumed that $\partial\Omega = \Gamma_D \cup \Gamma_N$, $\Gamma_D \cap \Gamma_N = \emptyset$, Γ_D is open in $\partial\Omega$, and ν is the exterior unit normal of $\partial\Omega$.

7.1 Symmetrization and entropy

In this section we introduce a change of unknowns which symmetrizes the energy-transport model (7.1)-(7.7) in the sense that the lower-order terms $-\nabla V \cdot J_1$ and $-\mathcal{D}_{i1} q \nabla V / k_B T$ are absent in the new formulation. Moreover, we define a so-called entropy functional which is non-increasing in time and which relates the old and the new variables.

Definition 7.1 *The variables*

$$u_1 = \frac{q\mu}{k_B T}, \quad u_2 = -\frac{1}{k_B T}$$

are called (primal) entropy variables; *the variables*

$$w_1 = \frac{q(\mu - V)}{k_B T} = u_1 + qV u_2, \quad w_2 = -\frac{1}{k_B T} = u_2$$

are called dual entropy variables.

We set $u = (u_1, u_2)$ and $w = (w_1, w_2)$. In the entropy variables the electron and energy density can be formulated as

$$\rho_1(u) \quad := \quad n = 2\pi e^{u_1} \int_0^\infty \sqrt{\gamma(\varepsilon)} \gamma'(\varepsilon) e^{\varepsilon u_2} \, d\varepsilon,$$

$$\rho_2(u) \quad := \quad \mathcal{E} = 2\pi e^{u_1} \int_0^\infty \sqrt{\gamma(\varepsilon)} \gamma'(\varepsilon) e^{\varepsilon u_2} \varepsilon \, d\varepsilon.$$

In the dual entropy variables, the energy-transport model (7.1)-(7.5) can be written in a symmetric form.

Lemma 7.2 *The equations (7.1)-(7.5) are formally equivalent to*

$$\partial_t b_1(u) - div\, I_1 \quad = \quad 0, \tag{7.10}$$
$$\partial_t b_2(u) - div\, I_2 \quad = \quad W - qn\partial_t V, \tag{7.11}$$
$$\varepsilon_s \Delta V \quad = \quad q(n - C(x)),$$

where

$$b_1(u) = n, \quad b_2(u) = \mathcal{E} - qVn$$

and

$$I_1 \quad = \quad \frac{1}{q} J_1 = L_{11} \nabla w_1 + L_{12} \nabla w_2,$$
$$I_2 \quad = \quad J_2 - V J_1 = L_{21} \nabla w_1 + L_{22} \nabla w_2.$$

The diffusion coefficients L_{ij} are given by

$$L_{11} = \mathcal{D}_{11}, \quad L_{12} = L_{21} = \mathcal{D}_{12} - \mathcal{D}_{11} qV, \quad L_{22} = \mathcal{D}_{22} - 2\mathcal{D}_{21} qV + \mathcal{D}_{11}(qV)^2.$$

Moreover, the new diffusion matrix $L = L(\mu, T) = (L_{ij})_{ij=1}^2 \in \mathbb{R}^{6 \times 6}$ is symmetric in the sense $L_{12} = L_{21}$ and positive definite if the functions

$$\frac{\partial \varepsilon}{\partial k_1}, \ \frac{\partial \varepsilon}{\partial k_2}, \ \frac{\partial \varepsilon}{\partial k_3}, \ \varepsilon \frac{\partial \varepsilon}{\partial k_1}, \ \varepsilon \frac{\partial \varepsilon}{\partial k_2}, \ and \ \varepsilon \frac{\partial \varepsilon}{\partial k_3} \tag{7.12}$$

are linearly independent.

Proof. A simple computation gives, by (7.4)-(7.5),

$$
\begin{aligned}
qI_1 &= \mathcal{D}_{11}\nabla u_1 + \mathcal{D}_{12}\nabla u_2 + \mathcal{D}_{11}u_2 q\nabla V \\
&= \mathcal{D}_{11}\nabla(u_1 + u_2 qV) + (\mathcal{D}_{12} - \mathcal{D}_{11}qV)\nabla u_2 \\
&= L_{11}\nabla w_1 + L_{12}\nabla w_2, \\
I_2 &= \mathcal{D}_{21}\nabla u_1 + \mathcal{D}_{22}\nabla u_2 + \mathcal{D}_{21}u_2 q\nabla V - VJ_1 \\
&= (\mathcal{D}_{21} - \mathcal{D}_{11}qV)\nabla(u_1 + u_2 qV) + (\mathcal{D}_{22} - 2\mathcal{D}_{21}qV + \mathcal{D}_{11}(qV)^2)\nabla u_2 \\
&\quad + \mathcal{D}_{11}qV\nabla(u_1 + u_2 qV) + qV(\mathcal{D}_{21} - \mathcal{D}_{11}qV)\nabla u_2 - qVI_1 \\
&= L_{21}\nabla w_1 + L_{22}\nabla w_2,
\end{aligned}
$$

since Proposition 5.8 implies $\mathcal{D}_{12} = \mathcal{D}_{21}$ and hence $L_{12} = L_{21}$. This also proves (7.10) and (7.11) since

$$
\begin{aligned}
\partial_t b_2(u) - \operatorname{div} I_2 &= \partial_t(\mathcal{E} - qVn) - \operatorname{div}(J_2 - VJ_1) \\
&= W - \nabla V \cdot J_1 - qn\partial_t V - qV\partial_t n + \nabla V \cdot J_1 + V\operatorname{div} J_1 \\
&= W - qn\partial_t V.
\end{aligned}
$$

It remains to prove the positive definiteness of L. We compute

$$L = \begin{pmatrix} \mathrm{Id} & -qV\mathrm{Id} \\ 0 & \mathrm{Id} \end{pmatrix}^{\top} \begin{pmatrix} \mathcal{D}_{11} & \mathcal{D}_{12} \\ \mathcal{D}_{21} & \mathcal{D}_{22} \end{pmatrix} \begin{pmatrix} \mathrm{Id} & -qV\mathrm{Id} \\ 0 & \mathrm{Id} \end{pmatrix}. \tag{7.13}$$

Here, Id is the unit matrix in $\mathbb{R}^{3 \times 3}$. The assumptions on the linear independency of $(\partial \varepsilon / \partial k_i, \ \varepsilon \partial \varepsilon / \partial k_i)_i$ ensure, by Proposition 5.8, that $\mathcal{D}$ is positive definite. Thus, L is positive definite since the matrix

$$\begin{pmatrix} \mathrm{Id} & -qV\mathrm{Id} \\ 0 & \mathrm{Id} \end{pmatrix}$$

is invertible. $\qquad\square$

The symmetrization property is related to the existence of an entropy functional. In fact, both properties are equivalent (see [20, 32] for details). Before we introduce the entropy, we need to discuss the thermal equilibrium state defined by $J_1 = J_2 = 0$. This implies

$$0 = \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = L\nabla \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}.$$

From now on we assume that the functions in (7.12) are linearly independent. Then L is symmetric and positive definite, hence invertible, and the above equation yields

$$w_1 = \text{const.} \quad \text{and} \quad w_2 = \text{const.} \quad \text{in } \Omega.$$

Therefore, $T = \text{const.}$ and $\mu - V = \text{const.}$ in Ω. As V is only determined up to an additive constant, we can choose this constant such that $\mu - V = 0$ or $w_1 = 0$ in Ω. Since $T = T_L$ on a part of the boundary, $T = T_L$ in Ω. By (7.8), this implies $W(\mu, T) = 0$ in Ω for all μ. Then we obtain from (7.1), (7.2) that $\partial_t n = 0$ and $\partial_t \mathcal{E} = 0$.

We notice that this and the assumption of vanishing total space charge on the Dirichlet boundary part (i.e. $n = C$ on Γ_D) determines the thermal state uniquely. Indeed, from (7.6) or

$$n = N e^{q\mu/k_B T_L} \quad \text{with } N = 2\pi \int_0^\infty \sqrt{\gamma(\varepsilon)} \gamma'(\varepsilon) e^{-\varepsilon/k_B T_L} \, d\varepsilon > 0$$

in thermal equilibrium we obtain

$$n e^{-q\mu V/k_B T_L} = N \quad \Longleftrightarrow \quad V = \frac{k_B T_L}{q} \ln \frac{n}{N}.$$

On the boundary Γ_D it holds

$$V = U_T \ln \frac{n}{N} = U_T \ln \frac{C(x)}{N},$$

where $U_T = k_B T_L / q$ and thus, the thermal equilibrium state (μ_e, T_e, V_e) is the unique solution of

$$\varepsilon_s \Delta V_e = q(N e^{V_e/U_T} - C(x)) \quad \text{in } \Omega, \tag{7.14}$$

$$V_e = U_T \ln \frac{C(x)}{N} \quad \text{on } \Gamma_D, \qquad \nabla V_e \cdot \nu = 0 \quad \text{on } \Gamma_N \tag{7.15}$$

and $\mu_e = V_e$, $T_e = L_L$ in Ω.

The *entropy density* is now defined by

$$\eta(u) = \rho(u) \cdot (u - u_e) - \chi(u) + \chi(u_e),$$

where

$$\rho(u) = \begin{pmatrix} n \\ \mathcal{E} \end{pmatrix}, \quad u_e = \begin{pmatrix} u_{e1} \\ u_{e2} \end{pmatrix} = \begin{pmatrix} q\mu_e/k_B T_L \\ -1/k_B T_L \end{pmatrix}$$

and $\chi \colon \mathbb{R}^2 \to \mathbb{R}$ is a function satisfying $\chi'(u) = \rho(u)$. More precisely, $-\eta$ is the *physical* entropy density relative to the thermal equilibrium state u_e. The *entropy* is the integral over the entropy density:

$$S_0(t) = \int_\Omega \eta(x,t) \, dx.$$

Example 7.3 (Parabolic band approximation)
Under the assumptions of Example 5.10 we can express the energy density in terms of n and T. Since, by Example 5.10,

$$n = N_c(k_B T)^{3/2} e^{q\mu/k_B T}, \quad \mathcal{E} = \frac{3}{2} k_B T n,$$

where $N_c = 2(2\pi m^*/\hbar^2)^{3/2}$, we have

$$u_1 = \ln \frac{n}{N_c} - \frac{3}{2} \ln(k_B T).$$

The equilibrium entropy variables are

$$u_{e1} = \ln \frac{n_e}{N_c} - \frac{3}{2} \ln(k_B T_L), \quad u_{e2} = -\frac{1}{k_B T_L},$$

where

$$n_e = N_c(k_B T_L)^{3/2} e^{q\mu_e/k_B T_L}$$

and $\mu_e = V_e$ is the (unique) solution of (7.14)-(7.15). A simple computation verifies that $\chi(u) = n$. Therefore

$$
\begin{aligned}
\eta &= \begin{pmatrix} n \\ \frac{3}{2} k_B T n \end{pmatrix} \cdot \begin{pmatrix} \ln(n/n_e) - \frac{3}{2} \ln(T/T_L) \\ (T - T_L)/k_B T_L T \end{pmatrix} - n + n_e \\
&= n \left(\ln \frac{n}{n_e} - \frac{3}{2} \ln \frac{T}{T_L} \right) + \frac{3}{2} \frac{nT}{T_L} - \frac{5}{2} n + n_e.
\end{aligned}
$$

Notice that the entropy vanishes in thermal equilibrium:

$$\eta\big|_{n=n_e, \ T=T_L} = 0.$$

The *free energy* of the energy-transport system is the sum of the entropy and the electric energy:

$$S(t) = \int_\Omega \left(\eta(u) + \frac{1}{2\varepsilon_s T_L} |\nabla(V - V_e)|^2 \right)(t) \, dx.$$

Sometimes, also the function $S(t)$ is called entropy. It satisfies the so-called *entropy inequality:*

Proposition 7.4 *Assume that the boundary data is in thermal equilibrium, i.e.*

$$\mu_D = \mu_e, \quad T_D = T_L, \quad V_D = V_e \quad \text{on } \Gamma_D \times (0, \infty),$$

that (7.8) holds and that the functions in (7.12) are linearly independent. Then

$$S(t) + \int_0^t \int_\Omega (\nabla w)^\top L (\nabla w) \, dx \, dt \leq S(0).$$

The assumptions of the proposition imply that L is symmetric, positive definite. Hence the above integral which is called *entropy dissipation term* is non-negative. The entropy inequality shows that the entropy is non-increasing. Under suitable assumptions on L, $\rho(u)$ and W it is possible to show that $S(t)$ converges exponentially fast to zero as $t \to \infty$ [19]. As we have not specified our notation of solution, computations have to be understood in a formal way. The arguments are made rigorous in [19].

Proof. We compute the time derivative of $S(t)$. Since $\chi'(u) = \rho(u)$ and therefore

$$
\begin{aligned}
\partial_t \eta(u) &= \partial_t \rho(u) \cdot (u - u_e) - \rho(u) \cdot \partial_t u - \chi'(u) \cdot \partial_t u \\
&= \partial_t \rho(u) \cdot (u - u_e),
\end{aligned}
$$

we obtain, by integrating by parts,

$$
\begin{aligned}
\partial_t S &= \int_\Omega \left[\partial_t n (u_1 - u_{e1}) + \partial_t \mathcal{E}(u_2 - u_{e2}) + \frac{1}{\varepsilon_s T_L} \partial_t \nabla V \cdot \nabla(V - V_e) \right] dx \\
&= \int_\Omega \Big[\partial_t n (u_1 + qVu_2 - (u_{e1} + qV_e u_{e2})) + \partial_t (\mathcal{E} - qVn)(u_2 - e_{e2}) \\
&\qquad - \partial_t n q (V u_2 - V_e u_{e2}) + q \partial_t (Vn)(u_2 - u_{e2}) - \frac{1}{T_L} \partial_t n (V - V_e) \Big] dx \\
&= \int_\Omega \big[\operatorname{div} I_1 (w_1 - w_{e1}) + \operatorname{div} I_2 (w_2 - w_{e2}) + W(u_2 - u_{e2}) \\
&\qquad - qn \partial_t V (u_2 - u_{e2}) - q \partial_t n ((V - V_e) u_2 + V_e (u_2 - u_{e2})) \\
&\qquad + qn \partial_t V (u_2 - u_{e2}) + q \partial_t n V(u_2 - u_{e2}) + q \partial_t n (V - V_e) u_{e2} \big] \, dx \\
&\qquad \text{(using (7.10) and (7.11))} \\
&= \int_\Omega \left[-I_1 \cdot \nabla w_1 - I_2 \cdot \nabla w_2 + W \frac{T - T_L}{k_B T T_L} \right] dx \\
&\qquad \text{(since } w_{e1} = 0, \ w_{e2} = \text{const.)} \\
&\leq -\int_\Omega (\nabla w)^\top L (\nabla w) \, dx.
\end{aligned}
$$

Integrating over $(0, t)$ gives the result. $\qquad\qquad\square$

7.2 A drift-diffusion formulation

In the previous section we have derived a formulation of the energy-transport model using the dual entropy variables w_1 and w_2. The advantage of this formulation is that first-order terms like $-\nabla V \cdot J_1$ vanish. This is also an advantage from a numerical point of view. Indeed, the symmetric operator $-\operatorname{div}(L\nabla w)$ can be discretized using standard methods and the discrete nonlinear system can be solved via the Newton method. However, the use of the Newton method has the

disadvantage that the Jacobian of a (2×2)-matrix has to be computed which makes the system less flexible (for instance, when changing the energy band) and computationally quite costly. One way to overcome this disadvantage is to look for another formulation of the system which allows for a decoupling of the equations like in the drift-diffusion model (see Section 6.4).

In this section we show that the energy-transport model can be formulated as a drift-diffusion system which allows for the use of the numerical methods developed for the classical drift-diffusion model. For this, we turn back to the definition of the particle current density J_1 and energy current density J_2 of Section 5.5:

$$J_i = \int_{\mathbb{R}} D(\varepsilon) \left(\nabla_x F + q \nabla_x V \frac{\partial F}{\partial \varepsilon} \right) \varepsilon^{i-1} \, d\varepsilon, \quad i = 1, 2,$$

where the diffusion matrix is assumed to be independent of x and F is given by the Maxwell-Boltzmann statistics:

$$F(\varepsilon) = e^{\varepsilon u_2 + u_1}$$

with $u_1 = q\mu/k_B T$ and $u_2 = -1/k_B T$. Then $\partial F/\partial \varepsilon = (-1/k_B T)F$ and

$$J_i = \nabla_x \int_{\mathbb{R}} D(\varepsilon) F(\varepsilon) \varepsilon^{i-1} \, d\varepsilon - \frac{q \nabla_x V}{k_B T} \int_{\mathbb{R}} D(\varepsilon) F(\varepsilon) \varepsilon^{i-1} \, d\varepsilon.$$

Setting

$$g_1 \;\; = \;\; \mathcal{D}_{11} = \int_{\mathbb{R}} D(\varepsilon) F(\varepsilon) \, d\varepsilon,$$

$$g_2 \;\; = \;\; \mathcal{D}_{12} = \int_{\mathbb{R}} D(\varepsilon) F(\varepsilon) \varepsilon \, d\varepsilon,$$

we obtain

$$J_1 = \nabla g_1 - \frac{q \nabla V}{k_B T} g_1, \quad J_2 = \nabla g_2 - \frac{q \nabla V}{k_B T} g_2.$$

In this formulation the energy-transport model simplifies since it is no longer strongly nonlinear in the sense that the equation for n (for $\mathcal{E}$, respectively) only depends on second-order derivatives of g_1 (of g_2) and *not* also of g_2 (of g_1). Notice that the equation for n in the (w_1, w_2)-formulation contains second-order derivatives of w_1 *and* w_2. Clearly, the (g_1, g_2)-formulation contains first-order terms like $-\nabla V \cdot J_1$ but they can be handled numerically (see [21, 29]). Moreover, we have to ensure that the variables T, μ, n and $\mathcal{E}$ can be determined uniquely from g_1, g_2. This will be proved in the following.

The temperature can be computed from the nonlinear equation

$$\frac{g_2}{g_1} = \frac{\int_0^\infty D(\varepsilon) e^{-\varepsilon/k_B T} \varepsilon \, d\varepsilon}{\int_0^\infty D(\varepsilon) e^{-\varepsilon/k_B T} \, d\varepsilon} =: f(k_B T).$$

The following lemma states under which assumptions the function f can be inverted.

Lemma 7.5 *Assume that the functions (7.12) are linearly independent. Then*

$$f'(k_BT) = \frac{\det L}{(k_BTg_1)^2} > 0 \quad \forall T > 0.$$

Proof. Since

$$
\begin{aligned}
\frac{\partial g_1}{\partial (k_BT)} &= \frac{1}{(k_BT)^2}\int_0^\infty D(\varepsilon)e^{-\varepsilon/k_BT}\varepsilon \, d\varepsilon = \frac{\mathcal{D}_{12}}{(k_BT)^2}, \\
\frac{\partial g_2}{\partial (k_BT)} &= \frac{1}{(k_BT)^2}\int_0^\infty D(\varepsilon)e^{-\varepsilon/k_BT}\varepsilon^2 \, d\varepsilon = \frac{\mathcal{D}_{22}}{(k_BT)^2},
\end{aligned}
$$

and $g_1 = \mathcal{D}_{11}$, $g_2 = \mathcal{D}_{12}$, we obtain

$$
\begin{aligned}
f'(k_BT) &= \frac{1}{g_1^2}\left(g_1\frac{\partial g_2}{\partial(k_BT)} - g_2\frac{\partial g_1}{\partial(k_BT)} \right) \\
&= \frac{1}{(k_BTg_1)^2}(\mathcal{D}_{11}\mathcal{D}_{22} - \mathcal{D}_{12}^2) \\
&= \frac{\det \mathcal{D}}{(k_BTg_1)^2} \\
&= \frac{\det L}{(k_BTg_1)^2} \\
&> 0,
\end{aligned}
$$

using (7.13) and Proposition 5.8(3). □

The above lemma shows that for given g_1 and g_2, the temperature

$$k_BT(g_1, g_2) = f^{-1}(g_2/g_1)$$

is well defined. Notice that the value $g_1 = 0$ is not possible since this would imply $\mathcal{D}_{11} = 0$ and $\det \mathcal{D} = -\mathcal{D}_{12}^2 \leq 0$—contradiction.

We still have to derive formulas for n, $\mathcal{E}$ and μ as functions of g_1 and g_2. From now on, we assume spherically symmetric stricty monotone energy bands. For notational convenience we set

$$
\begin{aligned}
Q_i(k_BT) &= 2\pi\int_0^\infty \sqrt{\gamma(\varepsilon)}\gamma'(\varepsilon)e^{-\varepsilon/k_BT}\varepsilon^i \, d\varepsilon, \quad i = 0, 1, \\
P_i(k_BT) &= \int_0^\infty D(\varepsilon)e^{-\varepsilon/k_BT}\varepsilon^i \, d\varepsilon, \quad i = 0, 1.
\end{aligned}
$$

Then, by (7.6)-(7.7),

$$n = e^{q\mu/k_BT}Q_0(k_BT), \quad \mathcal{E} = e^{q\mu/k_BT}Q_1(k_BT)$$

and

$$g_1 = e^{q\mu/k_BT}P_0(k_BT), \quad g_2 = e^{q\mu/k_BT}P_1(k_BT).$$

Observing the relations

$$\frac{n}{g_1} = \frac{Q_0(k_B T)}{P_0(k_B T)}, \quad \frac{\mathcal{E}}{g_2} = \frac{Q_1(k_B T)}{P_1(k_B T)},$$

we can compute, for given g_1, g_2,

$$n(g_1, g_2) = g_1 \frac{Q_0(k_B T(g_1, g_2))}{P_0(k_B T(g_1, g_2))}, \quad \mathcal{E} = g_2 \frac{Q_1(k_B T(g_1, g_2))}{P_1(k_B T(g_1, g_2))}.$$

The chemical potential μ can be expressed in terms of g_1, g_2 by using the expression $n = \exp(q\mu/k_B T)Q_0(k_B T)$:

$$
\begin{aligned}
\mu(g_1, g_2) &= \frac{k_B T}{q} \ln \frac{Q_0(k_B T)}{n(g_1, g_2)} = \frac{k_B T}{q} \ln \frac{Q_0(k_B T)P_0(k_B T)}{g_1 Q_0(k_B T)} \\
&= \frac{k_B T(g_1, g_2)}{q} \ln \frac{P_0(k_B T(g_1, g_2))}{g_1}.
\end{aligned}
$$

Thus we can summarize the energy -transport model in the (g_1, g_2)-formulation:

$$
\begin{aligned}
\partial_t n(g_1, g_2) - \frac{1}{q}\operatorname{div} J_1 &= 0, \\
\partial_t \mathcal{E}(g_1, g_2) - \operatorname{div} J_2 &= -\nabla V \cdot J_1 + W(\mu(g_1, g_2), T(g_1, g_2)), \\
J_1 &= \nabla g_1 - \frac{q\nabla V}{k_B T(g_1, g_2)} g_1, \\
J_2 &= \nabla g_2 - \frac{q\nabla V}{k_B T(g_1, g_2)} g_2, \\
\varepsilon_s \Delta V &= q(n(g_1, g_2) - C(x)).
\end{aligned}
$$

We have derived three different formulations of the energy-transport equations: the formulation in the (primal) entropy variables (u_1, u_2) (or, equivalently, μ, T); the formulation in the dual entropy variables (w_1, w_2); and in the variables (g_1, g_2). Table 7.1 summarizes the advantages and disadvantages of these formulations.

Formulation	Advantages	Disadvantages
(primal) entropy variables (u_1, u_2)	thermodynamical variables	drift terms, full diffusion matrix
dual entropy variables (w_1, w_2)	thermodynamical variables, no drift terms	full diffusion matrix
variables (g_1, g_2)	drift terms	diagonal diffusion matrix

Figure 7.1: Comparison of different formulations of the energy-transport model.

7.3 A non-parabolic band approximation

We compute the electron density $n(\mu, T)$, the energy density $\mathcal{E}(\mu, T)$, the diffusion matrices $\mathcal{D}_{ij}(\mu, T)$ and the relaxation term $W(\mu, T)$ in the case of Kane's non-parabolic band approximation [31]. We impose the following hypotheses (see [21]):

- The energy-band diagram is given by

$$\varepsilon(1 + \alpha\varepsilon) = \frac{\hbar^2 |k|^2}{2m^*}, \quad k \in \mathbb{R}^d, \ \alpha > 0.$$

- The relaxation time is given as in Example 5.5 by

$$\tau(\varepsilon) = \frac{1}{\phi(\varepsilon)N(\varepsilon)},$$

  where we choose $\phi(\varepsilon) = \phi_0 \varepsilon^\beta$ with $\phi_0 > 0$, $\beta > -2$.

- The relaxation term is given by the Fokker-Planck approximation (see Section 5.5):

$$W(\mu, T) = \int_0^\infty \frac{\partial}{\partial\varepsilon}\left[A(\varepsilon)\left(e^{(q\mu-\varepsilon)/k_B T} + k_B T_L \frac{\partial}{\partial\varepsilon}e^{(q\mu-\varepsilon)/k_B T}\right)\right]\varepsilon\,d\varepsilon,$$

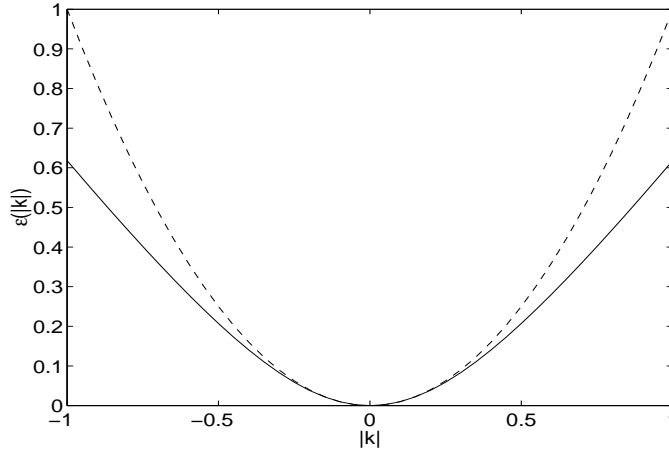 where $A(\varepsilon) = \phi(\varepsilon)N(\varepsilon)^2$.

Figure 7.2: Illustration of the energy band $\varepsilon(|k|)$ in the parabolic case (broken line) and the non-parabolic case (solid line) with parameters $\alpha = 0.25$, $\hbar/2m^* = 1$.

The parameter α is called *non-parabolicity parameter*. The difference to parabolic bands is shown in Figure 7.2. The function γ defined by $|k|^2 = \gamma(\varepsilon(|k|))$ reads as follows:

$$\gamma(\varepsilon) = \frac{2m^*}{\hbar^2}\varepsilon(1 + \alpha\varepsilon), \quad \varepsilon \geq 0.$$

Recall that the density of states is given by $N(\varepsilon) = 2\pi\sqrt{\gamma(\varepsilon)}\gamma'(\varepsilon)$. With the above hypotheses we are able to compute n, $\mathcal{E}$, $\mathcal{D}$ and W.

Direct computations yield (see the definitions (7.6), (7.7) and (7.9))

$$
\begin{aligned}
n &= 2\pi e^{q\mu/k_B T}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\int_0^\infty \sqrt{\varepsilon(1+\alpha\varepsilon)}\,(1+2\alpha\varepsilon)e^{-\varepsilon/k_B T}\,d\varepsilon \\
&= 2\pi e^{q\mu/k_B T}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}(k_B T)^{3/2} \\
&\quad \times \int_0^\infty \sqrt{1+\alpha k_B T u}\,(1+2\alpha k_B T u)u^{1/2}e^{-u}\,du \\
&= q_{1/2}(\alpha k_B T)(k_B T)^{3/2}e^{q\mu/k_B T},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{E} &= 2\pi e^{q\mu/k_B T}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\int_0^\infty \sqrt{1+\alpha\varepsilon}\,(1+2\alpha\varepsilon)\varepsilon^{3/2}e^{-\varepsilon/k_B T}\,d\varepsilon \\
&= q_{3/2}(\alpha k_B T)(k_B T)^{5/2}e^{q\mu/k_B T} \\
&= \frac{q_{3/2}(\alpha k_B T)}{q_{1/2}(\alpha k_B T)}k_B T n,
\end{aligned}
$$

where

$$
q_\beta(\alpha k_B T) = 2\pi\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\int_0^\infty \sqrt{1+\alpha k_B T u}\,(1+2\alpha k_B T u)u^\beta e^{-u}\,du.
$$

For the computation of $\mathcal{D}_{ij}$ we start from the formula (5.79) derived in Example 5.9 (identifying $\mathcal{D}_{ij}$ and $\mathcal{D}_{ij}\cdot\mathrm{Id}$)

$$
\begin{aligned}
\mathcal{D}_{ij} &= \frac{4q}{3\hbar^2}e^{q\mu/k_B T}\int_0^\infty \frac{\gamma(\varepsilon)\varepsilon^{i+j-2}}{\phi(\varepsilon)\gamma'(\varepsilon)^2}\,e^{-\varepsilon/k_B T}\,d\varepsilon \\
&= \frac{4q}{3\hbar^2}\frac{\hbar^2}{2m^*\phi_0}e^{q\mu/k_B T}\int_0^\infty \frac{1+\alpha\varepsilon}{(1+2\alpha\varepsilon)^2}\,\varepsilon^{i+j-\beta-1}e^{-\varepsilon/k_B T}\,d\varepsilon \\
&= e^{q\mu/k_B T}(k_B T)^{i+j-\beta}\,p_{i+j}(\alpha k_B T) \\
&= \frac{p_{i+j}(\alpha k_B T)}{q_{1,2}(\alpha k_B T)}\,(k_B T)^{i+j-\beta-3/2}\,n.
\end{aligned}
$$

where

$$
p_{i+j}(\alpha k_B T) = \frac{2q}{3m^*\phi_0}\int_0^\infty \frac{1+\alpha k_B T u}{(1+2\alpha k_B T u)^2}\,u^{i+j-\beta-1}\,du.
$$

Finally, we calculate the relaxation term, using the results of Example 5.11,

$$
\begin{aligned}
W &= -4\pi^2\phi_0 \left(\frac{2m^*}{\hbar^2}\right)^3 e^{q\mu/k_BT} \left(1 - \frac{T_L}{T}\right) \int_0^\infty \varepsilon^\beta \gamma(\varepsilon)\gamma'(\varepsilon)^2 e^{-\varepsilon/k_BT}\, d\varepsilon \\
&= -4\pi^2\phi_0 \left(\frac{2m^*}{\hbar^2}\right)^3 e^{q\mu/k_BT} \left(1 - \frac{T_L}{T}\right) (k_BT)^{\beta+2} \\
&\quad \times \int_0^\infty (1 + \alpha k_BTu)(1 + 2\alpha k_BTu)^2 u^{1+\beta} e^{-u}\, du \\
&= -4\pi^2\phi_0 \left(\frac{2m^*}{\hbar^2}\right)^3 e^{q\mu/k_BT} (k_BT)^{\beta+1} k_B(T - T_L)[\Gamma(\beta + 2) \\
&\quad + 5\alpha k_BT\,\Gamma(\beta + 3) + 8(\alpha k_BT)^2\,\Gamma(\beta + 4) + 4(\alpha k_BT)^3\,\Gamma(\beta + 5)] \\
&= -\frac{k_B(T - T_L)n}{\tau(T)},
\end{aligned}
$$

where Γ denotes the Gamma function defined by

$$
\Gamma(s) = \int_0^\infty u^{s-1} e^{-u}\, du, \quad s > 0,
$$

and $\tau(T)$ is a temperature-dependent relaxation time:

$$
\begin{aligned}
\frac{1}{\tau(T)} &= 2\pi\phi_0 \left(\frac{2m^*}{\hbar^2}\right)^{3/2} \frac{(k_BT)^{\beta-1/2}}{q_{1/2}(\alpha k_BT)} [\Gamma(\beta + 2) + 5\alpha k_BT\,\Gamma(\beta + 3) \\
&\quad + 8(\alpha k_BT)^2\,\Gamma(\beta + 4) + 4(\alpha k_BT)^3\,\Gamma(\beta + 5)].
\end{aligned}
$$

In the parabolic band case $\alpha = 0$ and the Chen model $\beta = 1/2$ we recover the expressions for n, $\mathcal{E}$, $\mathcal{D}$ and W from Examples 5.10 and 5.11.

8 From Kinetic to Quantum Hydrodynamic Models

8.1 The quantum hydrodynamic equations: first derivation

In Chapter 1 we have sketched the (formal) equivalence of the single-electron Schrödinger equation with electrostatic potential $V = V(x,t)$

$$i\hbar\partial_t\psi = -\frac{\hbar^2}{2m}\Delta\psi - qV\psi, \quad x \in \mathbb{R}^d,\ t > 0,$$

and the zero-temperature quantum hydrodynamic equations

$$\partial_t n - \frac{1}{q}\mathrm{div}\, J = 0,$$

$$\partial_t J - \frac{1}{q}\mathrm{div}\left(\frac{J \otimes J}{n}\right) + \frac{q^2}{m}n\nabla V + \frac{\hbar^2 q}{2m^2}n\nabla\left(\frac{\Delta\sqrt{n}}{\sqrt{n}}\right) = 0, \quad x \in \mathbb{R}^d,\ t > 0,$$

via the Madelung transform $\psi = \sqrt{n}\exp(imS/\hbar)$, where

$$n = |\psi|^2, \quad J = -\frac{\hbar q}{m}\mathrm{Im}(\overline{\psi}\nabla\psi)$$

are the electron density and electron current density, respectively.

In this section we consider an electron ensemble which is represented by a *mixed state* (see Section 4.1). A mixed quantum mechanical state consists of a sequence of single states with occupation probabilities $\lambda_k \geq 0$ ($k \in \mathbb{N}$) for the k-th state described by the Schrödinger equations

$$i\hbar\partial_t\psi_k = -\frac{\hbar^2}{2m}\Delta\psi_k - qV\psi_k, \quad x \in \mathbb{R}^d,\ t > 0, \tag{8.1}$$
$$\psi_k(x,0) = \psi_{I,k}(x), \quad x \in \mathbb{R}^d.$$

The occupation probabilities satisfy $\sum_{k=0}^{\infty} \lambda_k = 1$. We write the initial datum of the k-th state in the form

$$\psi_{I,k} = \sqrt{n_{I,k}}\exp(imS_{I,k}/\hbar). \tag{8.2}$$

The electron density n_k and the electron current density J_k of the k-th single state are given by

$$n_k = |\psi_k|^2, \quad J_k = -\frac{\hbar q}{m}\mathrm{Im}(\overline{\psi}_k\nabla\psi_k).$$

For the fluiddynamical formulation we use the Madelung transform

$$\psi_k = \sqrt{n_k}\exp(imS_k/\hbar). \tag{8.3}$$

Clearly, we have to assume that $n_k > 0$ for all $k \geq 0$. With this transform the current density now reads

$$J_k = -qn_k\nabla S_k. \tag{8.4}$$

We wish to find evolution equations for n_k and J_k. Setting (8.3) into (8.1) gives, after division of $\exp(imS_k/\hbar)$,

$$\frac{i\hbar}{2}\frac{\partial_t n_k}{\sqrt{n_k}} - m\sqrt{n_k}\partial_t S_k = -\frac{\hbar^2}{2m}\left(\Delta\sqrt{n_k} + \frac{2im}{\hbar}\nabla\sqrt{n_k}\cdot\nabla S_k - \frac{m^2}{\hbar^2}\sqrt{n_k}|\nabla S_k|^2\right.$$
$$\left. + \frac{im}{\hbar}\sqrt{n_k}\Delta S_k\right) - qV\sqrt{n_k}. \tag{8.5}$$

The imaginary part of this equation is

$$\partial_t n_k = -2\sqrt{n_k}\nabla\sqrt{n_k}\cdot\nabla S_k - n_k\Delta S_k = -\text{div}\,(n_k\nabla S_k)$$

or, using (8.4),

$$\partial_t n_k - \frac{1}{q}\text{div}\,J_k = 0. \tag{8.6}$$

The real part of (8.5) reads

$$\partial_t S_k = \frac{\hbar^2}{2m^2}\frac{\Delta\sqrt{n_k}}{\sqrt{n_k}} - \frac{1}{2}|\nabla S_k|^2 + \frac{q}{m}V.$$

In order to find an evolution equation for $J_k = -qn_k\nabla S_k$, we take the gradient of the above equation and multiply the resulting equation by $-qn_k$:

$$-qn_k\partial_t\nabla S_k = -\frac{\hbar^2 q}{2m^2}n_k\nabla\left(\frac{\Delta\sqrt{n_k}}{\sqrt{n_k}}\right) + \frac{q}{2}n_k\nabla|\nabla S_k|^2 - \frac{q^2}{m}n_k\nabla V.$$

Since

$$-qn_k\partial_t\nabla S_k = \partial_t J_k + q\partial_t n_k\nabla S_k = \partial_t J_k + (\text{div}\,J_k)\nabla S_k$$

and

$$\frac{q}{2}n_k\nabla|\nabla S_k|^2 = q\text{div}\,(n_k\nabla S_k \otimes \nabla S_k) - q\text{div}\,(n_k\nabla S_k)\nabla S_k$$
$$= \frac{1}{q}\text{div}\left(\frac{J_k \otimes J_k}{n_k}\right) + (\text{div}\,J_k)\nabla S_k,$$

the above equation can be written as

$$\partial_t J_k - \frac{1}{q}\text{div}\left(\frac{J_k \otimes J_k}{n_k}\right) + \frac{q^2}{2}n_k\nabla V + \frac{\hbar^2 q}{2m^2}n_k\nabla\left(\frac{\Delta\sqrt{n_k}}{\sqrt{n_k}}\right) = 0. \tag{8.7}$$

Here, $J_k \otimes J_k$ is a tensor (or matrix) with components $J_{k,i}J_{k,j}$ for $i,j = 1,\ldots,d$.

The total carrier density n and current density J of the mixed state are given by

$$n = \sum_{k=0}^{\infty} \lambda_k n_k, \quad J = \sum_{k=0}^{\infty} \lambda_k J_k.$$

In the following, we will derive evolution equations for n and J from (8.6) and (8.7). Multiplication of (8.6) and (8.7) by λ_k and summation over k yields

$$\partial_t n - \frac{1}{q} \operatorname{div} J = 0$$

and

$$\partial_t J - \frac{1}{q} \sum_{k=0}^{\infty} \lambda_k \operatorname{div} \left(\frac{J_k \otimes J_k}{n_k} \right) + qn\nabla V + \frac{\hbar^2 q}{2m^2} \sum_{k=0}^{\infty} \lambda_k n_k \nabla \left(\frac{\Delta\sqrt{n_k}}{\sqrt{n_k}} \right) = 0. \quad (8.8)$$

In order to rewrite the second and fourth term on the left-hand side of this equation, we introduce the "current" velocities

$$u_c = -\frac{J}{qn}, \quad u_{c,k} = -\frac{J_k}{qn_k}$$

and the "osmotic" velocities

$$u_{os} = \frac{\hbar}{2m} \nabla \log n, \quad u_{os,k} = \frac{\hbar}{2m} \nabla \log n_k.$$

The notion "osmotic" comes from the fact that

$$P_k = \frac{\hbar^2 q}{4m^2} n_k (\nabla \otimes \nabla) \log n_k$$

can be interpreted as a non-diagonal pressure tensor since

$$\frac{\hbar^2 q}{2m^2} n_k \nabla \left(\frac{\Delta\sqrt{n_k}}{\sqrt{n_k}} \right) = \operatorname{div} P_k.$$

Then

$$-\frac{1}{q} \sum_{k=0}^{\infty} \lambda_k \operatorname{div} \left(\frac{J_k \otimes J_k}{n_k} \right) = -q \sum_{k=0}^{\infty} \operatorname{div} \left(\lambda_k n_k u_{c,k} \otimes u_{c,k} \right)$$

$$= -q \sum_{k=0}^{\infty} \lambda_k \operatorname{div} \left(n_k (u_{c,k} - u_c) \otimes (u_{c,k} - u_c) + 2n_k u_{c,k} \otimes u_c \right)$$

$$+ q \operatorname{div} \left(nu \otimes u \right)$$

$$= -q \sum_{k=0}^{\infty} \operatorname{div} \left(\lambda_k n_k (u_{c,k} - u_c) \otimes (u_{c,k} - u_c) + \frac{2\lambda_k}{q^2} \frac{J_k \otimes J}{n} \right)$$

$$+ \frac{1}{q} \operatorname{div} \left(\frac{J \otimes J}{n} \right)$$

$$= -\frac{1}{q} \operatorname{div} \left(\frac{J \otimes J}{n} \right) - \operatorname{div} \left(n\theta_c \right),$$

145

where

$$\theta_c = q \sum_{k=0}^{\infty} \lambda_k \frac{n_k}{n} (u_{c,k} - u_c) \otimes (u_{c,k} - u_c)$$

is called the current temperature. Furthermore,

$$\frac{\hbar^2 q}{2m^2} \sum_{k=0}^{\infty} \lambda_k n_k \nabla \left(\frac{\Delta\sqrt{n_k}}{\sqrt{n_k}} \right)$$

$$= \frac{\hbar^2 q}{4m^2} \sum_{k=0}^{\infty} \lambda_k \mathrm{div} \left[(\nabla \otimes \nabla) n_k - \frac{\nabla n_k \otimes \nabla n_k}{n_k} \right]$$

$$= \frac{\hbar^2 q}{4m^2} \sum_{k=0}^{\infty} \lambda_k \mathrm{div} \left[(\nabla \otimes \nabla) n_k + n_k \frac{\nabla n \otimes \nabla n}{n^2} \right.$$

$$\left. - n_k \left(\frac{\nabla n_k}{n_k} - \frac{\nabla n}{n} \right) \otimes \left(\frac{\nabla n_k}{n_k} - \frac{\nabla n}{n} \right) - 2 \frac{\nabla n \otimes \nabla n_k}{n} \right]$$

$$= \frac{\hbar^2 q}{4m^2} \mathrm{div} \left((\nabla \otimes \nabla) n - \frac{\nabla n \otimes \nabla n}{n} \right) - \mathrm{div}\, (n\theta_{\mathrm{os}})$$

$$= \frac{\hbar^2 q}{2m^2} n \nabla \left(\frac{\Delta\sqrt{n}}{\sqrt{n}} \right) - \mathrm{div}\, (n\theta_{\mathrm{os}}),$$

where

$$\theta_{\mathrm{os}} = q \sum_{k=0}^{\infty} \lambda_k \frac{n_k}{n} (u_{\mathrm{os},k} - u_{\mathrm{os}}) \otimes (u_{\mathrm{os},k} - u_{\mathrm{os}})$$

is termed osmotic temperature.

Hence, the momentum equation (8.8) becomes

$$\partial_t J - \frac{1}{q} \mathrm{div} \left(\frac{J \otimes J}{n} \right) - \mathrm{div}\, (n\theta) + qn\nabla V + \frac{\hbar^2 q}{2m^2} n \nabla \left(\frac{\Delta\sqrt{n}}{\sqrt{n}} \right) = 0, \qquad (8.9)$$

where $\theta = \theta_c + \theta_{\mathrm{os}}$ is called *temperature tensor*. Equation (8.9) and the mass conservation equation

$$\partial_t n - \frac{1}{q} \mathrm{div}\, J = 0$$

are referred to as the *quantum hydrodynamic equations*.

The temperature tensor cannot be expressed in terms of n and J without further assumptions, and as in classical fluid dynamics, we need a *closure condition* to obtain a closed set of equations (similarly as in Section 4.2). In the literature, the following condition has been used: The temperature is a scalar (times the identity matrix),

$$\theta = T \cdot \mathrm{Id},$$

and the temperature scalar T satisfies either

- $T = T_0$, where $T_0 > 0$ is a constant (the so-called *isothermal case*), or

- $T = T(n) = T_0 n^{\alpha-1}$ depends on the carrier density, where $T_0 > 0$ and $\alpha > 1$ are constants (the so-called *isentropic case*).

In these cases the (classical) pressure $P = n\theta$ becomes

$$P(n) = T_0 n^\alpha \cdot \mathrm{Id}$$

and $\operatorname{div} P = \operatorname{div}(T_0 n^\alpha)$, with $\alpha = 1$ in the isothermal case and $\alpha > 1$ in the isentropic case.

8.2 The quantum hydrodynamic equations: second derivation

In this section, the quantum hydrodynamic equations are derived as a set of nonlinear conservation laws by a moment expansion of a Wigner-Boltzmann equation and an expansion of the thermal equilibrium Wigner distribution function up to order $O(\hbar^2)$. We proceed similarly as in [24].

We start with a Wigner-Boltzmann equation of the form (see (4.25), (4.26))

$$\partial_t w + \frac{\hbar}{m} k \cdot \nabla_x w + \frac{q}{\hbar}\theta[V]w = \frac{1}{\tau}\operatorname{div}_k(kw) + \frac{k_B T_0 m}{\tau \hbar^2}\Delta_k w, \qquad (8.10)$$
$$x, k \in \mathbb{R}^d, \ t > 0,$$
$$w(x, k, 0) = w_I(x, k), \qquad x, k \in \mathbb{R}^d, \qquad (8.11)$$

where $d \geq 1$ is the space dimension and $w(x, k, t)$ is the Wigner distribution function depending on the space variable x, the wave vector k and the time t. Notice that we have implicitly assumed a parabolic band approximation. The operator $\theta[V]$ is defined in the sense of pseudo-differential operators [42] as

$$(\theta[V]w)(x, k, t) = \frac{i}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{m}{\hbar} \left[V\left(x + \frac{\hbar}{2m}\eta, t\right) - V\left(x - \frac{\hbar}{2m}\eta, t\right) \right]$$
$$\times w(x, k, t) e^{-i(k-k')\cdot\eta} \, dk' \, d\eta,$$

where $V(x, t)$ is the electrostatic potential. The right-hand side of (8.10) models the interaction of the electrons with the phonons of the crystal lattice. The physical parameters are the reduced Planck constant $\hbar = h/2\pi$, the effective mass of the electrons m, the elementary charge q, the lattice temperature T_0, and the momentum relaxation time τ whose inverse is a measure of the strength of the coupling between the electrons and phonons.

The existence and uniqueness of local-in-time (so-called mild) solutions to the whole-space problem (8.10)-(8.11) has been shown in [7]. Existence of global-in-time solutions of the one-dimensional problem with periodic boundary conditions has been proved in [5].

The particle density n and current density J are related to the Wigner function by the formulas

$$n(x,t) = \int_{\mathbb{R}^d} w(x,k,t)\,dk \quad \text{and} \quad J(x,t) = -\frac{q\hbar}{m}\int_{\mathbb{R}^d} w(x,k,t)k\,dk. \qquad (8.12)$$

These formulas follow from Section 3.1 observing that in the parabolic band approximation, the Brillouin zone is extended to $\mathbb{R}^d$ and the mean velocity equals $v(k) = \hbar k/m$. Usually, the electrostatic potential is self-consistently produced by the electrons moving in the semiconductor crystal with fixed charged background ions of density $C(x)$:

$$\operatorname{div}(\varepsilon_s \nabla V) = q(n - C(x)), \quad x \in \mathbb{R}^d,$$

where ε_s denotes the permittivity of the material.

In order to derive macroscopic equations from (8.10) we apply a moment method. The first moments are defined by

$$
\begin{aligned}
\langle 1 \rangle &= n = \int_{\mathbb{R}^d} w(x,k,t)\,dk, \\
\langle k_j \rangle &= \int_{\mathbb{R}^d} w(x,k,t)k_j\,dk, \\
\langle k_j k_\ell \rangle &= \int_{R^d} w(x,k,t)k_j k_\ell\,dk,
\end{aligned}
$$

with $1 \le j, \ell \le d$. Integrating (8.10) over $\mathbb{R}^d$ with respect to k, we obtain

$$
\begin{aligned}
\partial_t \langle 1 \rangle + \frac{\hbar}{m}\operatorname{div}_x\langle k \rangle + \frac{q}{\hbar}\int_{\mathbb{R}^d} \theta[V]w\,dk &= \frac{1}{\tau}\int_{\mathbb{R}^d}\left(\operatorname{div}_k(kw) + \frac{k_B T_0 m}{\tau \hbar^2}\Delta_k w\right)dk \\
&= 0.
\end{aligned}
$$

For the evaluation of the integral on the left-hand side we recall the Fourier transform

$$\hat{\phi}(x,\eta,t) = \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\phi(x,k,t)e^{ik\cdot\eta}\,dk$$

with inverse

$$\phi(x,k,t) = \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\hat{\phi}(x,\eta,t)e^{-ik\cdot\eta}\,d\eta.$$

Integrating the last equation over $\mathbb{R}^d$ with respect to k gives the formula

$$\hat{\phi}(x,0,t) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\hat{\phi}(x,\eta,t)e^{-ik\cdot\eta}\,d\eta\,dk. \qquad (8.13)$$

With this expression we get

$$
\begin{aligned}
\int_{\mathbb{R}^d} \theta[V]w\, dk &= \frac{i}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{m}{\hbar} \left[V\left(x + \frac{\hbar}{2m}\eta, t\right) - V\left(x - \frac{\hbar}{2m}\eta, t\right) \right] \\
&\quad \times \hat{w}(x, \eta, t) e^{-ik\cdot\eta}\, d\eta\, dk \\
&= i(2\pi)^{d/2} \frac{m}{\hbar} \left[V\left(x + \frac{\hbar}{2m}\eta, t\right) - V\left(x - \frac{\hbar}{2m}\eta, t\right) \right]_{\eta=0} \hat{w}(x, 0, t) \\
&= 0,
\end{aligned}
$$

and therefore

$$
\partial_t \langle 1 \rangle + \frac{\hbar}{m}\, \mathrm{div}_x \langle k \rangle = 0
$$

or, with (8.12),

$$
\partial_t n - \frac{1}{q}\, \mathrm{div}_x J = 0. \tag{8.14}
$$

This equation expresses the conservation of mass.

Now we multiply (8.10) with k_j and integrate over $\mathbb{R}^d$ with respect to k. Since

$$
\begin{aligned}
\int_{\mathbb{R}^d} &\left(\mathrm{div}_k(kw) + \frac{k_B T_0 m}{\tau \hbar^2} \Delta_k w \right) k_j\, dk \\
&= -\int_{R^d} \left((k \cdot \nabla_k k_j)w + \frac{k_B T_0 m}{\tau \hbar^2} (\nabla_k k_j) \cdot (\nabla_k w) \right) dk \\
&= -\int_{\mathbb{R}^d} \left(k_j w + \frac{k_B T_0 m}{\tau \hbar^2} \frac{\partial w}{\partial k_j} \right) dk \\
&= -\langle k_j \rangle
\end{aligned}
$$

and, using integration by parts and (8.13),

$$
\begin{aligned}
\int_{\mathbb{R}^d} &\theta[V]w k_j\, dk \\
&= \frac{im}{(2\pi)^{d/2}\hbar} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left[V\left(x + \frac{\hbar}{2m}\eta, t\right) - V\left(x - \frac{\hbar}{2m}\eta, t\right) \right] \\
&\quad \times \hat{w}(x, \eta, t) i\frac{\partial}{\partial \eta_j} e^{-ik\cdot\eta}\, d\eta\, dk \\
&= -(2\pi)^{d/2} \frac{m}{\hbar} \frac{\partial}{\partial \eta_j} \left[V\left(x + \frac{\hbar}{2m}\eta, t\right) - V\left(x - \frac{\hbar}{2m}\eta, t\right) \right]_{\eta=0} \hat{w}(x, 0, t) \\
&\quad - (2\pi)^{d/2} \frac{m}{\hbar} \left[V\left(x + \frac{\hbar}{2m}\eta, t\right) - V\left(x - \frac{\hbar}{2m}\eta, t\right) \right]_{\eta=0} \frac{\partial \hat{w}}{\partial \eta_j}(x, 0, t) \\
&= -\frac{\partial V}{\partial x_j}(x, t) \int_{\mathbb{R}^d} w(x, k, t)\, dk \\
&= -\frac{\partial V}{\partial x_j} \langle 1 \rangle,
\end{aligned}
$$

149

this yields

$$\partial_t \langle k_j \rangle + \frac{\hbar}{m} \sum_{\ell=1}^{d} \frac{\partial}{\partial x_\ell} \langle k_j k_\ell \rangle - \frac{q}{\hbar} \frac{\partial V}{\partial x_j} \langle 1 \rangle = -\frac{1}{\tau} \langle k_j \rangle$$

or

$$\partial_t J_j - \frac{q\hbar^2}{m^2} \sum_{\ell=1}^{d} \frac{\partial}{\partial x_\ell} \langle k_j k_\ell \rangle + \frac{q^2}{m} \frac{\partial V}{\partial x_j} n = -\frac{J_j}{\tau}. \tag{8.15}$$

The system (8.14) and (8.15) has to be closed by expressing the term $\langle k_j k_\ell \rangle$ in terms of the primal variables n and J. As in the case of classical kinetic theory (see Chapter 5), we achieve the closure by assuming that the Wigner function w is close to a wave vector displaced equilibrium density such that

$$w(x, k, t) = w_e\Big(x, k - \frac{m}{\hbar} u(x, t), t\Big),$$

where $u(x, t)$ is some group velocity and

$$w_e(x, k, t) \tag{8.16}$$

$$= A(x, t) \exp\Big(-\frac{\hbar^2}{2mk_B T} |k|^2 + \frac{qV}{k_B T}\Big) \Big[1 + \hbar^2 \Big\{ \frac{q}{8m(k_B T)^2} \Delta_x V \tag{8.17}$$

$$+ \frac{q^2}{24m(k_B T)^3} |\nabla_x V|^2 - \frac{q^2}{24(k_B T)^3} \sum_{j,\ell=1}^{d} v_j(k) v_\ell(k) \frac{\partial^2 V}{\partial x_j \partial x_\ell} \Big\} + O(\hbar^4) \Big],$$

where we recall that $v(k) = \hbar k/m$. The form (8.16) is derived from an $O(\hbar^2)$ approximation of the thermal equilibrium density first given by Wigner [43]. The electron temperature $T = T_0$ is here assumed to be constant (see the following section for non-constant temperature).

The symmetry of w_e with respect to k implies that the odd order moments of w_e vanish. Therefore

$$\langle 1 \rangle = \int_{\mathbb{R}^d} w_e\Big(x, k - \frac{m}{\hbar} u(x, t), t\Big) dk = \int_{\mathbb{R}^d} w_e(x, \eta, t) d\eta,$$

$$\langle k_j \rangle = \int_{\mathbb{R}^d} \Big(\eta_j + \frac{m}{\hbar} u_j\Big) w_e(x, \eta, t) d\eta = \frac{m}{\hbar} n u_j,$$

$$\hbar^2 \langle k_j k_\ell \rangle = \hbar^2 \int_{\mathbb{R}^d} \Big(\eta_j + \frac{m}{\hbar} u_j\Big)\Big(\eta_\ell + \frac{m}{\hbar} u_\ell\Big) w_e(x, \eta, t) d\eta$$

$$= m^2 n u_j u_\ell + \hbar^2 \int_{\mathbb{R}^d} \eta_j \eta_\ell w_e(x, \eta, t) d\eta. \tag{8.18}$$

This implies $n = \langle 1 \rangle$ and $J = -qnu$.

We now derive an $O(\hbar^2)$ approximation for n. Since

$$\int_{\mathbb{R}} x^2 e^{-x^2/2} dx = \int_{\mathbb{R}} e^{-x^2/2} \cdot 1 dx = \sqrt{2\pi},$$

$$\int_{\mathbb{R}} x^4 e^{-x^2/2} dx = 3 \int_{\mathbb{R}} e^{-x^2/2} x^2 dx = 3\sqrt{2\pi},$$

150

we obtain

$$\int_{\mathbb{R}^d} e^{-\hbar^2|\eta|^2/2mk_BT}\,d\eta = (mk_BT)^{d/2}\hbar^{-d}\int_{\mathbb{R}^d} e^{-|z|^2/2}\,dz = (2\pi mk_BT)^{d/2}\hbar^{-d}$$

and

$$\begin{aligned}
\int_{\mathbb{R}^d} \eta_j\eta_\ell e^{-\hbar^2|\eta|^2/2mk_BT}\,d\eta &= (mk_BT)^{d/2}\hbar^{-d}(mk_BT\hbar^{-2})\int_{\mathbb{R}^d} z_jz_\ell e^{-|z|^2/2}\,dz \\
&= (2\pi mk_BT)^{d/2}mk_BT\hbar^{-d-2}\delta_{j\ell},
\end{aligned}$$

and hence, with $C = A(x,t)(2\pi mk_BT)^{d/2}\hbar^{-d}$,

$$\begin{aligned}
n &= \int_{\mathbb{R}^d} w_e(x,\eta,t)\,d\eta \\
&= Ce^{qV/k_BT}\left(1 + \frac{q\hbar^2}{8m(k_BT)^2}\Delta_x V + \frac{q^2\hbar^2}{24m(k_BT)^3}|\nabla_x V|^2\right. \\
&\quad \left. -\frac{q^2\hbar^4}{24m^2(k_BT)^3}\frac{mk_BT}{\hbar^2}\sum_{j,\ell=1}^d \frac{\partial^2 V}{\partial x_j\partial x_\ell}\delta_{j\ell}\right) + O(\hbar^4) \qquad (8.19) \\
&= Ce^{qV/k_BT}\left(1 + \frac{q\hbar^2}{12m(k_BT)^2}\Delta_x V + \frac{q^2\hbar^2}{24m(k_BT)^3}|\nabla_x V|^2\right) + O(\hbar^4).
\end{aligned}$$

In order to compute the last integral in (8.18), we have to evaluate

$$\begin{aligned}
\int_{\mathbb{R}^d} \eta_j\eta_\ell\eta_m\eta_n e^{-\hbar^2|\eta|/2mk_BT}\,d\eta &= (mk_BT)^{d/2+2}\hbar^{-d-4}\int_{\mathbb{R}^d} z_jz_\ell z_mz_n e^{-|z|^2/2}\,dz \\
&= (mk_BT)^{d/2+2}\hbar^{-d-4}\alpha(j,\ell,m,n),
\end{aligned}$$

where

$$\alpha(j,\ell,m,n) = (2\pi)^{d/2}\begin{cases} 3 & : & j=\ell=m=n \\ 1 & : & j=\ell\neq m=n \\ 1 & : & j=m\neq\ell=n \\ 1 & : & j=n\neq\ell=m \\ 0 & : & \text{else.} \end{cases}$$

This yields

$$\begin{aligned}
\hbar^2 &\int_{\mathbb{R}^d} \eta_j\eta_\ell w_e(x,\eta,t)\,d\eta \\
&= Ce^{qV/k_BT}\left[\left(mk_BT + \frac{q\hbar^2}{8k_BT}\Delta_x V + \frac{q^2\hbar^2}{24(k_BT)^2}|\nabla_x V|^2\right)\delta_{j\ell}\right. \\
&\quad \left. -\frac{q\hbar^2}{24k_BT}\sum_{m,n=1}^d \frac{\partial^2 V}{\partial x_m\partial x_n}\alpha(j,\ell,m,n)\right] + O(\hbar^4).
\end{aligned}$$

A simple computation shows that

$$\sum_{m,n=1}^{d} \frac{\partial^2 V}{\partial x_m \partial x_n} \alpha(j,\ell,m,n) = 2\frac{\partial^2 V}{\partial x_j \partial x_\ell} + \delta_{j\ell} \Delta_x V.$$

Hence

$$\hbar^2 \int_{\mathbb{R}^d} \eta_j \eta_\ell w_e(x,\eta,t) \, d\eta = Ce^{qV/k_B T}\left[\left(mk_B T + \frac{q\hbar^2}{12 k_B T}\Delta_x V \right. \right. \tag{8.20}$$
$$\left. \left. +\frac{q^2\hbar^2}{24(k_B T)^2}|\nabla_x V|^2\right)\delta_{j\ell} - \frac{q\hbar^2}{12 k_B T}\frac{\partial^2 V}{\partial x_j \partial x_\ell}\right] + O(\hbar^4).$$

We substitute (see (8.19))

$$Ce^{qV/k_B T}\left(mk_B T + \frac{q\hbar^2}{12 k_B T}\Delta_x V + \frac{q^2\hbar^2}{24(k_B T)^2}|\nabla_x V|^2\right) = mnk_B T + O(\hbar^4)$$

and

$$Ce^{qV/k_B T} = n + O(\hbar^2) \tag{8.21}$$

into (8.20) to conclude from (8.18)

$$\hbar^2\langle k_j k_\ell \rangle = m^2 n u_j u_\ell + mnk_B T\delta_{j\ell} - \frac{q\hbar^2}{12 k_B T}n\frac{\partial^2 V}{\partial x_j x_\ell} + O(\hbar^4). \tag{8.22}$$

The potential V is usually the sum of the electrostatic potential and an external potential modeling heterostructures (i.e. sandwiches of different materials). As the external potential can be discontinuous at the interfaces between two materials, also V may be discontinuous. Therefore, we wish to substitute the derivative $\partial^2 V/\partial x_j \partial x_\ell$ by an expression which only depends on n and its derivatives. From Taylor's formula $\log(x+\varepsilon) = \log x + x/\varepsilon + O(\varepsilon^2)$ for $x, \varepsilon > 0$ we obtain from (8.21) (with a slight abuse of notation)

$$\log n = \log(e^{qV/k_B T} + O(\hbar^2)) + \log C = \log e^{qV/k_B T} + \log C + O(\hbar^2)$$

and therefore,

$$\frac{\partial^2 \log n}{\partial x_j \partial x_\ell} = \frac{\partial^2}{\partial x_j \partial x_\ell}\left(\log\left(A(2\pi m k_B T)^{d/2}\right) + \frac{qV}{k_B T}\right) + O(\hbar^2).$$

We assume that $A(x,t)$ is slowly varying with respect to x such that we can approximate

$$\frac{\partial^2 \log n}{\partial x_j \partial x_\ell} = \frac{q}{k_B T}\frac{\partial^2 V}{\partial x_j \partial x_\ell} + O(\hbar^2).$$

For "$\hbar \to 0$", this corresponds to the classical thermal equilibrium expression $n = \text{const.}\ e^{qV/k_B T}$. We obtain finally from (8.22)

$$\hbar^2 \langle k_j k_\ell \rangle = m^2 n u_j u_\ell + m n k_B T \delta_{j\ell} - \frac{\hbar^2}{12} n \frac{\partial^2 \log n}{\partial x_j \partial x_\ell} + O(\hbar^4) \qquad (8.23)$$

and

$$\frac{q\hbar^2}{m^2} \sum_{\ell=1}^{d} \frac{\partial}{\partial x_\ell} \langle k_j k_\ell \rangle$$

$$= q \sum_{\ell=1}^{d} \frac{\partial}{\partial x_\ell}(n u_j u_\ell) + \frac{q k_B}{m} \frac{\partial}{\partial x_j}(nT) - \frac{\hbar^2 q}{12 m^2} \sum_{\ell=1}^{d} \frac{\partial}{\partial x_\ell}\left(n \frac{\partial^2 \log n}{\partial x_j \partial x_\ell} \right)$$

$$= \frac{1}{q} \sum_{\ell=1}^{d} \frac{\partial}{\partial x_\ell}\left(\frac{J_j J_\ell}{n} \right) + \frac{q k_B T}{m} \frac{\partial n}{\partial x_j} - \frac{\hbar^2 q}{6 m^2} n \frac{\partial}{\partial x_j}\left(\frac{\Delta_x \sqrt{n}}{\sqrt{n}} \right),$$

since $J = -qnu$.

With (8.14), (8.15) and the Poisson equation, the *(isothermal) quantum hydrodynamic equations* become

$$\partial_t n - \frac{1}{q} \text{div}_x J = 0, \qquad (8.24)$$

$$\partial_t J - \frac{1}{q} \text{div}_x \left(\frac{J \otimes J}{n} \right) - \frac{q k_B T}{m} \nabla_x n + \frac{\hbar^2 q}{6 m^2} n \nabla_x \left(\frac{\Delta_x \sqrt{n}}{\sqrt{n}} \right)$$

$$+ \frac{q^2}{m} n \nabla_x V = -\frac{J}{\tau}, \qquad (8.25)$$

$$\text{div}(\varepsilon_s \nabla_x V) = q(n - C(x)). \qquad (8.26)$$

Notice that the coefficient of the quantum term is $\hbar^2 q / 6m^2$ compared to the coefficient $\hbar^2 q / 2m^2$ in (8.9) obtained from the mixed-state approach of Section 8.1. The factor $1/3$ can be interpreted as a statistical factor coming from thermal averaging (see [25]).

8.3   Extensions

In this section we derive two extensions of the quantum hydrodynamic equations: a quantum hydrodynamic model including an evolution equation for the particle energy and a quantum hydrodynamic model taking into account viscous (or dissipative) effects.

Quantum hydrodynamic equations with energy. In the previous subsection we have assumed that the temperatue, which appears in the approximation of the thermal equilibrium Wigner function (8.16), is constant. We suppose

now that the temperature is time and space dependent, i.e. $T = T(x, t)$. Then the quantum hydrodynamic equations (8.24)-(8.26) are still valid but the pressure term in (8.25) has now to be written as

$$\frac{q}{m} \nabla_x (nT).$$

However, we also need an equation for the variable T. We derive this equation by calculating the next moment equation. For this we introduce the moments

$$\langle |k|^2 \rangle = \int_{\mathbb{R}^d} w(x, k, t)|k|^2 \, dk,$$

$$\langle k|k|^2 \rangle = \int_{\mathbb{R}^d} w(x, k, t)k|k|^2 \, dk.$$

Multiply (8.10) with $|k|^2$ and integrate over $\mathbb{R}^d$ with respect to k to obtain

$$\partial_t \langle |k|^2 \rangle + \frac{\hbar}{m} \text{div}_x \langle k|k|^2 \rangle + \frac{q}{\hbar} \int_{\mathbb{R}^d} \theta[V]w|k|^2 \, dk \qquad (8.27)$$

$$= \frac{1}{\tau} \int_{R^d} \left(\text{div}_k(kw) + \frac{k_B T_0 m}{\hbar^2} \Delta_k w \right) |k|^2 \, dk.$$

We first compute the integrals. For the integral on the right-hand side we get, after integration by parts,

$$\frac{1}{\tau} \int_{R^d} \left(\text{div}_k(kw) + \frac{k_B T_0 m}{\hbar^2} \Delta_k w \right) |k|^2 \, dk$$

$$= \frac{2}{\tau} \int_{\mathbb{R}^d} \left(-w|k|^2 + \frac{d k_B T_0 m}{\hbar^2} w \right) \, dk$$

$$= \frac{2}{\tau} \left(-\langle |k|^2 \rangle + \frac{d k_B T_0 m}{\hbar^2} \langle 1 \rangle \right).$$

Similarly as in the previous subsection, a computation gives

$$\frac{q}{\hbar} \int_{\mathbb{R}^d} \theta[V]w|k|^2 \, dk = \frac{2q}{\hbar} \nabla_x V \cdot \langle k \rangle = -\frac{2m}{\hbar^2} \nabla_x V \cdot J.$$

The two moments are calculated using the wave vector displaced equilibrium density (8.27). From (8.23) follows

$$\hbar^2 \langle |k|^2 \rangle = m^2 n|u|^2 + dmn k_B T - \frac{\hbar^2}{12} n \Delta \log n + O(\hbar^4).$$

We introduce now the *energy*

$$e = \frac{m}{2}|u|^2 + \frac{d}{2}k_B T - \frac{\hbar^2}{24m} \Delta \log n, \qquad (8.28)$$

consisting of the kinetic, thermal and quantum energy parts, and the *stress tensor*

$$P_{j\ell} = -nk_B T \delta_{j\ell} + \frac{\hbar^2}{12m} n \frac{\partial^2 \log n}{\partial x_j \partial x_\ell}, \quad j, \ell = 1, \ldots, d. \qquad (8.29)$$

Then $\hbar^2 \langle |k|^2 \rangle = 2me + O(\hbar^4)$ and tedious computations lead to

$$\begin{aligned}
\langle k|k|^2 \rangle &= \int_{\mathbb{R}^d} w_e(x, k - mu/\hbar, t)k|k|^2 \, dk \\
&= \frac{2m^2}{\hbar^2}(une - P \cdot u) + O(\hbar^2),
\end{aligned}$$

where $P = (P_{j\ell})_{j\ell}$. Thus we can write (8.27), up to order $O(\hbar^2)$, as

$$\partial_t(ne) + \mathrm{div}_x (une - P \cdot u) - \nabla_x V \cdot J = -\frac{1}{\tau}(2ne - dk_B T_0 n).$$

Using $u = -J/qn$, we obtain the *full quantum hydrodynamic equations*

$$\partial_t n - \frac{1}{q}\mathrm{div}_x J = 0,$$

$$\partial_t J - \frac{1}{q}\mathrm{div}_x \left(\frac{J \otimes J}{n} \right) + \frac{qk_B}{m}\nabla_x(nT) - \frac{\hbar^2 q}{6m^2}n\nabla_x \left(\frac{\Delta_x \sqrt{n}}{\sqrt{n}} \right) + \frac{q^2}{m}n\nabla_x V = -\frac{J}{\tau},$$

$$\partial_t(ne) - \frac{1}{q}\mathrm{div}_x \left(Je - \frac{P \cdot J}{n} \right) - \nabla_x V \cdot J = -\frac{2n}{\tau}\left(e - \frac{d}{2}k_B T_0 \right),$$

$$\mathrm{div}_x(\varepsilon_s \nabla_x V) = q(n - C(x)),$$

where e and $P = (P_{j\ell})$ are defined by (8.28) and (8.29), respectively.

The viscous quantum hydrodynamic equations. The derivation of the quantum hydrodynamic model is based on the Fokker-Planck operator

$$L_{FP}w = \frac{1}{\tau}\left(\mathrm{div}_k(kw) + \frac{k_B T_0 m}{\hbar^2}\Delta_k w \right)$$

which models interactions of electrons with a heat bath consisting of an ensemble of harmonic oscillators [15]. In [7] a more general Fokker-Planck operator has been proposed:

$$L_{FP}w = \frac{1}{\tau}\left(\mathrm{div}_k(kw) + \frac{k_B T_0 m}{\hbar^2}\Delta_k w + \frac{\hbar^2}{k_B T_0 m}\Delta_x w \right). \qquad (8.30)$$

We wish to derive the corresponding quantum hydrodynamic model in the case of constant temperature. For this we only need to compute the first moments of the last term in (8.30). It holds

$$\begin{aligned}
\int_{\mathbb{R}^d} \Delta_x w \, dk &= \Delta_x \langle 1 \rangle = \Delta_x n, \\
\int_{\mathbb{R}^d} \Delta_x w k_j \, dk &= \hbar \Delta_x \langle k_j \rangle = -\frac{m}{q}\Delta_x J_j, \quad j = 1, \ldots, d.
\end{aligned}$$

We interpret these terms as viscous terms and call the corresponding equations the *viscous quantum hydrodynamic model:*

$$\partial_t n - \frac{1}{q}\text{div}_x J = \frac{\hbar^2}{\tau k_B T_0 m}\Delta_x n,$$

$$\partial_t J - \frac{1}{q}\text{div}_x \left(\frac{J \otimes J}{n}\right) + \frac{qT}{m}\nabla_x n - \frac{\hbar^2 q}{6m^2}n\nabla_x\left(\frac{\Delta_x\sqrt{n}}{\sqrt{n}}\right) + \frac{q^2}{m}n\nabla_x V$$

$$= \frac{\hbar^2}{\tau k_B T_0 m}\Delta_x J - \frac{J}{\tau},$$

$$\text{div}(\varepsilon_s\nabla_x V) = q(n - C(x)), \quad x \in \Omega, \ t > 0.$$

This viscous terms have the important property that they *regularize* the equations in the following sense. Define the *total energy*

$$E(t) = \int_\Omega \left(\frac{\hbar^2 q}{6m^2}|\nabla\sqrt{n}|^2 + \frac{qk_B T_0}{m}n(\log n - 1) + \frac{\varepsilon_s}{2q}|\nabla V|^2 + \frac{|J|^2}{2n}\right)\,dx,$$

consisting of the quantum energy, entropy, electric energy, and kinetic energy. If $\Omega \subset \mathbb{R}^d$ is bounded and periodic boundary conditions are used or if $\Omega = \mathbb{R}^d$, a tedious (formal) computation shows that

$$\frac{dE}{dt} + \int_\Omega \left(\frac{\hbar^2 q}{3m^2}\frac{\hbar^2}{\tau k_B T_0 m}(\Delta\sqrt{n})^2 + \frac{4\hbar^2}{\tau k_B T_0 m}|\nabla\sqrt{n}|^2 + \frac{1}{\tau}\frac{|J|^2}{n}\right)\,dx$$

$$+ \int_\Omega \frac{\hbar^2}{\tau k_B T_0 m}\frac{1}{n^3}\left(\nabla n \cdot J - n\sum_{i=1}^d |\nabla J_i|\right)^2\,dx \leq 0. \qquad (8.31)$$

Without viscous terms the variation of the energy reads

$$\frac{dE}{dt} + \int_\Omega \left(\frac{\hbar^2 q}{\tau 3m^2}|\nabla\sqrt{n}|^2 + \frac{1}{\tau}\frac{|J|^2}{n}\right)\,dx = 0.$$

Hence, the viscous terms yield an additional term in the energy production, essentially given by $(\Delta\sqrt{n})^2$.

References

[1] R. Adler, A. Smith, and R. Longini. *Introduction to Semiconductor Physics*. John Wiley & Sons, New York, 1964.

[2] A. M. Anile and S. Pennisi. Thermodynamic derivation of the hydrodynamic model for charge transport in semiconductors. *Phys. Rev. B* 46 (1992), 13186-13193.

[3] A. M. Anile and V. Romano. Hydrodynamic modeling of charge carrier transport in semiconductors. *Meccanica* 35 (2000), 249-296.

[4] L. Arlotti and N. Bellomo. On the Cauchy problem for the Boltzmann equation. In N. Bellomo (ed.), *Lecture Notes on the Mathematical Theory of the Boltzmann Equation*, pages 1-64. World Scientific, Singapore, 1995.

[5] A. Arnold, J. A. Carrillo, and E. Dhamo. On the periodic Wigner-Poisson-Fokker-Planck system. To appear in *J. Math. Anal. Appl.*, 2002.

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

[7] A. Arnold, J.L. López, P. Markowich, and J. Soler. An analysis of quantum Fokker-Planck models: a Wigner function approach. Preprint, TU Berlin, Germany, 2000.

[8] N. Ashcroft and N. Mermin. *Solid State Physics*. Sanners College, Philadelphia, 1976.

[9] H. Babovsky. *Die Boltzmann-Gleichung*. Teubner, Stuttgart, 1998.

[10] G. Baccarani and M. Wordeman. An investigation on steady-state velocity overshoot in silicon. *Solid-State Electr.* 29 (1982), 970-977.

[11] N. Ben Abdallah and P. Degond. On a hierarchy of macroscopic models for semiconductors. *J. Math. Phys.* 37 (1996), 3308-3333.

[12] 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.

[13] K. Bløtekjær. Transport equations for electrons in two-valley semiconductors. *IEEE Trans. Electr. Devices* 17 (1970), 38-47.

[14] K. Brennan. *The Physics of Semiconductors*. Cambridge University Press, Cambridge, 1999.

[15] A. Caldeira and A. Leggett. Path integral approach to quantum Brownian motion. *Physica A* 121 (1983), 587-616.

[16] C. Cercignani. *The Boltzmann Equation and Its Applications*. Springer, Berlin, 1988.

[17] 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. Letters* 13 (1992), 26-28.

[18] P. Degond. Mathematical modelling of microelectronics semiconductor devices. Preprint, Université Paul Sabatier, Toulouse, France.

[19] P. Degond, S. Génieys and A. Jüngel. A system of parabolic equations in nonequilibrium thermodynamics including thermal and electrical effects. *J. Math. Pures Appl.* 76 (1997), 991-1015.

[20] P. Degond, S. Génieys and A. Jüngel. Symmetrization and entropy inequality for general diffusion equations. *C. R. Acad. Sci. Paris* 325 (1997), 963-968.

[21] P. Degond, A. Jüngel, and P. Pietra. Numerical discretization of energy-transprt model for semiconductors with non-parabolic band structure. *SIAM J. Sci. Comp.* 22 (2000), 986-1007.

[22] S. de Groot, P. Mazur. *Non-equilibrium Thermodynamics*, Dover, New York, 1984.

[23] R. DiPerna and P.L. Lions. On the Cauchy problem for the Boltzmann equation: global existence and weak stability. *Ann. Math.* 130 (1989), 321-366.

[24] C. Gardner. The quantum hydrodynamic model for semiconductor devices. *SIAM J. Appl. Math.* 54 (1994), 409-427.

[25] C. Gardner and C. Ringhofer. Smooth quantum potential for the hydrodynamic model. *Phys. Rev. E* 53 (1996), 157-167.

[26] A. Gnudi, D. Ventura, G. Baccarani, and F. Odeh. Two-dimensional MOS-FET simulations by means of a multidimensional spherical harmonic expansion of the Boltzmann transport equation. *Solic State Electr.* 36 (1993), 575-581.

[27] 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.

[28] H. Grahn. *Introduction to Semiconductor Physics.* World Scientific, Singapore, 1999.

[29] S. Holst, A. Jüngel and P. Pietra. A mixed finite-element discretization of the energy-transport equations for semiconductors. To appear in *SIAM J. Sci. Comp.*, 2002.

[30] R. Hudson. When is the Wigner quasiprobability nonnegative?, *Reports on Math. Phys.* 6 (1974), 249-252.

[31] E. Kane. Band structure of Indium-Antimonide. *J. Phys. Chem. Solids* 1 (1957), 249-261.

[32] S. Kawashima and Y. Shizuta. On the normal form of the symmetric hyperbolic-parabolic systems associated with the conservation laws. *Tohoku Math. J., II. Ser.* 40 (1988), 449-464.

[33] P. Markowich. *The Stationary Semiconductor Device Equations.* Springer, Vienna, 1986.

[34] P. Markowich, C. Ringhofer, and C. Schmeiser. *Semiconductor Equations.* Springer, Vienna, 1990.

[35] I. Müller and T. Ruggeri. *Rational Extended Thermodynamics.* Springer, Berlin, 1998.

[36] F. Poupaud. On a system of nonlinear Boltzmann equations of semiconductor physics. *SIAM J. Appl. Math.* 50 (1990), 1593-1606.

[37] F. Poupaud. Diffusion approximation of the linear semiconductor Boltzmann equaiton: analysis of boundary layers. *Asymp. Anal.* 4 (1991), 293-317.

[38] F. Poupaud and C. Ringhofer. Quantum hydrodynamic models in semiconductor crystals. *Appl. Math. Lett.* 8 (1995), 55-59.

[39] C. Ringhofer. Computational methods for semiclassical and quantum transport in semiconductor devices. *Acta Numerica* (1997), 485-521.

[40] C. Schmeiser and A. Zwirchmayr. Elastic and drift-diffusion limits of electron-phonon interaction in semiconductors. *Math. Models Meth. Appl. Sci.* 8 (1998), 37-53.

[41] K. Seeger. *Semiconductor Physics. An Introduction.* 6th edition, Springer, Berlin, 1997.

[42] M. Taylor. *Pseudodifferential Operators.* Princeton University Press, Princeton, 1981.

[43] E. Wigner. On the quantum correction for thermodynamic equilibrium. *Phys. Rev.* 40 (1932), 749-759.

[44] E. Zeidler. *Nonlinear Functional Analysis and its Applications.* Volume I. Springer, New York, 1986.