Chapter 19
Mathematical Model of Semiconductor Devices

19.1 Introduction

The chapter describes the reasoning that leads from the single-particle Schrödinger equation for an electron in a crystal to the mathematical model of semiconductor devices. The latter is a set of equations describing the evolution in space and time of a number of average quantities of interest: with reference to the electrons of the conduction band or holes of the valence band, such quantities are the concentration, average velocity, current density, average kinetic energy, and so on. The model of semiconductor devices has different levels of complexity depending on the trade-off between the information that one needs to acquire about the physical behavior of the device under investigation and the computational cost of the system of differential equations to be solved. In fact, the possible models are hierarchically ordered from the drift-diffusion model, which is the simplest one, to the hydrodynamic model, and so on. In essence, these models are different approaches to the problem of solving, in a more or less simplified form, the Boltzmann Transport Equation. Those described in this chapter are the most widely adopted in the commercial simulation programs used by semiconductor Companies. Other important methods, that are not addressed in this book, are the Monte Carlo method and the spherical-harmonics expansion.

The steps leading to the mathematical model of semiconductor devices start with a form of the single-particle Schrödinger equation based on the equivalent Hamiltonian operator, where it is assumed that the external potential energy is a small perturbation superimposed to the periodic potential energy of the nuclei; this leads to a description of the collisionless electron's dynamics in terms of canonically-conjugate variables, that are the expectation values of the wave packet's position and momentum. The dynamics of the Hamiltonian type makes it possible to introduce the statistical description of a many-electron system, leading to the semiclassical Boltzmann Transport Equation. After working out the collision operator, the perturbative approximation is considered; the simplified form of the transport equation thus found is tackled by means of the moments method, whence the hydrodynamic and drift-diffusion versions of the model are derived. A detailed analysis of the derivation of the electron and hole mobility in the parabolic-band approximation is provided. Then, the semiconductor model is coupled with Maxwell's equation, and

M. Rudan, *Physics of Semiconductor Devices,*
DOI 10.1007/978-1-4939-1151-6_19

the applicability of the quasi-static approximation is discussed. The typical boundary conditions used in the analysis of semiconductor devices are shown, and an example of analytical solution of the one-dimensional Poisson equation is given.

The complements discuss the analogy between the equivalent Hamiltonian operator and the corresponding Hamiltonian function introduced in an earlier chapter, provide a detailed description of the closure conditions of the models, and illustrate the Matthiessen's rule for the relaxation times. Finally, a short summary of the approximations leading to the derivation of the semiconductor model is given.

19.2 Equivalent Hamiltonian Operator

The separation procedure outlined in Sects. 16.2 through 16.5 has led to the single-electron Schrödinger Equation (17.40), namely, $[-\hbar^2/(2\,m)\,\nabla^2 + V]\,w = E\,w$, where the nuclei are kept fixed and the force acting onto the electron derives from a potential energy having the periodicity of the direct lattice: $V(\mathbf{r}+\mathbf{l}) = V(\mathbf{r})$, with $\mathbf{l}$ given by (17.1); as mentioned in Sects. 16.4, 16.5, the external forces are absent ($U_{\text{ext}} = 0$).

Thanks to the periodicity of the Hamiltonian operator of (17.40) it is possible to recast the time-independent Schrödinger equation into a different form as shown below. The procedure is based on the analogy with the Schrödinger equation for a free particle, $-\nabla^2\,w = k^2\,w$ with $k^2 = 2\,m\,E/\hbar^2$. One notes in fact that the left hand side is obtained by replacing $\mathbf{k}$ with $-\mathrm{i}\,\mathrm{grad}$ in the right hand side; this is just another form of the transformation of momentum into the momentum operator (Sect. 8.5), which in the present case yields the whole Hamiltonian operator because for a free particle the energy is purely kinetic. This type of transformation can be pursued also for the Schrödinger equation with a periodic potential energy by observing that the eigenvalues $E_i(\mathbf{k})$, for each branch i, are periodic within the reciprocal, scaled lattice, $E_i(\mathbf{k}+\mathbf{g}) = E_i(\mathbf{k})$ (Sect. 17.6), hence they can be expanded in terms of the direct-lattice translational vectors $\mathbf{l}$ (Sect. 17.3):

$$E_i(\mathbf{k}) = \sum_{\mathbf{l}} E_{i\mathbf{l}}\,\exp(\mathrm{i}\,\mathbf{l}\cdot\mathbf{k}) \qquad E_{i\mathbf{l}} = E_i(\mathbf{l}) = \frac{1}{\tau_g}\int_{\tau_g} E_i(\mathbf{k})\,\exp(-\mathrm{i}\,\mathbf{l}\cdot\mathbf{k})\,\mathrm{d}^3k. \tag{19.1}$$

The eigenvalue $E_i(\mathbf{k})$ is now transformed into an operator by letting $\mathbf{k} \leftarrow -\mathrm{i}\,\mathrm{grad}$; remembering (17.158), this yields

$$E_i(\mathbf{k}) \leftarrow E_i(-\mathrm{i}\,\mathrm{grad}) = \sum_{\mathbf{l}} E_{i\mathbf{l}}\,\exp(\mathbf{l}\cdot\mathrm{grad}) = \sum_{\mathbf{l}} E_{i\mathbf{l}}\,\mathcal{T}_{\mathbf{l}}. \tag{19.2}$$

The form of operator (19.2) is purely kinetic, as the space coordinates do not appear in it. The shape of the potential energy V whence $E_i(\mathbf{k})$ originates is embedded in the coefficients $E_{i\mathbf{l}}$ of expansion (19.1). To complete the procedure one must show that operator (19.2) has the same eigenvalues and eigenfunctions as the original operator

of $[-\hbar^2/(2\,m)\,\nabla^2 + V]\,w = E\,w$. Applying $E_i(-\,\mathrm{i}\,\mathrm{grad})$ to a Bloch function $w_{i\mathbf{k}}$ yields, using the first relation in (19.1) and the periodicity[1] of $\zeta_{i\mathbf{k}}$ (Sect. 17.6),

$$\sum_{\mathbf{l}} E_{i\mathbf{l}}\,\mathcal{T}_{\mathbf{l}}\,\zeta_{i\mathbf{k}}(\mathbf{r})\,\exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}) = \zeta_{i\mathbf{k}}(\mathbf{r})\sum_{\mathbf{l}} E_{i\mathbf{l}}\,\exp[\mathrm{i}\,\mathbf{k}\cdot(\mathbf{r}+\mathbf{l})] =$$

$$= \zeta_{i\mathbf{k}}(\mathbf{r})\,\exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{r})\sum_{\mathbf{l}} E_{i\mathbf{l}}\,\exp(\mathrm{i}\,\mathbf{k}\cdot\mathbf{l}) = w_{i\mathbf{k}}(\mathbf{r})\,E_i(\mathbf{k}). \tag{19.3}$$

The result

$$E_i(-\,\mathrm{i}\,\mathrm{grad})\,w_{i\mathbf{k}} = E_i(\mathbf{k})\,w_{i\mathbf{k}} \tag{19.4}$$

shows that, for each branch i, the purely kinetic operator $E_i(-\,\mathrm{i}\,\mathrm{grad})$ has the same eigenvalues and eigenfunctions as the Hamiltonian whence $E_i(\mathbf{k})$ originates. For this reason, operator $E_i(-\mathrm{i}\,\mathrm{grad})$ is called *equivalent Hamiltonian*. In summary, when the potential energy is periodic it is possible to directly reconstruct an equivalent operator by letting $\mathbf{k} \leftarrow -\mathrm{i}\,\mathrm{grad}$, in the same way as for a free particle. In this respect, the latter case may be viewed as a limiting condition of the former one, obtained by extending the period of the potential energy to infinity.[2] Another similarity between the free-particle case and that of a periodic potential energy is that the Hamiltonian operator is purely kinetic (albeit, in the latter case, at the cost of a more complicate form of the kinetic term).

19.2.1 Electron Dynamics

As illustrated in Sect. 17.6.1, the general solution of the time-dependent Schrödinger equation

$$\left[-\frac{\hbar^2}{2\,m}\,\nabla^2 + V(\mathbf{r})\right]\psi = \mathrm{i}\,\hbar\,\frac{\partial\psi}{\partial t}, \tag{19.5}$$

for a particle subjected to a periodic potential energy V, is (17.51), or its approximation using the wave packet of branch $i = n$,

$$\psi(\mathbf{r},t) \simeq \zeta_{n0}\,\exp(\mathrm{i}\,\Phi_{n0})\sum_{\mathbf{k}} c_{n\mathbf{k}}\,\exp[\mathrm{i}\,(\mathbf{r}-\mathbf{u}_n\,t)\cdot(\mathbf{k}-\mathbf{k}_0)]\,, \tag{19.6}$$

with $\Phi_{n0} = \mathbf{k}_0\cdot\mathbf{r} - \omega_{n0}\,t$.

---

[1] Like in Sect. 17.6.1, the periodic part of the Bloch function is indicated with $\zeta_{i\mathbf{k}}$ to avoid confusion with the group velocity.

[2] This method of reconstructing the operator from the eigenvalues was anticipated in Sect. 17.6.8.

Now, consider the case where an external,[3] non periodic potential energy $U(\mathbf{r})$ is added to the periodic one; the Hamiltonian operator becomes $-\hbar^2/(2\,m)\,\nabla^2+V+U$, yielding the eigenvalue equation

$$\left[-\frac{\hbar^2}{2\,m}\,\nabla^2 + V(\mathbf{r}) + U(\mathbf{r})\right] w'_{\mathbf{q}} = E'_{\mathbf{q}}\, w'_{\mathbf{q}}, \tag{19.7}$$

with $\mathbf{q}$ the label of the new eigenvalues. As the set $w'_{\mathbf{q}}$ is complete, a possible expansion of ψ is

$$\psi = \sum_{\mathbf{q}} c'_{\mathbf{q}}\, w'_{\mathbf{q}}\, \exp(-\mathrm{i}\, E'_{\mathbf{q}}\, t/\hbar). \tag{19.8}$$

However, expansion (19.8) is inconvenient because the Hamiltonian operator in (19.7) is not periodic; as a consequence, the properties of the eigenvalues and eigenfunctions typical of the periodic case are lost. A more suitable expansion[4] is found by using the eigenfunctions of the Hamiltonian operator corresponding to $U = 0$, namely, the Bloch functions; in this case the coefficients of the expansion depend on time:

$$\psi = \sum_{i\mathbf{k}} a_{i\mathbf{k}}(t)\, w_{i\mathbf{k}} = \sum_{i\mathbf{k}} c_{i\mathbf{k}}(t)\, w_{i\mathbf{k}}\, \exp(-\mathrm{i}\,\omega_{i\mathbf{k}}\, t), \tag{19.9}$$

where $c_{i\mathbf{k}} = a_{i\mathbf{k}}\, \exp(\mathrm{i}\,\omega_{i\mathbf{k}}\, t) \to$ const as $U \to 0$. This form is more convenient because it holds also in the case where the external potential energy depends on time, $U = U(\mathbf{r}, t)$. The approximate expression (19.6) of the wave function becomes

$$\psi(\mathbf{r}, t) \simeq \zeta_{n0}\, \exp(\mathrm{i}\,\Phi_{n0})\, A, \qquad A = \sum_{\mathbf{k}} c_{n\mathbf{k}}(t)\, \exp[\mathrm{i}\,(\mathbf{r} - \mathbf{u}_n\, t)\cdot(\mathbf{k} - \mathbf{k}_0)], \tag{19.10}$$

with $|\psi|^2 = |\zeta_{n0}|^2\, |A|^2$ and $\int_\Omega |\psi|^2\, \mathrm{d}^3 r = 1$. As ζ_{n0} is a rapidly-varying function of $\mathbf{r}$, the physical information about the dynamics of the wave packet is given by $|A|^2$.

So far, the only approximation in (19.10) is the use of a single branch n of the dispersion relation. On the other hand it must be observed that in the sum $V + U$ the first term has the periodicity of the lattice, namely, it varies rapidly in space, whereas the external potential energy is typically a slowly-varying function; in fact, it is due to the application of external generators and/or to the presence of a non-uniform distribution of charge within the material.[5] Also, the field associated to U is weak,

[3] For the sake of simplicity, suffix "ext" is dropped from the symbol of the external energy.

[4] The approach is the same as that used for treating the time-dependent perturbation theory (compare with 14.4).

[5] The field produced by non uniformities in the local charge density, which is present also in an equilibrium condition if the dopant distribution is not spatially constant (compare with Sect. 18.5), is classified as "external" because it can be treated as a perturbation. Instead, rapid variations of the physical properties of the material, like those that typically occur at interfaces, can not be treated using the perturbative method and require the solution of the Schrödinger equation without approximations.

so that it does not influence the form of V. This leads to the idea of treating U as a perturbation superimposed to the periodic Hamiltonian operator $-\hbar^2/(2\,m)\,\nabla^2 + V$. Using this approximation, the Hamiltonian operator of the perturbed problem is rewritten as

$$-\frac{\hbar^2}{2\,m}\,\nabla^2 + V(\mathbf{r}) + U(\mathbf{r},t) \simeq E_n(-\,\mathrm{i}\,\mathrm{grad}) + U(\mathbf{r},t), \tag{19.11}$$

where index n reminds one that the eigenfunctions of only the nth branch are used in the expansion. The approximation inherent in (19.11) consists in using the properties of the unperturbed problem in the perturbed case; in fact, the functional dependence of $E_n(-\,\mathrm{i}\,\mathrm{grad})$ on $-\mathrm{i}\,\mathrm{grad}$ derives from the unperturbed eigenvalues $E_n(\mathbf{k})$. Remembering that $E_n(-\,\mathrm{i}\,\mathrm{grad})$ is purely kinetic, (19.11) is similar to the Hamiltonian operator of a particle subjected only to the external potential U. The approximate form of the time-dependent Schrödinger equation then reads

$$\left[E_n(-\,\mathrm{i}\,\mathrm{grad}) + U\right]\,\psi = \mathrm{i}\,\hbar\frac{\partial\psi}{\partial t}. \tag{19.12}$$

19.2.2 Expectation Values—Crystal Momentum

The solution of (19.12) consists in determining the coefficients $c_{n\mathbf{k}}(t)$ of (19.10); this can be tackled by the method illustrated in Sect. 14.2, namely, by reducing (19.12) to a system of coupled differential equations in the unknowns $c_{n\mathbf{k}}$. More interesting it is to use (19.12) for calculating the expectation values of position and momentum; remembering that $\psi \simeq \exp(\mathrm{i}\,\Phi_{n0})\,\zeta_{n0}\,A$ is normalized to unity, one readily finds $\langle\mathbf{r}\rangle = \langle\psi|\mathbf{r}|\psi\rangle = \mathbf{r}_0$, where $\mathbf{r}_0$ denotes the center of the wave packet in the position space. As for momentum, it is $\langle\mathbf{p}\rangle = \langle\psi|-\mathrm{i}\,\hbar\mathrm{grad}|\psi\rangle$, namely, using $\Phi_{n0} = \mathbf{k}_0\cdot\mathbf{r} - \omega_{n0}\,t$,

$$\begin{aligned}\langle\mathbf{p}\rangle &= -\mathrm{i}\,\hbar\int_\Omega \psi^*\left[\mathrm{i}\,\mathbf{k}_0\,\psi + \exp(\mathrm{i}\,\Phi_{n0})\,\mathrm{grad}(\zeta_{n0}\,A)\right]\mathrm{d}^3r = \\ &= \hbar\,\mathbf{k}_0\int_\Omega |\psi|^2\,\mathrm{d}^3r - \mathrm{i}\,\hbar\int_\Omega (\zeta_{n0}A)^*\,\mathrm{grad}(\zeta_{n0}\,A)\,\mathrm{d}^3r.\end{aligned} \tag{19.13}$$

The first term at the right hand side of (19.13) yields $\hbar\,\mathbf{k}_0$ due to normalization. Letting $\zeta_{n0}\,A = a + \mathrm{i}\,b$, one finds in the second term $(\zeta_{n0}\,A)^*\,\mathrm{grad}(\zeta_{n0}\,A) = (1/2)\,\mathrm{grad}(a^2+b^2) + \mathrm{i}\,(a\,\mathrm{grad}b - b\,\mathrm{grad}a)$. The contribution of $\mathrm{grad}(a^2+b^2)$ to the integral is zero due to normalization; since A is slowly varying and normalizable, while ζ_{n0} oscillates rapidly, it follows

$$\langle\mathbf{p}\rangle = \hbar\,\mathbf{k}_0 + \hbar\int_\Omega (a\,\mathrm{grad}b - b\,\mathrm{grad}a)\;\mathrm{d}^3r \simeq \hbar\,\mathbf{k}_0, \tag{19.14}$$

with $\mathbf{k}_0$ the center of the wave packet in the $\mathbf{k}$ space. The product $\hbar\,\mathbf{k}_0$ is called *crystal momentum*. As for the time derivatives of $\langle\mathbf{r}\rangle$ and $\langle\mathbf{p}\rangle$ one finds, from (17.52),

$$\dot{\mathbf{r}}_0 = \mathbf{u}_n = \frac{1}{\hbar}\left(\mathrm{grad}_{\mathbf{k}} E_n\right)_{\mathbf{k}_0} = \frac{\mathrm{i}}{\hbar}\sum_{\mathbf{l}} \mathbf{l}\, E_{i\mathbf{l}}\, \exp\left(\mathrm{i}\,\mathbf{l}\cdot\mathbf{k}_0\right), \tag{19.15}$$

where the last expression derives from (19.1). For the time derivative of momentum one preliminarily observes that $E_n(-\mathrm{i}\,\mathrm{grad})$ commutes with the gradient operator; in fact,

$$E_n(-\mathrm{i}\,\mathrm{grad})\,\mathrm{grad}\psi = \sum_{\mathbf{l}} E_{n\mathbf{l}}\mathcal{T}_{\mathbf{l}}\,\mathrm{grad}\psi(\mathbf{r},t) = \sum_{\mathbf{l}} E_{n\mathbf{l}}\,\mathrm{grad}\psi(\mathbf{r}+\mathbf{l},t), \tag{19.16}$$

$$\mathrm{grad}\,E_n(-\mathrm{i}\,\mathrm{grad})\psi = \mathrm{grad}\sum_{\mathbf{l}} E_{n\mathbf{l}}\,\psi(\mathbf{r}+\mathbf{l},t) = \sum_{\mathbf{l}} E_{n\mathbf{l}}\,\mathrm{grad}\psi(\mathbf{r}+\mathbf{l},t). \tag{19.17}$$

Then, using definition (10.24) of the time derivative of an expectation value,[6] and remembering that the operator associated to $\mathbf{p}$ is $-\mathrm{i}\,\hbar\,\mathrm{grad}$, one finds

$$\hbar\,\dot{\mathbf{k}}_0 = \frac{\mathrm{d}\langle\mathbf{p}\rangle}{\mathrm{d}t} = \langle\psi|\left[E_n(-\mathrm{i}\,\mathrm{grad}) + U\right]\mathrm{grad} - \mathrm{grad}\left[E_n(-\mathrm{i}\,\mathrm{grad}) + U\right]|\psi\rangle. \tag{19.18}$$

Moreover it is $U\,\mathrm{grad}\psi - \mathrm{grad}(U\,\psi) = -\psi\,\mathrm{grad}U$, so that (19.18) eventually reduces to

$$\hbar\dot{\mathbf{k}}_0 = \frac{\mathrm{d}\langle\mathbf{p}\rangle}{\mathrm{d}t} = -\int_{\Omega} |\psi|^2\,\mathrm{grad}U\,\mathrm{d}^3r. \tag{19.19}$$

As U is slowly-varying in space, Ehrenfest approximation (10.33) applies, whence

$$\hbar\,\dot{\mathbf{k}}_0 = \frac{\mathrm{d}\langle\mathbf{p}\rangle}{\mathrm{d}t} \simeq -\left(\mathrm{grad}U\right)_{r_0}. \tag{19.20}$$

Introducing the function $H_n(\mathbf{r}_0,\mathbf{k}_0,t) = E_n(\mathbf{k}_0) + U(\mathbf{r}_0,t)$, one finds that (19.15) and (19.20) are equivalent, respectively, to

$$\dot{x}_{i\,0} = \frac{\partial H_n}{\partial(\hbar\,k_{i\,0})}, \qquad \hbar\,\dot{k}_{i\,0} = -\frac{\partial H_n}{\partial x_{i\,0}}, \qquad i = 1,2,3. \tag{19.21}$$

Relations (19.21) are of paramount importance in solid-state theory. They show in fact that within a periodic lattice the dynamics of the expectation values of a wave packet, subjected to an external potential energy that varies slowly in space, is described by Hamilton equations (compare with (1.42)), where $\mathbf{r}_0 = \langle\mathbf{r}\rangle$ and $\hbar\,\mathbf{k}_0 = \langle\mathbf{p}\rangle$ play the role of position and momentum, respectively. It follows that H_n is a Hamiltonian

[6] Definition (10.24) could be used also for deriving (19.15).

function proper. Another important observation is that the time variations of the wave packet's momentum are due to the external force only; as a consequence, if $U = \text{const}$ one has $\hbar\,\dot{\mathbf{k}}_0 = 0$, namely, the crystal momentum is a constant of motion.

A further insight into the structure of H_n is obtained by calculating the work exerted onto the wave packet by the external force $-\text{grad}_{\mathbf{r}_0} U = \hbar\,\dot{\mathbf{k}}_0$ during an elementary time $\mathrm{d}t$:

$$\mathrm{d}W = \hbar\,\dot{\mathbf{k}}_0 \cdot \mathrm{d}\mathbf{r}_0 = \hbar\,\dot{\mathbf{k}}_0 \cdot \mathbf{u}_n\,\mathrm{d}t = \hbar\,\mathbf{u}_n \cdot \mathrm{d}\mathbf{k}_0 = \left(\text{grad}_{\mathbf{k}} E_n\right)_{\mathbf{k}_0} \cdot \mathrm{d}\mathbf{k}_0 = \mathrm{d}E_n. \tag{19.22}$$

The work equals the variation of E_n; it follows that E_n, apart from an additive constant, is the kinetic energy of the wave packet. In turn, U is the potential energy which, as mentioned above, derives from the external force only. If the force acting on the electron is due to an electric field, then $U = -q\,\varphi$; this justifies the modified form (18.54) of the Fermi–Dirac statistics to be used when an electric field is present.[7] In the more general case where a magnetic field is also acting on the electron, $\hbar\,\delta\dot{\mathbf{k}}$ is given by the Lorentz force

$$\hbar\,\delta\dot{\mathbf{k}} = \mathbf{F} = -q\,(\mathbf{E} + \mathbf{u}_n \wedge \mathbf{B})\,, \tag{19.23}$$

and the Hamiltonian operator in (19.12) must be modified accordingly (compare with (9.19)).

It is important to remark again that the description of the wave packet's dynamics given in this section holds when the force is a weak perturbation with respect to the unperturbed situation. As a consequence, the description does not apply when the electron undergoes a collision; in fact, the force acting during a collision is strong and can not be treated as a perturbation.

19.2.3 Dynamics in the Parabolic-Band Approximation

When the wave packet is centered onto a wave vector $\mathbf{k}_0$ near the ath minimum of the conduction band, the diagonal expansion of $E_n(\mathbf{k})$ yields (17.57). Dropping the branch index n and letting $\mathbf{k} = \mathbf{k}_0$, $\delta k_i = k_{i0} - k_{ia}$ yields

$$E_e = E(\mathbf{k}_0) - E_C \simeq \frac{1}{2} \sum_{i=1}^{3} \frac{\hbar^2}{m_{ia}} (k_{i0} - k_{ia})^2 = \frac{1}{2}\,\hbar\,\delta\mathbf{k} \cdot \left(\hat{m}_a\right)^{-1} \hbar\,\delta\mathbf{k} \geq 0, \tag{19.24}$$

with $(\hat{m}_a)^{-1}$ given by (17.58). Expression (19.24) bears a strong similarity with the kinetic energy of the classical case. The same comment applies to the expression of group velocity given by (17.61), namely,

$$\mathbf{u} = \left(\hat{m}_a\right)^{-1} \hbar\,\delta\mathbf{k}. \tag{19.25}$$

[7] More comments about the analogy with the perturbation theory in the classical case are made in Sect. 19.6.1.

Replacing (19.25) into (19.24) yields $E_e = (1/2)\,\hat{m}_a \mathbf{u}\cdot\mathbf{u}$. When the expectation value $\hbar\,\mathbf{k}_0$ of momentum coincides with $\hbar\,\mathbf{k}_a$, corresponding to an absolute minimum E_C of the conduction band, it is $E_e = 0$. Such a value is also the minimum of the positive-definite quadratic form at the right hand side of (19.24). This shows that E_e is the kinetic energy of the electron, and allows one to identify E_C as the additive constant mentioned above.

In general, the relation between force and acceleration within a crystal is anisotropic. For the sake of simplicity consider the case of the parabolic-band approximation; the time derivative of (19.25) then yields

$$\dot{\mathbf{u}} = \left(\hat{m}_a\right)^{-1} \hbar\,\delta\dot{\mathbf{k}} = \left(\hat{m}_a\right)^{-1} \mathbf{F}. \tag{19.26}$$

If the entries of the mass tensor are different from each other, the acceleration is not parallel to the force; the physical reason for this is easily understood if one thinks that the forces due to the crystal structure are embedded in the mass tensor through the second derivatives of $E(\mathbf{k})$. The mass tensor becomes a scalar only if the branch E is isotropic: $\hat{m}_a = m_a\,\mathcal{I}$, with $\mathcal{I}$ the identity tensor. More comments about this issue are made in Sect. 19.6.2.

The analysis for the valence band is similar. Again, the branch index n is dropped and symbols $\mathbf{k} = \mathbf{k}_0$, $\delta k_i = k_{i0} - k_{ia}$ are used,[8] to find (17.64), namely,

$$E_h = E_V - E(\mathbf{k}_0) \simeq \frac{1}{2}\sum_{i=1}^{3} \frac{\hbar^2}{m_{ia}}\,(k_{i0} - k_{ia})^2 = \frac{1}{2}\,\hbar\,\delta\mathbf{k}\cdot(\hat{m}_a)^{-1}\,\hbar\,\delta\mathbf{k} \geq 0, \tag{19.27}$$

with $(\hat{m}_a)^{-1}$ given by (17.63), $m_{ia} > 0$. For the group velocity one finds

$$\mathbf{u} = \frac{1}{\hbar}\left(\mathrm{grad}_{\mathbf{k}} E\right)_{\mathbf{k}_0} = -\left(\hat{m}_a\right)^{-1} \hbar\,\delta\mathbf{k}. \tag{19.28}$$

The work exerted onto the wave packet by the external force $-\mathrm{grad}_{\mathbf{r}_0} U = \hbar\,\dot{\mathbf{k}}_0$ during an elementary time $\mathrm{d}t$ is

$$\mathrm{d}W = \hbar\,\dot{\mathbf{k}}_0 \cdot \mathbf{u}\,\mathrm{d}t = \mathrm{d}\left[-\sum_{i=1}^{3} \frac{\hbar^2}{2\,m_{ia}}\,\delta k_i^2\right] = \mathrm{d}E, \tag{19.29}$$

which, again, shows that E is the kinetic energy of the electron apart from an additive constant. The negative signs in (19.27) and (19.28) make the discussion of the valence-band case somewhat awkward; however, the difficulty is readily eliminated if one refers to holes instead of electrons. For example, consider the case of an electron whose expectation value of momentum, initially equal to $\hbar\,\mathbf{k}_a$, is brought by the action of an external field to some other value $\hbar\,\mathbf{k}_0'$ in the vicinity of $\hbar\,\mathbf{k}_a$. For

[8] For Si, Ge, and GaAs it is $k_{ia} = 0$ (Sect. 17.6.5).

this transition to occur it is implied that the initial state $\mathbf{k}_a$ is occupied[9] and the final state $\mathbf{k}_0'$ is empty. As a consequence of (19.27), E changes from E_V to $E(\mathbf{k}_0') < E_V$, namely, it decreases during the time interval Δt during which the energy variation occurs; hence, the external field has exerted in Δt a negative work onto the electron, in fact, energy has been absorbed from the electron by the field. If a hole is considered instead, the initial and final states of the transition exchange roles; however, from the standpoint of the energy balance nothing changes, namely, the field still absorbs energy from the particle. It follows that the hole's energy must decrease due to the transition: this is possible only if the energy axis associated to the hole is reversed with respect to that of the electron, so that, apart from an additive constant, the hole's kinetic energy is $-E$. From this point on, the reasoning becomes identical to that outlined above for the electron of the valence band: using (19.27), when the expectation value $\hbar\,\mathbf{k}_0$ of momentum coincides with $\hbar\,\mathbf{k}_a$, corresponding to an absolute maximum E_V of the valence band, it is $E_h = 0$. Such a value is also the minimum of the positive-definite quadratic form at the right hand side of (19.27). This shows that E_h is the kinetic energy of the hole, and allows one to identify E_V as the additive constant.

19.3 Dynamics in the Phase Space

The theory outlined in Sects. 19.2, 19.2.1, and 19.2.2 has led to the conclusion that the dynamics of the expectation values of a wave packet describing an electron's motion, subjected to an external potential energy that varies slowly in space, is described by the Hamilton equations (19.21) where $\langle\mathbf{r}\rangle$ and $\langle\mathbf{p}\rangle$ play the role of position and momentum.

For a system made of a large number of electrons, the description of the dynamics of the individual wave packets is impossible from the practical standpoint. In this case one resorts to the same device as that used in Sect. 6.2 for a system of classical particles, namely, the distribution function. Being the formal apparatus identical to that of Sect. 6.2, only the relevant differences will be remarked. The μ-type phase space is defined here by the variables

$$\mathbf{s} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ k_1 \\ k_2 \\ k_3 \end{bmatrix}, \qquad \mathbf{e} = \begin{bmatrix} \partial H/\partial k_1 \\ \partial H/\partial k_2 \\ \partial H/\partial k_3 \\ -\partial H/\partial x_1 \\ -\partial H/\partial x_2 \\ -\partial H/\partial x_3 \end{bmatrix}, \tag{19.30}$$

[9] For the sake of simplicity, spin is not considered here.

(compare with (1.57)) so that the distribution function[10] reads $f = f(\mathbf{r}, \mathbf{k}, t)$. Note that the units of f are different from those of the classical distribution function. For the latter, in fact, it is $[f_\mu] = (\mathrm{J\ s})^{-3}$, so that the product $f_\mu\, \mathrm{d}^3 r\, \mathrm{d}^3\, p$ is dimensionless (compare with (6.1)); in the present case, instead, both $f\, \mathrm{d}^3 r\, \mathrm{d}^3 k$ and $\mathrm{d}^3 r\, \mathrm{d}^3 k$ are dimensionless, hence the distribution function itself is dimensionless.

The system considered for the investigation is that of the electrons belonging to the conduction band. Remembering the first relation in (6.3), the concentration and average velocity of such electrons are given by

$$n(\mathbf{r}, t) = \iiint_{-\infty}^{+\infty} f(\mathbf{r}, \mathbf{k}, t)\, \mathrm{d}^3 k, \qquad \mathbf{v}(\mathbf{r}, t) = \frac{1}{n} \iiint_{-\infty}^{+\infty} \mathbf{u}(\mathbf{k})\, f(\mathbf{r}, \mathbf{k}, t)\, \mathrm{d}^3 k, \tag{19.31}$$

with $\mathbf{u}$ the electron's group velocity. In the equilibrium condition it is $f^{\mathrm{eq}} = Q\, P$, where the Fermi–Dirac statistics P depends on $\mathbf{k}$ only through $E(\mathbf{k})$, namely, it is even with respect to $\mathbf{k}$. In turn, $\mathbf{u} = (1/\hbar)\, \mathrm{grad}_{\mathbf{k}} E$ is odd, so that the whole integrand in the second definition of (19.31) is odd. As the integration domain is symmetric with respect to $\mathbf{k} = 0$, it is $\mathbf{v}^{\mathrm{eq}} = 0$ as should be. In a non-equilibrium condition it is $f = Q\, \Phi$, with $\Phi(\mathbf{r}, \mathbf{k})$ the occupation probability of a state. If the band is completely filled, then $\Phi = 1$, and the electron flux $n\, \mathbf{v}$ becomes proportional to the integral of $\mathbf{u}$; as the latter is odd, the flux vanishes: this explains why a completely filled band does not contribute to the material's conduction, as anticipated in Sect. 17.6.5.

The Boltzmann collisionless equation in the $\mathbf{r}, \mathbf{k}$ space is derived in the same manner as for (6.28); it reads

$$\frac{\partial f}{\partial t} + \dot{\mathbf{r}} \cdot \mathrm{grad}_{\mathbf{r}} f + \dot{\mathbf{k}} \cdot \mathrm{grad}_{\mathbf{k}} f = 0. \tag{19.32}$$

The effects of collisions may be grouped into two classes: the collisions of the first class induce transitions that change the number of electrons of the band. Such transitions are the generations and recombinations introduced in Sect. 17.6.6, where the initial state of the electron belongs to the conduction band and the final one belongs to the valence band, or vice versa.[11] The transitions of this class are collectively called *inter-band transitions*.

The collisions of the second class are those where the initial and final state belong to the same band, and are called *intra-band transitions*; they do not change the number of electrons of the band. The distinction between the two classes is useful because the inter-band transitions exhibit characteristic times that are much larger than those of the intra-band transitions. In turn, the intra-band transitions are further divided into two subclasses: the *intra-valley transitions*, where the initial and final

[10] Suffix μ is dropped to distinguish this distribution function from that of the classical case.

[11] In addition to this one must also consider the trapping-detrapping phenomena involving localized states. So far, only the localized states due to dopants have been considered (Sect. 18.4); other types of localized states are introduced in Chap. 20.

states are in the vicinity of the same extremum of the band, and the *inter-valley transitions*, where the initial and final state are in the vicinity of different extrema.[12]

Within each class, the transitions are further grouped depending on the entity with which the collision occurs; typical examples of collisions are those with phonons, impurities, defects, and photons. Like in the classical case, collisions are not accounted for in the derivation of (19.32), where the effect of only the slowly-varying external potential energy is present; the further time change of f due to collisions is more conveniently kept separate from that of the external potential energy. Also, it is assumed that the system under consideration is dilute, so that each wave packet spends a relatively large fraction of time without suffering any collision; in other terms, the time during which an electron is subjected to the external field is much longer than that involved in a collision. For this reason it is preferable to write the Boltzmann equation, when the collisions are accounted for, as

$$\frac{\partial f}{\partial t} + \mathbf{u} \cdot \mathrm{grad}_{\mathbf{r}} f - \frac{q}{\hbar} \left(\mathbf{E} + \mathbf{u} \wedge \mathbf{B}\right) \cdot \mathrm{grad}_{\mathbf{k}} f = C \tag{19.33}$$

(compare with (6.29) and (6.31)). To derive (19.33), the expression (19.23) of the Lorentz force acting on the electron is used, after dropping index n from the group velocity. Term C embeds the forces acting during the collisions; such forces are short ranged and much more intense than those due to the external field; as a consequence, the Ehrenfest approximation (10.33) does not apply, so that a full quantum-mechanical approach is necessary to treat the collision term.

In conclusion, the relations involving the expectation values at the left hand side of (19.33) are formally identical to those of the classical case, whereas the right hand side is calculated by quantum methods. The form (19.33) of the Boltzmann Transport Equation (BTE) is also called *semiclassical*.

19.3.1 Collision Term

The left hand side of (19.33) equals $\mathrm{d}f/\mathrm{d}t$, whence the right hand side is the rate of change of f due to collisions. As $f(\mathbf{r}, \mathbf{k}, t)\, \mathrm{d}^3 r\ \mathrm{d}^3 k$ is the number of electrons of the conduction band that at time t belong to the elementary volume $\mathrm{d}^3 r\ \mathrm{d}^3 k$ centered on $(\mathbf{r}, \mathbf{k})$, the rate of change can be expressed as

$$C = C_{\mathrm{in}} - C_{\mathrm{out}}, \tag{19.34}$$

where $C_{\mathrm{in}}\, \mathrm{d}^3 r\ \mathrm{d}^3 k$ is the number of electrons entering $\mathrm{d}^3 r\ \mathrm{d}^3 k$ per unit time, due to collisions, and $C_{\mathrm{out}}\, \mathrm{d}^3 r\ \mathrm{d}^3 k$ is the number of electrons leaving $\mathrm{d}^3 r\ \mathrm{d}^3 k$ per unit time, due to collisions. To illustrate the reasoning it is convenient to refer to Fig. 19.1,

[12] Here the extrema are the minima of the conduction band. The region near an extremum of a band is also called *valley*.

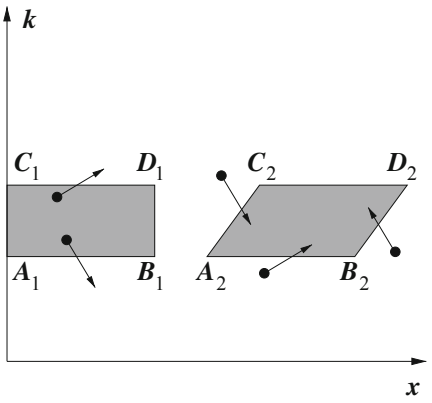

Fig. 19.1 Example of the time evolution of a phase-space domain in a one-dimensional case. The situation with no external force is considered

where a one-dimensional case is illustrated using the x, k coordinates. Instead of an elementary volume $\mathrm{d}x\,\mathrm{d}k$, a finite, rectangular cell is considered, whose position at time t_1 is fixed by the vertices A_1, B_1, C_1, and D_1. For simplicity it is assumed that no external force is present ($U = \text{const}$), so that the crystal momentum is conserved. In particular, the vertices' momenta at $t = t_1$ are $\hbar\, k_m = \hbar\, k(A_1) = \hbar\, k(B_1)$ and $\hbar\, k_M = \hbar\, k(C_1) = \hbar\, k(D_1)$. The corresponding positions are $x(A_1) = x(C_1) = 0$, $x_0 = x(B_1) = x(D_1)$. Letting m^* indicate the effective mass, the position of the vertices at a subsequent time $t_2 = t_1 + \Delta t$ is

$$x(A_2) = \frac{\hbar k_m}{m^*}\,\Delta t,\;\; x(B_2) = x_0 + x(A_2),\;\; x(C_2) = \frac{\hbar k_M}{m^*}\,\Delta t,\;\; x(D_2) = x_0 + x(C_2),$$

this giving rise to the parallelogram also shown in Fig. 19.1. If no collisions occur, the electrons inside the parallelogram at $t = t_2$ are the same as those that were inside the rectangle at $t = t_1$; in contrast, when collisions occur, some electrons leave the rectangle without reaching the parallelogram (hence, $C_{\text{out}} \neq 0$), while the parallelogram is reached by other electrons that originally did not belong to the rectangle ($C_{\text{in}} \neq 0$). This is schematically indicated by the arrows in Fig. 19.1. In general it is $C_{\text{out}} \neq C_{\text{in}}$, so that $\mathrm{d}f/\mathrm{d}t \neq 0$. The description is the same also in the case when the external force is present, the difference being that the trajectories in the phase space are not rectilinear and the deformation of the domain is more complicate.

To give the analysis a more formal aspect it is necessary to determine how the population of the elementary domain $\mathrm{d}^6 s = \mathrm{d}^3 r\,\mathrm{d}^3 k$ evolves in the elementary time interval $\mathrm{d}t$. To begin, one introduces the *scattering probability per unit time and unit phase volume*, S, from an initial state to a final state of the phase space.[13] The initial

[13] The units of S are $[S] = \mathrm{s}^{-1}$.

(final) state is indicated by the first (second) pair of arguments of S, namely,

$$S\left(\mathbf{r},\mathbf{k}\rightarrow\mathbf{r}',\mathbf{k}'\right)\,\mathrm{d}^3r'\,\mathrm{d}^3k' \tag{19.35}$$

is the probability per unit time that an electron scatters from $(\mathbf{r},\mathbf{k})$ to the elementary volume $\mathrm{d}^3r'\,\mathrm{d}^3k'$ centered at $(\mathbf{r}',\mathbf{k}')$. Then, let $\mathrm{d}N_{\mathrm{in}}=C_{\mathrm{in}}\,\mathrm{d}^6s$, $\mathrm{d}N_{\mathrm{out}}=C_{\mathrm{out}}\,\mathrm{d}^6s$, and $\mathbf{s}=(\mathbf{r},\mathbf{k})$, $\mathbf{s}'=(\mathbf{r}',\mathbf{k}')$. The number $\mathrm{d}N_{\mathrm{in}}$ is determined by observing that the electrons contributing to it are those that initially belong to elementary phase-space volumes, say, d^6s', different from d^6s. The population of d^6s' at time t is $f(\mathbf{s}',t)\,\mathrm{d}^6s'$; if the latter is multiplied by the scattering probability per unit time from $\mathbf{s}'$ to d^6s, given by $S\left(\mathbf{s}'\rightarrow\mathbf{s}\right)\,\mathrm{d}^6s$, the unconditional number of transitions from d^6s' to d^6s is obtained. The actual number of such transitions is then found by remembering that electrons are fermions, so that transitions towards d^6s are possible only if the final states are empty; in other terms, the unconditional number of $\mathbf{s}'\rightarrow\mathbf{s}$ transitions must be multiplied by $1-\Phi(\mathbf{s},t)$, where $\Phi(\mathbf{s},t)$ is the probability that the final state is full. Finally, the contributions of all elementary volumes d^6s' must be added up, to find[14]

$$\mathrm{d}N_{\mathrm{in}}=\int_{\mathbf{s}'}\left[f(\mathbf{s}',t)\,\mathrm{d}^6s'\right]\,\left[S\left(\mathbf{s}'\rightarrow\mathbf{s}\right)\,\mathrm{d}^6s\right]\,\left[1-\Phi(\mathbf{s},t)\right]. \tag{19.36}$$

The derivation of $\mathrm{d}N_{\mathrm{out}}$ is similar; one obtains

$$\mathrm{d}N_{\mathrm{out}}=\int_{\mathbf{s}'}\left[f(\mathbf{s},t)\,\mathrm{d}^6s\right]\,\left[S\left(\mathbf{s}\rightarrow\mathbf{s}'\right)\,\mathrm{d}^6s'\right]\,\left[1-\Phi(\mathbf{s}',t)\right]. \tag{19.37}$$

The collision term $C=C_{\mathrm{in}}-C_{\mathrm{out}}$ is now determined by subtracting (19.37) from (19.36) and dividing the result by d^6s. This shows that C is the sum of two terms; the first one is linear with respect to f and reads

$$\int_{\mathbf{s}'}\left[f(\mathbf{s}',t)\,S\left(\mathbf{s}'\rightarrow\mathbf{s}\right)-f(\mathbf{s},t)\,S\left(\mathbf{s}\rightarrow\mathbf{s}'\right)\right]\,\mathrm{d}^6s'. \tag{19.38}$$

As for the second term, one must preliminarily observe that $f=Q\,\Phi$, with $Q=1/(4\,\pi^3)$ the density of states in the phase space, (17.50); then, the second term of C turns out to be quadratic with respect to Φ or f:

$$Q\int_{\mathbf{s}'}\Phi(\mathbf{s},t)\,\Phi(\mathbf{s}',t)\,\left[S\left(\mathbf{s}\rightarrow\mathbf{s}'\right)-S\left(\mathbf{s}'\rightarrow\mathbf{s}\right)\right]\,\mathrm{d}^6s'. \tag{19.39}$$

[14] For the sake of conciseness, in Sects. 19.3.1, 19.3.2, and 19.3.3 the six-fold integrals over $\mathrm{d}^3r'\,\mathrm{d}^3k'$ and the three-fold integrals over d^3k' are indicated with $\int_{\mathbf{s}'}$ and $\int_{\mathbf{k}'}$, respectively.

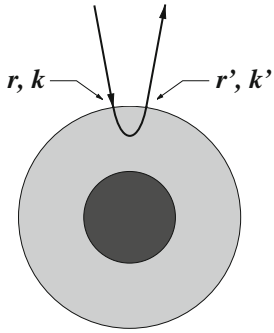

Fig. 19.2 Qualitative picture of a collision between an electron and a negatively-ionized impurity. The latter is schematically represented by the black circle, whereas the gray area indicates the screening region. The initial and final state of the electron are indicated with $(\mathbf{r}, \mathbf{k})$ and $(\mathbf{r}', \mathbf{k}')$, respectively

19.3.2 Point-Like Collisions

The two summands (19.38), (19.39) in the expression of C are substantially simplified thanks to the property that the collisional forces, albeit very strong, are short ranged; as a consequence, whereas the momentum of the colliding electron may undergo a large change due to the collision, the electron's position changes little. The issue is illustrated with the aid of Fig. 19.2, that schematically describes an electron collision with a negatively-ionized dopant. The latter is represented by the black circle, whereas the gray region around it indicates the positive charge attracted by the negative ion; such a positive charge acts like an *electric screen* that tends to neutralize the ion. As a consequence of the screen, the decay of the electrostatic potential acting between the ion and the incoming electron, when the relative distance increases, is much stronger than in the pure Coulomb case.[15] In practice, one can assume that the electron-ion repulsion is non negligible only when the electron is inside the screen. This makes the dynamics of the interaction rather different from that of the pure Coulomb case, treated in Sect. 3.8. As shown in the figure, the final momentum $\mathbf{k}'$ may differ largely from the initial one, $\mathbf{k}$; in contrast, considering the atomic scale of the phenomenon, the final position $\mathbf{r}'$ may be thought of as coinciding with the initial one, $\mathbf{r}$. To vest this observation with mathematical form, considering that the scattering probability in (19.38) and (19.39) undergoes an integration over $\mathbf{r}'$, one lets[16]

$$S\left(\mathbf{s} \to \mathbf{s}'\right) = S_0\left(\mathbf{r}, \mathbf{k} \to \mathbf{k}'\right) \delta(\mathbf{r}' - \mathbf{r}). \tag{19.40}$$

Another important consequence of the above discussion is that, although the duration of the interaction is very short, the force acting on the electron due the interaction is much stronger than the external forces, to the extent that the effects of

[15] The derivation and treatment of the screened Coulomb interaction are carried out in Sects. 20.6.4 and 14.7, respectively.

[16] Note that the units of S_0 are different from those of S: in fact, $[S_0] = \mathrm{cm}^3/\mathrm{s}$. Examples of calculations of phonon scattering and ionized-impurity scattering are given in Sects. 20.5.1 and 20.5.2, respectively.

the latter can be neglected during the interaction itself. It follows that S and S_0 do not depend on the external forces; this greatly simplifies the analysis of S_0. Inserting (19.40) into (19.38) yields, for the first part of C,

$$\int_{\mathbf{k}'} \left[f(\mathbf{r},\mathbf{k}',t)\, S_0\left(\mathbf{r},\mathbf{k}' \to \mathbf{k}\right) - f(\mathbf{r},\mathbf{k},t)\, S_0\left(\mathbf{r},\mathbf{k} \to \mathbf{k}'\right) \right] \mathrm{d}^3k' \,; \tag{19.41}$$

in turn, the second part (19.39) becomes

$$Q \int_{\mathbf{k}'} \Phi(\mathbf{r},\mathbf{k},t)\, \Phi(\mathbf{r},\mathbf{k}',t) \left[S_0\left(\mathbf{r},\mathbf{k} \to \mathbf{k}'\right) - S_0\left(\mathbf{r},\mathbf{k}' \to \mathbf{k}\right) \right] \mathrm{d}^3k'. \tag{19.42}$$

The term S_0 is typically calculated using the first-order perturbation theory (Sect. 14.3), which shows that the transition probability is invariant upon reversal of the initial and final states. It follows that the quantity in brackets in (19.42) vanishes, so that (19.41) is in fact the only contribution to C. The latter is recast in a more compact form by defining the *relaxation time* $\tau(\mathbf{r},\mathbf{k})$ and the *collision operator* $\tilde{f}(\mathbf{r},\mathbf{k})$ such that

$$\frac{1}{\tau} = \int_{\mathbf{k}'} S_0\left(\mathbf{r},\mathbf{k} \to \mathbf{k}'\right) \mathrm{d}^3k', \qquad \tilde{f} = \frac{\int_{\mathbf{k}'} f(\mathbf{r},\mathbf{k}',t)\, S_0\left(\mathbf{r},\mathbf{k}' \to \mathbf{k}\right) \mathrm{d}^3k'}{\int_{\mathbf{k}'} S_0\left(\mathbf{r},\mathbf{k} \to \mathbf{k}'\right) \mathrm{d}^3k'}, \tag{19.43}$$

to find $C = C_{\mathrm{in}} - C_{\mathrm{out}} = (\tilde{f} - f)/\tau$. The BTE (19.33) thus becomes

$$\frac{\partial f}{\partial t} + \mathbf{u} \cdot \mathrm{grad}_{\mathbf{r}} f - \frac{q}{\hbar} \left(\mathbf{E} + \mathbf{u} \wedge \mathbf{B}\right) \cdot \mathrm{grad}_{\mathbf{k}} f = -\frac{f - \tilde{f}}{\tau}. \tag{19.44}$$

In the derivation of (19.44) no distinction is made between the inter-band and intra-band transitions. For a dilute system one can assume that the transitions of the two types are uncorrelated, so that the corresponding probabilities are additive: $S_0 = S_{0b} + S_{0v}$, where index b (v) stands for "inter-band" ("intra-band"). As a consequence,

$$\frac{1}{\tau} = \frac{1}{\tau_b} + \frac{1}{\tau_v}, \qquad \frac{1}{\tau_b} = \int_{\mathbf{k}'} S_{0b}\, \mathrm{d}^3k', \qquad \frac{1}{\tau_v} = \int_{\mathbf{k}'} S_{0v}\, \mathrm{d}^3k'. \tag{19.45}$$

For the semiconductors of interest, the relaxation times defined in (19.45) differ by several orders of magnitude (e.g., in electronic-grade silicon[17] is $\tau_b > 10^{-6}$ s, $\tau_v < 10^{-12}$ s). This makes the intra-band transitions dominant ($\tau = \tau_b\, \tau_v/(\tau_b + \tau_v) \simeq \tau_v$); one exception exists though, where the effects of the intra-band transitions cancel each other exactly, so that the inter-band transitions only are left. Such an exception is discussed in Sect. 19.4.

[17] The semiconductor's purification degree necessary for manufacturing integrated circuit is called *electronic grade*; it indicates that the ratio between the concentration of impurities (different from dopants) and that of the semiconductor atoms is smaller than 10^{-9}. Lower-quality materials, with a ratio smaller than 10^{-6}, are used in the fabrication of solar cells; in this case the purification degree is called *solar grade*.

19.3.3 Perturbative Form of the BTE

The Boltzmann Transport equation (19.44) is an integral-differential equation in the phase space and time, in the unknown f. The kernel of the integral part is S_0, while the equation coefficients are $\mathbf{E}(\mathbf{r}, t)$, $\mathbf{B}(\mathbf{r}, t)$, $\tau(\mathbf{r}, \mathbf{k})$, and $\mathbf{u}(\mathbf{k}) = (1/\hbar)\,\mathrm{grad}_{\mathbf{k}} H$. In equilibrium f becomes $f^{\mathrm{eq}} = Q\,P$, with $P(\mathbf{r}, \mathbf{k})$ the Fermi–Dirac statistics; as shown in Sect. 19.2.2, P depends on position if the semiconductor is not uniform. Moreover it is $\mathrm{d}f^{\mathrm{eq}}/\mathrm{d}t = 0$; hence, to make the collision term to vanish at equilibrium, it must be $\tilde{f}^{\mathrm{eq}} = f^{\mathrm{eq}}$ (detailed-balance principle, Sect. 6.5).

In general, the solution of the BTE is quite a difficult task. The issue of effective solution methods for this equation will be introduced later; however, a solution procedure is outlined here which, although seldom used in practical cases, has the advantage of providing a simplified form of (19.44), upon which a number of models for semiconductor-device analysis are based. The procedure consists in setting up the iterative scheme

$$\frac{\mathrm{d}f^{(m+1)}}{\mathrm{d}t} = -\frac{f^{(m+1)} - \tilde{f}^{(m)}}{\tau}, \qquad \tilde{f}^{(m)} = \frac{\int_{\mathbf{k}'} f^{(m)}\, S_0\, \mathrm{d}^3k'}{\int_{\mathbf{k}'} S_0\, \mathrm{d}^3k'}. \tag{19.46}$$

with m the iteration index. In this way the $\tilde{f}^{(m)}$ term at the right hand side of the first equation in (19.46) is known from the previous iteration, so that the integral-differential equation is transformed into a number of differential-only equations. If convergence occurs, the iterations are brought to an end when a suitable norm $||f^{(m+1)} - f^{(m)}||$ is smaller than a prescribed value.

To start the procedure it is reasonable to choose, for the approximation of order zero, the equilibrium distribution: $f^{(0)} = f^{\mathrm{eq}}$; from the detailed-balance principle it follows $\tilde{f}^{(0)} = f^{\mathrm{eq}}$. The first step of the iteration procedure then yields $f^{(1)}$, called *first-perturbation solution*. In many cases of practical interest, the material or device under investigation is sufficiently close to the equilibrium condition to make $f^{(1)}$ an acceptable solution; one then stops the iterations at the first-perturbation solution and takes $f \simeq f^{(1)}$. This is equivalent to solving the *perturbative form of the BTE*

$$\frac{\partial f}{\partial t} + \mathbf{u} \cdot \mathrm{grad}_{\mathbf{r}} f - \frac{q}{\hbar}\,(\mathbf{E} + \mathbf{u} \wedge \mathbf{B}) \cdot \mathrm{grad}_{\mathbf{k}} f = -\frac{f - f^{\mathrm{eq}}}{\tau}. \tag{19.47}$$

It is interesting to comment on the form of (19.47). The first term at the left hand side differs from zero only if the distribution function depends explicitly on time; hence, it vanishes in a non-equilibrium condition if the latter is of the steady-state type. The second term vanishes if the distribution function is independent of the spatial coordinates, hence it describes a contribution to electron transport that originates from a spatial non uniformity. For this reason this term is called *diffusive term* (compare with Sect. 23.3). The third term vanishes if the external force is missing, hence it originates from the action of the external forces on the electrons and, for this reason, is called *drift term*. At the right hand side of (19.47), the magnitude of the relaxation time influences the amount by which the distribution

function departs from equilibrium; to better show this, one recasts (19.47) as

$$f = f^{\mathrm{eq}} - \tau\,\mathcal{L}f, \qquad \mathcal{L} = \frac{\partial}{\partial t} + \mathbf{u}\cdot\mathrm{grad}_{\mathbf{r}} - \frac{q}{\hbar}\,(\mathbf{E}+\mathbf{u}\wedge\mathbf{B})\cdot\mathrm{grad}_{\mathbf{k}}, \tag{19.48}$$

with $\mathcal{L}$ the *Liouvillian operator*. If $\tau \to 0$, then $f \to f^{\mathrm{eq}}$; this shows that the perturbative solution is in fact acceptable if the relaxation time is sufficiently small.[18]

A final comment refers to the spatially-uniform case, where $f = f(\mathbf{k}, t)$, $\tau = \tau(\mathbf{k})$; then, (19.47) simplifies to

$$\frac{\partial f}{\partial t} - \frac{q}{\hbar}\,(\mathbf{E}+\mathbf{u}\wedge\mathbf{B})\cdot\mathrm{grad}_{\mathbf{k}} f = -\frac{f - f^{\mathrm{eq}}}{\tau}. \tag{19.49}$$

If the fields are set to zero at some instant of time, say, $t = 0$, (19.49) reduces to a differential equation with respect to time only, with $\mathbf{k}$ a parameter; thus, for each $\mathbf{k}$ the solution approaches the equilibrium distribution[19] according to the law

$$f = f^{\mathrm{eq}} + (f_{t=0} - f^{\mathrm{eq}})\,\exp(-t/\tau). \tag{19.50}$$

In passing, this result explains why τ is called "relaxation time".

19.4 Moments Expansion of the BTE

The *moments expansion of the BTE* is a transformation method that has successfully been applied to the analysis of semiconductor devices. It reduces the original equation to a set of partial-differential equations in the $\mathbf{r}, t$ space; the number of such equations can be adapted to the type of information to be acquired about the device under investigation. More specifically, applying the moments method to the BTE and truncating the series of moments at a suitable order, one extracts from the BTE a hierarchically-ordered set of models, ranging from the simplest to the more complicate ones ([90–92], and references therein).

The BTE is an equation in the $\mathbf{r}, \mathbf{k}, t$ space. The basic idea of the moments method is that, for the description of carrier-transport phenomena in semiconductor devices, it is often sufficient to rely on equations defined over the $\mathbf{r}, t$ space alone; in fact, for practical applications the information about the distribution of the crystal momentum is less important. The equations in the $\mathbf{r}, t$ space are extracted from the BTE by multiplying the latter by suitable functions $\alpha(\mathbf{k})$ and integrating the result over the $\mathbf{k}$ space. The integration saturates the $\mathbf{k}$ coordinates and, as shown in Sect. C.6, provides an equation in the $\mathbf{r}, t$ space.[20] Remembering that the electron dynamics

[18] On the other hand, in a collisionless case it is $S_0 \to 0$ whence, from (19.43), it follows $\tau \to \infty$. In this situation there is no limit to the departure of f from f^{eq}.

[19] Compare with the discussion carried out in Sect. 6.6.3.

[20] As indicated in Sect. C.6, term "moment" is specifically used when α is a polynomial in $\mathbf{k}$. As the dependence of α on $\mathbf{k}$ is not specified yet, it is implied that the form of α is such that the integrals in (19.51) converge.

using the equivalent Hamiltonian operator is described by expanding the electron's wave function in terms of the Bloch functions (Sect. 19.2.1), the integration over the **k** space is in fact limited to the first Brillouin zone. On the other hand, the typical behavior of the distribution function at the boundary Γ of the first Brilluoin zone is such that $\alpha(\mathbf{k})\, f(\mathbf{r},\mathbf{k},t) \to 0$ when $\mathbf{k} \to \mathbf{k}_\Gamma$. This amount to assuming that there are no electrons at the boundary.[21] From a practical standpoint, the hypothesis has the same effect as that of replacing the first Brillouin zone with an infinite domain and assuming that the distribution function in a non-equilibrium condition vanishes at infinity in the same exponential-like fashion as it does at equilibrium, where it becomes proportional to the Fermi–Dirac statistics [86, 89]. With these premises, using the general form (19.33) of the equation, the moment of the BTE with respect to α reads

$$\iiint_{-\infty}^{+\infty} \alpha \left[\frac{\partial f}{\partial t} + \mathbf{u} \cdot \operatorname{grad}_{\mathbf{r}} f - \frac{q}{\hbar} \left(\mathbf{E} + \mathbf{u} \wedge \mathbf{B} \right) \cdot \operatorname{grad}_{\mathbf{k}} f \right] \mathrm{d}^3 k = \iiint_{-\infty}^{+\infty} \alpha\, C\, \mathrm{d}^3 k . \tag{19.51}$$

As the BTE is a continuity equation of the distribution function in the phase space, (19.51) is expected to be a continuity equation in the **r** space; in fact, as shown below, it is the continuity equation of the product $n\,\overline{\alpha}$, where n is the concentration of the electrons in the conduction band, given by the first relation in (19.31), and

$$\overline{\alpha}(\mathbf{r},t) = \frac{1}{n} \iiint_{-\infty}^{+\infty} \alpha(\mathbf{k})\, f(\mathbf{r},\mathbf{k},t)\, \mathrm{d}^3 k \tag{19.52}$$

is the average of α over the **k** space. The continuity equation is derived below by working out separately the different terms appearing in (19.51).

Time Derivative

The derivation of this term is readily accomplished by observing that $\alpha\, \partial f/\partial t = \partial(\alpha\, f)/\partial t$, whence

$$\iiint_{-\infty}^{+\infty} \alpha\, \frac{\partial f}{\partial t}\, \mathrm{d}^3 k = \frac{\partial}{\partial t} \iiint_{-\infty}^{+\infty} \alpha\, f\, \mathrm{d}^3 k = \frac{\partial}{\partial t} \left(n\, \overline{\alpha} \right) . \tag{19.53}$$

This shows that (19.51) is the continuity equation of $n\,\overline{\alpha}$, as anticipated.

Diffusion Term

To calculate this terms one starts with the relation $\alpha\, \mathbf{u} \cdot \operatorname{grad}_{\mathbf{r}} f = \operatorname{div}_{\mathbf{r}}(\alpha\, \mathbf{u}\, f)$, that derives from the second identity in (A.16) and from the fact that $\alpha\, \mathbf{u}$ does not depend on **r**. Thus,

$$\iiint_{-\infty}^{+\infty} \alpha\, \mathbf{u} \cdot \operatorname{grad}_{\mathbf{r}} f\, \mathrm{d}^3 k = \operatorname{div}_{\mathbf{r}} \iiint_{-\infty}^{+\infty} \alpha\, \mathbf{u}\, f\, \mathrm{d}^3 k = \operatorname{div}_{\mathbf{r}} \left(n\, \overline{\alpha\, \mathbf{u}} \right) . \tag{19.54}$$

[21] In the case of the conduction band of germanium, the minima are at the boundary (Sect. 17.6.5), which makes the hypothesis inconsistent as it stands; to perform the integration one must shift the origin of the **k** space and exploit the periodicity of the band structure. The hypothesis that the distribution function vanishes at the boundary of the first Brillouin zone is made also in the application of the moments method to the holes of the valence band.

Drift Term

The term containing the electric field is treated starting from the identity

$$\alpha\,\mathbf{E}\cdot\mathrm{grad}_{\mathbf{k}} f = \mathbf{E}\cdot\mathrm{grad}_{\mathbf{k}}(\alpha\, f) - f\,\mathbf{E}\cdot\mathrm{grad}_{\mathbf{k}}\alpha. \tag{19.55}$$

The integral of the first term at the right hand side of (19.55) vanishes due to identity (A.26) and to the asymptotic behavior of f described earlier. As a consequence,

$$-\iiint_{-\infty}^{+\infty} \alpha\,\mathbf{E}\cdot\mathrm{grad}_{\mathbf{k}} f\,\mathrm{d}^3k = \mathbf{E}\cdot\iiint_{-\infty}^{+\infty} f\,\mathrm{grad}_{\mathbf{k}}\alpha\,\mathrm{d}^3k = \mathbf{E}\cdot n\,\overline{\mathrm{grad}_{\mathbf{k}}\alpha}. \tag{19.56}$$

The term containing the magnetic induction is treated more easily by rewriting the mixed product as $\mathrm{grad}_{\mathbf{k}} f\cdot\alpha\,\mathbf{u}\wedge\mathbf{B} = \mathrm{grad}_{\mathbf{k}} f\wedge\alpha\,\mathbf{u}\cdot\mathbf{B}$ (compare with (A.31)) and using the first identity in (A.35), to find

$$\mathrm{grad}_{\mathbf{k}} f\wedge\alpha\,\mathbf{u}\cdot\mathbf{B} = \mathrm{rot}_{\mathbf{k}}(f\,\alpha\,\mathbf{u})\cdot\mathbf{B} - f\,\mathrm{rot}_{\mathbf{k}}(\alpha\,\mathbf{u})\cdot\mathbf{B}. \tag{19.57}$$

The integral of $\mathrm{rot}_{\mathbf{k}}(f\,\alpha\,\mathbf{u})\cdot\mathbf{B}$ over $\mathbf{k}$ vanishes due to identity (A.38) and to the asymptotic behavior of f; this yields

$$-\iiint_{-\infty}^{+\infty} \alpha\,\mathbf{u}\wedge\mathbf{B}\cdot\mathrm{grad}_{\mathbf{k}} f\,\mathrm{d}^3k = \mathbf{B}\cdot\iiint_{-\infty}^{+\infty} f\,\mathrm{rot}_{\mathbf{k}}(\alpha\,\mathbf{u})\,\mathrm{d}^3k. \tag{19.58}$$

In turn, identity (A.35) transforms the integrand at the right hand side of (19.58) as $f\,\mathrm{rot}_{\mathbf{k}}(\alpha\,\mathbf{u}) = f\,\alpha\,\mathrm{rot}_{\mathbf{k}}\mathbf{u} + f\,\mathrm{grad}_{\mathbf{k}}\alpha\wedge\mathbf{u}$ where, thanks to the definition (17.52) of the group velocity, it is $\mathrm{rot}_{\mathbf{k}}\mathbf{u} = (1/\hbar)\,\mathrm{rot}_{\mathbf{k}}\mathrm{grad}_{\mathbf{k}} H = 0$. Thus, (19.58) becomes

$$-\iiint_{-\infty}^{+\infty} \alpha\,\mathbf{u}\wedge\mathbf{B}\cdot\mathrm{grad}_{\mathbf{k}} f\,\mathrm{d}^3k = \mathbf{B}\cdot\iiint_{-\infty}^{+\infty} f\,\mathrm{grad}_{\mathbf{k}}\alpha\wedge\mathbf{u}\,\mathrm{d}^3k =$$

$$= \iiint_{-\infty}^{+\infty} f\,\mathrm{grad}_{\mathbf{k}}\alpha\cdot\mathbf{u}\wedge\mathbf{B}\,\mathrm{d}^3k = n\,\overline{\mathrm{grad}_{\mathbf{k}}\alpha\cdot\mathbf{u}\wedge\mathbf{B}}. \tag{19.59}$$

In (19.59), the term containing the magnetic induction does not contribute to the moment if α or f depends on $\mathbf{k}$ through energy alone, $\alpha = \alpha(H)$ or $f = f(H)$. In fact, in the first case it is $\mathrm{grad}_{\mathbf{k}}\alpha = (\mathrm{d}\alpha/\mathrm{d}H)\,\mathrm{grad}_{\mathbf{k}} H$, whence

$$\mathrm{grad}_{\mathbf{k}}\alpha\cdot\mathbf{u}\wedge\mathbf{B} = \frac{\mathrm{d}\alpha}{\mathrm{d}H}\,\hbar\,\mathbf{u}\cdot\mathbf{u}\wedge\mathbf{B} = 0. \tag{19.60}$$

The same calculation holds when $f = f(H)$, starting from the integrand at the left hand side of (19.59).

Collision Term

Here it is convenient to distinguish between the inter-band and intra-band transitions, introduced in Sects. 19.3, 19.3.2. Thus, the collision term is written $C = C_b + C_v$, where as above suffix b (v) stands for "inter-band" ("intra-band"). This yields

$$\iiint_{-\infty}^{+\infty} \alpha\,C\,\mathrm{d}^3k = W_b + W_v, \qquad W_{b(v)}[\alpha] = \iiint_{-\infty}^{+\infty} \alpha\,C_{b(v)}\,\mathrm{d}^3k, \tag{19.61}$$

where the functional symbol reminds one that $W_{b(v)}$ is determined by the form of α.

19.4.1 Moment Equations

Adding up (19.53), (19.54), (19.56), (19.59), and (19.61) after multiplying the drift terms by $q/\hbar$ provides the explicit form of (19.51), that reads

$$\frac{\partial}{\partial t}(n\,\overline{\alpha}) + \mathrm{div}_{\mathbf{r}}(n\,\overline{\alpha\,\mathbf{u}}) + \frac{q}{\hbar}\,n\,\overline{\mathrm{grad}_{\mathbf{k}}\alpha\cdot(\mathbf{E}+\mathbf{u}\wedge\mathbf{B})} = W_b[\alpha] + W_v[\alpha]. \quad (19.62)$$

A simple reasoning shows that the equations of the form (19.62) that are obtained from different choices of α are coupled with each other. To show this one takes for simplicity the one-dimensional case and lets $\alpha = c\,k^m$, with m a positive integer and c a constant. Also, the parabolic-band approximation is assumed to hold, so that $\mathbf{u}$ is a linear function of k. It follows that the time derivative in (19.62) contains the moment of order m of f, while the diffusion term (due to the product $\alpha\,\mathbf{u}$) contains the moment of order $m+1$; in turn, the summand proportional to $\mathbf{E}$ in the drift term contains the moment of order $m-1$ due to the derivative of α, while the summand proportional to $\mathbf{B}$ contains the moment of order m. These considerations are sufficient to show that the equation whose unknown is the moment of order m is coupled with those whose unknowns are the moment of order $m-1$ and $m+1$.

In the typical applications of the moments expansion a finite set of equations is considered, starting from the lowest-order moment $m = 0$ up to some order m_0. The system of equations thus obtained is indeterminate because, due to the coupling mentioned above, the number of equations is m_0 whereas the number of unknown moments appearing in them is m_0+1. To make the system determinate it is then necessary to add an extra condition, that is found by prescribing an approximate form of the (m_0+1)th moment.[22] Such a prescription reduces the number of unknown moments to m_0 and makes the system of differential equations determinate; for this reason it is called *closure condition*. The typical choice for the closure condition is to approximate the (m_0+1)th moment using the equilibrium distribution.

19.4.1.1 Moment of Order Zero

The *moment of order zero* is obtained by letting $\alpha = 1$ in (19.62), whose left hand side becomes $\partial n/\partial t + \mathrm{div}_{\mathbf{r}}(n\,\overline{\mathbf{u}})$; the electron concentration n is the moment of order zero and $\overline{\mathbf{u}} = \mathbf{v}$ is the average velocity, as defined in (19.31). If the zero-order moment of the collision term does not introduce further unknowns, the equation's unknowns are two: n and $\mathbf{v}$. The form of the left hand side shows that integrating the zero-order moment of the BTE over an arbitrary volume Ω of the $\mathbf{r}$ space provides the balance equation for the number of electrons of the conduction band (compare

[22] The choice of the highest-order moment as the function to be approximated is reasonable in view of the analysis of the moments method carried out in Sect. C.6. In fact, as the moments are the coefficients of a converging Taylor series, they become smaller and smaller as the order increases; thus, the error due to approximating the highest-order coefficient is expected to be the smallest.

with Sect. 23.2):

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega} n\ \mathrm{d}\Omega + \int_{\Sigma} n\,\mathbf{v}\cdot\mathbf{s}\ \mathrm{d}\Sigma = \int_{\Omega} W_b[1]\,\mathrm{d}\Omega + \int_{\Omega} W_v[1]\,\mathrm{d}\Omega, \tag{19.63}$$

where Σ is the boundary of Ω and $\mathbf{s}$ the unit vector normal to Σ, oriented in the outward direction. The second integral at the right hand side of (19.63) does not contribute to the electrons' balance; in fact, it describes transitions that do not influence the number of electrons of the band because both the initial and final state belong to it. On the other hand, due to the arbitrariness of Ω, the integral vanishes only if $W_v[1] = 0$. This result shows that the zero-order moment of the intra-band transitions vanishes,[23] so that the only transitions of importance for the zero-order moment, despite being dominated by much larger relaxation times, are the inter-band ones. This is the exception anticipated in Sect. 19.3.2. The zero-order moment for the electrons of the conduction band then reads

$$\frac{\partial n}{\partial t} + \mathrm{div}_{\mathbf{r}}\,(n\,\mathbf{v}) = W_b[1]. \tag{19.64}$$

The form of the inter-band term $W_b[1]$, which is not relevant for the analysis in hand, is worked out in Chap. 20.

19.4.1.2 General Form of the Higher-Order Moments

As shown above, the contribution of the intra-band transitions vanishes for $\alpha = 1$. In contrast, it becomes dominant for other choices of α; in fact, in such cases the intra-band transitions do not cancel out any more and their scattering rates turn out to be much higher than those of the inter-band transitions. This allows one to adopt an approximation for $W_b[\alpha]$, namely,

$$W_b[\alpha] = \iiint_{-\infty}^{+\infty} \alpha\, C_b\, \mathrm{d}^3k \simeq \overline{\alpha} \iiint_{-\infty}^{+\infty} C_b\, \mathrm{d}^3k. \tag{19.65}$$

In other terms it is assumed that, since the contribution of $W_b[\alpha]$ to the collision term is small when $\alpha \neq 1$, the error introduced by (19.65) is negligible. Expanding the time derivative in (19.62), and using (19.64), (19.65), yields

$$n\,\frac{\partial \overline{\alpha}}{\partial t} + \mathrm{div}_{\mathbf{r}}\,(n\,\overline{\alpha}\,\overline{\mathbf{u}}) - \overline{\alpha}\,\mathrm{div}_{\mathbf{r}}\,(n\,\mathbf{v}) + \frac{q}{\hbar}\,n\,\overline{\mathrm{grad}_{\mathbf{k}}\alpha\cdot(\mathbf{E}+\mathbf{u}\wedge\mathbf{B})} = W_v[\alpha], \tag{19.66}$$

where only the intra-band transitions appear. Due to its simpler form, (19.66) will be used in the following to derive the balance equations with $\alpha \neq 1$.

[23] A similar reasoning is used to explain (20.16).

19.4.1.3 Moments of Order One, Two, and Three

The *moment of order one* of the BTE is found by letting $\alpha = u_i$ with $i = 1, 2, 3$ in (19.66); this yields the continuity equation for the ith component of the average velocity of the electrons, $\overline{u}_i = v_i$:

$$n \frac{\partial v_i}{\partial t} + \mathrm{div}_{\mathbf{r}} (n \overline{u_i \mathbf{u}}) - v_i \mathrm{div}_{\mathbf{r}} (n \mathbf{v}) + \frac{q}{\hbar} n \overline{\mathrm{grad}_{\mathbf{k}} u_i \cdot (\mathbf{E} + \mathbf{u} \wedge \mathbf{B})} = W_v[u_i]. \tag{19.67}$$

To proceed it is necessary to introduce the definition of average kinetic energy and average flux of the electrons' kinetic energy,[24]

$$w(\mathbf{r}, t) = \frac{1}{n} \iiint_{-\infty}^{+\infty} E_e(\mathbf{k}) \, f(\mathbf{r}, \mathbf{k}, t) \, \mathrm{d}^3 k, \tag{19.68}$$

$$\mathbf{b}(\mathbf{r}, t) = \frac{1}{n} \iiint_{-\infty}^{+\infty} E_e(\mathbf{k}) \, \mathbf{u}(\mathbf{k}) \, f(\mathbf{r}, \mathbf{k}, t) \, \mathrm{d}^3 k, \tag{19.69}$$

with $E_e = E(\mathbf{k}) - E_C$. Then, the *moment of order two* of the BTE is found by letting $\alpha = E_e$ in (19.66); this yields the continuity equation for the average kinetic energy of the electrons, (19.68). In the derivation, the term containing the magnetic induction vanishes due to (19.60); using the definition (17.52) of the group velocity, the equation reads

$$n \frac{\partial w}{\partial t} + \mathrm{div}_{\mathbf{r}} (n \mathbf{b}) - w \, \mathrm{div}_{\mathbf{r}} (n \mathbf{v}) + q \, n \, \mathbf{v} \cdot \mathbf{E} = W_v[E_e]. \tag{19.70}$$

The *moment of order three* of the BTE is found by letting $\alpha = E_e \, u_i$ with $i = 1, 2, 3$ in (19.66); this yields the continuity equation for the average flux of the electrons' kinetic energy, (19.69); the equation reads

$$n \frac{\partial b_i}{\partial t} + \mathrm{div}_{\mathbf{r}} \left(n \overline{E_e \, u_i \, \mathbf{u}} \right) - b_i \, \mathrm{div}_{\mathbf{r}} (n \mathbf{v}) + \frac{q}{\hbar} n \overline{\mathrm{grad}_{\mathbf{k}} (E_e \, u_i) \cdot (\mathbf{E} + \mathbf{u} \wedge \mathbf{B})} = W_v[E_e \, u_i]. \tag{19.71}$$

The choices $\alpha = 1$, $\alpha = u_i$, $\alpha = E_e$, and $\alpha = E_e \, u_i$ are such that each moment equation provides the balance relation of a dynamic quantity of interest: number of electrons, average velocity, average kinetic energy, average flux of the kinetic energy; the even-order moments yield a scalar equation, whereas the odd-order moments yield a vector equation.

[24] In the equilibrium condition the product $E_e \, f^{\mathrm{eq}}$ is even with respect to $\mathbf{k}$. In turn, $\mathbf{u} = (1/\hbar) \, \mathrm{grad}_{\mathbf{k}} E$ is odd, so that $\mathbf{b}^{\mathrm{eq}} = 0$. Compare with the similar comment made about the average velocity in (19.31).

19.4.2 Hierarchical Models

The order-one moment (19.67) contains the new unknown $\overline{u_i\,\mathbf{u}}$ besides n and $\mathbf{v}$ already present in (19.64); the order-two moment (19.70) contains again n and $\mathbf{v}$, and the new unknowns w, $\mathbf{b}$. The order-three moment contains n, $\mathbf{v}$, $\mathbf{b}$, and the new unknown $\overline{E_e\,u_i\,\mathbf{u}}$. The drift terms and the collision terms, depending on their form, may, or may not introduce extra unknowns; even if they don't, the number of unknowns listed above exceeds that of the equations. It is worth anticipating that the finite set of balance equations indicated in Sect. 19.4.1 is obtained by taking the equations in pairs: specifically, the first pair is made of the balance equations of order zero and one, (19.64) and (19.66), that are collectively termed *drift-diffusion model*; in this case, the three unknowns n, $\mathbf{v}$, and $\overline{u_i\,\mathbf{u}}$ are reduced to two by the closure condition (Sect. 19.4.1), that consists in replacing the highest-order moment $\overline{u_i\,\mathbf{u}}$ with its equilibrium expression. A more elaborate model is obtained by taking the first two pairs, namely, the balance equations of order zero through three, (19.64), (19.66), (19.70), and (19.71), that are collectively termed *hydrodynamic model*; the six unknowns n, $\mathbf{v}$, and $\overline{u_i\,\mathbf{u}}$, w, $\mathbf{b}$, $\overline{E_e\,u_i\,\mathbf{u}}$ are reduced to five by prescribing the closure condition, then to four by determining a relation between the second-order moments $\overline{u_i\,\mathbf{u}}$ and w.

By this procedure one constructs a set of hierarchically-ordered models of increasing complexity. The type of model adopted in practical applications depends on the trade-off between the information that one needs to acquire about the physical behavior of the device under investigation and the computational cost of the system of differential equations to be solved. To date, the moments method has been investigated up to order 6 [45], and has been extended to order 21 using a scheme based on Legendre polynomial expansion [59]; the standard implementations in the commercial simulation programs used by semiconductor Companies adopt the hydrodynamic model.[25]

The balance equations derived so far are still rather cumbersome in view of the application to the analysis of semiconductor devices. A number of simplifications are illustrated below, which eventually lead to the standard form of the hydrodynamic model [38, 40, 89, 93]. To begin, one considers the time derivatives at the left hand side of the balance equations. Such derivatives differ from zero only if the distribution function depends explicitly on time, which typically happens when time-dependent boundary conditions are imposed to the device under investigation. In the practical cases, the maximum frequency of the electric signals applied to a device or an integrated circuit is lower by many orders of magnitude than the inverse relaxation times associated to the intra-band transitions; this makes it possible to neglect the time derivatives of v_i, w, and b_i. A quasi-static approximation[26] is thus assumed

[25] Comprehensive reviews of the solution methods for the BTE are in [55, 56] as far as the Monte Carlo method is concerned, and in [49] for deterministic methods.

[26] A similar reasoning is used to treat the time derivative of the vector potential when the semiconductor equations are coupled with the Maxwell equations (Sect. 19.5.4).

in the continuity Eqs. (19.67), (19.70), and (19.71). The argument leading to this approximation does not apply to the case of (19.64) because only the inter-band transitions take place there, whose relaxation times are much longer than those of the intra-band transitions and, in many cases, also than the inverse maximum frequency of the external signal. As a consequence, the term $\partial n/\partial t$ in (19.64) must be retained when the boundary conditions depend on time.

As a second approximation, one adopts the parabolic-band approximation; this implies that in a non-equilibrium condition the electrons of the conduction band still occupy energy states in the vicinity of the absolute minima. Such a condition is in general fulfilled as shown below.[27] Letting a indicate one of the absolute minima of the conduction band, and using (19.24) after dropping suffix "0", yields

$$u_i = \frac{\hbar\,(k_i - k_{ia})}{m_{ia}}, \quad \mathrm{grad}_{\mathbf{k}}\, u_i = \frac{\hbar}{m_{ia}}\,\mathbf{i}_i, \quad \mathrm{grad}_{\mathbf{k}}(E_e\, u_i) = \hbar\left(\frac{E_e}{m_{ia}}\,\mathbf{i}_i + u_i\,\mathbf{u}\right), \tag{19.72}$$

with $\mathbf{i}_i$ the unit vector of the ith axis. From now on, the equations derived from the parabolic-band approximation refer to the ath valley. In principle, the electron concentration, average velocity, and the other averages should be indicated with n_a, $\mathbf{v}_a$, and so on; this is not done here to avoid complicacies in the notation. The suffix will be introduced in Sect. 19.5.2, where the contributions of the valleys are added up. This comment does not apply to the moment of order zero, (19.64), because its derivation does not entail any simplifying hypothesis.

With these premises, one manipulates the terms $\overline{u_i\,\mathbf{u}}$, $E_e\,\overline{u_i\,\mathbf{u}}$ by introducing the auxiliary quantity $\mathbf{c} = \mathbf{u} - \mathbf{v}$, called *random velocity*. Clearly it is $\overline{c}_i = \overline{u}_i - v_i = 0$, so that $\overline{u_i\,\mathbf{u}} = v_i\,\mathbf{v} + \overline{u_i\,\mathbf{u}}$; it follows

$$\mathrm{div}_{\mathbf{r}}(n\,\overline{u_i\,\mathbf{u}}) = \mathrm{div}_{\mathbf{r}}(n\,\overline{c_i\,\mathbf{c}}) + v_i\,\mathrm{div}_{\mathbf{r}}(n\,\mathbf{v}) + n\,\mathbf{v}\cdot\mathrm{grad}_{\mathbf{r}}\, v_i. \tag{19.73}$$

The last term at the right hand side of (19.73) is called *convective term*. In the typical operating conditions of the semiconductor devices this term can be neglected (refer to [86] and the comments below). Replacing into (19.67) the simplified form of (19.73) along with the second relation of (19.72) yields

$$\mathrm{div}_{\mathbf{r}}\,(n\,m_{ia}\,\overline{c_i\,\mathbf{c}}) + q\,n\;\;(\mathbf{E} + \mathbf{v}\wedge\mathbf{B})_i = m_{ia}\,W_v[u_i]. \tag{19.74}$$

The latter equation contains the unknowns n and $\mathbf{v}$ already present in (19.64), and the new unknown $\overline{c_i\,\mathbf{c}}$. The drift term does not introduce extra unknowns. Note that $\overline{c_i\,\mathbf{c}}$ is actually made of three vectors, so that it may be thought of as a symmetric 3×3 tensor with components $\overline{c_i\,c_j}$, $i, j = 1, 2, 3$. To give it a more compact form, after observing that $m_{ia}\,\overline{c_i\,\mathbf{c}}$ has the units of an energy, one defines the *electron-temperature tensor* of components T_{ij} such that

$$k_B\,T_{ij} = m_{ia}\,\overline{c_i\,c_j}, \qquad n\,k_B\,T_{ij} = \iiint_{-\infty}^{+\infty} m_{ia}\,c_i\,c_j\,f\,\mathrm{d}^3k, \tag{19.75}$$

[27] The adoption of the parabolic-band approximation may be avoided at the cost of redefining the carrier temperature and introducing more relaxation times [108].

with k_B the Boltzmann constant. Letting $\mathbf{T}_i$ be the vector of entries T_{i1}, T_{i2}, T_{i3}, so that $k_B\,\mathbf{T}_i = m_{ia}\,\overline{c_i\,\mathbf{c}}$, one finds for the ith component of the moment of order one

$$\mathrm{div}_{\mathbf{r}}\,(n\,k_B\,\mathbf{T}_i) + q\,n\;\;(\mathbf{E} + \mathbf{v} \wedge \mathbf{B})_i = m_{ia}\,W_v[u_i]. \tag{19.76}$$

In the equilibrium condition the electron-temperature tensor reduces to the product of a scalar coefficient times the identity tensor; in the limit of the Boltzmann distribution, the scalar coefficient identifies with the lattice temperature (Sect. 19.6.4), this providing an estimate of the modulus $|\mathbf{c}|$ of the random velocity. The modulus $|\mathbf{v}|$ of the average velocity in a non-equilibrium condition can be estimated as well, basing upon the current density and carrier concentration of the devices' operating conditions. It is found that in typical situations it is $|\mathbf{v}| \ll |\mathbf{c}|$, so that the average motion of the carriers in a non-equilibrium condition can be thought of as that of a slowly-drifting fluid. This justifies the neglect of the convective term in (19.73), and also allows one to neglect $v_i\,\mathbf{v}$ with respect to $\overline{c_i\,\mathbf{c}}$ when these terms appear in the same expression.

The simplifications used in (19.67) apply in the same manner to the moment of order three, (19.71); in fact it is $\overline{E_e\,u_i\,\mathbf{u}} = \overline{E_e\,u_i\,\mathbf{c}} + \overline{E_e\,u_i}\;\mathbf{v}$, whence

$$\mathrm{div}_{\mathbf{r}}(n\,\overline{E_e\,u_i\,\mathbf{u}}) = \mathrm{div}_{\mathbf{r}}(n\,\overline{E_e\,u_i\,\mathbf{c}}) + b_i\,\mathrm{div}_{\mathbf{r}}(n\,\mathbf{v}) + n\,\mathbf{v}\cdot\mathrm{grad}_{\mathbf{r}}b_i\,. \tag{19.77}$$

Using the third relation of (19.72) in the drift term of (19.71) transforms the latter into $(q/m_{ia})\,n\;[(w\,\mathbf{i}_i + m_{ia}\,v_i\,\mathbf{v} + m_{ia}\,\overline{c_i\,\mathbf{c}})\cdot\mathbf{E} + \mathbf{b}\wedge\mathbf{B}\cdot\mathbf{i}_i + m_{ia}\,\overline{u_i\,\mathbf{u}\cdot\mathbf{u}\wedge\mathbf{B}}]$, where the mixed product vanishes due to the repeated factor, and $m_{ia}\,v_i\,\mathbf{v}$ is negligible as shown above. Replacing (19.77) into (19.71) after neglecting the time derivative and the convective term, yields

$$\mathrm{div}_{\mathbf{r}}\left(n\,m_{ia}\,\overline{E_e\,u_i\,\mathbf{c}}\right) + q\,n\;\;[(w\,\mathbf{i}_i + k_B\,\mathbf{T}_i)\cdot\mathbf{E} + \mathbf{b}\wedge\mathbf{B}\cdot\mathbf{i}_i] = m_{ia}\,W_v[E_e\,u_i]. \tag{19.78}$$

The relation between the second-order moments necessary to reduce the number of unknowns is now determined starting from the expression of E_e in the parabolic-band approximation. Using the first relation in (19.72) yields $E_e = (1/2)\sum_{i=1}^{3} m_{ia}\,u_i^2$ whence, from (19.68),

$$w = \frac{1}{2}\sum_{i=1}^{3} m_{ia}\,v_i^2 + \frac{3}{2}\,k_B\,T_e, \qquad T_e = \frac{T_{11}+T_{22}+T_{33}}{3}, \tag{19.79}$$

with T_e the *electron temperature*. The two summands of w in (19.79) are also called *convective part* and *thermal part* of the average kinetic energy, respectively. The same reasoning that has led to the neglect of $m_{ia}\,v_i\,\mathbf{v}$ with respect to $m_{ia}\,c_i\,\mathbf{c}$ also shows that the thermal part is dominant with respect to the convective part, so that $w \simeq (3/2)\,k_B\,T_e$. Moreover, it can be assumed that the electron-temperature tensor retains the same structure of the equilibrium case, so that

$$\begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix} \simeq \begin{bmatrix} T_{11} & 0 & 0 \\ 0 & T_{22} & 0 \\ 0 & 0 & T_{33} \end{bmatrix} \simeq T_e(\mathbf{r},t)\,\mathcal{I}, \tag{19.80}$$

with $\mathcal{I}$ the identity tensor. As a consequence, $(w\,\mathbf{i}_i + k_B\,\mathbf{T}_i)\cdot\mathbf{E} = (5/2)\,k_B\,T_e\,\mathbf{i}_i\cdot\mathbf{E}$ and $\mathrm{div}_{\mathbf{r}}(n\,k_B\,\mathbf{T}_i) = \partial(n\,k_B\,T_e)/\partial x_i$. The latter is the ith component of $\mathrm{grad}_{\mathbf{r}}\,(n\,k_B\,T_e)$. In summary, the balance equations for the moments of order one, two, and three read

$$\frac{\partial\,(n\,k_B\,T_e)}{\partial x_i} + q\,n\;\;(\mathbf{E} + \mathbf{v}\wedge\mathbf{B})_i = m_{ia}\,W_v[u_i], \tag{19.81}$$

$$\mathrm{div}_{\mathbf{r}}\,(n\,\mathbf{b}) - (3/2)\,k_B\,T_e\,\mathrm{div}_{\mathbf{r}}\,(n\,\mathbf{v}) + q\,n\;\;\mathbf{v}\cdot\mathbf{E} = W_v[E_e], \tag{19.82}$$

$$\mathrm{div}_{\mathbf{r}}\left(n\,m_{ia}\,\overline{E_e\,u_i\,\mathbf{c}}\right) + q\,n\;\;[(5/2)\,k_B\,T_e\,\mathbf{E} + \mathbf{b}\wedge\mathbf{B}]_i = m_{ia}\,W_v[E_e\,u_i]. \tag{19.83}$$

19.4.2.1 Macroscopic Relaxation Times of the Higher-Order Moments

As remarked in Sect. (19.4.1), the collision terms of the moments of order higher than zero account for the intra-band transitions only. These terms are worked out here using the perturbative form of the BTE (Sect. 19.3.3); this approach is coherent with the other approximations from which the balance Eqs. (19.81–19.83) derive. The collision term of (19.82) then becomes

$$W_v[E_e] = -\iiint_{-\infty}^{+\infty} E_e\,\frac{f - f^{\mathrm{eq}}}{\tau}\,\mathrm{d}^3k, \tag{19.84}$$

with $\tau \simeq \tau_v$ (Sect. 19.3.2). The equilibrium part is worked out by defining the *energy-relaxation time* τ_w such that

$$\iiint_{-\infty}^{+\infty} E_e\,\frac{f^{\mathrm{eq}}}{\tau_v}\,\mathrm{d}^3k = \frac{1}{\tau_w}\iiint_{-\infty}^{+\infty} E_e\,f^{\mathrm{eq}}\,\mathrm{d}^3k = \frac{n^{\mathrm{eq}}\,w^{\mathrm{eq}}}{\tau_w} \simeq \frac{3}{2}\,\frac{n^{\mathrm{eq}}\,k_B T_e^{\mathrm{eq}}}{\tau_w}, \tag{19.85}$$

where the definitions (19.68, 19.79) of the electrons' average kinetic energy and temperature are used. The left hand side of (19.85) does not vanish because the integrand is positive definite. The non-equilibrium part of $W_v[E_e]$ is approximated as

$$\iiint_{-\infty}^{+\infty} E_e\,\frac{f}{\tau_v}\,\mathrm{d}^3k \simeq \frac{1}{\tau_w}\iiint_{-\infty}^{+\infty} E_e\,f\,\mathrm{d}^3k = \frac{n\,w}{\tau_w} \simeq \frac{3}{2}\,\frac{n\,k_B T_e}{\tau_w}, \tag{19.86}$$

based on the observation that, due to the smallness of the intra-band relaxation time τ_v, the distribution function departs little from the equilibrium one (Sect. 19.3.3).

The derivation of the analogues of τ_w for the collision terms of (19.81) and (19.83) is somewhat more complicate. In fact, in most semiconductors, among which Si, Ge, and GaAs, the relaxation time τ_v is even with respect to $\mathbf{k}$ [57], which makes the integrals of $u_i\,f^{\mathrm{eq}}/\tau_v$ and $E_e\,u_i\,f^{\mathrm{eq}}/\tau_v$ to vanish because the integrand is odd. To overcome the difficulty one expands $f - f^{\mathrm{eq}}$ into a Taylor series with respect to a parameter and truncates the series in such a way as to retain the first summand which

is odd with respect to $\mathbf{k}$. For instance, letting λ be the parameter[28] and assuming that the first-order term of the expansion is odd, one lets $f - f^{\mathrm{eq}} \simeq (\mathrm{d}f/\mathrm{d}\lambda)^{\mathrm{eq}}\,\lambda$ whence

$$\iiint_{-\infty}^{+\infty} u_i \,\frac{(\mathrm{d}f/\mathrm{d}\lambda)^{\mathrm{eq}}}{\tau_v}\,\mathrm{d}^3k = \frac{1}{\tau_{pi}} \iiint_{-\infty}^{+\infty} u_i\,(\mathrm{d}f/\mathrm{d}\lambda)^{\mathrm{eq}}\,\mathrm{d}^3k, \tag{19.87}$$

$$\iiint_{-\infty}^{+\infty} E_e\,u_i \,\frac{(\mathrm{d}f/\mathrm{d}\lambda)^{\mathrm{eq}}}{\tau_v}\,\mathrm{d}^3k = \frac{1}{\tau_{bi}} \iiint_{-\infty}^{+\infty} E_e\,u_i\,(\mathrm{d}f/\mathrm{d}\lambda)^{\mathrm{eq}}\,\mathrm{d}^3k, \tag{19.88}$$

with τ_{pi} and τ_{bi} the *momentum-relaxation time* and *relaxation time of the energy flux*, respectively.[29] Due to their definitions, τ_{pi} and τ_{bi} are diagonal tensors. However, as their degree of anisotropy is small, they are approximated by scalar quantities, $\tau_{pi} \simeq \tau_p$ and $\tau_{bi} \simeq \tau_b$. Investigations about the relaxation times have been carried out by different techniques, specifically, the spherical-harmonics expansion method to determine the dependence on the average energy [90, 93], and the Monte Carlo method to study the anisotropy properties [11, 12, 43].

Using (19.87, 19.88) along with the definitions (19.31, 19.69) of the average velocity and average flux of kinetic energy finally yields

$$\iiint_{-\infty}^{+\infty} \frac{u_i}{\tau_v}\left(\frac{\mathrm{d}f}{\mathrm{d}\lambda}\,\lambda\right)^{\mathrm{eq}}\mathrm{d}^3k = \frac{n\,v_i}{\tau_p}, \quad \iiint_{-\infty}^{+\infty} \frac{E_e\,u_i}{\tau_v}\left(\frac{\mathrm{d}f}{\mathrm{d}\lambda}\,\lambda\right)^{\mathrm{eq}}\mathrm{d}^3k = \frac{n\,b_i}{\tau_b}. \tag{19.89}$$

19.5 Hydrodynamic and Drift-Diffusion Models

In Sect. 19.4 the moments method has been applied to derive a set of balance equations; the general form of the latter has successively been modified by introducing a number of simplifications: among them is the parabolic-band approximation, due to which, as indicated in Sect. 19.4.2, a set of equations restricted to the ath valley of the conduction band is obtained. In order to recover the equations for the whole band, it is necessary to add up the single-valley contributions. The procedure is the same for the hydrodynamic and drift-diffusion models; it will be worked out explicitly only for the simpler case of the drift-diffusion model.

[28] Typically the parameter used in this procedure is the electric field [57]. An expansion truncated to the first order is coherent with the first-order perturbation approach.

[29] The term "momentum" for τ_{pi} derives from the observation that the continuity equation for the ith component of the average velocity v_i of the electrons, (19.67), may also be thought of as the continuity equation for the ith component of the average momentum, $m_{ia}\,v_i$. In turn, τ_{bi} is also called *heat-relaxation time*.

19.5.1 HD Model

As anticipated in Sect. 19.4.2, the hydrodynamic (HD) model is obtained by taking the balance equations of order zero through three, (19.64), (19.81), (19.82), and (19.83), and imposing the closure condition onto the fourth-order moment. For simplicity, the latter is considered in the non-degenerate case whence, from (19.159), it is $m_{ia}\ \left(\overline{E_e\, u_i\, \mathbf{c}}\right)^{\text{eq}} \simeq \mathbf{i}_i\ (5/2)\ (k_B\, T)^2$. Letting $W = W_b[1]$ and using (19.85), (19.86), (19.89) yields

$$\frac{\partial n}{\partial t} + \text{div}_{\mathbf{r}}\,(n\,\mathbf{v}) = W, \qquad \frac{\partial\,(n\,k_B\,T_e)}{\partial x_i} + q\,n\ \ (\mathbf{E} + \mathbf{v} \wedge \mathbf{B})_i = -\frac{m_{ia}}{\tau_p}\,n\,v_i, \tag{19.90}$$

$$\text{div}_{\mathbf{r}}\,(n\,\mathbf{b}) - \frac{3}{2}\,k_B\,T_e\,\text{div}_{\mathbf{r}}\,(n\,\mathbf{v}) + q\,n\ \mathbf{v}\cdot\mathbf{E} = -\frac{3}{2}\,\frac{k_B}{\tau_w}\left[n\,T_e - (n\,T_e)^{\text{eq}}\right], \tag{19.91}$$

$$\frac{5}{2}\,(k_B\,T)^2\,\frac{\partial n}{\partial x_i} + q\,n\ \left(\frac{5}{2}\,k_B\,T_e\,\mathbf{E} + \mathbf{b} \wedge \mathbf{B}\right)_i = -\frac{m_{ia}}{\tau_b}\,n\,b_i. \tag{19.92}$$

which constitute a system of first-order, partial-differential equations in the unknowns n, $\mathbf{v}$, T_e, and $\mathbf{b}$. In general, the model's equations are to be solved over a volume that encloses the device under investigation; the boundary conditions that typically apply are discussed in Sect. 19.5.6. Two of the equations are scalar (namely, (19.91) and the first one in (19.90)), while the other two are vector equations. The system is non linear because the unknowns are multiplied by each other.[30] Note, however, that the second equation in (19.90) is linear with respect to the components of $\mathbf{v}$; the latter can be extracted and replaced into the two scalar equations. The same procedure is applicable to (19.92), which is linear with respect to the components of $\mathbf{b}$. After the replacements are completed, the system reduces to two scalar equations of the second order. An example is given in Sect. 19.5.5, with reference to the simpler case of the drift-diffusion model. Due to the components m_{ia} of the effective-mass tensor, the vector equations are anisotropic; however, when the contributions of the different valleys are combined together, the anisotropy cancels out (the explicit calculation is provided for the drift-diffusion model below).

The qualitative analysis of the model carried out above implies that the electric field and magnetic induction are known, so that they are embedded in the model's coefficients. In fact, this is not true, because the fields are influenced by the distribution of electric charge and current density that are some of the model's unknowns. For this reason, as shown below, the hydrodynamic equations, and the drift-diffusion ones as well, must be coupled with the Maxwell equations.

[30] Also, the generation-recombination term W embeds non-linear dependencies on some of the unknowns, Chap. 20.

19.5.2 DD Model

The drift-diffusion (DD) model is obtained by taking the balance equations of order zero and one, (19.90), and imposing the closure condition onto the second-order moment. For simplicity, the latter is considered in the non-degenerate case, whence $T_e^{\mathrm{eq}} = T$ (Sect. 19.6.4); the model thus reads

$$\frac{\partial n}{\partial t} + \mathrm{div}_{\mathbf{r}}\,(n\,\mathbf{v}) = W, \qquad k_B\,T\,\frac{\partial n_a}{\partial x_i} + q\,n_a\;(\mathbf{E} + \mathbf{v}_a \wedge \mathbf{B})_i = -\frac{m_{ia}}{\tau_p}\,n_a\,v_{ia}. \tag{19.93}$$

As indicated in Sect. 19.4.2, the first equation in (19.93) refers to the whole conduction band because its derivation did not entail any simplifying hypothesis; in contrast, the second equation refers to the ath minimum of the band due to the parabolic-band approximation. This explains the index attached to n and to the average velocity; the momentum-relaxation time, instead, does not depend on the valley [90]. As noted above, the dependence on the components v_{ia} is linear, which makes it possible to express them in terms of the other functions. In fact, it is more convenient to extract, instead of $\mathbf{v}_a$, the electron-current density of the ath minimum; remembering (4.21) and (4.22), the latter is given by $\mathbf{J}_a = -q\,n_a\,\mathbf{v}_a$. Then, the second equation in (19.93) is recast as

$$J_{ia} = J'_{ia} - \frac{q\,\tau_p}{m_{ia}}\;(\mathbf{J}_a \wedge \mathbf{B})_i\,, \qquad J'_{ia} = k_B\,T\,\frac{q\,\tau_p}{m_{ia}}\,\frac{\partial n_a}{\partial x_i} + \frac{q\,\tau_p}{m_{ia}}\,q\,n_a\,(\mathbf{E})_i, \tag{19.94}$$

with $J'_{ia} = J_{ia}(\mathbf{B} = 0)$. Letting $\mu_{ia} = q\,\tau_p/m_{ia}$, the matrix form of (19.94) reads

$$\begin{bmatrix} J_{a1} \\ J_{a2} \\ J_{a3} \end{bmatrix} = \begin{bmatrix} J'_{a1} \\ J'_{a2} \\ J'_{a3} \end{bmatrix} - \begin{bmatrix} \mu_{a1} & 0 & 0 \\ 0 & \mu_{a2} & 0 \\ 0 & 0 & \mu_{a3} \end{bmatrix} \begin{bmatrix} J_{a2}\,B_3 - J_{a3}\,B_2 \\ J_{a3}\,B_1 - J_{a1}\,B_3 \\ J_{a1}\,B_2 - J_{a2}\,B_1 \end{bmatrix}, \tag{19.95}$$

equivalent to

$$\begin{bmatrix} 1 & \mu_{a1}\,B_3 & -\mu_{a1}\,B_2 \\ -\mu_{a2}\,B_3 & 1 & \mu_{a2}\,B_1 \\ \mu_{a3}\,B_2 & -\mu_{a3}\,B_1 & 1 \end{bmatrix} \begin{bmatrix} J_{a1} \\ J_{a2} \\ J_{a3} \end{bmatrix} = \begin{bmatrix} J'_{a1} \\ J'_{a2} \\ J'_{a3} \end{bmatrix}. \tag{19.96}$$

The diagonal tensor $\hat{\mu}_a$ of entries μ_{ia} is called *mobility tensor* of the ath valley. Note that the product of a mobility by a magnetic induction is dimensionless. Letting $\mathbf{M}_a = \mu_{a1}\,\mu_{a2}\,\mu_{a3}\,(\hat{\mu}_a)^{-1}\,\mathbf{B}$, the components of the current density are found by solving the algebraic system (19.96), where the determinant of the matrix is

$$D_M = 1 + \mu_{a1}\,\mu_{a2}\,\mu_{a3}\left(\frac{B_1^2}{\mu_{1a}} + \frac{B_2^2}{\mu_{2a}} + \frac{B_3^2}{\mu_{3a}}\right) = 1 + \mathbf{B}\cdot\mathbf{M}_a. \tag{19.97}$$

The components of $\mathbf{J}_a$ are finally found to be

$$D_M\, J_{ai} = J'_{ai} + \mu_{ai}\, \left(\mathbf{B} \wedge \mathbf{J}'_a\right)_i + \left(\mathbf{M}_a \cdot \mathbf{J}'_a\right)\, B_i. \tag{19.98}$$

In typical situations the modulus of the magnetic induction is small; hence terms that are quadratic in the components of $\mathbf{B}$ may be neglected. This yields the approximate form

$$J_{ai} \simeq J'_{ai} + \mu_{ai}\, \left(\mathbf{B} \wedge \mathbf{J}'_a\right)_i . \tag{19.99}$$

The electron current density of the whole conduction band is thus found as

$$\mathbf{J} = \sum_{a=1}^{M_C} \mathbf{J}_a = \sum_{a=1}^{M_C} \left[k_B\, T\, \hat{\mu}_a\, \mathrm{grad} n_a + \hat{\mu}_a\, q\, n_a\, \mathbf{E} + \hat{\mu}_a\, \left(\mathbf{B} \wedge \mathbf{J}'_a\right)\right]. \tag{19.100}$$

In the perturbative approach followed here, it can be assumed that the total electron concentration n equally distributes[31] over the valleys, $n_a = n/M_C$. The first two summands at the right hand side of (19.100) then yield

$$\mathbf{J}' = \sum_{a=1}^{M_C} \mathbf{J}'_a = q\, \hat{\mu}_n\, n\, \mathbf{E} + q\, \hat{D}_n\, \mathrm{grad} n, \tag{19.101}$$

where the diagonal tensors $\hat{\mu}_n$, $\hat{D}_n$ are defined as

$$\hat{\mu}_n = \frac{1}{M_C} \sum_{a=1}^{M_C} \hat{\mu}_a, \qquad \hat{D}_n = \frac{k_B\, T}{q}\, \hat{\mu}_n. \tag{19.102}$$

They are called *electron-mobility tensor* and *electron-diffusivity tensor*, respectively. The second relation in (19.102), that states that diffusivity and mobility are proportional through $k_B\, T/q$, is called *Einstein relation*.[32] The form of $\hat{\mu}_n$ is specified on a case-by-case basis, depending on the semiconductor under consideration. Taking silicon by way of example ($M_C = 6$), the mass tensor is obtained from (17.82–17.84); thus, the mobility tensor $\hat{\mu}_a$ has one of the following forms:

$$\begin{bmatrix} \mu_l & 0 & 0 \\ 0 & \mu_t & 0 \\ 0 & 0 & \mu_t \end{bmatrix}, \quad \begin{bmatrix} \mu_t & 0 & 0 \\ 0 & \mu_l & 0 \\ 0 & 0 & \mu_t \end{bmatrix}, \quad \begin{bmatrix} \mu_t & 0 & 0 \\ 0 & \mu_t & 0 \\ 0 & 0 & \mu_l \end{bmatrix}, \tag{19.103}$$

[31] From this assumption and from (19.100) it also follows $\mathbf{J} = -q \sum_{a=1}^{M_C} n_a\, \mathbf{v}_a = -q\, (n/6) \sum_{a=1}^{M_C} \mathbf{v}_a$, whence $\mathbf{J} = -q\, n\, \mathbf{v}$ with $\mathbf{v} = (1/6) \sum_{a=1}^{M_C} \mathbf{v}_a$.

[32] The relation derives from Einstein's investigation on the Brownian motion [35] and has therefore a broader application. In a semiconductor it holds within the approximations of parabolic bands and non-degenerate conditions.

with $\mu_l = q\,\tau_p/m_l$, $\mu_t = q\,\tau_p/m_t$. The first form in (19.103) applies to the two minima belonging to axis k_1, and so on; thus, the electron-mobility tensor (19.102) is found to be

$$\hat{\mu}_n = \frac{1}{6}\left(2\begin{bmatrix}\mu_l & 0 & 0\\ 0 & \mu_t & 0\\ 0 & 0 & \mu_t\end{bmatrix} + 2\begin{bmatrix}\mu_t & 0 & 0\\ 0 & \mu_l & 0\\ 0 & 0 & \mu_t\end{bmatrix} + 2\begin{bmatrix}\mu_t & 0 & 0\\ 0 & \mu_t & 0\\ 0 & 0 & \mu_l\end{bmatrix}\right) = \mu_n\,\mathcal{I}, \tag{19.104}$$

with $\mathcal{I}$ the identity tensor and $\mu_n = (\mu_l + 2\,\mu_t)/3$ the *electron mobility*. The second definition in (19.102) then yields $\hat{D}_n = D_n\,\mathcal{I}$, with $D_n = (k_B\,T/q)\mu_n$ the *electron diffusivity* or *electron-diffusion coefficient*. From these results and (19.101) one derives

$$\mathbf{J}' = q\,\mu_n\,n\,\mathbf{E} + q\,D_n\,\mathrm{grad}n, \qquad \mathbf{J}'_a = \frac{1}{\mu_n\,M_C}\,\hat{\mu}_a\,\mathbf{J}'. \tag{19.105}$$

From this, after a somewhat lengthy calculation, the last term at the right hand side of (19.100) is found to be

$$\sum_{a=1}^{M_C}\hat{\mu}_a\,\left(\mathbf{B}\wedge\mathbf{J}'_a\right) = a_n\,\mu_n\,\mathbf{B}\wedge\mathbf{J}', \qquad a_n = \frac{\mu_t\,(\mu_t + 2\,\mu_l)}{3\,\mu_n^2}. \tag{19.106}$$

As anticipated in the qualitative discussion about the HD model, despite the fact that each vector equation is anisotropic, when the contributions of the different valleys are combined together the anisotropy cancels out. From the definition of the electron mobility $\mu_n = (\mu_l + 2\,\mu_t)/3$ one may also extract a scalar effective mass $m_n = q\,\tau_p/\mu_n$, that fulfills $1/m_n = (2/m_t + 1/m_l)/3$. Using the room-temperature values taken from Table 17.4 yields, for silicon, $m_n/m_0 \simeq 0.26$. By the same token one finds $a_n \simeq 2.61$.

In the next sections, the current density of the electrons in the conduction band will be used in equations involving also the current density of the holes in the valence band; for this reason it is necessary to use different symbols. Specifically, $\mathbf{J}_n$ for the former and $\mathbf{J}_p$ for the latter; with this provision, the above calculation yields

$$\mathbf{J}_n = q\,\mu_n\,n\,\mathbf{E} + q\,D_n\,\mathrm{grad}n + q\,a_n\,\mu_n\,\mathbf{B}\wedge(\mu_n\,n\,\mathbf{E} + D_n\,\mathrm{grad}n)\,, \tag{19.107}$$

which is called *drift-diffusion transport equation*. Thus the DD model for the electrons of the conduction band is given by (19.107) along with the first equation in (19.93); the latter is rewritten here as

$$\frac{\partial n}{\partial t} - \frac{1}{q}\,\mathrm{div}(\mathbf{J}_n) = W_n, \tag{19.108}$$

where a specific symbol for the generation-recombination term has been introduced as well.

19.5.3 DD Model for the Valence Band

The transport models illustrated so far are applicable to the valence band as well; here, the DD model will be worked out. Remembering the discussion of Sect. 19.2.3 about the dynamics in the parabolic-band approximation, the model is described in terms of the concentration and current density of holes. The two quantities are defined by adapting the corresponding expression for electrons, (19.31), as shown below. Letting $f = Q\,\Phi$, with $Q = 1/(4\,\pi^3)$ the density of states in the phase space and Φ the occupation probability, the hole concentration is

$$p(\mathbf{r},t) = \iiint_{-\infty}^{+\infty} Q\ (1-\Phi)\ \mathrm{d}^3k. \tag{19.109}$$

In turn, the hole current density is defined starting from the definition of the electron current density of the valence band. The latter is similar to (19.31), the difference being that the integration in (19.31) is restricted to the branch of $E(\mathbf{k})$ belonging to the conduction band, whereas the integration in (19.110) below is restricted to one of the two branches of the valence band:

$$\mathbf{J}_a = -q \iiint_{-\infty}^{+\infty} \mathbf{u}(\mathbf{k})\, Q\, \Phi(\mathbf{r},\mathbf{k},t)\, \mathrm{d}^3k. \tag{19.110}$$

Letting $\Phi = 1 - (1-\Phi)$ transforms (19.110) into

$$\mathbf{J}_a = q \iiint_{-\infty}^{+\infty} \mathbf{u}\, Q(1-\Phi)\, \mathrm{d}^3k - q \iiint_{-\infty}^{+\infty} \mathbf{u}\, Q\, \mathrm{d}^3k, \tag{19.111}$$

where the second integral vanishes because $\mathbf{u}$ is odd with respect to $\mathbf{k}$. As a consequence, the current density of the branch under consideration may also be thought of as given by the motion of the empty states (holes), having the group velocity $\mathbf{u}(\mathbf{k})$ and the positive charge q. Moreover, one defines the average velocity of holes using $Q\,(1-\Phi)$ as weighing function, to find

$$\mathbf{J}_a = q\ p_a\, \frac{\int_{\mathbf{k}} \mathbf{u}\, Q\,(1-\Phi)\, \mathrm{d}^3k}{\int_{\mathbf{k}} Q\,(1-\Phi)\, \mathrm{d}^3k} = q\ p_a\ \mathbf{v}_a \tag{19.112}$$

where the definition (19.109) of the hole concentration has been specified for the branch under consideration, and the short-hand notation $\int_{\mathbf{k}}$ has been used.

Given the above definitions, the derivation of the drift-diffusion model for holes follows the same pattern as for the electrons. Remembering the description of the band structure given in Sect. 17.6.5, for the valence band index a ranges over h and l; moreover, due to the isotropy of each branch deriving from the parabolic-band approximation (compare with 17.78), the effective mass is scalar. Then, the equivalent of (19.94) read, in vector form,

$$\mathbf{J}_a = \mathbf{J}'_a + \frac{q\,\tau_{pa}}{m_a}\,\mathbf{J}_a \wedge \mathbf{B}\,, \qquad \mathbf{J}'_a = -k_B\,T\,\frac{q\,\tau_{pa}}{m_a}\,\mathrm{grad}\,p_a + \frac{q\,\tau_{pa}}{m_a}\,q\,p_a\,\mathbf{E}\,, \tag{19.113}$$

where index a is attached also to the momentum-relaxation time because the two branches are different. Still due to such a difference, the holes do not distribute equally over the branches; the contribution of the drift and diffusion components then read, respectively,

$$\frac{q\,\tau_{ph}}{m_{hh}}\,q\,p_h\,\mathbf{E}+\frac{q\,\tau_{pl}}{m_{hl}}\,q\,p_l\,\mathbf{E}=q\,(\mu_{ph}+\mu_{pl})\,p\,\mathbf{E}, \tag{19.114}$$

$$-k_B\,T\,\frac{q\,\tau_{ph}}{m_{hh}}\,\mathrm{grad}\,p_h-k_B\,T\,\frac{q\,\tau_{pl}}{m_{hl}}\,\mathrm{grad}\,p_l=-k_B\,T\,(\mu_{ph}+\mu_{pl})\,\mathrm{grad}\,p, \tag{19.115}$$

where

$$\mu_{ph}=\frac{q\,\tau_{ph}}{m_{hh}}\,\frac{p_h}{p},\qquad \mu_{pl}=\frac{q\,\tau_{pl}}{m_{hl}}\,\frac{p_l}{p}. \tag{19.116}$$

An approximate expression of μ_{ph}, μ_{ph} is obtained by replacing the concentrations with the corresponding equilibrium values $p_h = N_{Vh}\,\Phi_{1/2}(\xi_h)$, $p_l = N_{Vl}\,\Phi_{1/2}(\xi_h)$, with

$$N_{Vh}=2\,M_V\left(\frac{m_{hh}}{2\,\pi\,\hbar^2}\,k_B\,T\right)^{3/2},\qquad N_{Vl}=2\,M_V\left(\frac{m_{hl}}{2\,\pi\,\hbar^2}\,k_B\,T\right)^{3/2} \tag{19.117}$$

(compare with (18.8)), whence, using $p = p_h + p_l$,

$$\mu_{ph}\simeq\frac{q\,\tau_{ph}\,m_{hh}^{1/2}}{m_{hh}^{3/2}+m_{hl}^{3/2}},\qquad \mu_{pl}\simeq\frac{q\,\tau_{pl}\,m_{hl}^{1/2}}{m_{hh}^{3/2}+m_{hl}^{3/2}}. \tag{19.118}$$

Then, $\mathbf{J}_a$ is extracted from the first relation in (19.113), whose matrix form is

$$\begin{bmatrix} 1 & \mu_a\,B_3 & -\mu_a\,B_2 \\ -\mu_a\,B_3 & 1 & \mu_a\,B_1 \\ \mu_a\,B_2 & -\mu_a\,B_1 & 1 \end{bmatrix}\begin{bmatrix} J_{a1} \\ J_{a2} \\ J_{a3} \end{bmatrix}=\begin{bmatrix} J'_{a1} \\ J'_{a2} \\ J'_{a3} \end{bmatrix}, \tag{19.119}$$

$\mu_a = q\,\tau_{pa}/m_a$. The determinant of the matrix in (19.119) is $D_M = 1+\mu_a^2\,B^2$. Still considering the case where $\mathbf{B}$ is weak, one finds

$$\mathbf{J}_a\simeq\mathbf{J}'_a-\mu_a\,\mathbf{B}\wedge\mathbf{J}'_a. \tag{19.120}$$

In turn, the contribution of the last term at the right hand side of the above yields

$$-a_p\,\mathbf{B}\wedge\left(q\,\mu_p\,p\,\mathbf{E}-q\,D_p\,\mathrm{grad}\,p\right),\qquad \mu_p=\mu_{ph}+\mu_{pl}, \tag{19.121}$$

with μ_p the *hole mobility*. In turn, the *hole diffusivity* (or *hole-diffusion coefficient*) and the dimensionless parameter a_p are given by

$$D_p=\frac{k_B\,T}{q}\,\mu_p,\qquad a_p=\frac{1}{\mu_p^2}\left(\frac{q\,\tau_{ph}}{m_{hh}}\,\mu_{ph}+\frac{q\,\tau_{pl}}{m_{hl}}\,\mu_{pl}\right). \tag{19.122}$$

Putting (19.114), (19.115), and (19.121) together finally provides the *drift-diffusion transport equation for the holes*,

$$\mathbf{J}_p = q\,\mu_p\,p\,\mathbf{E} - q\,D_p\,\mathrm{grad}\,p - q\,a_p\,\mu_p\,\mathbf{B} \wedge \left(\mu_p\,p\,\mathbf{E} - D_p\,\mathrm{grad}\,p\right). \tag{19.123}$$

Thus, the DD model for the holes of the valence band is given by (19.123) along with the balance equation for the holes' number, that reads

$$\frac{\partial p}{\partial t} + \frac{1}{q}\,\mathrm{div}(\mathbf{J}_p) = W_p. \tag{19.124}$$

19.5.4 Coupling with Maxwell's Equations

As anticipated in Sect. 19.5.1, as the electromagnetic field is influenced by the distribution of charge and current density, it is necessary to couple the equations describing the charge transport (in the form, e.g., of the hydrodynamic or drift-diffusion model) with the Maxwell equations. For this, one inserts the total charge density ρ and current density $\mathbf{J}$ into the right hand sides of (4.19); considering that there are different groups of charges and currents, one uses (4.22), where the charge density is given by (18.53), namely,[33]

$$\rho = q\,(p - n + N), \qquad N = N_D^+ - N_A^-. \tag{19.125}$$

In turn, the current density reads

$$\mathbf{J} = \mathbf{J}_p + \mathbf{J}_n = \rho_p\,\mathbf{v}_p + \rho_n\,\mathbf{v}_n = q\,p\,\mathbf{v}_p - q\,n\,\mathbf{v}_n, \tag{19.126}$$

with $\mathbf{J}_n$ and $\mathbf{J}_p$ given by (19.107) and (19.123), respectively. As noted in Sect. 18.5, the material's permittivity must be used here instead of vacuum's; as a consequence, the relation between electric displacement and field reads $\mathbf{D} = \varepsilon_{\mathrm{sc}}\,\mathbf{E}$.

One notes that the $\mathbf{E}$ and $\mathbf{B}$ fields are the sum of two contributions: the first one derives from the internal charge and current-density distribution as mentioned above, while the second one derives from external sources, e.g., voltage or current generators connected to the device or integrated circuits, or electric and magnetic fields present in the environment. In general, the internal contribution to $\mathbf{B}$ is negligible and is not considered in semiconductor devices or integrated circuit; it follows that $\mathbf{B}$ is to be accounted for in (19.107) and (19.123) only when it derives from external sources[34] and, due to this, it must be thought of as a prescribed function of $\mathbf{r}$ and t.

[33] Equation (18.53) is the definition of charge density in a semiconductor; as a consequence it holds in general, not only in the equilibrium condition considered in Sect. 18.5. In fact, it can readily be extended to account for charges trapped in energy states different from those of the dopants (Sect. 20.2.2).

[34] A typical example is found when a semiconductor device or circuit is used as a magnetic-field sensor or in specific measurement setups, like in the Hall-voltage measurement (Sect. 25.4).

With these premises, the analysis will continue here after letting $\mathbf{B} = 0$. Despite this simplification, the continuity and transport Eqs. (19.108), (19.107) and (19.123), (19.124) must be coupled with the whole set of Maxwell equations: in fact, the expression of the electric field in terms of the potentials is given by the second relation in (4.26), namely, $\mathbf{E} = -\text{grad}\varphi - \partial\mathbf{A}/\partial t$; as a consequence, in a dynamic condition both the scalar and vector potential must be determined. In a steady-state or equilibrium condition, instead, the expression of the electric field reduces to $\mathbf{E} = -\text{grad}\varphi$; in this case it is sufficient to couple the semiconductor model with the first equation in (4.19) only.

The presence of the vector potential $\mathbf{A}$ makes the model more complicate; thus, it is useful to ascertain whether, in the typical operating conditions, the time derivative $\partial\mathbf{A}/\partial t$ in the expression of $\mathbf{E}$ should be kept or not. As noted above, the derivative differs from zero only if the boundary conditions (e.g., the applied voltages) vary with time. To associate a characteristic time to a boundary condition one takes the period associated to the maximum frequency of the boundary condition's spectrum, $\tau_{\min} = 1/\nu_{\max}$; then, one compares $\tau_{\min}$ with the time Δt necessary for the electromagnetic perturbation produced by the boundary condition to propagate to a position internal to the semiconductor. If d is the distance between a point on the boundary and an internal point, the propagation time can be estimated to be $\Delta t = d/u_f$, with u_f the radiation's phase velocity corresponding to $\nu_{\max}$. If it happens that $\Delta t \ll \tau_{\min}$, the propagation is practically instantaneous, namely, the electromagnetic field at the internal point is consistent with the boundary condition existing at the same instant of time; as a consequence, the boundary condition is thought of as stationary, and the $\partial\mathbf{A}/\partial t$ derivative is neglected. This is called *quasi-static approximation*; the condition of its applicability is summarized as[35]

$$\Delta t = \frac{d}{u_f} \ll \tau_{\min} = \frac{1}{\nu_{\max}}, \qquad \nu_{\max} \ll \frac{u_f}{d}. \tag{19.127}$$

To estimate the condition one must fix the value of d; as a conservative choice one takes the channel length of the MOSFET transistors of the old generations, $d \approx 10^{-7}$ m. Using $u_f \approx 10^8$ m s^{-1} yields $\nu_{\max} \ll 10^{15}$ Hz, which is amply fulfilled in the present state-of-the-art integrated circuits. Note that the condition is even better verified in the last-generation devices, whose channel length is shorter than 10^{-7} m.

The choice of the channel length in the above estimate is dictated by the fact that the channel is the active region of the device. Choosing, instead, d as the (much larger) thickness of the silicon wafer would not make sense, because the phenomena taking place in the wafer's bulk are relatively unimportant. Other distances within an integrated circuit are larger by orders of magnitude than the value of d considered in the estimate; for instance, the diameter of the integrated circuit itself is of the order of 10^{-2} m, hence the quasi-static approximation is not applicable to the case of two devices placed, e.g., at opposite corners of a chip and connected by a line. In fact,

[35] A similar reasoning is used to treat the time derivative of v_i, w, and b_i in the derivation of the BTE's moments of order larger than zero (Sect. 19.4.2).

the propagation of signals along the lines connecting different devices on a chip is modeled using the whole set of Maxwell equations.[36]

19.5.5 Semiconductor-Device Model

Thanks to the quasi-static approximation, the equations describing the semiconductor, in the drift-diffusion case and with $\mathbf{B} = 0$, read

$$\mathrm{div}\mathbf{D} = q\,(p - n + N), \qquad \mathbf{D} = -\varepsilon_{\mathrm{sc}}\,\mathrm{grad}\varphi, \tag{19.128}$$

$$\frac{\partial n}{\partial t} - \frac{1}{q}\,\mathrm{div}\mathbf{J}_n = W_n, \qquad \mathbf{J}_n = q\,\mu_n\,n\,\mathbf{E} + q\,D_n\,\mathrm{grad}n, \tag{19.129}$$

$$\frac{\partial p}{\partial t} + \frac{1}{q}\,\mathrm{div}\mathbf{J}_p = W_p, \qquad \mathbf{J}_p = q\,\mu_p\,p\,\mathbf{E} - q\,D_p\,\mathrm{grad}p. \tag{19.130}$$

As outlined in Chap. 24, insulating layers play an essential role in the fabrication of integrated circuits; it is then necessary to extend the model to incorporate also the description of such layers. This is easily accomplished by observing that mobile charges are absent in an insulator, so that the balance equations for the number of particles and the transport equations reduce to identities, $0 = 0$. The model for the insulators then reduces to Poisson's equation only. The right hand side of the latter does not necessarily vanish because, as indicated in Sect. 24.1, contaminants may enter the insulator during the fabrication steps; some of these contaminants may ionize and act as fixed charges so that, letting N_{ox} be their density, the model for the insulator reads[37]

$$\mathrm{div}\mathbf{D} = q\,N_{\mathrm{ox}}, \qquad \mathbf{D} = -\varepsilon_{\mathrm{ox}}\,\mathrm{grad}\varphi, \qquad n = p = 0, \qquad \mathbf{J}_n = \mathbf{J}_p = 0. \tag{19.131}$$

The set of Eqs. (19.128–19.131) is commonly called *semiconductor-device model*. It is made of partial-differential equations of the first order, in the unknowns φ, n, p and $\mathbf{D}$, $\mathbf{J}_n$, $\mathbf{J}_p$. The coefficients are μ_n, μ_p, $D_n = (k_B\,T/q)\,\mu_n$, $D_p = (k_B\,T/q)\,\mu_p$. In turn, N, W_n, W_p are either known functions, or are expressed in terms of the unknowns themselves. Some of the equations are non linear because the unknowns are multiplied by each other. Each equation on the left in (19.128–19.131) contains the divergence of a vector; in turn, the expression of the vector is given by the corresponding equation on the right, in terms of the scalar unknowns (in fact it is $\mathbf{E} = -\mathrm{grad}\varphi$). It follows that, by introducing the expressions of $\mathbf{D}$, $\mathbf{J}_n$, and $\mathbf{J}_p$

[36] The progressive device scaling from one generation to the next is in general associated to an increase in the size of the chips. Due to this, the constraints on the circuit's speed are rather imposed by the lines connecting the devices than by the devices themselves.

[37] The insulator's permittivity is indicated with $\varepsilon_{\mathrm{ox}}$ because, in the examples shown later, silicon dioxide (SiO_2) is used as the reference insulator.

into the divergence operator, each pair of first-order equation is transformed into a single, second-order equation. This observation is useful in view of the application of numerical methods to the solution of the semiconductor-device model.

Remembering the derivation of the transport model, the terms W_n, W_p in (19.129), (19.130) are due to the generation-recombination phenomena. Specifically, W_n is the difference between the number of electrons entering the conduction band, and of those leaving it, per unit volume and time; in turn, W_p is the difference between the number of holes entering the valence band, and of those leaving it, per unit volume and time.[38] For this reason, they are also called *net generation rates*.

As mentioned in Sect. 19.3, the transitions of a given class are further grouped depending on the entity with which the particle's collision occurs. As far as the net generation rates are concerned, it is customary to separate the contribution of the phonon collisions from those of the other types (e.g., electron–electron collisions, electron–photon collisions, and so on); in fact, unless the device is kept at a very low temperature, the phonon collisions are the most important ones. Thus, the net generation rates are recast as

$$W_n = G_n - U_n, \qquad W_p = G_p - U_p, \tag{19.132}$$

where U_n, U_p describe the transitions due to phonon collisions, while G_n, G_p describe those of the other types. The minus signs in (19.132) come from the fact that U_n is defined as the difference between the number of electrons leaving the conduction band, and of those entering it, because of phonon collisions, per unit volume and time; similarly, U_p is defined as the difference between the number of holes leaving the valence band, and of those entering it, because of phonon collisions, per unit volume and time. The terms used for U_n, U_p are *net thermal recombination rates*, those for G_n, G_p are *net non-thermal generation rates*.

Another comment about the semiconductor-device model concerns the drift terms in (19.129), (19.130). The latter can be recast as $\mathbf{J}_n^{\mathrm{dr}} = \sigma_n \mathbf{E}$ and $\mathbf{J}_p^{\mathrm{dr}} = \sigma_p \mathbf{E}$, where

$$\sigma_n = q\,\mu_n\,n, \qquad \sigma_p = q\,\mu_p\,p \tag{19.133}$$

are the *electron conductivity* and *hole conductivity*, respectively. From $\mathbf{J} = \mathbf{J}_n + \mathbf{J}_p$, in a uniform material one obtains $\mathbf{J} = \sigma\,\mathbf{E}$, that is, *Ohm's law*, with

$$\sigma = \sigma_n + \sigma_p = q\,\left(\mu_n\,n + \mu_p\,p\right). \tag{19.134}$$

19.5.6 Boundary Conditions

In practical applications, the equations of the semiconductor-device model, (19.128) through (19.131), are solved over a closed domain whose boundary[39] is indicated

[38] The units are $[W_n, W_p] = \mathrm{m}^{-3}\ \mathrm{s}^{-1}$.

[39] A two- or three-dimensional case is considered. In the one-dimensional case the boundary reduces to the two points enclosing the segments over which the equations are to be solved.

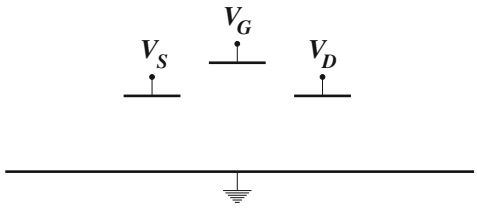

Fig. 19.3 MOS structure used to discuss the boundary conditions for the mathematical model of semiconductor devices. Only the conducting boundaries are shown. Note that the vertical scale of the drawing is not realistic

here with Γ. The boundary is partitioned into portions, some of which, indicated with $\Gamma_{i1}, \Gamma_{i2}, \ldots$, are *insulating boundaries*, namely, they can not be crossed by electrons or holes; the remaining portions, $\Gamma_{c1}, \Gamma_{c2}, \ldots$, can be crossed by the carriers and are termed *conducting boundaries*.

Considering that the domain over which the equations are solved is in general a part of a much larger domain enclosing an integrated circuit, some flexibility exists as for the choice of Γ. Thanks to this, it is possible to select $\Gamma_{i1}, \Gamma_{i2}, \ldots$ such that the normal component of the vector unknowns vanishes there; in other terms, letting $\mathbf{s}$ be the unit vector normal to an insulating boundary at some point $\mathbf{r}$, it is

$$\mathbf{E} \cdot \mathbf{s} = 0, \qquad \frac{\partial \varphi}{\partial s} = 0, \qquad \mathbf{r} \in \Gamma_{i1}, \Gamma_{i2}, \ldots . \tag{19.135}$$

where $\partial/\partial s$ indicates the derivative normal to the boundary at $\mathbf{r}$. An example of how this can be accomplished is given in Figs. 19.3 and 19.4, representing the schematic cross-section of a MOSFET. In Fig. 19.3 only the conducting boundaries are shown, with V_S, V_G, and V_D indicating the voltage applied to the source, gate, and drain contact, respectively. The bulk contact is grounded, as shown by the line below; such a contact is the ground reference for the whole chip and, for this reason, extends laterally beyond the region occupied by the MOSFET under consideration. The insulating boundaries must now be selected in order to form a closed line that completes the boundary Γ; in principle, such a completion may be accomplished in different ways. Consider, however, the completion shown in the upper part of Fig. 19.4: as in general it is $V_S \neq V_G$, it is likely that one of the field lines in the region between the source and gate contacts coincides with the segment ab. As a consequence, choosing ab as the insulating boundary in that region guarantees that the component of $\mathbf{E}$ normal to such a boundary is zero, thus achieving the condition sought. The same reasoning applies to line cd. By this procedure one succeeds in prescribing the boundary conditions for the Poisson equation along the insulating boundaries. If, instead, the insulating boundaries were placed differently, like, e.g., in the lower part of Fig. 19.4, the component of $\mathbf{E}$ normal to the boundary would be different from zero; besides that, it would be impossible to determine it *a priori*, and Poisson's equation would become ill-posed.

Once the insulating boundaries are completed, the same condition as (19.135) is prescribed onto the current densities, namely, $\mathbf{J}_n \cdot \mathbf{s} = \mathbf{J}_p \cdot \mathbf{s} = 0$ whence, using the second equation in (19.129),

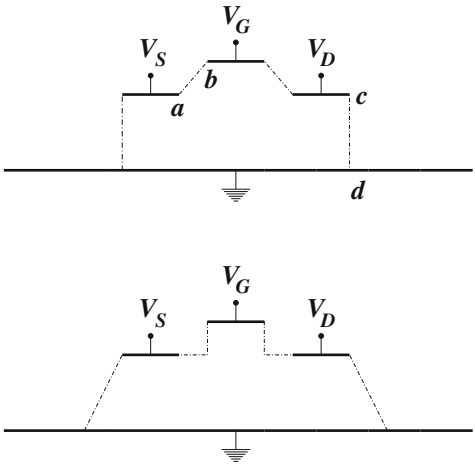

Fig. 19.4 The same structure as in Fig. 19.3, to which the insulating boundaries have been added (dash-dotted lines). The upper part of the figure shows the correct placement of the insulating boundaries, the lower part shows a wrong placement

$$-q\,\mu_n\,n\frac{\partial\varphi}{\partial s}+q\,D_n\frac{\partial n}{\partial s}=0,\qquad \mathbf{r}\in\Gamma_{i1},\Gamma_{i2},\ldots\,. \tag{19.136}$$

Combining the above with (19.135) yields $\partial n/\partial s = 0$; repeating the calculation for the holes finally shows that at the insulating boundaries the boundary condition is the same for all scalar unknowns φ, n, p:

$$\frac{\partial\varphi}{\partial s}=0,\qquad \frac{\partial n}{\partial s}=0,\qquad \frac{\partial p}{\partial s}=0,\qquad \mathbf{r}\in\Gamma_{i1},\Gamma_{i2},\ldots\,, \tag{19.137}$$

i.e., a boundary condition of the *homogeneous Neumann type*.

The conducting boundaries $\Gamma_{c1},\Gamma_{c2},\ldots$ are typically covered with metal layers or heavily-doped polycrystalline layers that provide the electric contacts to the device. Unless the operating condition of the device departs strongly from equilibrium, a contact is able to supply the amount of charge necessary to keep the equilibrium and charge-neutrality conditions in the semiconductor layer adjacent to it. Thus at each point $\mathbf{r}$ of this layer one takes $\rho = 0$ or, more specifically,

$$\rho_c = q\,(p_c - n_c + N_c) = 0,\qquad \mathbf{r}\in\Gamma_{c1},\Gamma_{c2},\ldots\,, \tag{19.138}$$

where index c indicates the conducting boundary. In most cases the metal or polycrystalline layer is connected to a voltage generator, so that the electric potential is prescribed, or to another part of the integrated circuit, so that the electric potential can be determined from a separate calculation. In these cases, the electric potential of the contact is known. From it, one derives the electric potential φ_c of the conducting boundary adjacent to the contact; in fact, when the departure from the equilibrium condition is not too strong, the difference between the electric potential of the conducting boundary and that of the contact does not depend on the current density that crosses the boundary, and equals the contact's work function.[40] The latter is experimentally known, this yielding φ_c. In conclusion, at a conducting boundary where the

[40] Examples of application of this concept are given in Sects. 21.2.2 and 22.2.

voltage is prescribed, the boundary condition is the same for all scalar unknowns: $\varphi = \varphi_c, n = n_c, p = p_c, \mathbf{r} \in \Gamma_{c1}, \Gamma_{c2}, \ldots$, i.e., a boundary condition of the *Dirichlet type*. Note that the quantities in parenthesis in (19.138) depend on φ_c (compare with (18.56) and (18.57)); as a consequence, they can be calculated only after the electric potential has been determined.

In some instances a current generator is connected to a contact; as a consequence, the voltage is not prescribed at the corresponding conducting boundary. However, such a voltage can be determined by observing that the flux of the current density across the boundary equals the generator's current. This provides the extra relation that keeps the well-posedness of the mathematical problem.[41]

19.5.7 Quasi-Fermi Potentials

The drift-diffusion transport equations, given by the second relation in (19.129), (19.130), can be recast in a monomial form by defining two auxiliary functions

$$\varphi_n(\mathbf{r}, t) = \varphi - \frac{k_B\, T}{q}\, \log\left(\frac{n}{n_i}\right), \qquad \varphi_p(\mathbf{r}, t) = \varphi + \frac{k_B\, T}{q}\, \log\left(\frac{p}{n_i}\right), \tag{19.139}$$

whose inversion yields

$$n = n_i\, \exp\left[\frac{q\,(\varphi - \varphi_n)}{k_B\, T}\right], \qquad p = n_i\, \exp\left[\frac{q\,(\varphi_p - \varphi)}{k_B\, T}\right]. \tag{19.140}$$

In the equilibrium limit, (19.140) must coincide with (18.62), namely, $\varphi_n \to \varphi_F$, $\varphi_p \to \varphi_F$. It follows that the auxiliary functions (19.139) are a formal generalization of the concept of Fermi potential; they have the advantage of keeping the exponential form of the expressions of n and p in the non-equilibrium case. For this reason, φ_n and φ_p are called *quasi-Fermi potentials* for electrons and holes, respectively.[42] From (19.140) one finds $(k_B\, T/q)\, \mathrm{grad} n = n\, \mathrm{grad}(\varphi - \varphi_n)$ and $(k_B\, T/q)\, \mathrm{grad}\, p = p\, \mathrm{grad}(\varphi_p - \varphi)$ which, replaced into the second relation of (19.129), (19.130), respectively, yield

$$\mathbf{J}_n = -q\, \mu_n\, n\, \mathrm{grad}\varphi + q\, \frac{k_B T_L}{q}\, \mu_n\, \mathrm{grad} n = -q\, \mu_n\, n\, \mathrm{grad}\varphi_n, \tag{19.141}$$

$$\mathbf{J}_p = -q\, \mu_p\, p\, \mathrm{grad}\varphi - q\, \frac{k_B\, T}{q}\, \mu_p\, \mathrm{grad}\, p = -q\, \mu_p\, p\, \mathrm{grad}\varphi_p. \tag{19.142}$$

---

[41] This outcome becomes immediately clear by applying a numerical-discretization method to the problem. In fact, the component of the current density normal to the contact depends on the electric potential of the contact itself; thus, the extra relation provided by the flux-conservation equation embeds the extra unknown φ_c.

[42] By some authors, φ_n and φ_p are called *Imref potentials*, where "Imref" is "Fermi" read from right to left [103].

One notes that the monomial forms (19.141), (19.142) thus achieved are similar to drift-diffusion equations where the drift term only is present. This result allows one to interpret $-\text{grad}\varphi_n$ and $-\text{grad}\varphi_p$ as effective fields acting on the electrons (or, respectively, holes) and incorporating both drift and diffusion effects. The monomial form is useful for describing unipolar devices, where one of the current densities $\mathbf{J}_n$, $\mathbf{J}_p$ dominates over the other and is essentially solenoidal (Sect. 22.6).

The definition of the quasi-Fermi potential given above is applicable only to drift-diffusion equations of the form (19.129), (19.130), that are valid within the approximations of parabolic bands and non-degenerate conditions. However, the concept of quasi-Fermi potential can be generalized by disposing, e.g., of the non-degeneracy hypothesis. For this, one starts from the equilibrium expressions of n and p, given by (18.56) and (18.57), respectively, and replaces E_F in the definitions of ξ_e, ξ_h with $E_F = E_{Fi} - q\,\varphi_F$ (compare with (18.26)); then, the Fermi potential φ_F is replaced with φ_n in the definition of ξ_e, and with φ_p in that of ξ_h. The non-equilibrium concentrations then read

$$n = N_C\,\Phi_{1/2}\,(\xi_e)\,, \qquad \xi_e = -\frac{E_C - E_{Fi}}{k_B\,T} + \frac{q\,(\varphi - \varphi_n)}{k_B\,T}, \tag{19.143}$$

and

$$p = N_V\,\Phi_{1/2}\,(\xi_h)\,, \qquad \xi_h = -\frac{E_{Fi} - E_V}{k_B\,T} + \frac{q\,(\varphi_p - \varphi)}{k_B\,T}. \tag{19.144}$$

19.5.8 Poisson Equation in a Semiconductor

In the equilibrium condition the concentrations of electrons (19.143) and holes (19.144) depend on the electric potential only. In turn, the ionized donor and acceptor concentrations depend on the electric potential and, possibly, on position if the dopant distributions are position dependent. In summary, the general form of the equilibrium charge concentration is $\rho = \rho(\varphi, \mathbf{r})$. The semiconductor-device model reduces to the Poisson equation alone because at equilibrium it is $\partial/\partial t = 0$, $\mathbf{J}_n = \mathbf{J}_p = 0$; the equation reads $-\varepsilon_{sc}\,\nabla^2\varphi = q\,(p - n + N) = \varrho(\varphi, \mathbf{r})$, which is a semi-linear partial differential equation (PDE).[43] If the explicit dependence on $\mathbf{r}$ is absent, $\rho = \rho(\varphi)$, and the problem is one dimensional, say, in the x direction, Poisson's equation can be solved analytically over a domain I where $\mathrm{d}\varphi/\mathrm{d}x \neq 0$. In

[43] A PDE of order s in the unknown φ is called *quasi-linear* if it is linear in the order-s derivatives of φ and its coefficients depend on the independent variables and the derivatives of φ of order $m < s$. A quasi-linear PDE where the coefficients of the order-s derivatives are functions of the independent variables alone, is called *semi-linear*. A PDE which is linear in the unknown function and all its derivatives, with coefficients depending on the independent variables alone, is called *linear*. PDE's not belonging to the classes above are *fully non-linear*.

fact, multiplying by $\mathrm{d}\varphi/\mathrm{d}x$ both sides of $-\varepsilon_{\mathrm{sc}}\,\mathrm{d}^2\varphi/\mathrm{d}x^2 = \rho(\varphi)$, one obtains

$$-\varepsilon_{\mathrm{sc}}\,\frac{\mathrm{d}^2\varphi}{\mathrm{d}x^2}\,\frac{\mathrm{d}\varphi}{\mathrm{d}x} = \rho(\varphi)\,\frac{\mathrm{d}\varphi}{\mathrm{d}x}, \qquad \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}\varphi}{\mathrm{d}x}\right)^2 = \frac{\mathrm{d}}{\mathrm{d}x}S(\varphi), \tag{19.145}$$

where $\mathrm{d}S/\mathrm{d}\varphi = -2\,\rho(\varphi)/\varepsilon_{\mathrm{sc}}$. Integrating (19.145) from x_0 to x, where both points belong to I, and letting $\varphi_0 = \varphi(x_0)$, yields

$$\left(\frac{\mathrm{d}\varphi}{\mathrm{d}x}\right)^2 = G^2(\varphi), \qquad G(\varphi) = \left[\left(\frac{\mathrm{d}\varphi}{\mathrm{d}x}\right)_0^2 + S(\varphi) - S(\varphi_0)\right]^{1/2}. \tag{19.146}$$

Separating the variables in (19.146) then provides

$$\mathrm{d}x = \pm\frac{\mathrm{d}\varphi}{G(\varphi)}, \qquad x = x_0 \pm \int_{\varphi_0}^{\varphi}\frac{\mathrm{d}\varphi}{G(\varphi)}, \tag{19.147}$$

namely, the inverse relation $x = x(\varphi)$. The choice of the sign is made on case-by-case basis (an example is given in Sect. 21.2). The approach can be extended to the non-equilibrium case if it happens that the electric potential depends on one independent variable, say, x, while the Poisson equation contains terms that depend also on independent variables different from x. In fact, the other variables can be considered as parameters during the integration with respect to x. An example of this case is given in Sect. 22.7.1.

19.6 Complements

19.6.1 Comments on the Equivalent Hamiltonian Operator

It has been observed in Sect. 19.2.2 that, in the description of the wave-packet dynamics based on the equivalent Hamiltonian operator and crystal momentum, that the time variations of the latter are due to the external force only; as a consequence, if $U = \text{const}$ one has $\hbar\,\dot{\mathbf{k}}_0 = 0$, namely, the crystal momentum is a constant of motion. The periodic part of the potential energy is incorporated with the kinetic part, to form an equivalent operator. Thus, the expression of the Hamiltonian operator is eventually made of two terms, the equivalent-kinetic part, where the space coordinates do not explicitly appear, and the potential part due to the external force only; from this standpoint it is similar to that of a particle *in vacuo*.

The same results were found in the analysis of the motion of a classical particle subjected to a periodic potential onto which a weak perturbation is superimposed (Sect. 3.11). The classical investigation makes it clear that the crystal momentum is in fact different from the actual momentum of the particle; also, the subsequent elaboration carried out in Sect. 3.12 shows that the concept of effective mass is not distinctive of Quantum Mechanics.

19.6.2 Special Cases of Anisotropy

It is interesting to note that for an anisotropic branch the acceleration $\dot{\mathbf{u}}$ may still be parallel to $\mathbf{F}$; this happens when the force is parallel to one of the coordinate axes. Taking by way of example a force parallel to the first coordinate axis, $\mathbf{F} = F\,\mathbf{i}_1$, it follows in fact

$$\dot{u}_1 = F/m_{1a}, \qquad \dot{u}_2 = 0, \qquad \dot{u}_3 = 0. \tag{19.148}$$

This seems to imply that a suitable choice of the reference is sufficient to make the acceleration parallel to the force. However, this is not so: as noted in Sect. 17.6.2, for (19.24) to hold, the reference in the $\mathbf{k}$ space must be chosen in such a way as to make the Hessian matrix of $E_n(\mathbf{k})$ diagonal; this, in turn, fixes the reference in the $\mathbf{r}$ space, because the two references are reciprocal to each other. As a consequence, if one rotates the $\mathbf{r}$ reference to align one of its axes with the force, the $\mathbf{k}$ reference rotates as well; as the mass tensor in the new reference is not necessarily diagonal, the anisotropy present in the old reference is maintained.

19.6.3 α-Moment at Equilibrium

The derivation of the moment with respect to α of the BTE has been shown in Sect. 19.4.1, leading to (19.62). In the equilibrium condition the distribution function f^{eq} is independent of t and depends on $\mathbf{k}$ through the Hamiltonian function H only; since the latter is even with respect to $\mathbf{k}$, the distribution function is even as well. In turn, as the transitions balance each other, the right hand side of (19.62) vanishes. The term with the magnetic induction vanishes as well (compare with (19.60)); in conclusion, in the equilibrium condition (19.62) reduces to

$$\mathrm{div}_{\mathbf{r}}\,(n\,\overline{\alpha\,\mathbf{u}})^{\mathrm{eq}} + \frac{q}{\hbar}\,\left(n\,\overline{\mathrm{grad}_{\mathbf{k}}\alpha}\cdot\mathbf{E}\right)^{\mathrm{eq}} = 0. \tag{19.149}$$

It is easily found that (19.149) yields the identity $0 = 0$ if α is even with respect to $\mathbf{k}$. This, on the contrary, is not true when α is odd; in this case the equilibrium condition consists in the balance between the diffusion term, due to the spatial non-uniformity of $(n\,\overline{\alpha\,\mathbf{u}})^{\mathrm{eq}}$, and the second term, proportional to the carrier concentration and linearly dependent on the electric field.

19.6.4 Closure Conditions

The closure conditions for the drift-diffusion and hydrodynamic model are derived in this section. The former consists in calculating (19.75) using the equilibrium distribution $f^{\mathrm{eq}} = Q\,P$, with $Q = 1/(4\,\pi^3)$ the density of states in the $\mathbf{r},\mathbf{k}$ space

and P the Fermi–Dirac statistics. In the equilibrium case it is $\mathbf{v} = 0$, whence $\mathbf{c} = \mathbf{u}$, and

$$n^{\mathrm{eq}}\, k_B\, T_{ij}^{\mathrm{eq}} = \frac{m_{ia}}{4\,\pi^3} \iiint_{-\infty}^{+\infty} \frac{u_i\, u_j}{\exp\left[(E_e + \zeta_e)/(k_B\, T)\right] + 1}\, \mathrm{d}^3 k, \tag{19.150}$$

with $E_e = E - E_C$ and $\zeta_e = E_C - q\,\varphi - E_F$ (compare with (18.54)). In the above, E_e is even with respect to all components of $\mathbf{k}$. For $j \neq i$ the integrand is odd with respect to k_i because $u_i = (1/\hbar)\,\partial E_e/\partial k_i$, and with respect to k_j because $u_j = (1/\hbar)\,\partial E_e/\partial k_j$; as a consequence it is $T_{ij} = 0$ for $j \neq i$, while

$$n^{\mathrm{eq}}\, k_B\, T_{ii}^{\mathrm{eq}} = \frac{m_{ia}}{4\,\pi^3} \iiint_{-\infty}^{+\infty} \frac{u_i^2}{\exp\left[(E_e + \zeta_e)/(k_B\, T)\right] + 1}\, \mathrm{d}^3 k, \tag{19.151}$$

namely, in the equilibrium condition the electron-temperature tensor is diagonal. For this result to hold, the parabolic-band approximation is not necessary.

In the parabolic-band approximation, (19.24) and (19.72) hold, and the temperature tensor in equilibrium is evaluated at the ath minimum of the conduction band by letting

$$\eta_i = \frac{\hbar\,\delta k_i}{\sqrt{2\,m_{ia}}}, \qquad \mathrm{d}^3 k = 2\,\frac{\sqrt{2}}{\hbar^3}\, m_{ea}^{3/2}\, \mathrm{d}^3 \eta, \qquad m_{ia}\, u_i^2 = 2\,\eta_i^2, \tag{19.152}$$

where the first relation is the Herring-Vogt transformation (17.66) and m_{ea} is defined in (17.68). Adding up over the minima and using (17.72) yields

$$n^{\mathrm{eq}}\, k_B\, T_{ii}^{\mathrm{eq}} = \frac{\sqrt{2}}{\pi^3\,\hbar^3} \iiint_{-\infty}^{+\infty} \frac{M_C\, m_e^{3/2}\, \eta_i^2}{\exp\left[(\eta^2 + \zeta_e)/(k_B\, T)\right] + 1}\, \mathrm{d}^3 \eta. \tag{19.153}$$

As the value of the integral in (19.153) does not depend on index i, it follows that the diagonal entries T_{ii}^{eq} are equal to each other; as a consequence, the common value of the three integrals can be replaced with $T_e^{\mathrm{eq}} = (T_{11}^{\mathrm{eq}} + T_{22}^{\mathrm{eq}} + T_{33}^{\mathrm{eq}})/3$, namely,

$$n^{\mathrm{eq}}\, k_B\, T_{ii}^{\mathrm{eq}} = \frac{\sqrt{2}}{3\,\pi^3\,\hbar^3} \iiint_{-\infty}^{+\infty} \frac{M_C\, m_e^{3/2}\, \eta^2}{\exp\left[(\eta^2 + \zeta_e)/(k_B T_L)\right] + 1}\, \mathrm{d}^3 \eta. \tag{19.154}$$

Turning to spherical coordinates (B.1) yields $\mathrm{d}^3\eta = \eta^2\,\mathrm{d}\eta\,\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi$, with $\eta^2 = \eta_1^2 + \eta_2^2 + \eta_3^2 = E_e$ and $\mathrm{d}\eta = 1/(2\sqrt{E_e})\,\mathrm{d}E_e$, $\eta^2\,\mathrm{d}\eta = \sqrt{E_e}\,\mathrm{d}E_e/2$. The integral over the angles equals $4\,\pi$ whence, using (C.104) with $\Gamma(1 + 3/2) = (3/2)\,\sqrt{\pi}/2$,

$$n^{\mathrm{eq}}\, k_B\, T_{ii}^{\mathrm{eq}} = k_B\, T\, N_C\, \Phi_{3/2}(\xi_e), \tag{19.155}$$

with $N_C = 2\,M_C\,[m_e\, k_B\, T/(2\,\pi\,\hbar^2)]^{3/2}$ the effective density of states (18.4), and $\xi_e = -\zeta_e/(k_B T)$. On the other hand, from (18.17) it is $n^{\mathrm{eq}} = N_C\,\Phi_{1/2}(\xi_e)$, whence

$$T_e^{\mathrm{eq}}(\mathbf{r}) = T\,\frac{\Phi_{3/2}(\xi_e)}{\Phi_{1/2}(\xi_e)}. \tag{19.156}$$

The dependence on position is due to $\xi_e = (q\,\varphi - E_C + E_F)/(k_B\,T)$. However, in the non-degenerate case the approximation (C.105) holds, $\Phi_\alpha(\xi_e) = \exp(\xi_e)$, and the electron-temperature tensor at equilibrium reduces to $T_e^{\rm eq} = T$.

Coming now to the hydrodynamic model, and remembering (19.83), the closure condition is found by calculating the equilibrium value of $n\,m_{ia}\,\overline{E_e\,u_i\,\mathbf{c}}$, that reads

$$n^{\rm eq}\,m_{ia}\,\left(\overline{E_e\,u_i\,\mathbf{c}}\right)^{\rm eq} = \frac{m_{ia}}{4\,\pi^3}\iiint_{-\infty}^{+\infty} E_e\,u_i\,\mathbf{c}\,P\,\mathrm{d}^3k, \tag{19.157}$$

with P the Fermi–Dirac statistics. The procedure is similar to that of the drift-diffusion model. At equilibrium one can replace $\mathbf{c}$ with $\mathbf{u}$; then, out of the three components of $\mathbf{u}$, only that of index i contributes to the integral, because those of index $j \neq i$ make the integrand odd with respect to both k_i and k_j. In other terms, the integrand in (19.157) is replaced with $E_e\,u_i^2\,P\,\mathbf{i}_i$, which differs from the integrand of (19.151) because of factor $E_e\mathbf{i}_i$; this shows that the tensor defined by (19.157) is diagonal as well.[44] When the transformation (19.152) is used, the right hand side of (19.157) becomes similar to (19.153), the only difference being the additional factor $\eta^2\,\mathbf{i}$ that derives from $E_e\mathbf{i}_i$. The next step, replacing the common value of the three integrals with one third of their sum, yields an expression similar to (19.154), the only difference being that η^2 is replaced with $\eta^4\,\mathbf{i}$. Transforming into spherical coordinates and inserting (C.104) with $\Gamma(1+5/2) = (5/2)\,\Gamma(1+3/2) = (15/4)\,\sqrt{\pi}/2$ finally yields

$$n^{\rm eq}\,m_{ia}\,\left(\overline{E_e\,u_i\,\mathbf{c}}\right)^{\rm eq} = \mathbf{i}_i\,\frac{5}{2}\,(k_B\,T)^2\,N_C\,\Phi_{5/2}(\xi_e). \tag{19.158}$$

It follows

$$m_{ia}\,\left(\overline{E_e\,u_i\,\mathbf{c}}\right)^{\rm eq} = \mathbf{i}_i\,\frac{5}{2}\,(k_B\,T)^2\,\frac{\Phi_{5/2}(\xi_e)}{\Phi_{1/2}(\xi_e)} \simeq \mathbf{i}_i\,\frac{5}{2}\,(k_B\,T)^2\,, \tag{19.159}$$

where the approximation holds in the non-degenerate case.

19.6.5 Matthiessen's Rule

The effects of the inter-band and intra-band transitions have been separated in Sect. 19.3.2 under the assumption that the two types are uncorrelated. The separation may further be pursued within each class, depending on the entity with which the collision occurs. Here the collisions leading to the intra-band transitions only are considered. With reference to (19.45), and assuming that the intra-band transitions are uncorrelated, one lets $S_{0v} = S_{0v}^{(1)} + S_{0v}^{(2)} + \ldots$, whence

$$\frac{1}{\tau_v} = \frac{1}{\tau_{v1}} + \frac{1}{\tau_{v2}} + \ldots\,, \qquad \frac{1}{\tau_{vj}} = \iiint_{-\infty}^{+\infty} S_{0v}^{(j)}(\mathbf{r},\mathbf{k}\to\mathbf{k}')\,\mathrm{d}^3k'. \tag{19.160}$$

[44] As above, for this result to hold the parabolic-band approximation is not necessary.

This way of combining the relaxation times, also called *Matthiessen's rule*, still holds in the definitions of the macroscopic relaxation times τ_p, τ_w and τ_b (Sect. 19.4.2), and in the definitions (19.104), (19.118) of electron and hole mobilities through τ_p.

19.6.6 Order of Magnitude of Mobility and Conductivity

As shown in Sect. 19.5.2, the electron and hole mobilities are expressed by relations of the form $\mu = q\,\tau_p/m^*$, where τ_p is the momentum-relaxation time and m^* an effective mass. To the purpose of estimating the order of magnitude it is not necessary to distinguish between electrons and holes. Considering the $T = 300$ K case and taking $\tau_p \approx 0.25 \times 10^{-12}$ s, $m^* \approx 0.4 \times 10^{-30}$ kg yields[45]

$$\mu = q\,\frac{\tau_p}{m^\star} \approx 1.60 \times 10^{-19}\ \mathrm{C} \times \frac{0.25 \times 10^{-12}\ \mathrm{s}}{0.4 \times 10^{-30}\ \mathrm{kg}} = 10^3 \quad \frac{\mathrm{cm}^2}{\mathrm{V\,s}}. \tag{19.161}$$

The diffusion coefficient at $T = 300$ K is estimated from the Einstein relation (19.102), $D = (k_B\,T/q)\,\mu$, where

$$\frac{k_B\,T}{q} \approx \frac{1.38 \times 10^{-23}\ (\mathrm{J/K}) \times 300\ \mathrm{K}}{1.60 \times 10^{-19}\ \mathrm{C}} = 26 \times 10^{-3} \quad \mathrm{V}. \tag{19.162}$$

One finds

$$D \approx 26 \times 10^{-3}\ \mathrm{V} \times 10^3\ \frac{\mathrm{cm}^2}{\mathrm{V\,s}} = 26 \quad \frac{\mathrm{cm}^2}{\mathrm{s}}. \tag{19.163}$$

To estimate the conductivity one takes by way of example the expression for electrons from (19.133), $\sigma_n = q\,\mu_n\,n$, where, due to the estimates above, it is $q\,\mu_n \approx 1.60 \times 10^{-19}\ \mathrm{C} \times 10^3\ \mathrm{cm}^2/(\mathrm{V\,s}) = 1.60 \times 10^{-16}\ \mathrm{cm}^2/\Omega$. For silicon at $T = 300$ K it is $n = n_i \approx 10^{10}\ \mathrm{cm}^{-3}$ (Table 18.2);[46] in comparison, when silicon is doped with a uniform donor concentration equal to, say, $N'_D = 10^{16}\ \mathrm{cm}^{-3}$, it is $n \simeq N'_D = 10^{16}\ \mathrm{cm}^{-3}$ (compare with (18.30)). In conclusion,

$$\sigma_n[\text{intrinsic}] \approx 10^{-6}\ (\Omega\,\mathrm{cm})^{-1}, \qquad \sigma_n[N'_D] \approx 1\ (\Omega\,\mathrm{cm})^{-1}. \tag{19.164}$$

As a further comparison, the estimates for an insulator and, respectively, a conductor, are made with $n \simeq 10^4\ \mathrm{cm}^{-3}$ and $n \simeq 10^{22}\ \mathrm{cm}^{-3}$, to find

$$\sigma_n[\text{insulator}] \approx 10^{-12}\ (\Omega\,\mathrm{cm})^{-1}, \qquad \sigma_n[\text{conductor}] \approx 10^{6}\ (\Omega\,\mathrm{cm})^{-1}. \tag{19.165}$$

[45] Mobility is traditionally expressed in $\mathrm{cm}^2/(\mathrm{V\,s})$ instead of $\mathrm{m}^2/(\mathrm{V\,s})$.

[46] The equilibrium concentrations are used in the estimates.

19.6.7 A Resumé of the Transport Model's Derivation

The number of steps that lead to the hydrodynamic or drift-diffusion model for semiconductors is quite large; thus, a brief summary is of use. The starting point is the single-particle Schrödinger equation for an electron in a crystal. To reach this stage of the theory a considerable amount of work has already been spent, necessary to reduce the many-particle problem to a tractable form (Chap. 16). When the external forces are absent, the single-particle equation is recast in a form based on the equivalent Hamiltonian operator, that exploits the periodicity of the lattice (this implies in turn that the nuclei are kept fixed in the equilibrium positions). Finally, the external forces are added, assuming that the external potential energy is a small perturbation; thanks to this hypothesis, it is possible to describe the collisionless motion of a single electron by means of a Hamiltonian function whose canonical variables are the expectation values of the wave packet's position and momentum (Sects. 19.2.1 and 19.2.2).

Basing on the Hamiltonian function thus found, the analysis shifts from the description of the single-particle to the statistical treatment of a system made of a large number of electrons, following the same pattern as in the classical case; this leads to the semiclassical BTE (Sect. 19.3), for which the general form of the collision term is worked out (Sect. 19.3.1). The latter is simplified, first, by considering point-like collisions, then by taking the perturbative form of the collision operator (Sects. 19.3.2 and 19.3.3).

The perturbative form of the BTE is treated with the moments method, that provides a hierarchical set of models, e.g., the drift-diffusion and the hydrodynamic model. Important approximations at this stage are, for all moments of order larger than zero, the neglect of the inter-band transitions, of the time derivatives, and of the convective terms. The models reach the final form thanks to the hypothesis of parabolic bands and the approximation of the relaxation-time tensors with scalar quantities.

Problems

19.1 In the expressions (19.115), (19.118) defining the hole mobility μ_p, assume that $\tau_{ph} \simeq \tau_{pl}$. Letting τ_p be the common value, determine the value of the normalized effective mass $\overline{m}_h/m_0$ to be used in $\mu_p = q\,\tau_p/\overline{m}_h$ for silicon at room temperature. Also, determine the value of parameter a_p in (19.122) in the same conditions.