# Chapter 6. Elements of Kinetics

*This chapter gives a brief introduction to the basic notions of physical kinetics. Its main focus is on the Boltzmann transport equation, especially within the simple relaxation-time approximation (RTA), which allows an approximate but reasonable and simple description of transport phenomena (such as the electric current and thermoelectric effects) in gases, including electron gases in metals and semiconductors.*

## 6.1. The Liouville theorem and the Boltzmann equation

Physical kinetics (not to be confused with "kinematics"!) is the branch of statistical physics that deals with systems out of thermodynamic equilibrium. Major effects addressed by kinetics include:

(i) for *autonomous* systems (those out of external fields): the transient processes (*relaxation*), that lead from an arbitrary initial state of a system to its thermodynamic equilibrium;

(ii) for systems in time-dependent (say, sinusoidal) external fields: the field-induced periodic oscillations of the system's variables; and

(iii) for systems in time-independent ("dc") external fields: dc transport.

In the last case, we are dealing with stationary ($\partial/\partial t = 0$ everywhere), but *non-equilibrium* situations, in which the effect of an external field, continuously driving the system out of equilibrium, is partly balanced by its simultaneous relaxation – the trend back to equilibrium. Perhaps the most important effect of this class is the dc current in conductors and semiconductors,[1] which alone justifies the inclusion of the basic notions of kinetics into any set of core physics courses.

The reader who has reached this point of the notes already has some taste of physical kinetics because the subject of the last part of Chapter 5 *was* the kinetics of a "Brownian particle", i.e. of a "heavy" system interacting with an environment consisting of many "lighter" components. Indeed, the equations discussed in that part – whether the Smoluchowski equation (5.122) or the Fokker-Planck equation (5.149) – are valid if the environment is in thermodynamic equilibrium, but the system of our interest is not necessarily so. As a result, we could use those equations to discuss such non-equilibrium phenomena as the Kramers problem of the metastable state's lifetime.

In contrast, this chapter is devoted to the more traditional subject of kinetics: systems of many *similar* particles – generally, interacting with each other but not too strongly, so the energy of the system still may be partitioned into a sum of single-particle components, with the interparticle interactions considered as a perturbation. Actually, we have already started the job of describing such a system at the beginning of Sec. 5.7. Indeed, in the absence of particle interactions (i.e. when it is unimportant whether the particle of our interest is "light" or "heavy"), the probability current densities in the coordinate and momentum spaces are given, respectively, by Eq. (5.142) and the first form of Eq. (5.143a), so the continuity equation (5.140) takes the form

$$\frac{\partial w}{\partial t} + \nabla_q \cdot (w\dot{\mathbf{q}}) + \nabla_p \cdot (w\dot{\mathbf{p}}) = 0 . \tag{6.1}$$

[1] This topic was briefly addressed in EM Chapter 4, avoiding its aspects related to thermal effects.

If similar particles do *not* interact, this equation for the single-particle probability density $w(\mathbf{q}, \mathbf{p}, t)$ is valid for each of them, and the result of its solution may be used to calculate any ensemble-average characteristic of the system as a whole.

Let us rewrite Eq. (1) in the Cartesian component form,

$$\frac{\partial w}{\partial t}+\sum_j\left[\frac{\partial}{\partial q_j}\left(w\dot{q}_j\right)+\frac{\partial}{\partial p_j}\left(w\dot{p}_j\right)\right]=0, \qquad (6.2)$$

where the index $j$ numbers all degrees of freedom of the particle under consideration, and assume that its motion (perhaps in an external, time-dependent field) may be described by a Hamiltonian function $\mathscr{H}(q_j, p_j, t)$. Plugging into Eq. (2) the Hamiltonian equations of motion:[2]

$$\dot{q}_j=\frac{\partial\mathscr{H}}{\partial p_j}, \qquad \dot{p}_j=-\frac{\partial\mathscr{H}}{\partial q_j}, \qquad (6.3)$$

we get

$$\frac{\partial w}{\partial t}+\sum_j\left[\frac{\partial}{\partial q_j}\left(w\frac{\partial\mathscr{H}}{\partial p_j}\right)-\frac{\partial}{\partial p_j}\left(w\frac{\partial\mathscr{H}}{\partial q_j}\right)\right]=0. \qquad (6.4)$$

After differentiation of both parentheses by parts, the equal mixed terms $w\partial^2\mathscr{H}/\partial q_j\partial p_j$ and $w\partial^2\mathscr{H}/\partial p_j\partial q_j$ cancel, and using Eq. (3) again, we get the so-called *Liouville theorem*[3]

$$\boxed{\frac{\partial w}{\partial t}+\sum_j\left(\frac{\partial w}{\partial q_j}\dot{q}_j+\frac{\partial w}{\partial p_j}\dot{p}_j\right)=0.} \qquad (6.5)$$

Liouville theorem

Since the left-hand side of this equation is just the full derivative of the probability density $w$ considered as a function of the generalized coordinates $q_j(t)$ of a particle, its generalized momenta components $p_j(t)$, and (possibly) time $t$,[4] the Liouville theorem (5) may be represented in a surprisingly simple form:

$$\frac{dw(\mathbf{q},\mathbf{p},t)}{dt}=0. \qquad (6.6)$$

Physically, this means that the elementary probability $dW = wd^3qd^3p$ to find a Hamiltonian particle in a small volume of the coordinate-momentum space [$\mathbf{q}$, $\mathbf{p}$], with its center moving in accordance to the deterministic law (3), does not change with time – see Fig. 1.

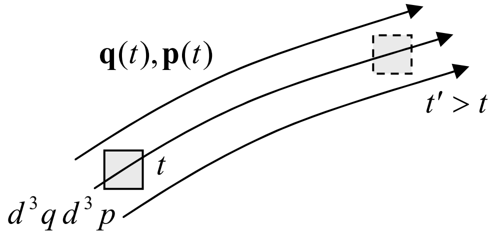


Fig. 6.1. The Liouville theorem's interpretation: probability's conservation at the system's motion flow through the [$\mathbf{q}$, $\mathbf{p}$] space.

[2] See, e.g., CM Sec. 10.1.

[3] Actually, this is just one of several theorems bearing the name of Joseph Liouville (1809-1882).

[4] See, e.g., MA Eq. (4.2).

At first glance, this fact may not look surprising because according to the fundamental Einstein relation (5.78), one needs non-Hamiltonian forces (such as the kinematic friction) to have diffusion. On the other hand, it is striking that the Liouville theorem is valid even for (Hamiltonian) systems with deterministic chaos,[5] in which the deterministic trajectories corresponding to slightly different initial conditions become increasingly mixed with time.

For an ideal gas of 3D particles, we may use the ordinary Cartesian coordinates $r_j$ (with $j = 1, 2, 3$) as the generalized coordinates $q_j$, so $p_j$ become the Cartesian components $mv_j$ of the usual (linear) momentum, and the elementary volume is just $d^3rd^3p$ – see Fig. 1. In this case, Eqs. (3) are just

$$\dot{r}_j = \frac{p_j}{m} \equiv v_j, \quad \dot{p}_j = \mathscr{F}_j, \tag{6.7}$$

where $\mathscr{F}$ is the force exerted on the particle, so the Liouville theorem may be rewritten as

$$\frac{\partial w}{\partial t} + \sum_{j=1}^{3} \left( v_j \frac{\partial w}{\partial r_j} + \mathscr{F}_j \frac{\partial w}{\partial p_j} \right) = 0, \tag{6.8}$$

and conveniently represented in the vector form

$$\frac{\partial w}{\partial t} + \mathbf{v} \cdot \nabla_r w + \boldsymbol{\mathscr{F}} \cdot \nabla_p w = 0. \tag{6.9}$$

Of course, the situation becomes much more complex if the particles interact. Generally, a system of $N$ similar particles in 3D space has to be described by the probability density being a function of $(6N + 1)$ arguments: $3N$ Cartesian coordinates, plus $3N$ momentum components, plus time. An analytical or numerical solution of any equation describing the time evolution of such a function for a typical system of $N \sim 10^{23}$ particles is evidently a hopeless task. Hence, any theory of realistic systems' kinetics has to rely on making reasonable approximations that would simplify the situation.

One of the most useful approximations (sometimes called *Stosszahlansatz* – German for the "collision-number assumption") was suggested by Ludwig Boltzmann for a gas of particles that move freely most of the time but interact during short time intervals, when a particle comes close to either an immobile scattering center (say, an impurity in a conductor's crystal lattice) or to another particle of the gas. Such brief *scattering events* may change the particle's momentum. Boltzmann argued that they may be still approximately described Eq. (9), with the addition of a special term (called the *scattering integral*) to its right-hand side:

Boltzmann transport equation

$$\boxed{\frac{\partial w}{\partial t} + \mathbf{v} \cdot \nabla_r w + \boldsymbol{\mathscr{F}} \cdot \nabla_p w = \left.\frac{\partial w}{\partial t}\right|_{\text{scattering}}.} \tag{6.10}$$

This is the *Boltzmann transport equation*, sometimes called just the "Boltzmann equation" for short. As will be discussed below, it may give a very reasonable description of not only classical but also quantum particles, though it evidently neglects the quantum-mechanical coherence/entanglement effects[6] – besides those that may be hidden inside the scattering integral.

[5] See, e.g., CM Sec. 9.3.

[6] Indeed, the quantum state coherence is described by off-diagonal elements of the density matrix, while the classical probability $w$ represents only the diagonal elements of that matrix. However, at least for the ensembles close to thermal equilibrium, this is a reasonable approximation – see the discussion in Sec. 2.1.

The concrete form of the scattering integral depends on the type of particle scattering. If the scattering centers do not belong to the ensemble under consideration (an example is given, again, by impurity atoms in a conductor), then the scattering integral may be expressed as an evident generalization of the master equation (4.100):

$$\frac{\partial w}{\partial t}\Big|_{\text{scatteering}} = \int d^3p' \left[\Gamma_{\mathbf{p}'\to\mathbf{p}} w(\mathbf{r},\mathbf{p}',t) - \Gamma_{\mathbf{p}\to\mathbf{p}'} w(\mathbf{r},\mathbf{p},t)\right], \tag{6.11}$$

where the physical sense of $\Gamma_{\mathbf{p}\to\mathbf{p}'}$ is the rate (i.e. the probability per unit time) for the particle to be scattered from the state with the momentum **p** into the state with the momentum **p** ’ – see Fig. 2.

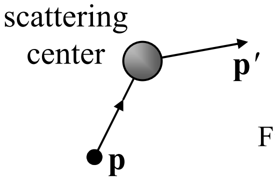


Fig. 6.2. A single-particle scattering event.

Most elastic interactions are *reciprocal*, i.e. obey the following relation (closely related to the reversibility of time in Hamiltonian systems): $\Gamma_{\mathbf{p}\to\mathbf{p}'} = \Gamma_{\mathbf{p}'\to\mathbf{p}}$, so Eq. (11) may be rewritten as[7]

$$\frac{\partial w}{\partial t}\Big|_{\text{scatteering}} = \int d^3p'\ \Gamma_{\mathbf{p}\to\mathbf{p}'} \left[w(\mathbf{r},\mathbf{p}',t) - w(\mathbf{r},\mathbf{p},t)\right]. \tag{6.12}$$

With such scattering integral, Eq. (10) stays linear in *w* but becomes an *integro-differential equation*, typically harder to solve analytically than differential equations.

The equation becomes even more complex if the scattering is due to the mutual interaction of the particle members of the system – see Fig. 3.

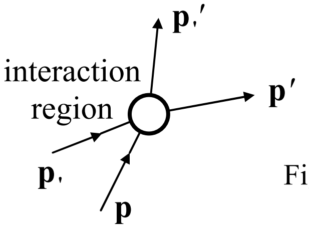


Fig. 6.3. A particle-particle scattering event.

In this case, the probability of a scattering event scales as a product of two single-particle probabilities, and the simplest reasonable form of the scattering integral is[8]

[7] One may wonder whether this approximation may work for Fermi particles, such as electrons, for whom the Pauli principle forbids scattering into the already occupied state, so for the scattering **p** → **p** ’, the term *w*(**r**, **p,** *t*) in Eq. (12) has to be multiplied by the probability [1 – *w*(**r**, **p** ’, *t*)] that the final state is available. This is a valid argument, but one should notice that if this modification has been done with both terms of Eq. (12), it becomes

$$\frac{\partial w}{\partial t}\Big|_{\text{scatteering}} = \int d^3p'\ \Gamma_{\mathbf{p}\to\mathbf{p}'} \left\{w(\mathbf{r},\mathbf{p}',t)\left[\mathbf{1} - w(\mathbf{r},\mathbf{p},t)\right] - w(\mathbf{r},\mathbf{p},t)\left[\mathbf{1} - w(\mathbf{r},\mathbf{p}',t)\right]\right\}.$$

Opening both square brackets, we see that the probability density products cancel, bringing us back to Eq. (12).

[8] This was the approximation used by L. Boltzmann to prove the famous *H-theorem*, stating that the entropy of the gas described by Eq. (13) may only grow (or stay constant) in time, *dS*/*dt* ≥ 0. Since the model is very approximate, that result does not seem too fundamental nowadays, despite all its historic significance.

$$\left.\frac{\partial w}{\partial t}\right|_{\text{scatteering}} = \int d^3 p' \int d^3 p_1 \begin{bmatrix} \Gamma_{\mathbf{p}'\to\mathbf{p},\ \mathbf{p}_1'\to\mathbf{p}_1} w(\mathbf{r},\mathbf{p}',t)w(\mathbf{r},\mathbf{p}_1',t) \\ -\Gamma_{\mathbf{p}\to\mathbf{p}',\ \mathbf{p}_1\to\mathbf{p}_1'} w(\mathbf{r},\mathbf{p},t)w(\mathbf{r},\mathbf{p}_1,t) \end{bmatrix}. \qquad (6.13)$$

The integration dimensionality in Eq. (13) takes into account the fact that due to the conservation of the total momentum at scattering,

$$\mathbf{p} + \mathbf{p}_1 = \mathbf{p}' + \mathbf{p}_1', \qquad (6.14)$$

one of the momenta is not an independent argument, so the integration in Eq. (13) may be restricted to a 6D *p*-space rather than the 9D one. For the reciprocal interaction, Eq. (13) may also be a bit simplified, but it still keeps Eq. (10) a *nonlinear* integro-differential transport equation, excluding such powerful solution methods as the Fourier expansion – which hinges on the linear superposition principle.

This is why most useful results based on the Boltzmann transport equation depend on its further simplifications, most notably the *relaxation-time approximation* – RTA for short.[9] This approximation is based on the fact that in the absence of spatial gradients ($\nabla = 0$), and external forces ($\mathscr{F} = 0$), in the thermal equilibrium, Eq. (10) yields

$$\frac{\partial w}{\partial t} = \left.\frac{\partial w}{\partial t}\right|_{\text{scattering}}, \qquad (6.15)$$

so the equilibrium probability distribution $w_0(\mathbf{r}, \mathbf{p}, t)$ has to turn any scattering integral to zero. Hence at a *small* deviation from the equilibrium,

$$\tilde{w}(\mathbf{r},\mathbf{p},t) \equiv w(\mathbf{r},\mathbf{p},t) - w_0(\mathbf{r},\mathbf{p},t) \to 0, \qquad (6.16)$$

the scattering integral should be proportional to the deviation $\tilde{w}$, and its simplest reasonable model is

Relaxation-time approximation (RTA)

$$\boxed{\left.\frac{\partial w}{\partial t}\right|_{\text{scatteering}} = -\frac{\tilde{w}}{\tau},} \qquad (6.17)$$

where $\tau$ is a phenomenological constant (which, according to Eq. (15), has to be positive for the system's stability) called the *relaxation time*. Its physical meaning will be more clear in the next section.

The relaxation-time approximation is quite reasonable if the angular distribution of the scattering rate is dominated by small angles between vectors **p** and **p***'* – as it is, for example, for the Rutherford scattering by a Coulomb center.[10] Indeed, in this case the two values of the function *w* participating in Eq. (12) are close to each other for most scattering events, so the loss of the second momentum argument (**p***'*) is not too essential. However, using the Boltzmann-RTA equation that results from combining Eqs. (10) and (17),

Boltzmann-RTA equation

$$\boxed{\frac{\partial w}{\partial t} + \mathbf{v}\cdot\nabla_r w + \mathscr{F}\cdot\nabla_p w = -\frac{\tilde{w}}{\tau},} \qquad (6.18)$$

we should always remember that this is just a phenomenological model, sometimes giving completely wrong results. For example, it prescribes the same time scale ($\tau$) to the relaxation of the net *momentum*

[9] Sometimes this approximation is called the "BGK model", after P. Bhatnager, E. Gross, and M. Krook who suggested it in 1954. (The same year, a similar model was considered by P. Welander.)

[10] See, e.g., CM Sec. 3.7.

of the system, and to its *energy* relaxation, while in many real systems, the latter process (that results from inelastic collisions) may be substantially longer. Naturally, in the following sections, I will describe only those applications of the Boltzmann-RTA equation that give a reasonable description of physical reality.

## 6.2. The Ohm law and the Drude formula

Despite its shortcomings, Eq. (18) is adequate for quite a few applications. Perhaps the most important of them is deriving the Ohm law for dc current in a "nearly-ideal" gas of charged particles, whose only important deviation from ideality is the rare scattering effects described by Eq. (17). As a result, in equilibrium it is described by the stationary probability $w_0$ of an ideal gas (see Sec. 3.1):

$$w_0(\mathbf{r},\mathbf{p},t) = \frac{g}{(2\pi\hbar)^3}\langle N(\varepsilon)\rangle, \tag{6.19}$$

where $g$ is the internal degeneracy factor (say, $g = 2$ for electrons due to their spin), and $\langle N(\varepsilon)\rangle$ is the average occupancy of a quantum state with momentum $\mathbf{p}$, that obeys either the Fermi-Dirac or the Bose-Einstein distribution:

$$\langle N(\varepsilon)\rangle = \frac{1}{\exp\{(\varepsilon-\mu)/T\}\pm 1}, \qquad \varepsilon = \varepsilon(\mathbf{p}). \tag{6.20}$$

(The following calculations will be valid, up to a point, for both statistics and hence, in the limit $\mu/T \to -\infty$, for a classical gas as well.)

Now let a uniform dc electric field $\mathscr{E}$ be applied to a uniform gas of similar particles with electric charge $q$, exerting the force $\mathscr{F} = q\mathscr{E}$ on each of them. Then the stationary solution of Eq. (18), with $\partial/\partial t = 0$, should also be stationary and spatially uniform ($\nabla_r = 0$), so this equation is reduced to

$$q\mathscr{E}\cdot\nabla_p w = -\frac{\tilde{w}}{\tau}. \tag{6.21}$$

Let us require the electric field to be relatively low, so that the perturbation $\tilde{w}$ it produces is relatively small, as required by our basic assumption (16).[11] Then on the left-hand side of Eq. (21), we can neglect that perturbation, by replacing $w$ with $w_0$, because that side already has a small factor ($\mathscr{E}$). As a result, this equation yields

$$\tilde{w} = -\tau q\mathscr{E}\cdot\nabla_p w_0 \equiv -\tau q\mathscr{E}\cdot(\nabla_p\varepsilon)\frac{\partial w_0}{\partial\varepsilon}, \tag{6.22}$$

where the second step implies isotropy of the parameters $\mu$ and $T$, i.e. their independence of the direction of the particle's momentum $\mathbf{p}$. But the gradient $\nabla_p\varepsilon$ is nothing else than the particle's velocity

[11] Since the scale of the fastest change of $w_0$ in the momentum space is of the order of $\partial w_0/\partial p = (\partial w_0/\partial\varepsilon)(d\varepsilon/dp) \sim (1/T)v$, where $v$ is the particle speed scale, the necessary condition of the linear approximation (22) is $e\mathscr{E}\tau << T/v$, i.e. if $e\mathscr{E}l << T$, where $l \equiv v\tau$ has the meaning of the effective mean free path. Since the left-hand side of the last inequality is just the average energy given to the particle by the electric field between two scattering events, the condition may be interpreted as the smallness of the gas' "overheating" by the applied field. However, another condition is also necessary – see the last paragraph of this section.

**v** – for a quantum particle, its group velocity.[12] (This fact is easy to verify for the isotropic and parabolic dispersion law, pertinent to classical particles moving in free space,

$$\varepsilon(\mathbf{p}) = \frac{p^2}{2m} \equiv \frac{p_1^2 + p_2^2 + p_3^2}{2m}. \qquad (6.23)$$

Indeed, in this case, the $j^{\text{th}}$ Cartesian components of the vector $\nabla_p \varepsilon$ is

$$\left(\nabla_p \varepsilon\right)_j \equiv \frac{\partial \varepsilon}{\partial p_j} = \frac{p_j}{m} = v_j, \qquad (6.24)$$

so $\nabla_p \varepsilon = \mathbf{v}$.) Hence, Eq. (22) may be rewritten as

$$\tilde{w} = -\tau q \mathscr{E} \cdot \mathbf{v} \frac{\partial w_0}{\partial \varepsilon}. \qquad (6.25)$$

Let us use this result to calculate the electric current density **j**. The contribution of each particle to the current density is $q\mathbf{v}$, so the total density is

$$\mathbf{j} = \int q\mathbf{v} w d^3p \equiv q \int \mathbf{v}\left(w_0 + \tilde{w}\right) d^3p. \qquad (6.26)$$

Since in the equilibrium state (with $w = w_0$), the current has to be zero, the integral of the first term in the parentheses has to vanish. For the integral of the second term, plugging in Eq. (25), and then using Eq. (19), we get

Sommerfeld theory's result

$$\boxed{\mathbf{j} = q^2 \tau \int \mathbf{v}(\mathscr{E} \cdot \mathbf{v})\left(-\frac{\partial w_0}{\partial \varepsilon}\right) d^3p = \frac{gq^2\tau}{(2\pi\hbar)^3} \int \mathbf{v}(\mathscr{E} \cdot \mathbf{v})\left[-\frac{\partial \langle N(\varepsilon)\rangle}{\partial \varepsilon}\right] d^2p_\perp dp_{||},} \qquad (6.27)$$

where $d^2p_\perp$ is the elementary area of the constant energy surface in the momentum space, while $dp_{||}$ is the momentum differential's component normal to that surface. The real power of this result[13] is that it is valid even for particles with an arbitrary dispersion law $\varepsilon(\mathbf{p})$ (which may be rather complicated, for example, for particles moving in space-periodic potentials[14]), and gives, in particular, a fair description of conductivity's anisotropy in crystals.

For free particles whose dispersion law is isotropic and parabolic, as in Eq. (23), the constant energy surface is a sphere of radius $p$, so $d^2p_\perp = p^2 d\Omega = p^2 \sin\theta d\theta d\varphi$, while $dp_{||} = dp$. In the spherical coordinates, with the polar axis directed along the electric field vector $\mathscr{E}$, we get $(\mathscr{E}\cdot\mathbf{v}) = \mathscr{E}\, v\cos\theta$. Now separating the vector **v** outside the parentheses into the component $v\cos\theta$ directed along the vector $\mathscr{E}$, and two perpendicular components, $v\sin\theta\cos\varphi$ and $v\sin\theta\sin\varphi$, we see that the integrals of the last two components over the angle $\varphi$ give zero. Hence, as we could expect, in the isotropic case the net current is directed along the electric field and obeys the linear *Ohm law*,

Ohm law

$$\boxed{\mathbf{j} = \sigma \mathscr{E},} \qquad (6.28)$$

[12] See, e.g., QM Sec. 2.1.

[13] It was obtained by Arnold Sommerfeld in 1927.

[14] See, e.g., QM Secs. 2.7, 2.8, and 3.4. (In this case, **p** should be understood as the quasimomentum rather than the genuine momentum.)

with a field-independent, scalar[15] *electric conductivity*

$$\sigma = \frac{gq^2\tau}{(2\pi\hbar)^3}\int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta d\theta \cos^2\theta \int_0^{\infty} p^2 dp\, v^2 \left[-\frac{\partial \langle N(\varepsilon)\rangle}{\partial \varepsilon}\right]. \qquad (6.29)$$

(Note that $\sigma$ is proportional to $q^2$ and hence does not depend on the particle charge sign.[16])

Since $\sin\theta d\theta$ is just $-d(\cos\theta)$, the integral over $\theta$ equals (2/3). The integral over $d\varphi$ is of course just $2\pi$, while that over $p$ may be readily transformed to one over the particle's energy $\varepsilon(\mathbf{p}) = p^2/2m$: $p^2 = 2m\varepsilon$, $v^2 = 2\varepsilon/m$, $p = (2m\varepsilon)^{1/2}$, so $dp = (m/2\varepsilon)^{1/2}d\varepsilon$, and $p^2dpv^2 = (2m\varepsilon)(m/2\varepsilon)^{1/2}d\varepsilon\,(2\varepsilon/m) \equiv (8m\varepsilon^3)^{1/2}d\varepsilon$. As a result, the conductivity equals

$$\sigma = \frac{gq^2\tau}{(2\pi\hbar)^3}\frac{4\pi}{3}\int_0^{\infty}\left(8m\varepsilon^3\right)^{1/2}\left[-\frac{\partial\langle N(\varepsilon)\rangle}{\partial\varepsilon}\right]d\varepsilon\ . \qquad (6.30)$$

Now we may work out the integral in Eq. (30) by parts, first rewriting $[-\partial\langle N(\varepsilon)\rangle/\partial\varepsilon]d\varepsilon$ as $-d[\langle N(\varepsilon)\rangle]$. Due to the fast (exponential) decay of the factor $\langle N(\varepsilon)\rangle$ at $\varepsilon \to \infty$, its product by the factor $(8m\varepsilon^3)^{1/2}$ vanishes at both integration limits, and we get

$$\begin{aligned}\sigma &= \frac{gq^2\tau}{(2\pi\hbar)^3}\frac{4\pi}{3}\int_0^{\infty}\langle N(\varepsilon)\rangle\, d\left[\left(8m\varepsilon^3\right)^{1/2}\right] \equiv \frac{gq^2\tau}{(2\pi\hbar)^3}\frac{4\pi}{3}(8m)^{1/2}\int_0^{\infty}\langle N(\varepsilon)\rangle\frac{3}{2}\varepsilon^{1/2}d\varepsilon \\ &\equiv \frac{q^2\tau}{m}\times\frac{gm^{3/2}}{\sqrt{2}\pi^2\hbar^3}\int_0^{\infty}\langle N(\varepsilon)\rangle\varepsilon^{1/2}d\varepsilon.\end{aligned} \qquad (6.31)$$

But according to Eq. (3.40), the last factor in this expression (following the × sign) is just the particle density $n \equiv N/V$, so Sommerfeld's result is reduced, for an arbitrary temperature and any particle statistics, to the very simple *Drude formula*,[17]

$$\boxed{\sigma = \frac{q^2\tau}{m}n\,,} \qquad (6.32)$$

Drude formula

which should be well familiar to the reader from an undergraduate physics course.

As a reminder, here is its simple classical derivation.[18] Let $\tau$ be the average time after the last scattering event that has caused particles to lose the deterministic component of their velocity, $\mathbf{v}_{\text{drift}}$, provided by the electric field $\mathscr{E}$ on the top of the particle's random thermal motion – which does not contribute to the net current. Using the 2nd Newton law to describe the particle's acceleration by the

[15] As Eq. (27) shows, if the dispersion law $\varepsilon(\mathbf{p})$ is anisotropic, the current density direction may be different from that of the electric field. In this case, conductivity should be described by a tensor $\sigma_{jj'}$, rather than a scalar. However, in most important conducting materials, the anisotropy is rather small – see, e.g., EM Table 4.1.

[16] This is why the Hall effect, which lacks such ambivalence (see, e.g., QM 3.2), is frequently used to determine the dominating type of charge carriers in semiconductors: electrons or holes, see Sec. 4 below.

[17] It was derived in 1900 by Paul Drude. Note that Drude also used the same arguments to derive a very simple (and very reasonable) approximation for the complex electric conductivity in the ac field of frequency $\omega$: $\sigma(\omega) = \sigma(0)/(1 - i\omega\tau)$, with $\sigma(0)$ given by Eq. (32); sometimes the name "Drude formula" is used for this expression. Let me leave its derivation, from the Boltzmann-RTA equation, for the reader's exercise.

[18] See also EM Sec. 4.2. Note that the frequently met definition of $\tau$ as the "the average time interval between two sequential scattering events" would lead to an extra factor of ½ in the expressions for $\langle\mathbf{v}_{\text{drift}}\rangle$ and $\sigma$.

field, $d\mathbf{v}/dt = q\mathscr{E}/m$, we get $\langle \mathbf{v}_{\text{drift}} \rangle = \tau q\mathscr{E}/m$. Multiplying this result by the particle's charge $q$ and density $n \equiv N/V$, we get the Ohm law $\mathbf{j} = \sigma\mathscr{E}$, with $\sigma$ given by Eq. (32).

Sommerfeld's derivation of the Drude formula poses an important conceptual question. The structure of Eq. (30) implies that the only quantum states contributing to the electric conductivity are those whose derivative $[-\partial\langle N(\varepsilon)\rangle/\partial\varepsilon]$ is significant. For the Fermi particles such as electrons, in the limit $T << \varepsilon_F$, these are the states at the very Fermi surface. On the other hand, Eq. (32) and the whole Drude reasoning, involve the density $n$ of *all* electrons. So, what exactly electrons are responsible for the conductivity: all of them, or only those at the Fermi surface? For the resolution of this paradox, let us return to Eq. (22) and analyze the physical meaning of that result. Let us compare it with the following model distribution:

$$w_{\text{model}} \equiv w_0(\mathbf{r}, \mathbf{p} - \tilde{\mathbf{p}}, t), \tag{6.33}$$

where $\tilde{\mathbf{p}}$ is some time-independent, small vector that describes a small shift of the unperturbed distribution $w_0$ as a whole, in the momentum space. Performing the Taylor expansion of Eq. (33) in this small parameter, and keeping only two leading terms, we get

$$w_{\text{model}} \approx w_0(\mathbf{r}, \mathbf{p}, t) + \tilde{w}_{\text{model}}, \qquad \text{with } \tilde{w}_{\text{model}} = -\tilde{\mathbf{p}} \cdot \nabla_p w_0(\mathbf{r}, \mathbf{p}, t). \tag{6.34}$$

Comparing the last expression with the first form of Eq. (22), we see that they coincide if

$$\tilde{\mathbf{p}} = q\mathscr{E}\,\tau \equiv \mathscr{F}\tau . \tag{6.35}$$

This means that Eq. (22) describes a small shift of the equilibrium distribution of all particles (in the momentum space) by $q\mathscr{E}\tau$ along the electric field's direction, justifying the cartoon shown in Fig. 4.

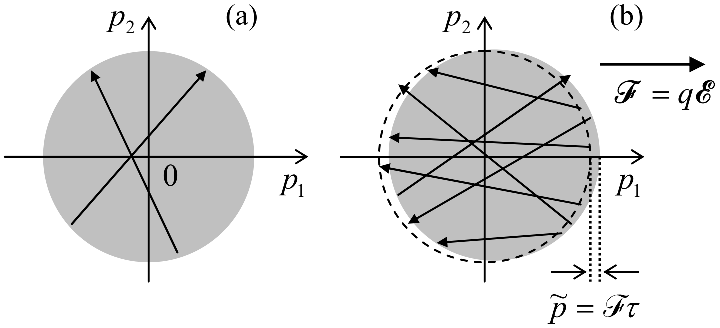


Fig. 6.4. Filling of momentum states by a degenerate electron gas: (a) in the absence and (b) in the presence of an external electric field $\mathscr{E}$. Arrows show representative scattering events.

At $\mathscr{E} = 0$, the system is in equilibrium, so the quantum states inside the Fermi sphere ($p < p_F$), are occupied, while those outside of it are empty – see Fig. 4a. Electron scattering events may happen only between states within a very thin layer ($| p^2/2m - \varepsilon_F | \sim T$) at the Fermi surface because only in this layer the states are partially occupied, so both components of the product $w(\mathbf{r}, \mathbf{p}, t)[1 - w(\mathbf{r}, \mathbf{p}', t)]$, mentioned in Sec. 1, do not vanish. These scattering events, on average, do not change the equilibrium probability distribution, because they are uniformly spread over the Fermi surface.

Now let the electric field be turned on instantly. Immediately it starts accelerating all electrons in its direction, i.e. the whole Fermi sphere starts moving in the momentum space, along the field's direction in the real space. For elastic scattering events (with $| \mathbf{p}' | = | \mathbf{p} |$), this creates an addition of occupied states at the leading edge of the accelerating sphere and an addition of free states on its trailing edge (Fig. 4b). As a result, now there are more scattering events bringing electrons from the leading edge to the trailing edge of the sphere than in the opposite direction. This creates the average backflow of the state occupancy in the momentum space. These two trends eventually cancel each other, and the Fermi sphere approaches a stationary (though not a thermally-equilibrium!) state, with the shift (35) from its equilibrium position.

Now Fig. 4b may be used to answer which of the two different interpretations of the Drude formula is correct, and the answer is: *either*. On one hand, we can look at the electric current as a result of the shift (35) of *all* electrons in the momentum space. On the other hand, each filled quantum state deep inside the sphere gives exactly the same contribution to the net current density as it did without the field. All these internal contributions to the net current cancel each other so the applied field changes the situation only at the Fermi surface. Thus it is equally legitimate to say that only the surface states are responsible for the non-zero net current.[19]

Let me also mention another paradox related to the Drude formula, which is often misunderstood (not only by students :-). As was emphasized above, $\tau$ is finite even at *elastic* scattering – that by itself does not change the total energy of the gas. The question is how can such scattering be responsible for the Ohmic resistivity $\rho \equiv 1/\sigma$, and hence for the Joule heat production, with the power density $\mathscr{p} = \mathbf{j}\cdot\mathscr{E} = \rho j^2$?[20] The answer is that the Drude/Sommerfeld formulas describe just the "bottleneck" of the Joule heat formation. In the scattering picture (Fig. 4b) the states filled by elastically scattered electrons are located above the (shifted) Fermi surface, and these electrons eventually need to relax onto it via some inelastic process, which releases their excessive energy in the form of heat (in a solid, described by phonons – see Sec. 2.6). The rate and other features of these inelastic phenomena do not participate in the Drude formula directly, but for keeping the theory valid (in particular, holding the probability distribution $w$ close to its equilibrium value $w_0$), their intensity has to be sufficient to avoid gas overheating by the applied field.[21]

One final comment is that the Sommerfeld theory of Ohmic conductivity, based on the Boltzmann-RTA equation (18), works very well for the electron gas in most conductors. The scheme shown in Fig. 4 helps to understand why: for degenerate Fermi gases the energies of all particles whose scattering contributes to transport properties, are close ($\varepsilon \approx \varepsilon_F$), and prescribing them all the same relaxation time $\tau$ is very reasonable. In contrast, in classical gases, with their relatively broad distribution of $\varepsilon$, some results given by Eq. (18) are valid only by the order of magnitude.

## 6.3. Electrochemical potential and the drift-diffusion equation

Now let us generalize our calculation to the case when the particle transport takes place in the presence of a time-independent spatial gradient of the probability distribution, $\nabla_r w \neq 0$, caused for example by that of the particle concentration $n = N/V$ (and hence, according to Eq. (3.40), of the

[19] So here, as it frequently happens in physics, formulas (or graphical sketches, such as Fig. 4b) give a clearer description of reality than words – the privilege lacked by many "scientific" disciplines that are rich with unending, shallow verbal debates. Note also that, as frequently happens in physics, the dual interpretation of $\sigma$ is expressed by two different but equal integrals (30) and (31), related by the integration-by-parts rule.

[20] This formula is probably self-evident, but if you need, you may revisit EM Sec. 4.4.

[21] In some poorly conducting materials, charge carrier overheating effects resulting in deviations from the Ohm law, i.e. from the linear relation (28) between $\mathbf{j}$ and $\mathscr{E}$, may be observed already at practicable electric fields.

chemical potential $\mu$), while still assuming that the temperature $T$ is constant. For this generalization, we should keep the second term on the left-hand side of Eq. (18). If the gradient of $w$ is sufficiently small, we can repeat the arguments of the last section and replace $w$ with $w_0$ in this term as well. With the applied electric field $\mathscr{E}$ represented as $(-\nabla\phi)$,[22] where $\phi$ is the electrostatic potential, Eq. (25) becomes

$$\tilde{w} = \tau\,\mathbf{v}\cdot\left(\frac{\partial w_0}{\partial\varepsilon}q\nabla\phi - \nabla w_0\right). \tag{6.36}$$

Since in any of the equilibrium distributions (20), $\langle N(\varepsilon)\rangle$ is a function of $\varepsilon$ and $\mu$ only in the combination $(\varepsilon-\mu)$, it obeys the following relation:

$$\frac{\partial\langle N(\varepsilon)\rangle}{\partial\mu} = -\frac{\partial\langle N(\varepsilon)\rangle}{\partial\varepsilon}. \tag{6.37}$$

Using it, the gradient of $w_0 \propto \langle N(\varepsilon)\rangle$ may be represented as[23]

$$\nabla w_0 = -\frac{\partial w_0}{\partial\varepsilon}\nabla\mu, \qquad \text{for } T = \text{const}, \tag{6.38}$$

so Eq. (36) becomes

$$\tilde{w} = \tau\frac{\partial w_0}{\partial\varepsilon}\mathbf{v}\cdot\left(q\nabla\phi + \nabla\mu\right) \equiv \tau\frac{\partial w_0}{\partial\varepsilon}\mathbf{v}\cdot\nabla\mu', \tag{6.39}$$

where the following sum,

Electro-chemical potential

$$\boxed{\mu' \equiv \mu + q\phi,} \tag{6.40}$$

is called the *electrochemical potential*. Now replicating the calculation of the electric current, carried out in the last section, we get the following generalization of the Ohm law (28):

$$\mathbf{j} = \sigma\left(-\nabla\mu'/q\right) \equiv \sigma\boldsymbol{\mathcal{E}}, \tag{6.41}$$

where the *effective electric field* $\boldsymbol{\mathcal{E}}$ is proportional to the gradient of the electrochemical potential, rather of the electrostatic potential:

Effective electric field

$$\boxed{\boldsymbol{\mathcal{E}} \equiv -\frac{\nabla\mu'}{q} = \mathscr{E} - \frac{\nabla\mu}{q}.} \tag{6.42}$$

The physics of this extremely important and general result[24] may be explained in two ways. First, let us have a look at the energy spectrum of a uniform degenerate Fermi gas confined in a volume of finite size. To ensure such confinement we need a piecewise-constant potential $U(\mathbf{r})$ – a "hard-wall, flat-bottom potential well" – see Fig. 5a. (For conduction electrons in a metal, such profile is provided

[22] Since we will not encounter $\nabla_p$ in the balance of this chapter, from this point on the subscript of the operator $\nabla_r$ is dropped for the notation brevity.

[23] Since we consider $w_0$ as a function of two *independent* arguments $\mathbf{r}$ and $\mathbf{p}$, taking its gradient, i.e. the differentiation of this function over $\mathbf{r}$, does not involve its differentiation over the kinetic energy $\varepsilon$ – which is a function of $\mathbf{p}$ only.

[24] Note that Eq. (42) does not include the phenomenological parameter $\tau$ of the relaxation-time approximation, signaling that it is much more general than the RTA. Indeed, this equality is based entirely on the relation between the second and third terms on the left-hand side of the general Boltzmann equation (10), rather than on any details of the scattering integral on its right-hand side.

by the positively charged ions of the crystal lattice, augmented by its screening by the conduction electrons.) The well should be of a sufficient depth $U_0 > \varepsilon_F \equiv \mu|_{T=0}$ to provide the confinement of the overwhelming majority of the particles, with energies below and somewhat above the Fermi level $\varepsilon_F$. This means that there should be a substantial energy gap,

$$\psi \equiv U_0 - \mu >> T, \qquad (6.43)$$

between the Fermi energy of a particle inside the well, and its potential energy $U_0$ outside the well. (The latter value of energy is usually called the *vacuum level.*) The difference defined by Eq. (43) is called the *workfunction*;[25] for most metals, it is between 4 and 5 eV, so the relation $\psi >> T$ is well fulfilled for room temperatures ($T \sim 0.025$ eV) – and actually for all temperatures below the metal's evaporation point.

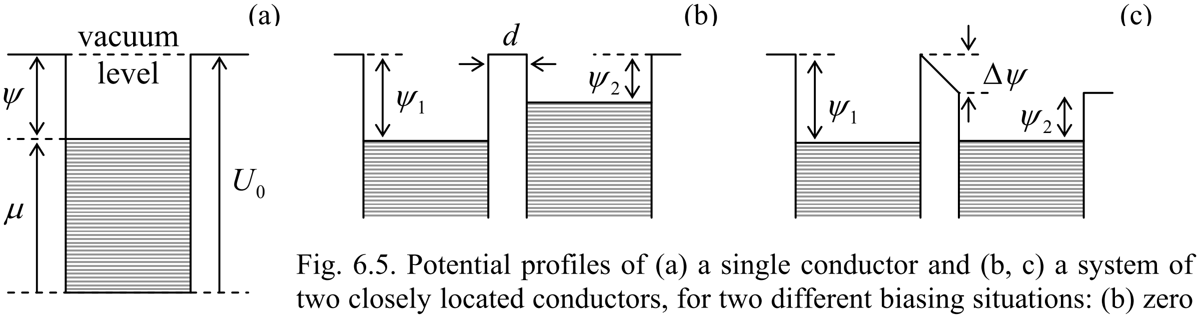


Fig. 6.5. Potential profiles of (a) a single conductor and (b, c) a system of two closely located conductors, for two different biasing situations: (b) zero electrostatic field (the "flat-band condition"), and (c) zero voltage $\Delta\mu'$.

Now let us consider two conductors with different values of $\psi$, separated by a small spatial gap $d$ – see Figs. 5b,c. Panel (b) shows the case when the electric field $\mathscr{E} = -\nabla\phi$ in the free-space gap between the conductors equals zero, i.e. their *electrostatic* potentials $\phi$ are equal.[26] If there is an opportunity for particles to cross the gap (e.g., by either the thermally-activated hopping *over* the potential barrier, discussed in Secs. 5.6-5.7, or the quantum-mechanical tunneling *through* it), there will be an average flow of particles from the conductor with the higher Fermi level to that with the lower Fermi level,[27] because the chemical equilibrium requires their equality – see Secs. 1.5 and 2.7. If the particles have an electric charge (as electrons do), the equilibrium will be automatically achieved by them recharging the effective capacitor formed by the conductors, until the electrostatic energy difference $q\Delta\phi$ reaches the value reproducing that of the workfunctions (Fig. 5c). So for the equilibrium potential difference[28] we may write

$$q\Delta\phi = \Delta\psi = -\Delta\mu. \qquad (6.44)$$

At this equilibrium, the electric field in the gap between the conductors is

[25] Sometimes it is also called "electron affinity", though this term is mostly used for atoms and molecules.

[26] In semiconductor physics and engineering, the situation shown in Fig. 5b is called the *flat-band condition*, because any electric field applied normally to a semiconductor's surface leads to the so-called energy *band bending* – see the next section.

[27] As measured from a common reference value, for example from the vacuum level – rather than from the bottom of an individual potential well as in Fig. 5a.

[28] In physics literature, it is usually called the *contact potential difference*, while in electrochemistry (for which it is one of the key notions), the term *Volta potential* is more common.

$$\mathscr{E} \equiv -\frac{\Delta\phi}{d}\mathbf{n} = \frac{\Delta\mu}{qd}\mathbf{n} = \frac{\nabla\mu}{q}; \qquad (6.45)$$

in Fig. 5c this field is clearly visible as the tilt of the electric potential profile. Comparing Eq. (45) with the definition (42) of the effective electric field $\mathcal{E}$, we see that the equilibrium, i.e. the absence of current through the potential barrier, is achieved exactly when $\mathcal{E} = 0$, in accordance with Eq. (41).

The electric field dichotomy, $\mathscr{E} \leftrightarrow \mathcal{E}$, raises a natural question: which of these fields we are speaking about in everyday and laboratory practice? Upon some contemplation, the reader should agree that most of our electric field measurements are done indirectly, by measuring corresponding voltages – with voltmeters. A vast majority of these instruments belong to the so-called *electrodynamic* variety, which is based on the measurement of a small current flowing through the voltmeter.[29] As Eq. (41) shows, such electrodynamic voltmeters measure the *electrochemical* potential difference $\Delta\mu'/q$. However, there exists a rare breed of *electrostatic* voltmeters (also called "electrometers") that measure the *electrostatic* potential difference $\Delta\phi$ between two conductors. One way to implement such an instrument is to use an ordinary, electrodynamic voltmeter, but with the reference point set at the flat-band condition (Fig. 5b) between the conductors. (This condition may be detected by vanishing electric charge on the adjacent surfaces of the conductors, and hence by the absence of its modulation in time if the distance between the surfaces is periodically modulated.)

Now let me return to Eq. (41) and make two very important remarks. First, it says that in the presence of an electric field, the current vanishes only if $\nabla\mu' = 0$, i.e. that the electrochemical potential $\mu'$, rather than the chemical potential $\mu$, has to be position-independent in a system in the thermodynamic (thermal, chemical, and electric) equilibrium of a conducting system. This result by no means contradicts the fundamental thermodynamic relations for $\mu$ discussed in Sec. 1.5, or the statistical relations involving $\mu$, which were discussed in Sec. 2.7 and beyond. Indeed, according to Eq. (40), $\mu'(\mathbf{r})$ is "merely" the chemical potential measured from the local value of the electrostatic energy $q\phi(\mathbf{r})$, and in all previous parts of the course, this energy was assumed to be constant throughout the system.

Second, note another interpretation of Eq. (41), which may be achieved by modifying Eq. (38) for the particular case of the classical gas. Indeed, the local density $n \equiv N/V$ of the gas obeys Eq. (3.32), which may be rewritten as

$$n(\mathbf{r}) = \text{const} \times \exp\left\{\frac{\mu(\mathbf{r})}{T}\right\}. \qquad (6.46)$$

Taking the spatial gradient of both sides of this relation (still at constant $T$), we get

$$\nabla n = \text{const} \times \frac{1}{T}\exp\left\{\frac{\mu}{T}\right\}\nabla\mu = \frac{n}{T}\nabla\mu, \qquad (6.47)$$

so $\nabla\mu = (T/n)\nabla n$, and Eq. (41), with $\sigma$ given by Eq. (32), may be recast as

$$\mathbf{j} = \sigma\left(-\frac{\nabla\mu'}{q}\right) \equiv \frac{q^2\tau}{m}n\left(-\nabla\phi - \frac{1}{q}\nabla\mu\right) \equiv q\frac{\tau}{m}\left(nq\mathscr{E} - T\nabla n\right). \qquad (6.48)$$

[29] The devices for such measurement may be based on the interaction between the measured current and a permanent magnet, as pioneered by A.-M. Ampère in the 1820s – see, e.g., EM Chapter 5. Such devices are sometimes called *galvanometers*, honoring another pioneer of electricity, Luigi Galvani.

The second term in the parentheses is a specific manifestation of the general Fick's law of diffusion $\mathbf{j}_w = D\nabla n$, already mentioned in Sec. 5.6. Hence the current density may be viewed as consisting of two independent parts: one due to particle *drift* induced by the "usual" electric field $\mathscr{E} = -\nabla\phi$, and another due to their *diffusion* – see Eq. (5.118) and its discussion. This is exactly the physics of the "mysterious" term $\nabla\mu$ in Eq. (42), though its simple form (48) is valid only in the classical limit.

Besides being very useful for applications, Eq. (48) also gives us a pleasant surprise. Namely, plugging it into the continuity equation for electric charge,[30]

$$\frac{\partial(qn)}{\partial t} + \nabla\cdot\mathbf{j} = 0\,, \tag{6.49}$$

we get (after the division of all terms by $q\tau/m$) the so-called *drift-diffusion equation*:[31]

Drift-diffusion equation

$$\boxed{\frac{m}{\tau}\frac{\partial n}{\partial t} = \nabla\left(n\nabla U\right) + T\nabla^2 n, \quad \text{with } U \equiv q\phi\,.} \tag{6.50}$$

Comparing it with Eq. (5.122), we see that the drift-diffusion equation is identical to the Smoluchowski equation,[32] provided that we parallel the ratio $\tau/m$ with the mobility $\mu_\text{m} = 1/\eta$ of the Brownian particle. Now using Einstein's relation (5.78), we see that the effective diffusion constant $D$ of the classical gas of similar particles is

$$D = \frac{\tau T}{m}\,. \tag{6.51a}$$

This important relation is more frequently represented in either of two other forms. First, since the rare scattering events we are considering do not change the statistics of the gas in thermal equilibrium, we may still use the Maxwell-distribution result (3.9) for the average-square velocity $\langle v^2\rangle$, to recast Eq. (51a) as

$$D = \frac{1}{3}\left\langle v^2\right\rangle\tau\,. \tag{6.51b}$$

One more popular form of the same relation uses the notion of the *mean free path* $l$, which may be defined as the average distance to be passed by a particle before its next scattering:

$$D = \frac{1}{3}l\left\langle v^2\right\rangle^{1/2}, \qquad \text{with } l \equiv \left\langle v^2\right\rangle^{1/2}\tau\,. \tag{6.51c}$$

In the forms (51b)-(51c), the result for $D$ makes more physical sense, because it may be readily derived (admittedly, with some uncertainty of the numerical coefficient) from simple kinematic arguments – the task left for the reader's exercise.

Note also that using Eq. (51a), Eq. (48) may be rewritten as an expression for the *particle flow density* $\mathbf{j}_n \equiv n\mathbf{j}_w = \mathbf{j}/q$:

$$\mathbf{j}_n = n\mu_\text{m} q\mathscr{E} - D\nabla n\,, \tag{6.52}$$

[30] If this relation is not obvious, please revisit EM Sec. 4.1.

[31] Sometimes this name is used for Eq. (52). One may also run into the term "convection-diffusion equation" for Eq. (50) with the replacement (51a).

[32] And hence, at negligible $\nabla U$, identical to the diffusion equation (5.116).

with the first term on the right-hand side describing particles' drift, and the second one, their diffusion. I will discuss the application of this equation to the most important case of non-degenerate ("quasi-classical") gases of electrons and holes in semiconductors, in the next section.

To complete this section, let me emphasize again that the mathematically similar drift-diffusion equation (50) and the Smoluchowski equation (5.122) describe different physical situations. Indeed, our (or rather Einstein and Smoluchowski's :-) treatment of the Brownian motion in Chapter 5 was based on a strong hierarchy of the system, consisting of a large "Brownian particle" in an environment of many smaller particles – "molecules". On the other hand, in this chapter, we are considering a gas of similar particles. Nevertheless, the equations describing the dynamics of their probability distribution, are the same – at least within the framework of the Boltzmann transport equation with the relaxation-time approximation (17) of the scattering integral. The origin of this similarity is the fact that Eq. (12) is clearly applicable to a Brownian particle as well, with each "scattering" event being the particle's hit by a random molecule of its environment. Since, due to the mass hierarchy, the particle momentum change at each such event is very small, the scattering integral has to be local, i.e. depend only on *w* at the same momentum **p** as the left-hand side of the Boltzmann equation, so the relaxation time approximation (17) is absolutely natural – indeed, more natural than for our current case of similar particles.

## 6.4. Charge carriers in semiconductors

Now let me demonstrate the application of the concepts discussed in the last section, first of all of the electrochemical potential, to understanding the basic kinetic properties of semiconductors and a few key semiconductor structures – which are the basis of most modern electronic and optoelectronic devices, and hence of all our IT civilization. For that, I will need to take a detour to discuss their equilibrium properties first.

I will use an approximate but reasonable picture in which the energy of the electron subsystem in a solid may be partitioned into the sum of the effective energies $\varepsilon$ of independent electrons. Quantum mechanics says[33] that in such periodic structures as crystals, the stationary state energy $\varepsilon$ of a particle interacting with the atomic lattice follows one of the periodic functions $\varepsilon_n(\mathbf{q})$ of the *quasimomentum* **q**, oscillating between two extreme values $\varepsilon_n|_{\min}$ and $\varepsilon_n|_{\max}$. These *allowed energy bands* are separated with *bandgaps*, of widths $\Delta_n \equiv \varepsilon_n|_{\min} - \varepsilon_{n-1}|_{\max}$, with no allowed states inside them. Semiconductors and insulators (dielectrics) are defined as such crystals that in equilibrium at $T = 0$, all electron states in several energy bands (with the highest of them called the *valence band*) are completely filled, $\langle N(\varepsilon_v)\rangle = 1$, while those in the upper bands, starting from the lowest, *conduction band*, are completely empty, $\langle N(\varepsilon_c)\rangle = 0$.[34, 35] Since the electrons follow the Fermi-Dirac statistics (2.115), this means that at $T \rightarrow 0$,

[33] See, e.g., QM Sec. 2.7 and 3.4, but a thorough knowledge of this material is not necessary for following discussions in this section. If the reader is not familiar with the notion of quasimomentum (alternatively called the "crystal momentum"), the following interpretation may be useful: **q** is the result of quantum averaging of the genuine electron momentum **p** over the crystal lattice period. In contrast to **p**, which is not conserved because of the electron's interaction with the lattice, **q** is an integral of motion – in the absence of other forces.

[34] This mapping of electrical properties of crystals onto their band structure was pioneered in 1931-32 by Alan H. Wilson.

[35] In insulators, the bandgap $\Delta$ is so large (e.g., ~9 eV in $SiO_2$) that the conduction band remains unpopulated in all practical situations, so the following discussion is only relevant for semiconductors, with their moderate bandgaps – such as 1.14 eV in the most important case of silicon at room temperature.

the Fermi energy $\varepsilon_F \equiv \mu(0)$ is located somewhere between the valence band's maximum $\varepsilon_v|_{\max}$ (usually called simply $\varepsilon_V$), and the conduction band's minimum $\varepsilon_c|_{\min}$ (called $\varepsilon_C$) – see Fig. 6.

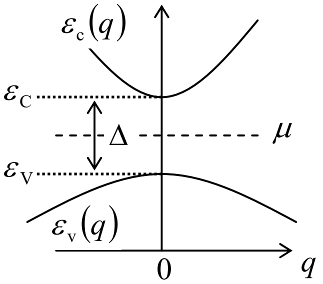


Fig. 6.6. Calculating $\mu$ in an intrinsic semiconductor.

Let us calculate the population of both branches $\varepsilon_n(\mathbf{q})$, and the chemical potential $\mu$ in equilibrium at $T > 0$. Since the functions $\varepsilon_n(\mathbf{q})$ are typically smooth, near the bandgap edges the dispersion laws $\varepsilon_c(\mathbf{q})$ and $\varepsilon_v(\mathbf{q})$ may be well approximated with quadratic parabolas. For our analysis, let us take the parabolas in the simplest, isotropic form, with origins at the same quasimomentum, taking it for the reference point:[36]

$$\varepsilon = \begin{cases} \varepsilon_C + q^2/2m_C, & \text{for } \varepsilon \geq \varepsilon_C, \\ \varepsilon_V - q^2/2m_V, & \text{for } \varepsilon \leq \varepsilon_V, \end{cases} \qquad \text{with } \varepsilon_C - \varepsilon_V \equiv \Delta. \tag{6.53}$$

The positive constants $m_C$ and $m_V$ are usually called the effective masses of, respectively, electrons and holes. (In a typical semiconductor, $m_C$ is a few times smaller than the free electron mass $m_e$, while $m_V$ is closer to $m_e$.)

Due to the similarity between the top line of Eq. (53) and the dispersion law (3.3) of free particles, we may reuse Eq. (3.40), with the appropriate particle mass $m$, the degeneracy factor $g$, and the energy origin, to calculate the full spatial density of the populated states (in semiconductor physics, called *electrons* in the narrow sense of the word):

$$n \equiv \frac{N_e}{V} = \int_{\varepsilon_C}^{\infty} \langle N(\varepsilon)\rangle g_3(\varepsilon)d\varepsilon \equiv \frac{g_C m_C^{3/2}}{\sqrt{2}\pi^2\hbar^3}\int_0^{\infty}\langle N(\tilde{\varepsilon}+\varepsilon_C)\rangle \tilde{\varepsilon}^{1/2}d\tilde{\varepsilon}, \tag{6.54}$$

where $\tilde{\varepsilon} \equiv \varepsilon - \varepsilon_C \geq 0$. Similarly, the density $p$ of "no-electron" excitations (called *holes*) in the valence band is the number of *unfilled* states in the band, and hence may be calculated as

$$p \equiv \frac{N_h}{V} = \int_{-\infty}^{\varepsilon_V}[1-\langle N(\varepsilon)\rangle]g_3(\varepsilon)d\varepsilon \equiv \frac{g_V m_V^{3/2}}{\sqrt{2}\pi^2\hbar^3}\int_0^{\infty}[1-\langle N(\varepsilon_V-\tilde{\varepsilon})\rangle]\tilde{\varepsilon}^{1/2}d\tilde{\varepsilon}, \tag{6.55}$$

where in this case, $\tilde{\varepsilon} \geq 0$ is defined as $(\varepsilon_V - \varepsilon)$. If the electrons and holes[37] are in the thermal and chemical equilibrium, the functions $\langle N(\varepsilon)\rangle$ in these two relations should follow the Fermi-Dirac

[36] It is easy (and hence is left for the reader's exercise) to verify that all equilibrium properties of charge carriers remain the same (with some effective values of $m_C$ and $m_V$) if $\varepsilon_c(\mathbf{q})$ and $\varepsilon_v(\mathbf{q})$ are arbitrary quadratic forms of the Cartesian components of the quasimomentum. A mutual displacement of the branches $\varepsilon_c(\mathbf{q})$ and $\varepsilon_v(\mathbf{q})$ in the quasimomentum space is also unimportant for statistical and most transport properties of the semiconductors, though it is very important for their optical properties – which I will not have time to discuss in any detail.

[37] The collective name for them in semiconductor physics is *charge carriers* – or just "carriers".

distribution (2.115) with the same temperature $T$ and the same chemical potential $\mu$. Moreover, in our current case of an undoped (*intrinsic*) semiconductor, these densities have to be equal,

$$n = p \equiv n_\mathrm{i}\,, \tag{6.56}$$

because if this *electroneutrality condition* was violated, the volume would acquire a non-zero electric charge density $\rho = e(p - n)$, which would result, in a bulk sample, in an extremely high electric field energy. From this condition and Eqs. (54)-(55), we get a system of two equations,

$$n_\mathrm{i} = \frac{g_\mathrm{C} m_\mathrm{C}^{3/2}}{\sqrt{2}\pi^2\hbar^3}\int_0^\infty \frac{\tilde{\varepsilon}^{1/2}d\tilde{\varepsilon}}{\exp\{(\tilde{\varepsilon} + \varepsilon_\mathrm{C} - \mu)/T\}+1} = \frac{g_\mathrm{V} m_\mathrm{V}^{3/2}}{\sqrt{2}\pi^2\hbar^3}\int_0^\infty \frac{\tilde{\varepsilon}^{1/2}d\tilde{\varepsilon}}{\exp\{(\tilde{\varepsilon} - \varepsilon_\mathrm{V} + \mu)/T\}+1}, \tag{6.57}$$

whose solution gives both the requested charge carrier density $n_\mathrm{i}$ and the Fermi level $\mu$.

For an arbitrary ratio $\Delta/T$, this solution may be found only numerically, but in most practical cases, this ratio is very large. (Again, for Si at room temperature, $\Delta \approx 1.14$ eV, while $T \approx 0.025$ eV.) In this case, we may use the same classical approximation as in Eq. (3.45), to reduce Eqs. (54) and (55) to simple expressions

$$n = n_\mathrm{C}\exp\left\{\frac{\mu - \varepsilon_\mathrm{C}}{T}\right\}, \qquad p = n_\mathrm{V}\exp\left\{\frac{\varepsilon_\mathrm{V} - \mu}{T}\right\}, \qquad \text{for } T \ll \Delta\,, \tag{6.58}$$

where the temperature-dependent parameters,

$$n_\mathrm{C} \equiv \frac{g_\mathrm{C}}{\hbar^3}\left(\frac{m_\mathrm{C}T}{2\pi}\right)^{3/2} \text{ and } n_\mathrm{V} \equiv \frac{g_\mathrm{V}}{\hbar^3}\left(\frac{m_\mathrm{V}T}{2\pi}\right)^{3/2}, \tag{6.59}$$

may be interpreted as the effective numbers of states (per unit volume) available for occupation in, respectively, the conduction and valence bands, in thermal equilibrium. For usual semiconductors (with $g_\mathrm{C} \sim g_\mathrm{V} \sim 1$, and $m_\mathrm{C} \sim m_\mathrm{V} \sim m_\mathrm{e}$), at room temperature, these numbers are of the order of $3\times10^{25}\mathrm{m}^{-3} \equiv 3\times10^{19}\mathrm{cm}^{-3}$. (Note that all results based on Eqs. (58) are only valid if both $n$ and $p$ are much lower than, respectively, $n_\mathrm{C}$ and $n_\mathrm{V}$.)

With the substitution of Eqs. (58), the system of equations (56) allows a straightforward solution:

$$\mu = \frac{\varepsilon_\mathrm{V} + \varepsilon_\mathrm{C}}{2} + \frac{T}{2}\left(\ln\frac{g_\mathrm{V}}{g_\mathrm{C}} + \frac{3}{2}\ln\frac{m_\mathrm{V}}{m_\mathrm{C}}\right), \qquad n_\mathrm{i} = (n_\mathrm{C}n_\mathrm{V})^{1/2}\exp\left\{-\frac{\Delta}{2T}\right\}. \tag{6.60}$$

Since in all practical materials the logarithms in the first of these expressions are never much larger than 1,[38] it shows that the Fermi level in intrinsic semiconductors never deviates much from the so-called *midgap value* $(\varepsilon_\mathrm{V} + \varepsilon_\mathrm{C})/2$ – see the (schematic) Fig. 6. In the result for $n_\mathrm{i}$, the last (exponential) factor is very small, so the equilibrium number of charge carriers is much lower than that of the atoms – for the most important case of silicon at room temperature, $n_\mathrm{i} \sim 10^{10}\mathrm{cm}^{-3}$. The exponential temperature dependence of $n_\mathrm{i}$ (and hence of the electric conductivity $\sigma \propto n_\mathrm{i}$) of intrinsic semiconductors is the basis of several applications, for example, simple *germanium resistance thermometers* efficient in the whole range from ~0.5K to ~100K. Another useful application of the same fact is the extraction of the bandgap

[38] Note that in the case of simple electron spin degeneracy ($g_\mathrm{V} = g_\mathrm{C} = 2$), the first logarithm vanishes altogether. However, in many semiconductors, the degeneracy is factored by the number of similar energy bands (e.g., six similar conduction bands in silicon), and the factor $\ln(g_\mathrm{V}/g_\mathrm{C})$ may slightly affect quantitative results.

of a semiconductor from the experimental measurement of the temperature dependence of $\sigma \propto n_\text{i}$ – frequently, in just two well-separated temperature points.

However, most applications require a much higher concentration of carriers. It may be increased quite dramatically by planting into a semiconductor a relatively small number of slightly different atoms – either *donors* (e.g., phosphorus atoms for Si) or *acceptors* (e.g., boron atoms for Si). Let us analyze the first opportunity, called *n-doping*, using the same simple energy band model (53). If the donor atom is only slightly different from those in the crystal lattice, it may be easily ionized – giving an additional electron to the conduction band and hence becoming a positive ion. This means that the effective ground state energy $\varepsilon_\text{D}$ of the additional electrons is just slightly below the conduction band edge $\varepsilon_\text{C}$ – see Fig. 7a.[39]

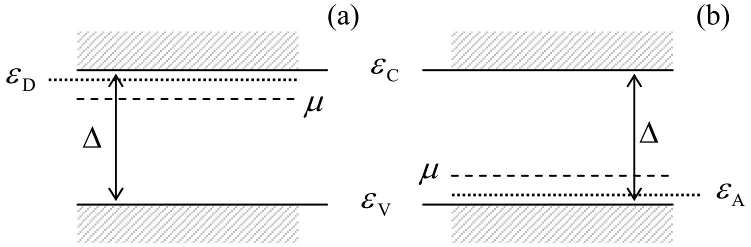


Fig. 6.7. The Fermi levels $\mu$ in (a) $n$-doped and (b) $p$-doped semiconductors. Hatching shows the ranges of unlocalized state energies.

Reviewing the arguments that have led us to Eqs. (58), we see that at relatively low doping, when the strong inequalities $n << n_\text{C}$ and $p << n_\text{V}$ still hold, these relations are not affected by the doping, so the concentrations of electrons and holes given by these equalities still obey a universal (doping-independent) relation following from Eqs. (58) and (60):[40]

$$np = n_\text{i}^2 . \qquad (6.61)$$

However, for a doped semiconductor, the electroneutrality condition looks differently from Eq. (56) because the total density of positive charges in a unit volume is not $p$, but rather ($p + n_+$), where $n_+$ is the density of positively-ionized ("activated") donor atoms, so the condition becomes

$$n = p + n_+ . \qquad (6.62)$$

If virtually all dopants are activated, as it is in most practical cases,[41] then we may take $n_+ = n_\text{D}$, where $n_\text{D}$ is the total concentration of donor atoms, i.e. their number per unit volume, and Eq. (62) becomes

$$n = p + n_\text{D} . \qquad (6.63)$$

Plugging in Eq. (61) in the form $p = n_\text{i}^2/n$, we get a simple quadratic equation for $n$, with the following physically acceptable (positive) solution:

[39] Note that in comparison with Fig. 6, here the (for most purposes, redundant) information on the $q$-dependence of the energies is collapsed, leaving the horizontal axis of such a *band-edge diagram* free for showing their possible spatial dependences – see Figs. 8, 10, and 11 below.

[40] Very similar relations may be met in the theory of chemical reactions (where it is called the *law of mass action*), and other disciplines – including such exotic examples as theoretical ecology.

[41] Let me leave it for the reader's exercise to prove that this assumption is always valid unless the doping density $n_\text{D}$ becomes comparable to $n_\text{C}$, and as a result, the Fermi level $\mu$ is shifted into a ~$T$-wide vicinity of $\varepsilon_\text{D}$.

$$n = \frac{n_\text{D}}{2} + \left( \frac{n_D^2}{4} + n_\text{i}^2 \right)^{1/2} . \tag{6.64}$$

This result shows that the doping affects $n$ (and hence $\mu = \varepsilon_\text{C} - T\ln(n_\text{C}/n)$ and $p = n_\text{i}^2/n$) only if the dopant concentration $n_\text{D}$ is comparable with, or higher than the intrinsic carrier density $n_\text{i}$ given by Eq. (60). For most applications, $n_\text{D}$ is made *much* higher than $n_\text{i}$; in this case Eq. (64) yields

$$n \approx n_\text{D} >> n_\text{i}, \qquad p = \frac{n_\text{i}^2}{n} \approx \frac{n_\text{i}^2}{n_\text{D}} << n, \qquad \mu \approx \mu_p \equiv \varepsilon_\text{C} - T \ln \frac{n_\text{C}}{n_\text{D}} . \tag{6.65}$$

Because of the reasons to be discussed very soon, modern electron devices require doping densities above $10^{18}$cm$^{-3}$, so the logarithm in Eq. (65) is not much larger than 1. This means that the Fermi level rises from the midgap to a position only slightly below the conduction band edge $\varepsilon_\text{C}$ – see Fig. 7a.

The opposite case of purely $p$-doping, with $n_\text{A}$ acceptor atoms per unit volume, and a small activation (negative ionization) energy $\varepsilon_\text{A} - \varepsilon_\text{V} << \Delta$,[42] may be considered absolutely similarly, using the electroneutrality condition in the form

$$n + n_- = p , \tag{6.66}$$

where $n_-$ is the number of activated (and hence negatively charged) acceptors. For the relatively high concentration ($n_\text{i} << n_\text{A} << n_\text{V}$), virtually all acceptors are activated, so $n_- \approx n_\text{A}$, Eq. (66) may be approximated as $n + n_\text{A} = p$, and the analysis gives the results dual to Eq. (65):

$$p \approx n_\text{A} >> n_\text{i}, \qquad n = \frac{n_\text{i}^2}{p} \approx \frac{n_\text{i}^2}{n_\text{A}} << p, \qquad \mu \approx \mu_n \equiv \varepsilon_\text{V} + T \ln \frac{n_\text{V}}{n_\text{A}} , \tag{6.67}$$

so in this case, the Fermi level is just slightly above the valence band edge (Fig. 7b), and the number of holes far exceeds that of electrons – again, in the narrow sense of the word. Let me leave the analysis of the simultaneous $n$- and $p$-doping (which enables, in particular, so-called *compensated semiconductors* with the sign-variable difference $n - p \approx n_\text{D} - n_\text{A}$) for the reader's exercise.

Now let us consider how a sample of a doped semiconductor (say, a $p$-doped one) responds to a static external electrostatic field $\mathscr{E}$ applied normally to its surface.[43] (In semiconductor integrated circuits, such a field is usually created by the voltage applied to a special highly-conducting *gate electrode* separated from the semiconductor surface by a thin insulating layer.) Assuming that the field penetrates into the sample by a distance $\lambda$ much larger than the crystal lattice period $a$ (the assumption to be verified *a posteriori*), we may calculate the distribution of the electrostatic potential $\phi$ using the macroscopic version of the Poisson equation.[44] Assuming that the semiconductor occupies the semi-space $x > 0$ and that $\boldsymbol{\mathscr{E}} = \mathbf{n}_x\mathscr{E}$, the equation reduces to the following 1D form[45]

[42] For the typical donors (P) and acceptors (B) in silicon, both ionization energies, $(\varepsilon_\text{C} - \varepsilon_\text{D})$ and $(\varepsilon_\text{A} - \varepsilon_\text{V})$, are close to 45 meV, i.e. are indeed much smaller than $\Delta \approx 1.14$ eV.

[43] A simplified version of this analysis was carried out in EM Sec. 2.1.

[44] See, e.g., EM Sec. 3.4.

[45] I am sorry for using, for the SI electric constant $\varepsilon_0$, the same Greek letter as for single-particle energies, but both notations are traditional, and the difference between these uses will be clear from the context.

$$\frac{d^2\phi}{dx^2} = -\frac{\rho(x)}{\kappa\varepsilon_0}. \tag{6.68}$$

Here $\kappa$ is the dielectric constant of the semiconductor matrix – excluding the dopants and charge carriers, which in this approach are treated as explicit ("stand-alone") charges, with the volumic density

$$\rho = e\left(p - n_- - n\right). \tag{6.69}$$

(As a sanity check, Eqs. (68)-(69) show that if $\mathscr{E} \equiv -d\phi/dx = 0$, then $\rho = 0$, bringing us back to the electroneutrality condition (66), and hence the "flat" band-edge diagrams shown in Figs. 7b and 8a.)

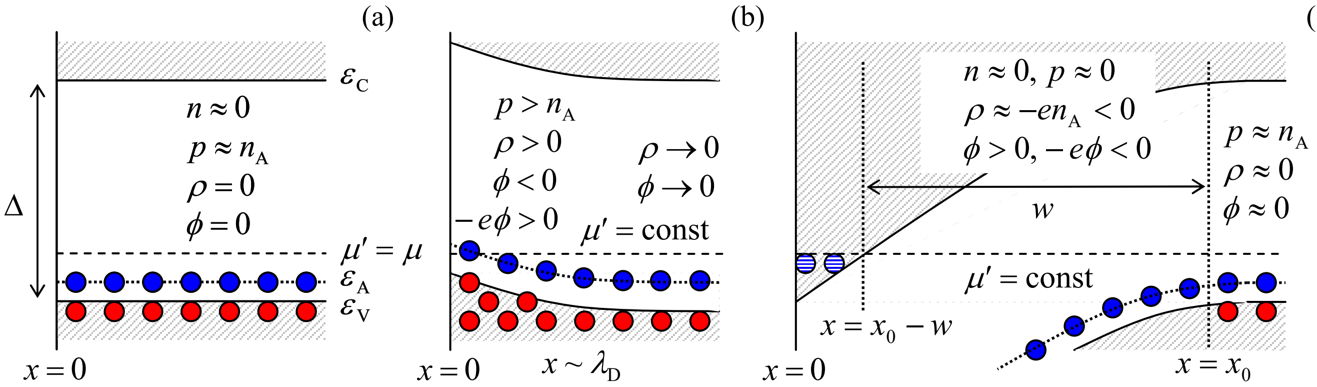


Fig. 6.8. The band-edge diagrams of the electric field penetration into a uniform *p*-doped semiconductor: (a) $\mathscr{E} = 0$, (b) $\mathscr{E} < 0$, and (c) $0 < \mathscr{E}_c < \mathscr{E}$. Solid red points depict positive charges; solid blue points, negative charges; and hatched blue points, possible electrons in the inversion layer – all very schematically.

In order to get a closed system of equations for the case $\mathscr{E} \neq 0$, we should take into account that the electrostatic potential $\phi \neq 0$, penetrating into the sample with the field,[46] adds the potential component $q\phi(x) = -e\phi(x)$ to the energy of each electron, and hence shifts the whole local system of single-electron energy levels "vertically" by this amount – down for $\phi > 0$, and up for $\phi < 0$. As a result, the field penetration leads to what is called *band bending* – see the band-edge diagrams schematically shown in Figs. 8b,c for two possible polarities of the applied field, which affects the distribution $\phi(x)$ via the boundary condition[47]

$$\frac{d\phi}{dx}(0) = -\mathscr{E}. \tag{6.70}$$

Note that the electrochemical potential $\mu'$ (which, in accordance with the discussion in Sec. 3, replaces the chemical potential in presence of the electric field),[48] has to stay constant through the system in equilibrium, keeping the electric current equal to zero – see Eq. (41). For arbitrary doping parameters,

[46] It is common (though not necessary) to select the energy reference so deep inside the semiconductor, $\phi = 0$; in what follows I will use this convention.

[47] Here $\mathscr{E}$ is the field *just inside* the semiconductor. The free-space field necessary to create it is $\kappa$ times larger – see, e.g., the same EM Sec. 3.4, in particular, Eq. (3.56).

[48] In semiconductor physics literature, the value of $\mu'$ is usually called the *Fermi level*, even in the absence of the degenerate Fermi sea typical for metals – cf. Sec. 3.3. In this section, I will follow this common terminology.

the system of equations (58) (with the replacements $\varepsilon_V \to \varepsilon_V - e\phi$, and $\mu \to \mu'$) and (68)-(70), plus the relation between $n_-$ and $n_A$ (describing the acceptor activation), does not allow an analytical solution. However, as was discussed above, in the most practical cases $n_A >> n_i$, we may use the approximate relations $n_- \approx n_A$ and $n \approx 0$ at virtually any values of $\mu'$ within the locally shifted bandgap $[\varepsilon_V - e\phi(x), \varepsilon_C - e\phi(x)]$, so the substitution of these relations, and the second of Eqs. (58), with the mentioned replacements, into Eq. (69) yields

$$\rho \approx en_V \exp\left\{\frac{\varepsilon_V - e\phi - \mu'}{T}\right\} - en_A \equiv en_A\left[\left(\frac{n_V}{n_A}\exp\left\{\frac{\varepsilon_V - \mu'}{T}\right\}\right)\exp\left\{-\frac{e\phi}{T}\right\} - 1\right]. \qquad (6.71)$$

The $x$-independent electrochemical potential (a.k.a. the Fermi level) $\mu'$ in this relation should be equal to the value of the chemical potential $\mu(x \to \infty)$ in the semiconductor's bulk, given by the last of Eqs. (67), which turns the expression in the parentheses into 1. With these substitutions, Eq. (68) becomes

$$\frac{d^2\phi}{dx^2} = -\frac{en_A}{\kappa\varepsilon_0}\left[\exp\left\{-\frac{e\phi}{T}\right\} - 1\right], \qquad \text{for } \varepsilon_V - e\phi(x) < \mu' < \varepsilon_C - e\phi(x). \qquad (6.72)$$

This nonlinear differential equation may be solved analytically, but in order to avoid a distraction by this (rather bulky) solution, let me first consider the case when the electrostatic potential is sufficiently small – either because the external field is small, or because we focus on the distances sufficiently far from the surface – see Fig. 8 again. In this case, in the Taylor expansion of the exponent in Eq. (72), with respect to small $\phi$, we may keep only two leading terms, turning it into a linear equation:

$$\frac{d^2\phi}{dx^2} = \frac{e^2 n_A}{\kappa\varepsilon_0 T}\phi, \qquad \text{i.e. } \frac{d^2\phi}{dx^2} = \frac{\phi}{\lambda_D^2}, \qquad \text{where } \lambda_D \equiv \left(\frac{\kappa\varepsilon_0 T}{e^2 n_A}\right)^{1/2}, \qquad (6.73)$$

with the well-known exponential solution, satisfying also the boundary condition $\phi \to 0$ at $x \to \infty$:

$$\phi = C\exp\left\{-\frac{x}{\lambda_D}\right\}, \qquad \text{at } e|\phi| << T. \qquad (6.74)$$

The constant $\lambda_D$ given by the last of Eqs. (73) is called the *Debye screening length*. It may be rather substantial; for example, at $T_K$ = 300K, even for the relatively high doping $n_A \approx 10^{18}\text{cm}^{-3}$ typical for modern silicon ($\kappa \approx 12$) integrated circuits, it is close to 4 nm – still much larger than the crystal lattice constant $a \sim 0.3$ nm, so the above analysis is indeed quantitatively valid. Note also that $\lambda_D$ does not depend on the charge's sign; hence it should be no large surprise that repeating our analysis for an $n$-doped semiconductor, we may find out that Eqs. (73)-(74) are valid for that case as well, with the only replacement $n_A \to n_D$.

If the applied field $\mathscr{E}$ is weak, Eq. (74) is valid in the whole sample, and the constant $C$ in it may be readily calculated using the boundary condition (70), giving

$$\phi\big|_{x=0} \equiv C = \lambda_D\mathscr{E} \equiv \left(\frac{\kappa\varepsilon_0 T}{e^2 n_A}\right)^{1/2}\mathscr{E}. \qquad (6.75)$$

This formula allows us to express the condition of validity of the linear approximation leading to Eq. (74), $e|\,\phi\,| << T$, in terms of the applied field:

$$\left|\mathscr{E}\right| << \mathscr{E}_{\max}, \qquad \text{with } \mathscr{E}_{\max} \equiv \frac{T}{e\lambda_D} \equiv \left(\frac{Tn_A}{\kappa\varepsilon_0}\right)^{1/2}; \tag{6.76}$$

in the above example, $\mathscr{E}_{\max} \sim 60$ kV/cm. On the lab scale, such field is not low at all (it is twice higher than the threshold of electric breakdown in the air at ambient conditions), but it may be sustained by many solid-state materials that are much less prone to breakdown.[49] This is why we should be interested in what happens if the applied field is higher than this value.

The semi-quantitative answer is relatively simple if the field is directed out of the $p$-doped semiconductor (in our nomenclature, $\mathscr{E} < 0$ – see Fig. 8b). As the valence band bends up by a few $T$, the local hole concentration $p(x)$, and hence the charge density $\rho(x)$, grow exponentially – see Eq. (71). Hence the effective local length of the nonlinear field's penetration, $\lambda_{ef}(x) \propto \rho^{-1/2}(x)$, shrinks exponentially. A detailed analysis of this effect using Eq. (72) does not make much sense, because as soon as $\lambda_{ef}(0)$ decreases to $\sim a$, the macroscopic Poisson equation (68) is no longer valid quantitatively. For typical semiconductors, this happens at the field that raises the edge $\varepsilon_V - e\phi(0)$ of the bent valence band at the sample's surface above the Fermi level $\mu'$. In this case, the valence-band electrons near the surface form a degenerate Fermi gas, with an "open" Fermi surface – essentially a metal, which a very small (atomic-size) *Thomas-Fermi screening length*:[50]

$$\lambda_{ef}(0) \sim \lambda_{TF} \equiv \left[\frac{\kappa\varepsilon_0}{e^2 g_3(\varepsilon_F)}\right]^{1/2}. \tag{6.77}$$

The effects taking place at the opposite polarity of the field, $\mathscr{E} > 0$, are much more interesting – and more useful for applications. Indeed, in this case, the band bending down leads to an exponential decrease of $\rho(x)$ as soon as the valence band edge $\varepsilon_V - e\phi(x)$ drops down by just a few $T$ below its unperturbed value $\varepsilon_V$. If the applied field is large enough, $\mathscr{E} > \mathscr{E}_c$ (as it is in the situation shown in Fig. 8c), it forms, on the left of such point $x_0$ the so-called *depletion layer*, of a certain width $w$. Within this layer, not only the electron density $n$ but the hole density $p$ as well are negligible, so the only substantial contribution to the charge density $\rho$ is given by the fully ionized acceptors: $\rho \approx -en_- \approx -en_A$, and Eq. (72) becomes very simple:

$$\frac{d^2\phi}{dx^2} = \frac{en_A}{\kappa\varepsilon_0} = \text{const}, \qquad \text{for } x_0 - w < x < x_0. \tag{6.78}$$

Let us use this equation to calculate the largest possible width $w$ of the depletion layer, and the critical value, $\mathscr{E}_c$, of the applied field necessary for this. (By definition, at $\mathscr{E} = \mathscr{E}_c$, the left boundary of the layer, where $\varepsilon_V - e\phi(x) = \varepsilon_C$, i.e. $e\phi(x) = \varepsilon_V - \varepsilon_A \equiv \Delta$, just touches the semiconductor surface: $x_0 - w = 0$, i.e. $x_0 = w$. (Figure 8c shows the case when $\mathscr{E}$ is slightly larger than $\mathscr{E}_c$.) For this, Eq. (78) has to be solved with the following boundary conditions:

$$\phi(0) = \frac{\Delta}{e}, \qquad \frac{d\phi}{dx}(0) = -\mathscr{E}_c, \qquad \phi(w) = 0, \qquad \frac{d\phi}{dx}(w) = 0. \tag{6.79}$$

[49] Even some amorphous thin-film insulators, such as properly grown silicon and aluminum oxides, can withstand fields up to ~10 MV/cm.

[50] As a reminder, the derivation of this formula was the task of Problem 3.14.

Note that the first of these conditions is strictly valid only if $T << \Delta$, i.e. at the assumption we have made from the very beginning, while the last two conditions are asymptotically correct only if $\lambda_D << w$ – the assumption we should not forget to check after the solution.

After all the undergraduate experience with projective motion problems, the reader certainly knows by heart that the solution of Eq. (78) is a quadratic parabola, so let me immediately write its final form satisfying the boundary conditions (79):

$$\phi(x) = \frac{en_A}{\kappa\varepsilon_0}\frac{(w-x)^2}{2}, \qquad \text{with } w = \left(\frac{2\kappa\varepsilon_0\Delta}{e^2 n_A}\right)^{1/2}, \quad \text{at } \mathscr{E}_c = \frac{2\Delta}{e\varepsilon_0 w}. \tag{6.80}$$

Comparing the result for $w$ with Eq. (73), we see that if our basic condition $T << \Delta$ is fulfilled, then $\lambda_D << w$, confirming the qualitative validity of the whole solution (80). For the same particular parameters as in the example before ($n_A \approx 10^{18}\text{cm}^{-3}$, $\kappa \approx 10$), and $\Delta \approx 1$ eV, Eqs. (80) give $w \approx 40$ nm and $\mathscr{E}_c \approx 600$ kV/cm – still a practicable field. (As Fig. 8c shows, to create it, we need a gate voltage only slightly larger than $\Delta/e$, i.e. close to 1 V for typical semiconductors.)

Figure 8c also shows that if the applied field exceeds this critical value, near the surface of the semiconductor the conduction band edge drops below the Fermi level. This is the so-called *inversion layer*, in which electrons with energies below $\mu'$ form a highly conductive degenerate Fermi gas. However, typical rates of electron tunneling from the bulk through the depletion layer are very low, so after the inversion layer has been created (say, by the gate voltage application), it may be only populated from another source – hence the hatched blue points in Fig. 8c. This is exactly the fact used in the workhorse device of semiconductor integrated circuits – the *field-effect transistor* (FET) – see Fig. 9.[51]

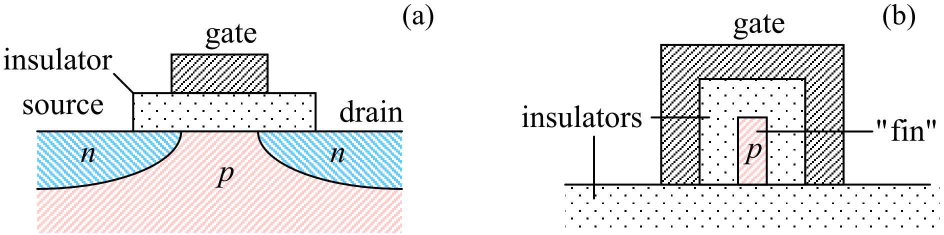


Fig. 6.9. Two main species of the $n$-FET: (a) the *bulk FET*, and (b) the *FinFET*. While on panel a, the current flow from the source to the drain is parallel to the plane of the drawing, on panel b, it is normal to the plane, with the $n$-doped source and drain contacting the thin "fin" from two sides off this plane.

In the "bulk" variety of this structure (Fig. 9a), a gate electrode overlaps a gap between two similar highly-$n$-doped regions near the surface, called *source* and *drain*, formed by $n$-doping inside a $p$-doped semiconductor. It should be more or less obvious (and will be shown in a moment) that in the absence of gate voltage, the electrons cannot pass through the $p$-doped region, so virtually no current flows between the source and the drain, even if a modest voltage is applied between these electrodes. However, if the gate voltage is positive and large enough to induce the electric field $\mathscr{E} > \mathscr{E}_c$ at the surface of the $p$-doped semiconductor, it creates the inversion layer as shown in Fig. 8c, and the electron current

[51] This device was invented (by Julius E. Lilienfeld) in 1930 but demonstrated experimentally only in the mid-1950s.

between the source and drain electrodes may readily flow through this *surface channel*. (Very unfortunately, in this course I would not have time/space for a detailed analysis of transport properties of this keystone electron device and have to refer the reader to special literature.[52])

Fig. 9a shows that another major (and virtually unavoidable) structure of semiconductor integrated circuits is the famous *p-n junction* – an interface between *p*- and *n*-doped regions. Let us analyze its simple model, in which the interface is in the plane $x = 0$, and the doping profiles $n_D(x)$ and $n_A(x)$ are step-like, making an abrupt jump at the interface:

$$n_A(x) = \begin{cases} n_A = \text{const}, & \text{at } x < 0, \\ 0, & \text{at } x > 0, \end{cases} \qquad n_D(x) = \begin{cases} 0, & \text{at } x < 0, \\ n_D = \text{const}, & \text{at } x > 0. \end{cases} \tag{6.81}$$

(This model is very reasonable for modern integrated circuits where the doping is performed by *implantation* using high-energy ion beams.)

To start with, let us assume that no voltage is applied between the *p*- and *n*-regions, so the system may be in thermodynamic equilibrium. In the equilibrium, the Fermi level $\mu'$ should be flat through the structure, and at $x \to -\infty$ and $x \to +\infty$, where $\phi \to 0$, the level structure has to approach the positions shown, respectively, on panels (a) and (b) of Fig. 7. In addition, the distribution of the electric potential $\phi(x)$, shifting the level structure vertically by $-e\phi(x)$, has to be continuous to avoid unphysical infinite electric fields. With that, we inevitably arrive at the band-edge diagram that is (schematically) shown in Fig. 10.

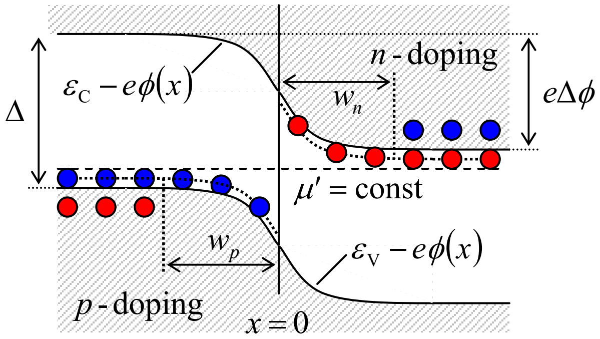


Fig. 6.10. The band-edge diagram of a *p-n* junction in thermodynamic equilibrium ($T$ = const, $\mu'$ = const). The notation is the same as in Figs. 7 and 8.

The diagram shows that the contact of differently doped semiconductors gives rise to a built-in electric potential difference $\Delta\phi$, equal to the difference of their values of $\mu$ in the absence of the contact – see Eqs. (65) and (67):[53]

$$e\Delta\phi \equiv e\phi(+\infty) - e\phi(-\infty) = \mu_n - \mu_p = \Delta - T \ln \frac{n_C n_V}{n_D n_A}, \tag{6.82}$$

[52] The classical monograph in this field is S. Sze, *Physics of Semiconductor Devices*, 2nd ed., Wiley 1981. (The 3rd edition, circa 2006, co-authored with K. Ng, is more tilted toward technical details.) I can also recommend a detailed textbook by R. Pierret, *Semiconductor Device Fundamentals*, 2nd ed., Addison Wesley, 1996.

[53] Frequently, Eq. (82) is also rewritten in the form $e\Delta\varphi = T \ln(n_D n_A / n_i^2)$. In view of the second of Eqs. (60), this equality is formally correct but may be misleading because the intrinsic carrier density $n_i$ is an exponential function of temperature and is physically irrelevant to this particular problem.

which is usually just slightly smaller than the bandgap. (Qualitatively, this is the same contact potential difference that was discussed, for the case of metals, in Sec. 3 – see Fig. 5.) The arising internal electrostatic field $\mathscr{E} = -d\phi/dx$ induces, in both semiconductors, depletion layers similar to that induced by an external field (Fig. 8c). Their widths $w_p$ and $w_n$ may also be calculated similarly, by solving the following boundary problem of electrostatics, mostly similar to that given by Eqs. (78)-(79):

$$\frac{d^2\phi}{dx^2} = \frac{e}{\kappa\varepsilon_0} \times \begin{cases} n_\mathrm{A}, & \text{for } -w_p < x < 0, \\ (-n_\mathrm{D}), & \text{for } 0 < x < +w_n, \end{cases} \tag{6.83}$$

$$\phi(w_n) = \phi(-w_p) + \Delta\phi, \quad \frac{d\phi}{dx}(w_n) = \frac{d\phi}{dx}(-w_p) = 0, \quad \phi(-0) = \phi(+0), \quad \frac{d\phi}{dx}(-0) = \frac{d\phi}{dx}(+0), \tag{6.84}$$

also exact only in the limit $T << \Delta$, $n_\mathrm{i} << n_\mathrm{D}$, $n_\mathrm{A}$. Its (easy) solution gives a result similar to Eq. (80):

$$\phi = \mathrm{const} + \begin{cases} en_\mathrm{A}(w_p + x)^2 / 2\kappa\varepsilon_0, & \text{for } -w_p < x < 0, \\ \Delta\phi - en_\mathrm{D}(w_n - x)^2 / 2\kappa\varepsilon_0, & \text{for } 0 < x < +w_n, \end{cases} \tag{6.85}$$

with expressions for $w_p$ and $w_n$ giving the following formula for the full depletion layer width:

$$w \equiv w_p + w_n = \left(\frac{2\kappa\varepsilon_0\Delta\phi}{en_\mathrm{ef}}\right)^{1/2}, \qquad \text{with } n_\mathrm{ef} \equiv \frac{n_\mathrm{A} n_\mathrm{D}}{n_\mathrm{A} + n_\mathrm{D}}, \quad \text{i.e. } \frac{1}{n_\mathrm{ef}} = \frac{1}{n_\mathrm{A}} + \frac{1}{n_\mathrm{D}}. \tag{6.86}$$

This expression is similar to that given by Eq. (80), so for typical highly doped semiconductors ($n_\mathrm{ef} \sim 10^{18}\mathrm{cm}^{-3}$) it gives for $w$ a similar estimate of a few tens nm.[54] Returning to Fig. 9a, we see that this scale imposes an essential limit on the reduction of bulk FETs (whose scaling down is at the heart of the well-known *Moore's law*),[55] explaining why such high doping is necessary. In the early 2010s, the problems with implementing even higher doping, plus issues with dissipated power management, have motivated the transition of advanced silicon integrated circuit technology from the bulk FETs to the *FinFET* (also called "double-gate", or "tri-gate", or "wrap-around-gate") variety of these devices, schematically shown in Fig. 9b, despite their essentially 3D structure and hence a more complex fabrication technology. In the FinFETs, the role of *p-n* junctions is reduced, but these structures remain an important feature of semiconductor integrated circuits.

Now let us have a look at the *p-n* junction in equilibrium from the point of view of Eq. (52). In the simple model we are considering now (in particular, at $T << \Delta$), this equation is applicable separately to the electron and hole subsystems, because in this model the gases of these charge carriers are classical in all parts of the system, and the *generation-recombination* processes[56] coupling these subsystems have relatively small rates – see below. Hence, for the electron subsystem, we may rewrite Eq. (52) as

$$j_n = n\mu_\mathrm{m} q\mathscr{E} - D_n \frac{\partial n}{\partial x}, \tag{6.87}$$

[54] Note that such $w$ is again much larger than $\lambda_\mathrm{D}$ – the fact that justifies the first two boundary conditions (84).

[55] Another important limit is quantum-mechanical tunneling through the gate insulator, whose thickness has to be scaled down in parallel with lateral dimensions of a FET, including its channel length.

[56] In the semiconductor physics lingo, the "carrier generation" event is the thermal excitation of an electron from the valence band to the conduction band, leaving a hole behind, while the reciprocal event of filling such a hole by a conduction-band electron is called the "carrier recombination".

where $q = -e$. Let us discuss how each term of the right-hand of this equality depends on the system's parameters. Because of the $n$-doping at $x > 0$, there are many more electrons in this part of the system. According to the Boltzmann distribution (58), some number of them,

$$n_> \propto \exp\left\{-\frac{e\Delta\phi}{T}\right\}, \tag{6.88}$$

have energies above the conduction band edge in the $p$-doped part (see Fig. 11a) and try to diffuse into this part through the depletion layer; this diffusion flow of electrons from the $n$-side to the $p$-side of the structure (in Fig. 11, from the right to the left) is described by the second term on the right-hand side of Eq. (87). On the other hand, the intrinsic electric field $\mathscr{E} = -\partial\phi/\partial x$ inside the depletion layer, directed as Fig. 11a shows, exerts on the electrons the force $\mathscr{F} = q\mathscr{E} \equiv -e\mathscr{E}$ pushing them in the opposite direction (from $p$ to $n$), is described by the first, "drift" term on the right-hand side of Eq. (87).[57]

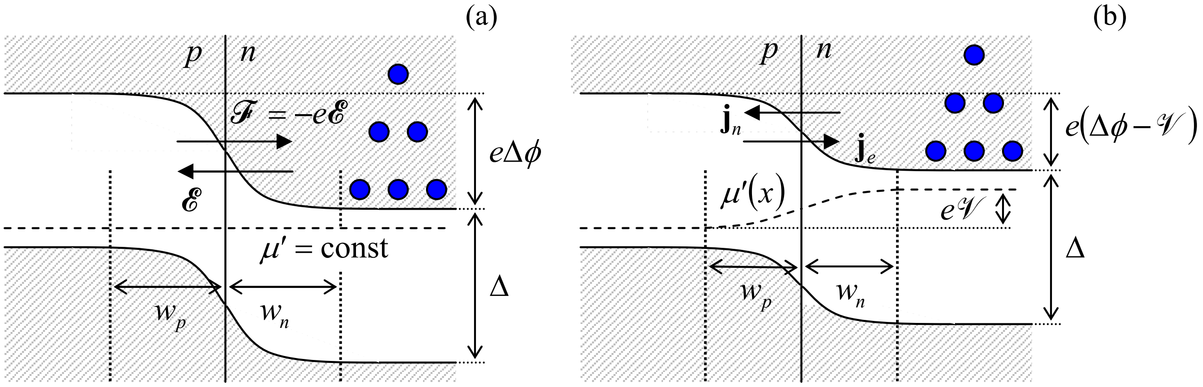


Fig. 6.11. Electrons in the conduction band of a $p$-$n$ junction at: (a) $\mathscr{V} = 0$, and (b) $\mathscr{V} > 0$. For clarity, other charges (of the holes and all ionized dopant atoms) are not shown.

The explicit calculation of these two flows[58] shows, unsurprisingly, that in the equilibrium, they are exactly equal and opposite, so $j_n = 0$, and such analysis does not give us any new information. However, the picture of two electron counter-flows, given by Eq. (87), enables a prediction of the functional dependence of $j_n$ on a modest external voltage $\mathscr{V}$, with $|\mathscr{V}| < \Delta\phi$, applied to the junction. Indeed, since the doped semiconductor regions outside the depletion layer are much more conductive

[57] Note that if an external photon with energy $\hbar\omega > \Delta$ generates an electron-hole pair somewhere inside the depletion layer, this electric field immediately drives its electron component to the right, and the hole component to the left, thus generating a pulse of electric current through the junction. This is the physical basis of the whole vast technological field of *photovoltaics*, currently strongly driven by the demand for renewable electric power. Due to the progress of this technology, the cost of solar power systems has dropped from ~\$300 per watt in the mid-1950s to ~\$1 per watt in 2020, and its global generation is now approaching $10^{15}$ watt-hours per year – though it is still below 2% of the electric power generated by all methods.

[58] I will not try to reproduce this calculation (which may be found in any of the semiconductor physics books mentioned above), because getting all its scaling factors right requires using some model of the recombination process, and in this course, there is just no time for its quantitative discussion. (However, see Eq. (93) below.)

than it, virtually all applied voltage (i.e. the difference of values of the electrochemical potential $\mu'$) drops across this layer, changing the total band edge shift – see Fig. 11b:[59]

$$e\Delta\phi \rightarrow e\Delta\phi + \Delta\mu' \equiv e\Delta\phi + q\mathscr{V} \equiv e(\Delta\phi - \mathscr{V}). \qquad (6.89)$$

This change results in an exponential change of the number of electrons able to diffuse into the $p$-side of the junction – cf. Eq. (88):

$$n_>(\mathscr{V}) \approx n_>(0)\exp\left\{\frac{e\mathscr{V}}{T}\right\}, \qquad (6.90)$$

and hence in a proportional change of the diffusion flow $j_n$ of electrons from the $n$-side to the $p$-side of the system, i.e. of the oppositely directed density of the electron current $j_e = -ej_n$ – see Fig. 11b.

On the other hand, the drift counter-flow of electrons is not altered too much by the applied voltage: though it does change the electrostatic field $\mathscr{E} = -\nabla\phi$ inside the depletion layer, and also the depletion layer width,[60] these changes are incremental, not exponential. As a result, the net density of the current carried by electrons may be approximately expressed as

$$j_e(\mathscr{V}) = j_{\text{diffusion}} - j_{\text{drift}} \approx j_e(0)\exp\left\{\frac{e\mathscr{V}}{T}\right\} - \text{const.} \qquad (6.91a)$$

As was discussed above, at $\mathscr{V} = 0$, the net current has to vanish, so the constant in Eq. (91a) has to equal $j_e(0)$, and we may rewrite this equality as

$$j_e(\mathscr{V}) = j_e(0)\left(\exp\left\{\frac{e\mathscr{V}}{T}\right\} - 1\right). \qquad (6.91b)$$

Now repeating this analysis for the current $j_h$ of the holes (the exercise highly recommended to the reader), we get a similar expression, with the *same* sign before $e\mathscr{V}$,[61] though with a different scaling factor, $j_h(0)$ instead of $j_e(0)$. As a result, the total electric current density obeys the famous *Shockley law*

$$j(\mathscr{V}) \equiv j_e(\mathscr{V}) + j_h(\mathscr{V}) = j(0)\left(\exp\left\{\frac{e\mathscr{V}}{T}\right\} - 1\right), \quad \text{with} \quad j(0) \equiv j_e(0) + j_h(0), \qquad (6.92)$$

describing the main $p$-$n$ junction's property as an *electric diode* – a two-terminal device passing the current more "readily" in one direction (from the $p$- to the $n$-terminal) than in the opposite one.[62]

[59] In our model, the positive sign of $\mathscr{V} \equiv \Delta\mu'/q \equiv -\Delta\mu'/e$ corresponds to the additional electric field, $-\nabla\mu'/q \equiv \nabla\mu'/e$, directed in the positive direction of the $x$-axis (in Fig. 11, from the left to the right), i.e. to the positive terminal of the voltage source connected to the $p$-doped semiconductor – which is the common convention.

[60] This change, schematically shown in Fig. 11b, may be readily calculated by making the replacement (89) in the first of Eqs. (86).

[61] This sign invariance may look strange, due to the opposite (positive) electric charge of the holes. However, this difference in the charge sign is compensated by the opposite direction of the hole diffusion – see Fig. 10. (Note also that the actual charge carriers in the valence band are still electrons, and the effective positive charge of holes is just a convenient representation of the specific dispersion law in this energy band, with a negative effective mass – see Fig. 6, the second line of Eq. (53), and a more detailed discussion of this issue in QM Sec. 2.8.)

Besides numerous practical applications in electrical and electronic engineering, diodes have very interesting statistical properties, in particular performing very non-trivial transformations of the spectra of deterministic and random signals. Very unfortunately, I would not have time for their discussion and have to refer the interested reader to the special literature.[63]

Still, before proceeding to our next (and last!) topic, let me give for the reader's reference, without proof, the expression for the scaling factor $j(0)$ in Eq. (92), which follows from a simple but broadly used model of the recombination process:

$$j(0) = en_{\mathrm{i}}^2\left(\frac{D_{\mathrm{e}}}{l_{\mathrm{e}}n_{\mathrm{A}}} + \frac{D_{\mathrm{h}}}{l_{\mathrm{h}}n_{\mathrm{D}}}\right). \qquad (6.93)$$

Here $l_{\mathrm{e}}$ and $l_{\mathrm{h}}$ are the characteristic lengths of diffusion of electrons and holes before their recombination, which may be expressed by Eq. (5.113), $l_{\mathrm{e}} = (2D_{\mathrm{e}}\tau_{\mathrm{e}})^{1/2}$ and $l_{\mathrm{h}} = (2D_{\mathrm{h}}\tau_{\mathrm{h}})^{1/2}$, with $\tau_{\mathrm{e}}$ and $\tau_{\mathrm{h}}$ being the characteristic times of recombination of the so-called *minority carriers*: of electrons in the $p$-doped part and of holes in the $n$-doped part of the structure. Since the recombination is an inelastic process, its times are typically rather long – of the order of $10^{-7}$s, i.e. much longer than the typical times of elastic scattering of the same carriers, which define their diffusion coefficients – see Eq. (51).

## 6.5. Heat transfer and thermoelectric effects

Now let us return to our analysis of kinetic effects using the Boltzmann-RTA equation, and extend it even further, to the effects of a non-zero (albeit small) temperature gradient. Again, since for any of the statistics (20), the average occupancy $\langle N(\varepsilon)\rangle$ is a function of just one combination of all its arguments, $\xi \equiv (\varepsilon - \mu)/T$, its partial derivatives obey not only Eq. (37) but also the following relation:

$$\frac{\partial\langle N(\varepsilon)\rangle}{\partial T} = -\frac{\varepsilon-\mu}{T^2}\frac{\partial\langle N(\varepsilon)\rangle}{\partial\xi} = \frac{\varepsilon-\mu}{T}\frac{\partial\langle N(\varepsilon)\rangle}{\partial\mu}. \qquad (6.94)$$

As a result, Eq. (38) is generalized as

$$\nabla w_0 = -\frac{\partial w_0}{\partial\varepsilon}\left(\nabla\mu + \frac{\varepsilon-\mu}{T}\nabla T\right), \qquad (6.95)$$

giving the following generalization of Eq. (39):

$$\tilde{w} = \tau\frac{\partial w_0}{\partial\varepsilon}\mathbf{v}\cdot\left(\nabla\mu' + \frac{\varepsilon-\mu}{T}\nabla T\right). \qquad (6.96)$$

Now, calculating the current density as in Sec. 3, we get the result that is traditionally represented as

$$\mathbf{j} = \sigma\left(-\frac{\nabla\mu'}{q}\right) + \sigma\mathcal{S}(-\nabla T), \qquad (6.97)$$

[62] Some metal-semiconductor junctions, called *Schottky diodes*, have similar rectifying properties (and may be better fitted for high-power applications than silicon $p$-$n$ junctions), but their properties are more complex because of the rather involved chemistry and physics of interfaces between different materials.

[63] See, e.g., the monograph by R. Stratonovich cited in Sec. 4.2.

where the constant $\mathcal{S}$, called the *Seebeck coefficient*[64] (or the "thermoelectric power", or just "thermopower") is given by the following relation:

Seebeck coefficient

$$\boxed{\sigma\mathcal{S} = \frac{gq\tau}{(2\pi\hbar)^3}\frac{4\pi}{3}\int_0^\infty \left(8m\varepsilon^3\right)^{1/2}\frac{(\varepsilon-\mu)}{T}\left[-\frac{\partial\langle N(\varepsilon)\rangle}{\partial\varepsilon}\right]d\varepsilon .} \qquad (6.98)$$

Working out this integral for the most important case of a degenerate Fermi gas, with $T << \varepsilon_F$, we have to be careful because the center of the sharp peak of the last factor under the integral coincides with the zero point of the previous factor, $(\varepsilon-\mu)/T$. This uncertainty may be resolved using the Sommerfeld expansion formula (3.59). Indeed, for a smooth function $f(\varepsilon)$ obeying Eq. (3.60), so $f(0) = 0$, we may use Eq. (3.61) to rewrite Eq. (3.59) as

$$\int_0^\infty f(\varepsilon)\left[-\frac{\partial\langle N(\varepsilon)\rangle}{\partial\varepsilon}\right]d\varepsilon = f(\mu) + \frac{\pi^2T^2}{6}\frac{d^2f(\varepsilon)}{d\varepsilon^2}\Big|_{\varepsilon=\mu} . \qquad (6.99)$$

In particular, for working out the integral (98), we may take $f(\varepsilon) \equiv (8m\varepsilon^3)^{1/2}(\varepsilon-\mu)/T$. (For this function, the condition $f(0) = 0$ is evidently satisfied.) Then $f(\mu) = 0$, $d^2f/d\varepsilon^2|_{\varepsilon=\mu} = 3(8m\mu)^{1/2}/T \approx 3(8m\varepsilon_F)^{1/2}/T$, and Eq. (98) yields

$$\sigma\mathcal{S} = \frac{gq\tau}{(2\pi\hbar)^3}\frac{4\pi}{3}\frac{\pi^2T^2}{6}\frac{3\left(8m\varepsilon_F\right)^{1/2}}{T} . \qquad (6.100)$$

Comparing the result with Eqs. (3.54) and (32), for the constant $\mathcal{S}$ we get a simple expression independent of $\tau$:[65]

$$\mathcal{S} = \frac{\pi^2}{2q}\frac{T}{\varepsilon_F} = \frac{c_V}{q}, \qquad \text{for } T << \varepsilon_F , \qquad (6.101)$$

where $c_V \equiv C_V/N$ is the heat capacity of the gas per unit particle, in this case given by Eq. (3.70).

In order to understand the physical meaning of the Seebeck coefficient, it is sufficient to consider a conductor carrying no current. For this case, Eq. (97) yields

Seebeck effect

$$\boxed{\nabla\left(\mu'/q + \mathcal{S}T\right) = 0 .} \qquad (6.102)$$

So, at these conditions, a temperature gradient creates a proportional gradient of the electrochemical potential $\mu'$, and hence the effective electric field $\mathscr{E}$ defined by Eq. (42). This is the *Seebeck effect*. Figure 12 shows the standard way of its measurement, using an ordinary (electrodynamic) voltmeter that measures the difference of $\mu'/e$ at its terminals, and a pair of junctions (in this context, called the *thermocouple*) of two materials with different coefficients $\mathcal{S}$.

[64] Named after Thomas Johann Seebeck who experimentally discovered, in 1822, the effect described by the second term in Eq. (97) – and hence by Eq. (103).

[65] Again, such independence hints that Eq. (101) has a broader validity than in our simple model of an isotropic gas. This is indeed the case: this result turns out to be valid for any form of the Fermi surface and for any dispersion law $\varepsilon(\mathbf{p})$. Note, however, that all calculations of this section are valid for the simplest RTA model in that $\tau$ is an energy-independent parameter; for real metals, a more accurate description of experimental results may be obtained by tweaking this model to take this dependence into account – see, e.g., Chapter 13 in the monograph by N. Ashcroft and N. D. Mermin, which was cited in Sec. 3.5.

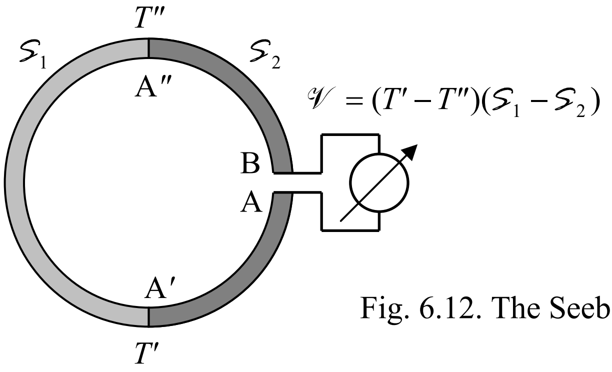


Fig. 6.12. The Seebeck effect in a thermocouple.

Integrating Eq. (102) around the loop from point *A* to point *B*, and neglecting the temperature drop across the voltmeter, we get the following simple expression for the thermally-induced difference of the electrochemical potential, usually called either the *thermoelectric power* or the “thermo e.m.f.”:

$$\begin{aligned}\mathscr{V} &\equiv \frac{\mu'_\mathrm{B}}{q} - \frac{\mu'_\mathrm{A}}{q} = \frac{1}{q}\int_\mathrm{A}^\mathrm{B} \nabla\mu' \cdot d\mathbf{r} = -\int_\mathrm{A}^\mathrm{B} \mathcal{S}\nabla T \cdot d\mathbf{r} = -\mathcal{S}_1 \int_{\mathrm{A}'}^{\mathrm{A}''} \nabla T \cdot d\mathbf{r} - \mathcal{S}_2 \left( \int_\mathrm{A}^{\mathrm{A}'} \nabla T \cdot d\mathbf{r} + \int_{\mathrm{A}''}^\mathrm{B} \nabla T \cdot d\mathbf{r} \right) \\ &= -\mathcal{S}_1\left(T'' - T'\right) - \mathcal{S}_2\left(T' - T''\right) \equiv (\mathcal{S}_1 - \mathcal{S}_2)(T' - T'')\,.\end{aligned} \tag{6.103}$$

(Note that according to Eq. (103), any attempt to measure such voltage across any two points of a *uniform* conductor would give results depending on the voltmeter wire materials, due to an unintentional gradient of temperature in them.)

Using thermocouples is a very popular, inexpensive method of temperature measurement – especially in the a-few-hundred-°C range, where gas- and fluid-based thermometers are not too practicable – if a 1°C-scale accuracy is sufficient. The *temperature responsivity* ($\mathcal{S}_1$ – $\mathcal{S}_2$) of a popular thermocouple, chromel-constantan,[66] is about 70 μV/°C. To understand why the typical values of $\mathcal{S}$ are so small, let us discuss the Seebeck effect’s physics. Superficially, it is very simple: the particles heated by an external source, diffuse from it toward the colder parts of the conductor, creating an electric current if they are electrically charged. However, this naïve argument neglects the fact that at $\mathbf{j} = 0$, there is no total flow of particles. For a more accurate interpretation, note that inside the integral (98), the Seebeck effect is described by the factor $(\varepsilon - \mu)/T$, which changes its sign at the Fermi surface, i.e. at the same energy where the term $[-\partial\langle N(\varepsilon)\rangle/\partial\varepsilon]$, describing the availability of quantum states for transport (due to their intermediate occupancy $0 < \langle N(\varepsilon)\rangle < 1$), reaches its peak. The only reason why that integral does not vanish completely, and hence $\mathcal{S} \neq 0$, is the growth of the first factor under the integral (which describes the density of available quantum states on the energy scale) with $\varepsilon$, so the hotter particles (with $\varepsilon > \mu$) are more numerous and hence carry more heat than the colder ones carry in the opposite direction.

The Seebeck effect is not the only result of a temperature gradient; the same diffusion of particles also causes the less subtle effect of *heat flow* from the region of higher *T* to that with lower *T*, i.e. the effect of *thermal conductivity*, well-known from our everyday practice. The density of this flow

[66] Both these materials are *alloys*, i.e. solid solutions: chromel is 10% chromium in 90% nickel, while constantan is 45% nickel and 55% copper.

(i.e. that of thermal energy) may be calculated similarly to that of the electric current – see Eq. (26), with the natural replacement of the electric charge $q$ of each particle with its thermal energy $(\varepsilon - \mu)$:

$$\mathbf{j}_\text{h} = \int (\varepsilon - \mu)\mathbf{v} w d^3 p . \qquad (6.104)$$

(Indeed, we may look at this expression is as at the difference between the total energy flow density, $\mathbf{j}_\varepsilon = \int \varepsilon \mathbf{v} w d^3 p$, and the product of the average energy needed to add a particle to the system ($\mu$) by the particle flow density, $\mathbf{j}_n = \int \mathbf{v} w d^3 p \equiv \mathbf{j}/q$.)[67] Again, at equilibrium ($w = w_0$) the heat flow vanishes, so $w$ in Eq. (104) may be replaced with its perturbation $\tilde{w}$ that already has been calculated – see Eq. (96). The substitution of that expression into Eq. (104), and its transformation exactly similar to the one performed above for the electric current $\mathbf{j}$, yields

$$\mathbf{j}_\text{h} = \sigma \Pi \left( -\frac{\nabla \mu'}{q} \right) + \kappa (-\nabla T), \qquad (6.105)$$

with the coefficients $\Pi$ and $\kappa$ given, in our approximation, by the following formulas:

Peltier coefficient

$$\boxed{\sigma \Pi = \frac{g q \tau}{(2\pi\hbar)^3} \frac{4\pi}{3} \int_0^\infty \left(8 m \varepsilon^3\right)^{1/2} (\varepsilon - \mu) \left[ -\frac{\partial \langle N(\varepsilon) \rangle}{\partial \varepsilon} \right] d\varepsilon ,} \qquad (6.106)$$

Thermal conductivity

$$\boxed{\kappa = \frac{g \tau}{(2\pi\hbar)^3} \frac{4\pi}{3} \int_0^\infty \left(8 m \varepsilon^3\right)^{1/2} \frac{(\varepsilon - \mu)^2}{T} \left[ -\frac{\partial \langle N(\varepsilon) \rangle}{\partial \varepsilon} \right] d\varepsilon .} \qquad (6.107)$$

Besides the missing factor $T$ in the denominator, the integral in Eq. (106) is the same as the one in Eq. (98), so the constant $\Pi$ (called the *Peltier coefficient*[68]), is simply and fundamentally related to the Seebeck coefficient:[69]

$\Pi$ vs. $\mathcal{S}$

$$\boxed{\Pi = \mathcal{S} T .} \qquad (6.108)$$

[67] An alternative explanation of the factor $(\varepsilon - \mu)$ in Eq. (104) is that according to Eqs. (1.37) and (1.56), for a uniform system of $N$ particles this factor is just $(E - G)/N \equiv (TS - PV)/N$. The full differential of the numerator is $TdS + SdT - PdV - VdP$, so in the absence of the mechanical work $d\mathcal{W} = -PdV$, and changes of temperature and pressure, it is just $TdS \equiv dQ$ – see Eq. (1.19).

[68] Named after Jean Charles Athanase Peltier who experimentally discovered, in 1834, the effect expressed by the first term in Eq. (105) – and hence by Eq. (112).

[69] This extremely simple relation (first discovered experimentally in 1854 by W. Thompson, a.k.a. Lord Kelvin) is frequently considered as the most prominent example of the so-called *Onsager's reciprocal relations* between kinetic coefficients, first suggested by L. Onsager in 1931. Unfortunately, the common derivation of these relations, reproduced in even very popular textbooks, assumes without proof that the mutual correlation function of statistical averages of thermodynamic variables have the same time-reversal symmetry as that of the underlying microscopic variables. As was argued, among others, by R. Zwanzig, *J. Chem. Phys*. **40**, 2527 (1964), this assumption may be plausibly justified for the processes that, by their physical nature, lack very fast fluctuations, such as the volume fluctuations discussed in Sec. 5.3, but not for those that feature them – see the discussion of pressure fluctuations in the same section, and the solution of Problem 5.15. Unfortunately, I would have no time/space for a sufficiently rigorous discussion of this interesting topic, and have to refer the reader to the corresponding literature including B. Coleman and C. Truesdell, *J. Chem. Phys*. **33**, 28 (1960), R. Zwanzig, *Annu. Rev. Phys. Chem*. **16**, 67 (1965), and U. Geigenmüller *et al*., *Physica A* **119**, 53 (1983).

On the other hand, the integral in Eq. (107) is different, but may be readily calculated\, for the most important case of a degenerate Fermi gas, using the Sommerfeld expansion in the form (99), with $f(\varepsilon) \equiv (8m\varepsilon^3)^{1/2}(\varepsilon - \mu)^2/T$, for which $f(\mu) = 0$ and $d^2f/d\varepsilon^2|_{\varepsilon=\mu} = 2(8m\mu^3)^{1/2}/T \approx 2(8m\varepsilon_F{}^3)^{1/2}/T$, so

$$\kappa = \frac{g\tau}{(2\pi\hbar)^3}\frac{4\pi}{3}\frac{\pi^2}{6}T^2\frac{2(8m\varepsilon_F^3)^{1/2}}{T} \equiv \frac{\pi^2}{3}\frac{n\tau T}{m}. \qquad (6.109)$$

Comparing the result with Eq. (32), we get the so-called *Wiedemann-Franz law*[70]

$$\boxed{\frac{\kappa}{\sigma} = \frac{\pi^2}{3}\frac{T}{q^2}.} \qquad (6.110)$$

Wiedemann-Franz law

This relation between the electric conductivity $\sigma$ and the *thermal conductivity* $\kappa$ is more general than our formal derivation might imply. Indeed, it may be shown that the Wiedemann-Franz law is also valid for an arbitrary anisotropy (i.e. an arbitrary Fermi surface shape) and, moreover, well beyond the relaxation-time approximation. (For example, it is also valid for the scattering integral (12) with an arbitrary angular dependence of rate $\Gamma$, provided that the scattering is elastic.) Experiments show that the law is well obeyed by most metals, but only at relatively low temperatures when the thermal conductance due to electrons is well above the one due to lattice vibrations, i.e. phonons – see Sec. 2.6. Moreover, in the context of the definition (105) of the coefficient $\kappa$, for a non-degenerate gas, Eq. (107) should be treated with the utmost care: for the most practicable measurements of thermal conductivity, it has to be modified. (Let me leave this issue for the reader's analysis.)

Now let us discuss the effects described by Eq. (105), starting from the less obvious, first term on its right-hand side. It describes the so-called *Peltier effect*, which may be measured in the loop geometry similar to that shown in Fig. 12, but now driven by an external voltage source – see Fig. 13.

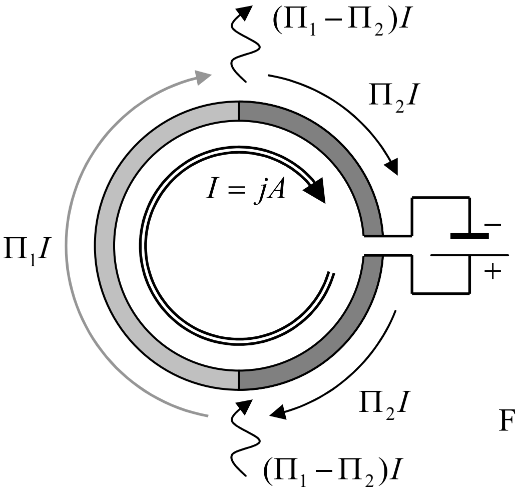


Fig. 6.13. The Peltier effect at $T$ = const.

[70] It was named after Gustav Wiedemann and Rudolph Franz who noticed the constancy of ratio $\kappa/\sigma$ for various materials, at the same temperature, as early as 1853. The direct proportionality of the ratio to the absolute temperature was noticed by Ludwig Lorenz in 1872. Due to his contribution, the Wiedemann-Franz law is frequently represented, in the SI temperature units, as $\kappa/\sigma = LT_K$, where the constant $L \equiv (\pi^2/3)k_B/e^2$, called the *Lorenz number*, is close to $2.45\times10^{-8}$W·Ω·K$^{-2}$. Theoretically, Eq. (110) was derived in 1928 by A. Sommerfeld.

The voltage drives a certain dc current $I = jA$ (where $A$ is the area of the conductor's cross-section), necessarily the same in the whole loop. However, according to Eq. (105), if materials 1 and 2 are different, the power $\mathscr{P} = j_{\rm h}A$ of the associated heat flow is different in the two parts of the loop.[71] Indeed, if the whole system is kept at the same temperature ($\nabla T = 0$), the integration of that relation over the cross-sections of each part yields

$$\mathscr{P}_{1,2} = \Pi_{1,2}A_{1,2}\sigma_{1,2}\left(-\frac{\nabla\mu'}{q}\right)_{1,2} = \Pi_{1,2}A_{1,2}j_{1,2} = \Pi_{1,2}I_{1,2} = \Pi_{1,2}I\,, \qquad (6.111)$$

where, at the second step, Eq. (41) for the electric current density has been used. This equality means that to sustain the constant temperature, the following power difference,

Peltier effect

$$\boxed{\Delta\mathscr{P} = (\Pi_1 - \Pi_2)I\,,} \qquad (6.112)$$

has to be extracted from one junction of the two materials (in Fig. 13, shown on the top), and inserted into the counterpart junction.

If a constant temperature is not maintained, the former junction is heated (in excess of the bulk, Joule heating), while the latter one is cooled, thus implementing a thermoelectric heat pump/refrigerator. Such *Peltier refrigerators* (also called "thermoelectric coolers") which require neither moving parts nor fluids, are very convenient for modest (by a few tens °C) cooling of relatively small components of various systems – from sensitive radiation detectors on mobile platforms (including spacecraft), all the way to cold drinks in vending machines. It is straightforward (and hence is left for the reader) to use the above formulas to show that the practical efficiency of active materials used in such thermoelectric refrigerators may be characterized by the following dimensionless figure-of-merit,

$$\mathrm{ZT} \equiv \frac{\sigma\mathscr{S}^2}{\kappa}T\,. \qquad (6.113)$$

For the best thermoelectric materials found so far, the values of ZT at room temperature are close to 2, providing the $\mathrm{COP}_{\rm cooling}$, defined by Eq. (1.69), of the order of 20% of the Carnot limit (1.70), i.e. a few times lower than that of traditional refrigerators using mechanical compressors. The search for composite materials (including those with nanoparticles) with higher values of ZT is an active field of applied solid-state physics.[72] Another currently explored idea in this field is to reduce $\kappa$ (and hence to increase ZT) radically by replacing the electron diffusion with their transfer through vacuum gaps.

Finally, let us discuss the second term of Eq. (105), in the absence of $\nabla\mu'$ (and hence of the electric current) giving

Fourier law

$$\boxed{\mathbf{j}_{\rm h} = -\kappa\nabla T,} \qquad (6.114)$$

This equality should be familiar to the reader because it describes the very common effect of *thermal conductivity*. Indeed, this linear relation[73] is much more general than the particular expression (107) for

[71] Let me emphasize that here we are discussing the heat *transferred* through a conductor, not the Joule heat *generated* in it by the current. (The latter effect is quadratic, rather than linear, in current, and hence is much smaller at $I \to 0$.)

[72] See, e.g., D. Rowe (ed.), *Thermoelectrics Handbook: Macro to Nano*, CRC Press, 2005.

[73] It was suggested (in 1822) by the same universal scientific genius J.-B. J. Fourier who has not only developed such a key mathematical tool as the Fourier series but also discovered what is now called the greenhouse effect!

$\kappa$: for sufficiently small temperature gradients it is valid for virtually any medium – for example, for insulators. (Table 6.1 gives typical values of $\kappa$ for most common and/or representative materials.) Due to its universality and importance, Eq. (114) has deserved its own name – the *Fourier law*.

Acting absolutely similarly to the derivation of other continuity equations, such as Eqs. (5.117) for the classical probability, and Eq. (49) for the electric charge,[74] let us consider the conservation of the aggregate variable corresponding to $\mathbf{j}_\mathrm{h}$ – the internal energy $E$ within a time-independent volume $V$. According to the basic Eq. (1.18), in the absence of media's expansion ($dV = 0$ and hence $d\mathscr{W} = 0$), the energy change[75] has only the thermal component, so its only cause may be the heat flow through its boundary surface $S$:

$$\frac{dE}{dt} = -\oint_S \mathbf{j}_\mathrm{h} \cdot d^2\mathbf{r} . \tag{6.115}$$

In the simplest case of thermally-independent heat capacity $C_V$, we may integrate Eq. (1.22) over temperature to write[76]

$$E = C_V T = \int_V c_V T d^3 r , \tag{6.116}$$

where $c_V$ is the volumic specific heat, i.e. the heat capacity per unit volume – see the rightmost column in Table 6.1.

Table 6.1. Approximate values of two major thermal coefficients of some materials at 20°C.

| Material | $\kappa$ (W·m$^{-1}$·K$^{-1}$) | $c_V$ (J·K$^{-1}$·m$^{-3}$) |
|---|---|---|
| Air[(a),(b)] | 0.026 | $1.2\times10^3$ |
| Teflon ($[C_2F_4]_n$) | 0.25 | $0.6\times10^6$ |
| Water[(b)] | 0.60 | $4.2\times10^6$ |
| Amorphous silicon dioxide | 1.1–1.4 | $1.5\times10^6$ |
| Undoped silicon | 150 | $1.6\times10^6$ |
| Aluminum[(c)] | 235 | $2.4\times10^6$ |
| Copper[(c)] | 400 | $3.4\times10^6$ |
| Diamond | 2,200 | $1.8\times10^6$ |

[(a)] At ambient pressure.

[(b)] In fluids (gases and liquids), heat flow may be much enhanced by temperature-gradient-induced turbulent circulation – *convection*, which is highly dependent on the system's geometry. The given values correspond to conditions preventing convection.

[(c)] In the context of the Wiedemann-Franz law (valid for metals only!), the values of $\kappa$ for Al and Cu correspond to the Lorenz numbers, respectively, $2.22\times10^{-8}$ W·Ω·K$^{-2}$ and $2.39\times10^{-8}$ W·Ω·K$^{-2}$, in a pretty impressive comparison with the universal theoretical value of $2.45\times10^{-8}$W·Ω·K$^{-2}$ given by Eq. (110).

[74] They are all similar to continuity equations for other quantities – e.g., the mass (see CM Sec. 8.3) and the quantum-mechanical probability (see QM Secs. 1.4 and 9.6).

[75] According to Eq. (1.25), in the case of negligible thermal expansion, it does not matter whether we speak about the internal energy $E$ or the enthalpy $H$.

[76] If the dependence of $c_V$ on temperature may be ignored only within a limited temperature interval, Eqs. (116) and (118) may be still used within that interval, for temperature deviations from some reference value.

Now applying to the right-hand side of Eq. (115) the divergence theorem,[77] and taking into account that for a time-independent volume, the full and partial derivatives over time are equivalent, we get

$$\int_V \left( c_V \frac{\partial T}{\partial t} + \nabla \cdot \mathbf{j}_h \right) d^3 r = 0 . \qquad (6.117)$$

This equality should hold for any time-independent volume $V$, which is possible only if the function under the integral equals zero at any point. Using Eq. (114), we get the following partial differential equation, called the *heat conduction equation* (or, rather inappropriately, the "heat equation"):

Heat conduction equation

$$\boxed{c_V(\mathbf{r}) \frac{\partial T}{\partial t} - \nabla \cdot \left[ \kappa(\mathbf{r}) \nabla T \right] = 0 ,} \qquad (6.118)$$

where the spatial arguments of the coefficients $c_V$ and $\kappa$ are spelled out to emphasize that this equation is valid even for nonuniform media. (Note, however, that Eq. (114) and hence Eq. (118) are valid only if the medium is *isotropic*.)

In a uniform medium, the thermal conductivity $\kappa$ may be taken out from the external spatial differentiation, and the heat conduction equation becomes mathematically similar to the diffusion equation (5.116), and also to the drift-diffusion equation (50) in the absence of drift ($\nabla U = 0$):

$$\frac{\partial T}{\partial t} = D_T \nabla^2 T, \qquad \text{with } D_T \equiv \frac{\kappa}{c_V} . \qquad (6.119)$$

This means, in particular, that the solutions of these equations, discussed earlier in this course (such as Eqs. (5.112)-(5.113) for the evolution of the delta-functional initial perturbation) are valid for Eq. (119) as well, with the only replacement $D \rightarrow D_T$. This is why I will leave a few other examples of the solution of this equation for the reader's exercise.

Another topic I have to leave for the reader's exercise is making estimates of the kinetic coefficients (such as the $\sigma$, $D$, and $\kappa$ discussed above, and also the *shear viscosity* $\eta$) of a nearly ideal classical gas[78] from simple kinematic arguments, and comparing the results with those following from the Boltzmann-RTA equation.

More generally, let me emphasize again that due to time/space restrictions, in this chapter I was able to barely scratch the surface of physical kinetics.[79]

## 6.6. Exercise problems

6.1. Use the Boltzmann equation in the relaxation-time approximation to derive the Drude formula for the complex ac conductivity $\sigma(\omega)$. Give a physical interpretation of the result's trend at high frequencies.

[77] I hope the reader knows it by heart by now, but if not – see, e.g., MA Eq. (12.2).

[78] Here the term "nearly ideal gas" means that its mean free path $l$ is so large that particle collisions do not significantly affect the basic statistical properties of the gas.

[79] A much more detailed coverage of this important part of physics may be found, for example, in the textbook by L. Pitaevskii and E. Lifshitz, *Physical Kinetics*, Butterworth-Heinemann, 1981. For a discussion of applied aspects of kinetics see, e.g., T. Bergman *et al.*, *Fundamentals of Heat and Mass Transfer*, 7th ed., Wiley, 2011.

6.2. At $t = 0$, similar particles were uniformly distributed in a plane layer of thickness $2a$:

$$n(x,0) = \begin{cases} n_0, & \text{for } -a \leq x \leq +a, \\ 0, & \text{otherwise.} \end{cases}$$

At $t > 0$, the particles are allowed to propagate by diffusion through an unlimited uniform medium. Use the variable separation method[80] to calculate the time evolution of the particle density distribution.

6.3. Solve the previous problem using an appropriate Green's function for the 1D version of the diffusion equation, and discuss the relative convenience of the results.

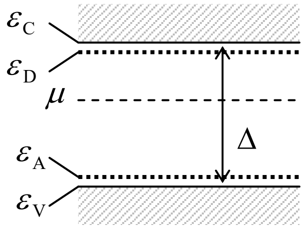
6.4. Particles with the same initial spatial distribution as in the two previous problems are now freed at $t = 0$ to propagate ballistically – without scattering. Calculate the time evolution of their density distribution at $t > 0$, provided that initially, the particles were in thermal equilibrium at temperature $T$. Compare the solution with that of the previous problem.

6.5.* Calculate the electric conductance of a narrow uniform conducting link between two bulk conductors, in the low-voltage and low-temperature limit, neglecting the electron interaction and scattering inside the link.

6.6. Calculate the effective capacitance (per unit area) of a broad plane sheet of a degenerate 2D electron gas, separated by an insulating gap of thickness $d$ from a well-conducting ground plane.

6.7. Give a quantitative description of the dopant atom ionization, which would be consistent with the conduction and valence band occupation statistics, using the same simple model of an $n$-doped semiconductor as in Sec. 4 (see Fig. 7a), and taking into account that the ground state of the dopant atom is typically doubly degenerate, due to two possible spin orientations of the bound electron. Use the results to verify Eq. (65), within the displayed limits of its validity.

6.8. Generalize the solution of the previous problem to the case when the $n$-doping of a semiconductor by $n_D$ donor atoms (per unit volume) is complemented with its simultaneous $p$-doping by $n_A$ acceptor atoms whose energy $\varepsilon_A - \varepsilon_V$ of activation, i.e. of accepting an additional electron and hence becoming a negative ion, is much lower than the bandgap $\Delta$ – see the figure on the right.



6.9. A nearly ideal classical gas of $N$ particles with mass $m$ was in thermal equilibrium at temperature $T$, in a closed container of volume $V$. At some moment, an orifice of a very small area $A$ is opened in one of the container's walls, allowing the particles to escape into the surrounding vacuum.[81] In the limit of very low density $n \equiv N/V$, use simple kinetic arguments to calculate the r.m.s. velocity of the escaped particles during the time period when the total number of such particles is still much smaller than $N$. Formulate the limits of validity of your results in terms of $V$, $A$, and the mean free path $l$.

[80] A detailed introduction to this method may be found, for example, in EM Sec. 2.5.

[81] In chemistry-related fields, this process is frequently called *effusion*.

6.10. For the system analyzed in the previous problem, calculate the rate of particle flow through the orifice – the so-called *effusion rate*. Discuss the limits of validity of your result.

6.11. Use simple kinematic arguments to estimate:

(i) the diffusion coefficient $D$,
(ii) the thermal conductivity $\kappa$, and
(iii) the *shear viscosity* $\eta$,

of a nearly ideal classical gas with mean free path $l$. Compare the result for $D$ with that calculated in Sec. 3 from the Boltzmann-RTA equation.

*Hint*: In fluid dynamics, the shear viscosity (frequently called simply "viscosity") is defined as the coefficient $\eta$ in the following relation:

$$\frac{d\mathscr{F}_{j'}}{dA_j} = \eta \frac{\partial v_{j'}}{\partial r_j}.$$

Here $d\mathscr{F}_{j'}$ is the $j'$ th Cartesian component of the elementary tangential force exerted by one part of a fluid, separated from its counterpart by an imaginary plane normal to some direction $\mathbf{n}_j$ (with $j \neq j'$, and hence $\mathbf{n}_j \perp \mathbf{n}_{j'}$), $dA_j$ is the elementary area of this interface, and $\mathbf{v}(\mathbf{r})$ is the fluid velocity's distribution.[82]

6.12. Use simple kinematic arguments to relate the mean free path $l$ in a nearly ideal classical gas, to the full cross-section $\sigma$ of mutual scattering of its particles.[83] Then use the result to express the thermal conductivity and the viscosity coefficient estimates made in the previous problem, in terms of $\sigma$.

6.13. Use the Boltzmann-RTA equation to calculate the thermal conductivity of a nearly ideal classical gas, measured in conditions when the applied thermal gradient does not create a net particle flow. Compare the result with that following from the simple kinetic arguments (Problem 11).

6.14. Use the Boltzmann-RTA equation to calculate the shear viscosity of a nearly ideal gas. Spell out the result in the classical limit, and compare it with the estimate made in the solution of Problem 11.

6.15. Use a simple model of a thermoelectric refrigerator ("cooler") based on the Peltier effect to analyze its efficiency. In particular, explain why the fraction ZT given by Eq. (6.113) of the lecture notes may be used as the figure-of-merit of materials for such devices.

6.16. Use the heat conduction equation (119) to calculate the amplitude of day-periodic temperature variations at depth $z$ under the surface of the soil with a temperature-independent specific heat $c_V$, thermal conductivity $\kappa$, and negligible thermal expansion. Assume that the incident heat flux is

[82] See, e.g., CM Eq. (8.56). Note the difference between the shear viscosity coefficient $\eta$ considered in this problem and the drag coefficient $\eta$ whose calculation was the task of Problem 3.2. Despite the similar (traditional) notation, and belonging to the same realm (kinematic friction), these coefficients have different definitions and even different dimensionalities.

[83] I am sorry for using the same letter for the cross-section as for the electric Ohmic conductivity. (Both notations are very traditional.) Let me hope this will not lead to confusion; the conductivity is not discussed in this problem.

a sinusoidal function of time, with amplitude $j_0$ per unit area. Estimate the temperature variation amplitude, at depth $z = 1$ m, for a typical dry soil, taking necessary parameters from a reliable source.

6.17. Use Eq. (119) to calculate the time evolution of temperature in the center of a uniform solid sphere of radius $R$, initially heated to a uniformly distributed temperature $T_{\text{ini}}$, and at $t = 0$ placed into a heat bath that gives the sphere's surface a constant temperature $T_0$.

6.18. Suggest a reasonable definition of the entropy production rate (per unit volume), and calculate this rate for stationary thermal conduction, assuming that it obeys the Fourier law, in a material with negligible thermal expansion. Give a physical interpretation of the result. Does the stationary temperature distribution in a sample correspond to the minimum of the total entropy production in it?