deflected by ion cores arranged on a *periodic* lattice because matter waves can propagate freely in a periodic structure, as a consequence of the mathematics treated in the following chapter. (b) A conduction electron is scattered only infrequently by other conduction electrons. This property is a consequence of the Pauli exclusion principle. By a **free electron Fermi gas**, we shall mean a gas of free electrons subject to the Pauli principle.

ENERGY LEVELS IN ONE DIMENSION

Consider a free electron gas in one dimension, taking account of quantum theory and of the Pauli principle. An electron of mass m is confined to a length L by infinite barriers (Fig. 2). The wavefunction $\psi_n(x)$ of the electron is a solution of the Schrödinger equation $\mathcal{H}\psi = \epsilon\psi$; with the neglect of potential energy we have $\mathcal{H} = p^2/2m$, where p is the momentum. In quantum theory p may be represented by the operator $-i\hbar\, d/dx$, so that

$$\mathcal{H}\psi_n = -\frac{\hbar^2}{2m}\frac{d^2\psi_n}{dx^2} = \epsilon_n\psi_n \ , \tag{1}$$

where ϵ_n is the energy of the electron in the orbital.

We use the term **orbital** to denote a solution of the wave equation for a system of only one electron. The term allows us to distinguish between an exact quantum state of the wave equation of a system of N interacting electrons and an approximate quantum state which we construct by assigning the N electrons to N different orbitals, where each orbital is a solution of a wave equation for one electron. The orbital model is exact only if there are no interactions between electrons.

The boundary conditions are $\psi_n(0) = 0$; $\psi_n(L) = 0$, as imposed by the infinite potential energy barriers. They are satisfied if the wavefunction is sinelike with an integral number n of half-wavelengths between 0 and L:

$$\psi_n = A \sin\left(\frac{2\pi}{\lambda_n}x\right) \ ; \qquad \tfrac{1}{2}n\lambda_n = L \ , \tag{2}$$

where A is a constant. We see that (2) is a solution of (1), because

$$\frac{d\psi_n}{dx} = A\left(\frac{n\pi}{L}\right)\cos\left(\frac{n\pi}{L}x\right) \ ; \qquad \frac{d^2\psi_n}{dx^2} = -A\left(\frac{n\pi}{L}\right)^2 \sin\left(\frac{n\pi}{L}x\right) \ ,$$

whence the energy ϵ_n is given by

$$\epsilon_n = \frac{\hbar^2}{2m}\left(\frac{n\pi}{L}\right)^2 \ . \tag{3}$$

We want to accommodate N electrons on the line. According to the **Pauli exclusion principle**, no two electrons can have all their quantum numbers

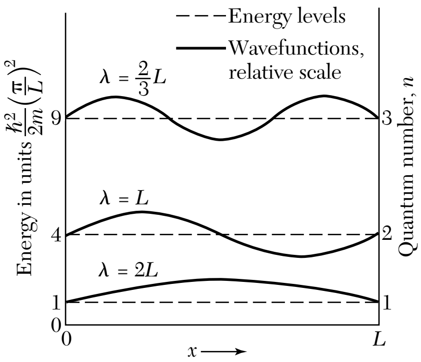


Figure 2 First three energy levels and wavefunctions of a free electron of mass m confined to a line of length L. The energy levels are labeled according to the quantum number n which gives the number of half-wavelengths in the wavefunction. The wavelengths are indicated on the wavefunctions. The energy ϵ_n of the level of quantum number n is equal to $(h^2/2m)(n/2L)^2$.

identical. That is, each orbital can be occupied by at most one electron. This applies to electrons in atoms, molecules, or solids.

In a linear solid the quantum numbers of a conduction electron orbital are n and m_s, where n is any positive integer and the magnetic quantum number $m_s = \pm\frac{1}{2}$, according to spin orientation. A pair of orbitals labeled by the quantum number n can accommodate two electrons, one with spin up and one with spin down.

If there are six electrons, then in the ground state of the system the filled orbitals are those given in the table:

n	m_s	Electron occupancy	n	m_s	Electron occupancy
1	↑	1	3	↑	1
1	↓	1	3	↓	1
2	↑	1	4	↑	0
2	↓	1	4	↓	0

More than one orbital may have the same energy. The number of orbitals with the same energy is called the **degeneracy.**

Let n_F denote the topmost filled energy level, where we start filling the levels from the bottom ($n = 1$) and continue filling higher levels with electrons until all N electrons are accommodated. It is convenient to suppose that N is an even number. The condition $2n_F = N$ determines n_F, the value of n for the uppermost filled level.

The **Fermi energy ϵ_F** is defined as the energy of the topmost filled level in the ground state of the N electron system. By (3) with $n = n_F$ we have in one dimension:

$$\epsilon_F = \frac{\hbar^2}{2m}\left(\frac{n_F\pi}{L}\right)^2 = \frac{\hbar^2}{2m}\left(\frac{N\pi}{2L}\right)^2 \quad . \tag{4}$$

EFFECT OF TEMPERATURE ON THE FERMI-DIRAC DISTRIBUTION

The ground state is the state of the N electron system at absolute zero. What happens as the temperature is increased? This is a standard problem in elementary statistical mechanics, and the solution is given by the Fermi-Dirac distribution function (Appendix D and *TP*, Chapter 7).

The kinetic energy of the electron gas increases as the temperature is increased: some energy levels are occupied which were vacant at absolute zero, and some levels are vacant which were occupied at absolute zero (Fig. 3). The **Fermi-Dirac distribution** gives the probability that an orbital at energy ϵ will be occupied in an ideal electron gas in thermal equilibrium:

$$\boxed{f(\epsilon) = \frac{1}{\exp[(\epsilon - \mu)/k_B T] + 1}\ .} \tag{5}$$

The quantity μ is a function of the temperature; μ is to be chosen for the particular problem in such a way that the total number of particles in the system comes out correctly—that is, equal to N. At absolute zero $\mu = \epsilon_F$, because in the limit $T \to 0$ the function $f(\epsilon)$ changes discontinuously from the value 1 (filled) to the value 0 (empty) at $\epsilon = \epsilon_F = \mu$. *At all temperatures* $f(\epsilon)$ *is equal to* $\frac{1}{2}$ *when* $\epsilon = \mu$, for then the denominator of (5) has the value 2.

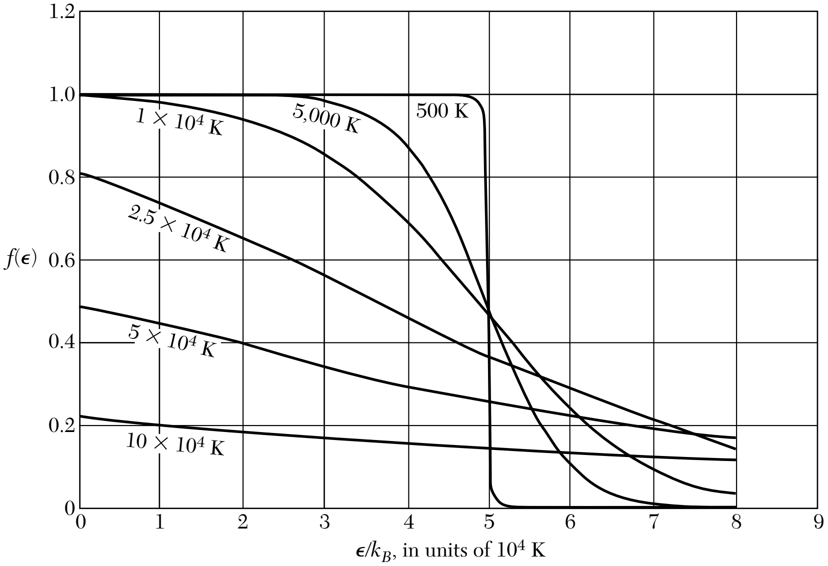


Figure 3 Fermi-Dirac distribution function (5) at the various labelled temperatures, for $T_F \equiv \epsilon_F/k_B = 50{,}000$ K. The results apply to a gas in three dimensions. The total number of particles is constant, independent of temperature. The chemical potential μ at each temperature may be read off the graph as the energy at which $f = 0.5$.

Table 1 Calculated free electron Fermi surface parameters for metals at room temperature

(Except for Na, K, Rb, Cs at 5 K and Li at 78 K)

Valency	Metal	Electron concentration, in cm^{-3}	Radius[a] parameter r_n	Fermi wavevector, in cm^{-1}	Fermi velocity, in cm s^{-1}	Fermi energy, in eV	Fermi temperature $T_F \equiv \epsilon_F/k_B$, in deg K
1	Li	4.70×10^{22}	3.25	1.11×10^{8}	1.29×10^{8}	4.72	5.48×10^{4}
	Na	2.65	3.93	0.92	1.07	3.23	3.75
	K	1.40	4.86	0.75	0.86	2.12	2.46
	Rb	1.15	5.20	0.70	0.81	1.85	2.15
	Cs	0.91	5.63	0.64	0.75	1.58	1.83
	Cu	8.45	2.67	1.36	1.57	7.00	8.12
	Ag	5.85	3.02	1.20	1.39	5.48	6.36
	Au	5.90	3.01	1.20	1.39	5.51	6.39
2	Be	24.2	1.88	1.93	2.23	14.14	16.41
	Mg	8.60	2.65	1.37	1.58	7.13	8.27
	Ca	4.60	3.27	1.11	1.28	4.68	5.43
	Sr	3.56	3.56	1.02	1.18	3.95	4.58
	Ba	3.20	3.69	0.98	1.13	3.65	4.24
	Zn	13.10	2.31	1.57	1.82	9.39	10.90
	Cd	9.28	2.59	1.40	1.62	7.46	8.66
3	Al	18.06	2.07	1.75	2.02	11.63	13.49
	Ga	15.30	2.19	1.65	1.91	10.35	12.01
	In	11.49	2.41	1.50	1.74	8.60	9.98
4	Pb	13.20	2.30	1.57	1.82	9.37	10.87
	Sn(w)	14.48	2.23	1.62	1.88	10.03	11.64

[a] The dimensionless radius parameter is defined as $r_n = r_0/a_H$, where a_H is the first Bohr radius and r_0 is the radius of a sphere that contains one electron.

The quantity μ is the **chemical potential** (*TP*, Chapter 5), and we see that at absolute zero the chemical potential is equal to the Fermi energy, defined as the energy of the topmost filled orbital at absolute zero.

The high energy tail of the distribution is that part for which $\epsilon - \mu \gg k_BT$; here the exponential term is dominant in the denominator of (5), so that $f(\epsilon) \cong \exp[(\mu - \epsilon)/k_BT]$. This limit is called the Boltzmann or Maxwell distribution.

FREE ELECTRON GAS IN THREE DIMENSIONS

The free-particle Schrödinger equation in three dimensions is

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\psi_{\mathbf{k}}(\mathrm{r}) = \epsilon_{\mathbf{k}}\,\psi_{\mathbf{k}}(\mathbf{r}) \ . \qquad (6)$$

If the electrons are confined to a cube of edge L, the wavefunction is the standing wave

$$\psi_n(\mathbf{r}) = A \sin(\pi n_x x/L)\sin(\pi n_y y/L)\sin(\pi n_z z/L) \ , \qquad (7)$$

where n_x, n_y, n_z are positive integers. The origin is at one corner of the cube.

It is convenient to introduce wavefunctions that satisfy periodic boundary conditions, as we did for phonons in Chapter 5. We now require the wavefunctions to be periodic in x, y, z with period L. Thus

$$\psi(x + L, y, z) = \psi(x, y, z) \ , \qquad (8)$$

and similarly for the y and z coordinates. Wavefunctions satisfying the free-particle Schrödinger equation and the periodicity condition are of the form of a traveling plane wave:

$$\boxed{\psi_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r}) \ ,} \qquad (9)$$

provided that the components of the wavevector **k** satisfy

$$k_x = 0 \ ; \quad \pm\frac{2\pi}{L} \ ; \quad \pm\frac{4\pi}{L} \ ; \ \ldots \ , \qquad (10)$$

and similarly for k_y and k_z.

Any component of **k** of the form $2n\pi/L$ will satisfy the periodicity condition over a length L, where n is a positive or negative integer. The components of k are the quantum numbers of the problem, along with the quantum number m_s for the spin direction. We confirm that these values of k_x satisfy (8), for

$$\begin{aligned}\exp[ik_x(x + L)] &= \exp[i2n\pi(x + L)/L] \\ &= \exp(i2n\pi x/L)\exp(i2n\pi) = \exp(i2n\pi x/L) = \exp(ik_x x) \ . \qquad (11)\end{aligned}$$

On substituting (9) in (6) we have the energy $\epsilon_{\mathbf{k}}$ of the orbital with wavevector **k**:

$$\epsilon_{\mathbf{k}} = \frac{\hbar^2}{2m} k^2 = \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2) \ . \tag{12}$$

The magnitude k of the wavevector is related to the wavelength λ by $k = 2\pi/\lambda$.

The linear momentum $\mathbf{p}$ may be represented in quantum mechanics by the operator $\mathbf{p} = -i\hbar\nabla$, whence for the orbital (9)

$$\mathbf{p}\psi_{\mathbf{k}}(\mathrm{r}) = -i\hbar\boldsymbol{\nabla}\psi_{\mathbf{k}}(\mathbf{r}) = \hbar\mathbf{k}\psi_{\mathbf{k}}(\mathbf{r}) \ , \tag{13}$$

so that the plane wave $\psi_{\mathbf{k}}$ is an eigenfunction of the linear momentum with the eigenvalue $\hbar\mathbf{k}$. The particle velocity in the orbital **k** is given by $\mathbf{v} = \hbar\mathbf{k}/m$.

In the ground state of a system of N free electrons, the occupied orbitals may be represented as points inside a sphere in **k** space. The energy at the surface of the sphere is the Fermi energy; the wavevectors at the Fermi surface have a magnitude k_F such that (Fig. 4):

$$\epsilon_F = \frac{\hbar^2}{2m} k_F^2 \ . \tag{14}$$

From (10) we see that there is one allowed wavevector—that is, one distinct triplet of quantum numbers k_x, k_y, k_z—for the volume element $(2\pi/L)^3$ of **k** space. Thus in the sphere of volume $4\pi k_F^3/3$ the total number of orbitals is

$$2 \cdot \frac{4\pi k_F^3/3}{(2\pi/L)^3} = \frac{V}{3\pi^2} k_F^3 = N \ , \tag{15}$$

where the factor 2 on the left comes from the two allowed values of the spin quantum number for each allowed value of **k**. Then (15) gives

$$k_F = \left(\frac{3\pi^2 N}{V}\right)^{1/3} \ , \tag{16}$$

which depends only on the particle concentration.

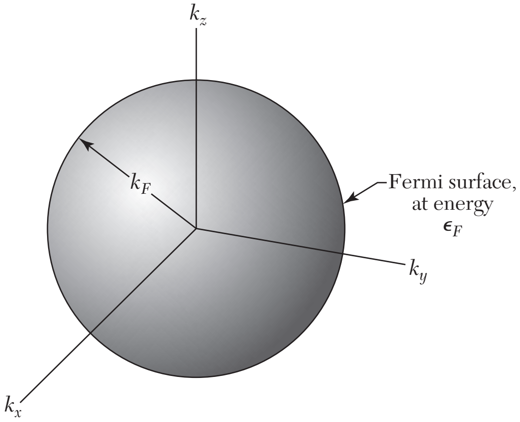


Figure 4 In the ground state of a system of N free electrons the occupied orbitals of the system fill a sphere of radius k_F, where $\epsilon_F = \hbar^2 k_F^2/2m$ is the energy of an electron having a wavevector k_F.

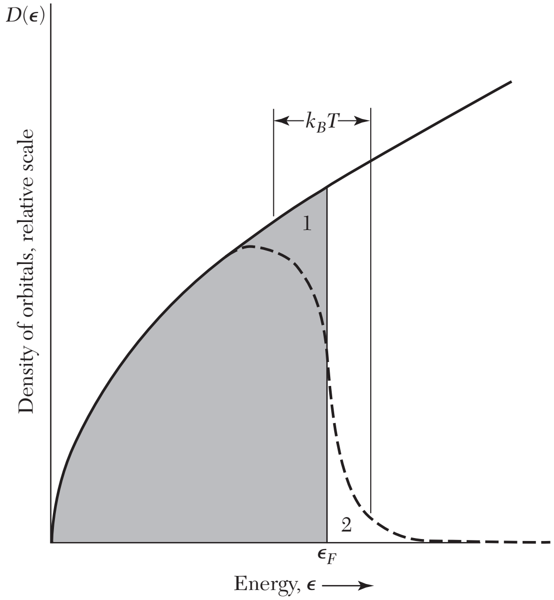


Figure 5 Density of single-particle states as a function of energy, for a free electron gas in three dimensions. The dashed curve represents the density $f(\epsilon, T)D(\epsilon)$ of filled orbitals at a finite temperature, but such that k_BT is small in comparison with ϵ_F. The shaded area represents the filled orbitals at absolute zero. The average energy is increased when the temperature is increased from 0 to T, for electrons are thermally excited from region 1 to region 2.

Using (14) and (16),

$$\boxed{\epsilon_F = \frac{\hbar^2}{2m}\left(\frac{3\pi^2 N}{V}\right)^{2/3} .} \tag{17}$$

This relates the Fermi energy to the electron concentration N/V. The electron velocity v_F at the Fermi surface is

$$v_F = \left(\frac{\hbar k_F}{m}\right) = \left(\frac{\hbar}{m}\right)\left(\frac{3\pi^2 N}{V}\right)^{1/3} . \tag{18}$$

Calculated values of k_F, v_F, and ϵ_F are given in Table 1 for selected metals; also given are values of the quantity T_F which is defined as ϵ_F/k_B. (The quantity T_F has nothing to do with the temperature of the electron gas!)

We now find an expression for the number of orbitals per unit energy range, $D(\epsilon)$, called the **density of states.**[1] We use (17) to obtain the total number of orbitals of energy $\leq\epsilon$:

$$N = \frac{V}{3\pi^2}\left(\frac{2m\epsilon}{\hbar^2}\right)^{3/2} , \tag{19}$$

so that the density of states (Fig. 5) is

$$\boxed{D(\epsilon) \equiv \frac{dN}{d\epsilon} = \frac{V}{2\pi^2}\cdot\left(\frac{2m}{\hbar^2}\right)^{3/2}\cdot\epsilon^{1/2} .} \tag{20}$$

---

[1] Strictly, $D(\epsilon)$ is the density of one-particle states, or density of orbitals.

This result may be expressed more simply by comparing (19) and (20) to obtain at ϵ

$$D(\epsilon) \equiv \frac{dN}{d\epsilon} = \frac{3N}{2\epsilon} \ . \tag{21}$$

Within a factor of the order of unity, the number of orbitals per unit energy range at the Fermi energy is the total number of conduction electrons divided by the Fermi energy, just as we would expect.

HEAT CAPACITY OF THE ELECTRON GAS

The question that caused the greatest difficulty in the early development of the electron theory of metals concerns the heat capacity of the conduction electrons. Classical statistical mechanics predicts that a free particle should have a heat capacity of $\frac{3}{2}k_B$, where k_B is the Boltzmann constant. If N atoms each give one valence electron to the electron gas, and the electrons are freely mobile, then the electronic contribution to the heat capacity should be $\frac{3}{2}Nk_B$, just as for the atoms of a monatomic gas. But the observed electronic contribution at room temperature is usually less than 0.01 of this value.

This important discrepancy distracted the early workers, such as Lorentz: How can the electrons participate in electrical conduction processes as if they were mobile, while not contributing to the heat capacity? The question was answered only upon the discovery of the Pauli exclusion principle and the Fermi distribution function. Fermi found the correct result and he wrote, "One recognizes that the specific heat vanishes at absolute zero and that at low temperatures it is proportional to the absolute temperature."

When we heat the specimen from absolute zero, not every electron gains an energy $\sim k_BT$ as expected classically, but only those electrons in orbitals within an energy range k_BT of the Fermi level are excited thermally, as in Fig. 5. This gives an immediate qualitative solution to the problem of the heat capacity of the conduction electron gas. If N is the total number of electrons, only a fraction of the order of T/T_F can be excited thermally at temperature T, because only these lie within an energy range of the order of k_BT of the top of the energy distribution.

Each of these NT/T_F electrons has a thermal energy of the order of k_BT. The total electronic thermal kinetic energy U is of the order of

$$U_{el} \approx (NT/T_F)k_BT \ . \tag{22}$$

The electronic heat capacity is given by

$$C_{el} = \partial U/\partial T \approx Nk_B(T/T_F) \tag{23}$$

and is directly proportional to T, in agreement with the experimental results discussed in the following section. At room temperature C_{el} is smaller crystals respond to applied fields as if endowed with negative or positive charges, $-e$ or $+e$, and herein lies the explanation of the negative and positive values of the Hall coefficient.

NEARLY FREE ELECTRON MODEL

On the free electron model the allowed energy values are distributed essentially continuously from zero to infinity. We saw in Chapter 6 that

$$\epsilon_{\mathbf{k}} = \frac{\hbar^2}{2m}(k_x^2 + k_y^2 + k_z^2) \ , \tag{1}$$

where, for periodic boundary conditions over a cube of side L,

$$k_x, k_y, k_z = 0 \ ; \quad \pm\frac{2\pi}{L} \ ; \quad \pm\frac{4\pi}{L} \ ; \ \ldots \ . \tag{2}$$

The free electron wavefunctions are of the form

$$\psi_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r}) \ ; \tag{3}$$

they represent running waves and carry momentum $\mathbf{p} = \hbar\mathbf{k}$.

The band structure of a crystal can often be explained by the nearly free electron model for which the band electrons are treated as perturbed only weakly by the periodic potential of the ion cores. This model answers almost all the qualitative questions about the behavior of electrons in metals.

We know that Bragg reflection is a characteristic feature of wave propagation in crystals. Bragg reflection of electron waves in crystals is the cause of energy gaps. (At Bragg reflection wavelike solutions of the Schrödinger equation do not exist, as in Fig. 2.) These energy gaps are of decisive significance in determining whether a solid is an insulator or a conductor.

We explain physically the origin of energy gaps in the simple problem of a linear solid of lattice constant a. The low energy portions of the band structure

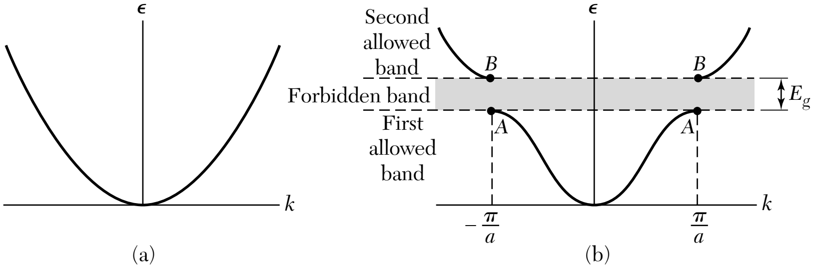


Figure 2 (a) Plot of energy ϵ versus wavevector k for a free electron. (b) Plot of energy versus wavevector for an electron in a monatomic linear lattice of lattice constant a. The energy gap E_g shown is associated with the first Bragg reflection at $k = \pm\pi/a$; other gaps are found at higher energies at $\pm n\pi/a$, for integral values of n.

are shown qualitatively in Fig. 2, in (a) for entirely free electrons and in (b) for electrons that are nearly free, but with an energy gap at $k = \pm\pi/a$. The Bragg condition $(\mathbf{k} + \mathbf{G})^2 = k^2$ for diffraction of a wave of wavevector **k** becomes in one dimension

$$k = \pm\tfrac{1}{2}G = \pm n\pi/a \ , \tag{4}$$

where $G = 2\pi n/a$ is a reciprocal lattice vector and n is an integer. The first reflections and the first energy gap occur at $k = \pm\pi/a$. The region in **k** space between $-\pi/a$ and π/a is the **first Brillouin zone** of this lattice. Other energy gaps occur for other values of the integer n.

The wavefunctions at $k = \pm\pi/a$ are not the traveling waves $\exp(i\pi x/a)$ or $\exp(-i\pi x/a)$ of free electrons. At these special values of k the wavefunctions are made up of equal parts of waves traveling to the right and to the left. When the Bragg reflection condition $k = \pm\pi/a$ is satisfied by the wavevector, a wave traveling to the right is Bragg-reflected to travel to the left, and vice versa. Each subsequent Bragg reflection will reverse the direction of travel of the wave. A wave that travels neither to the right nor to the left is a standing wave: it doesn't go anywhere.

The time-independent state is represented by standing waves. We can form two different standing waves from the two traveling waves

$$\exp(\pm i\pi x/a) = \cos(\pi x/a) \pm i\sin(\pi x/a),$$

so that the standing waves are

$$\begin{aligned} \psi(+) &= \exp(i\pi x/a) + \exp(-i\pi x/a) = 2\cos(\pi x/a) \ ; \\ \psi(-) &= \exp(i\pi x/a) - \exp(-i\pi x/a) = 2i\sin(\pi x/a) \ . \end{aligned} \tag{5}$$

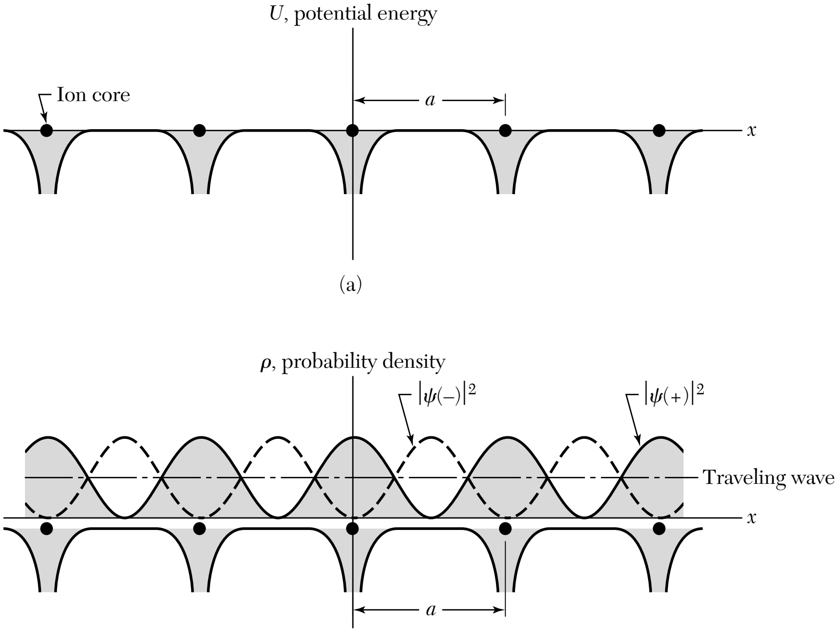
The standing waves are labeled $(+)$ or $(-)$ according to whether or not they change sign when $-x$ is substituted for x. Both standing waves are composed of equal parts of right- and left-directed traveling waves.

Origin of the Energy Gap

The two standing waves $\psi(+)$ and $\psi(-)$ pile up electrons at different regions, and therefore the two waves have different values of the potential energy in the field of the ions of the lattice. This is the origin of the energy gap. The probability density ρ of a particle is $\psi^*\psi = |\psi|^2$. For a pure traveling wave $\exp(ikx)$, we have $\rho = \exp(-ikx)\exp(ikx) = 1$, so that the charge density is constant. The charge density is not constant for linear combinations of plane waves. Consider the standing wave $\psi(+)$ in (5); for this we have

$$\rho(+) = |\psi(+)|^2 \propto \cos^2 \pi x/a \ .$$

This function piles up electrons (negative charge) on the positive ions centered at $x = 0, a, 2a, \ldots$ in Fig. 3, where the potential energy is lowest.



(a)

(b)

Figure 3 (a) Variation of potential energy of a conduction electron in the field of the ion cores of a linear lattice. (b) Distribution of probability density ρ in the lattice for $|\psi(-)|^2 \propto \sin^2 \pi x/a$; $|\psi(+)|^2 \propto \cos^2 \pi x/a$; and for a traveling wave. The wavefunction $\psi(+)$ piles up electronic charge on the cores of the positive ions, thereby lowering the potential energy in comparison with the average potential energy seen by a traveling wave. The wavefunction $\psi(-)$ piles up charge in the region between the ions, thereby raising the potential energy in comparison with that seen by a traveling wave. This figure is the key to understanding the origin of the energy gap.

Figure 3a pictures the variation of the electrostatic potential energy of a conduction electron in the field of the positive ion cores. The ion cores bear a net positive charge because the atoms are ionized in the metal, with the valence electrons taken off to form the conduction band. The potential energy of an electron in the field of a positive ion is negative, so that the force between them is attractive.

For the other standing wave $\psi(-)$ the probability density is

$$\rho(-) = |\psi(-)|^2 \propto \sin^2 \pi x/a \quad ,$$

which concentrates electrons away from the ion cores. In Fig. 3b we show the electron concentration for the standing waves $\psi(+)$, $\psi(-)$, and for a traveling wave.

When we calculate the average or expectation values of the potential energy over these three charge distributions, we find that the potential energy of $\rho(+)$ is lower than that of the traveling wave, whereas the potential energy of $\rho(-)$ is higher than the traveling wave. We have an energy gap of width E_g if the energies of $\rho(-)$ and $\rho(+)$ differ by E_g. Just below the energy gap at points A in Fig. 2 the wavefunction is $\psi(+)$, and just above the gap at points B the wavefunction is $\psi(-)$.

Magnitude of the Energy Gap

The wavefunctions at the Brillouin zone boundary $k = \pi/a$ are $\sqrt{2}\cos\pi x/a$ and $\sqrt{2}\sin\pi x/a$, normalized over unit length of line. Let us suppose that the potential energy of an electron in the crystal at point x is

$$U(x) = U\cos 2\pi x/a \ .$$

The first-order energy difference between the two standing wave states is

$$\begin{aligned} E_g &= \int_0^1 dx\, U(x)\,[|\psi(+)|^2 - |\psi(-)|^2] \\ &= 2\int dx\, U\cos(2\pi x/a)(\cos^2\pi x/a - \sin^2\pi x/a) = U \ . \end{aligned} \tag{6}$$

We see that the gap is equal to the Fourier component of the crystal potential.

BLOCH FUNCTIONS

F. Bloch proved the important theorem that the solutions of the Schrödinger equation for a periodic potential must be of a special form:

$$\boxed{\psi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r})\exp(i\mathbf{k}\cdot\mathbf{r})} \ , \tag{7}$$

where $u_{\mathbf{k}}(\mathbf{r})$ has the period of the crystal lattice with $u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r} + \mathbf{T})$. Here $\mathbf{T}$ is a translation vector of the lattice. The result (7) expresses the Bloch theorem:

> The eigenfunctions of the wave equation for a periodic potential are the product of a plane wave $\exp(i\mathbf{k}\cdot\mathbf{r})$ times a function $u_{\mathbf{k}}(\mathbf{r})$ with the periodicity of the crystal lattice.

A one-electron wavefunction of the form (7) is called a Bloch function and can be decomposed into a sum of traveling waves, as we see later. Bloch functions can be assembled into wave packets to represent electrons that propagate freely through the potential field of the ion cores.

We give now a restricted proof of the Bloch theorem, valid when ψ_k is nondegenerate; that is, when there is no other wavefunction with the same energy and wavevector as ψ_k. The general case will be treated later. We consider N identical lattice points on a ring of length Na. The potential energy is periodic in a, with $U(x) = U(x + sa)$, where s is an integer.

Let us be guided by the symmetry of the ring to look for solutions of the wave equation such that

$$\psi(x + a) = C\psi(x) \ , \tag{8}$$

where C is a constant. Then, on going once around the ring,

$$\psi(x + Na) = \psi(x) = C^N \psi(x) \quad ,$$

because $\psi(x)$ must be single-valued. It follows that C is one of the N roots of unity, or

$$C = \exp(i2\pi s/N) \quad ; \qquad s = 0, 1, 2, \; . \; . \; . \; . \; , N - 1 \quad . \tag{9}$$

We use (9) to see that

$$\psi(x) = u_k(x) \exp(i2\pi sx/Na) \tag{10}$$

satisfies (8), provided that $u_k(x)$ has the periodicity a, so that $u_k(x) = u_k(x + a)$. This is the Bloch result (7).

KRONIG-PENNEY MODEL

A periodic potential for which the wave equation can be solved in terms of elementary functions is the square-well array of Fig. 4. The wave equation is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + U(x)\psi = \epsilon\psi \quad , \tag{11}$$

where $U(x)$ is the potential energy and ϵ is the energy eigenvalue.

In the region $0 < x < a$ in which $U = 0$, the eigenfunction is a linear combination,

$$\psi = Ae^{iKx} + Be^{-iKx} \quad , \tag{12}$$

of plane waves traveling to the right and to the left, with energy

$$\epsilon = \hbar^2K^2/2m \quad . \tag{13}$$

In the region $-b < x < 0$ within the barrier the solution is of the form

$$\psi = Ce^{Qx} + De^{-Qx} \quad , \tag{14}$$

with

$$U_0 - \epsilon = \hbar^2Q^2/2m \quad . \tag{15}$$

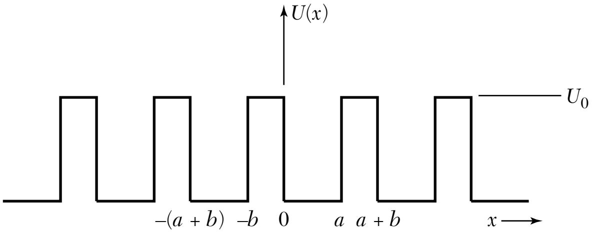


Figure 4 Square-well periodic potential as introduced by Kronig and Penney.

We want the complete solution to have the Bloch form (7). Thus the solution in the region $a < x < a + b$ must be related to the solution (14) in the region $-b < x < 0$ by the Bloch theorem:

$$\psi(a < x < a + b) = \psi(-b < x < 0)\, e^{ik(a+b)} \ , \tag{16}$$

which serves to define the wavevector k used as an index to label the solution.

The constants A, B, C, D are chosen so that ψ and $d\psi/dx$ are continuous at $x = 0$ and $x = a$. These are the usual quantum mechanical boundary conditions in problems that involve square potential wells. At $x = 0$,

$$A + B = C + D \ ; \tag{17}$$

$$iK(A - B) = Q(C - D) \ , \tag{18}$$

with Q from (14). At $x = a$, with the use of (16) for $\psi(a)$ under the barrier in terms of $\psi(-b)$,

$$Ae^{iKa} + Be^{-iKa} = (Ce^{-Qb} + De^{Qb})\, e^{ik(a+b)} \ ; \tag{19}$$

$$iK(Ae^{iKa} - Be^{-iKa}) = Q(Ce^{-Qb} - De^{Qb})\, e^{ik(a+b)} \ . \tag{20}$$

The four equations (17) to (20) have a solution only if the determinant of the coefficients of A, B, C, D vanishes, yielding

$$[(Q^2 - K^2)/2QK] \sinh Qb \sin Ka + \cosh Qb \cos Ka = \cos k(a + b) \ . \tag{21a}$$

It is rather tedious to obtain this equation.

The result is simplified if we represent the potential by the periodic delta function obtained when we pass to the limit $b = 0$ and $U_0 = \infty$ in such a way that $Q^2ba/2 = P$, a finite quantity. In this limit $Q \gg K$ and $Qb \ll 1$. Then (21a) reduces to

$$(P/Ka)\sin Ka + \cos Ka = \cos ka \ . \tag{21b}$$

The ranges of K for which this equation has solutions are plotted in Fig. 5, for the case $P = 3\pi/2$. The corresponding values of the energy are plotted in Fig. 6. Note the energy gaps at the zone boundaries. The wavevector k of the Bloch function is the important index, not the K in (12), which is related to the energy by (13). A treatment of this problem in wavevector space is given later in this chapter.

WAVE EQUATION OF ELECTRON IN A PERIODIC POTENTIAL

We considered in Fig. 3 the approximate form we expect for the solution of the Schrödinger equation if the wavevector is at a zone boundary, as at $k = \pm\pi/a$. We treat in detail the wave equation for a general potential, at general

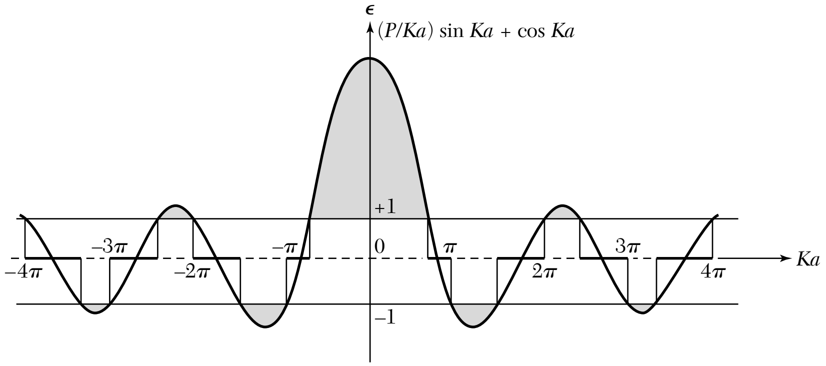


Figure 5 Plot of the function (P/Ka) sin Ka + cos Ka, for $P = 3\pi/2$. The allowed values of the energy e are given by those ranges of $Ka = (2m\epsilon/\hbar^2)^{1/2}a$ for which the function lies between ±1. For other values of the energy there are no traveling wave or Bloch-like solutions to the wave equation, so that forbidden gaps in the energy spectrum are formed.

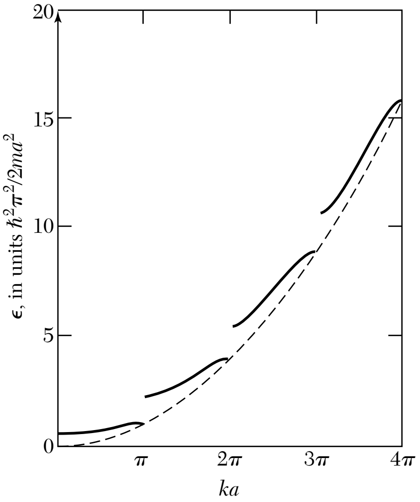


Figure 6 Plot of energy vs. wavenumber for the Kronig-Penney potential, with $P = 3\pi/2$. Notice the energy gaps at $ka = \pi, 2\pi, 3\pi \dots$.

values of k. Let $U(x)$ denote the potential energy of an electron in a linear lattice of lattice constant a. We know that the potential energy is invariant under a crystal lattice translation: $U(x) = U(x + a)$. A function invariant under a crystal lattice translation may be expanded as a Fourier series in the reciprocal lattice vectors G. We write the Fourier series for the potential energy as

$$U(x) = \sum_G U_G \, e^{iGx} \quad . \tag{22}$$

The values of the coefficients U_G for actual crystal potentials tend to decrease rapidly with increasing magnitude of G. For a bare coulomb potential U_G decreases as $1/G^2$.

We want the potential energy $U(x)$ to be a real function:

$$U(x) = \sum_{G>0} U_G(e^{iGx} + e^{-iGx}) = 2\sum_{G>0} U_G \cos Gx \ . \tag{23}$$

For convenience we have assumed that the crystal is symmetric about $x = 0$ and that $U_0 = 0$.

The wave equation of an electron in the crystal is $\mathcal{H}\psi = \epsilon\psi$, where $\mathcal{H}$ is the hamiltonian and ϵ is the energy eigenvalue. The solutions ψ are called eigenfunctions or orbitals or Bloch functions. Explicitly, the wave equation is

$$\left(\frac{1}{2m}p^2 + U(x)\right)\psi(x) = \left(\frac{1}{2m}p^2 + \sum_G U_G\, e^{iGx}\right)\psi(x) = \epsilon\psi(x) \ . \tag{24}$$

Equation (24) is written in the one-electron approximation in which the orbital $\psi(x)$ describes the motion of one electron in the potential of the ion cores and in the average potential of the other conduction electrons.

The wavefunction $\psi(x)$ may be expressed as a Fourier series summed over all values of the wavevector permitted by the boundary conditions, so that

$$\psi = \sum_k C(k)\, e^{ikx} \ , \tag{25}$$

where k is real. (We could equally well write the index k as a subscript on C, as in C_k.)

The set of values of k has the form $2\pi n/L$, because these values satisfy periodic boundary conditions over length L. Here n is any integer, positive or negative. We do not assume, nor is it generally true, that $\psi(x)$ itself is periodic in the fundamental lattice translation a. The translational properties of $\psi(x)$ are determined by the Bloch theorem (7).

Not all wavevectors of the set $2\pi n/L$ enter the Fourier expansion of any one Bloch function. If one particular wavevector k is contained in a ψ, then all other wavevectors in the Fourier expansion of this ψ will have the form $k + G$, where G is any reciprocal lattice vector. We prove this result in (29) below.

We can label a wavefunction ψ that contains a component k as ψ_k or, equally well, as ψ_{k+G}, because if k enters the Fourier expansion then $k + G$ may enter. The wavevectors $k + G$ running over G are a restricted subset of the set $2\pi n/L$, as shown in Fig. 7.

We shall usually choose as a label for the Bloch function that k which lies within the first Brillouin zone. When other conventions are used, we shall say so. This situation differs from the phonon problem for a monatomic lattice where there are no components of the ion motion outside the first zone. The electron problem is like the x-ray diffraction problem because like the electron wavefunction the electromagnetic field exists everywhere within the crystal and not only at the ions.

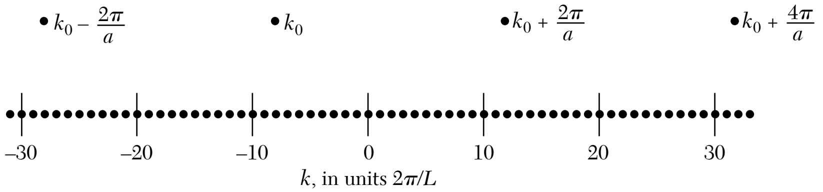


Figure 7 The lower points represent values of the wavevector $k = 2\pi n/L$ allowed by the periodic boundary condition on the wavefunction over a ring of circumference L composed of 20 primitive cells. The allowed values continue to $\pm\infty$. The upper points represent the first few wavevectors which may enter into the Fourier expansion of a wavefunction $\psi(x)$, starting from a particular wavevector $k = k_0 = -8(2\pi/L)$. The shortest reciprocal lattice vector is $2\pi/a = 20(2\pi/L)$.

To solve the wave equation, substitute (25) in (24) to obtain a set of linear algebraic equations for the Fourier coefficients. The kinetic energy term is

$$\frac{1}{2m}p^2\psi(x) = \frac{1}{2m}\left(-i\hbar\frac{d}{dx}\right)^2\psi(x) = -\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = \frac{\hbar^2}{2m}\sum_k k^2 C(k)\,e^{ikx} \ ;$$

and the potential energy term is

$$\left(\sum_G U_G e^{iGx}\right)\psi(x) = \sum_G\sum_k U_G e^{iGx} C(k)\,e^{ikx} \ .$$

The wave equation is obtained as the sum:

$$\sum_k \frac{\hbar^2}{2m}k^2 C(k)\,e^{ikx} + \sum_G\sum_k U_G C(k)\,e^{i(k+G)x} = \epsilon\sum_k C(k)\,e^{ikx} \ . \tag{26}$$

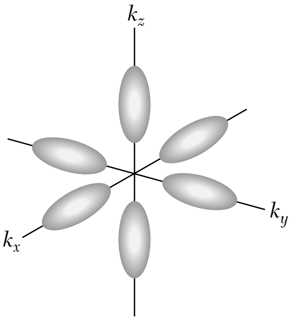
Each Fourier component must have the same coefficient on both sides of the equation. Thus we have the **central equation**

$$\boxed{(\lambda_k - \epsilon)C(k) + \sum_G U_G C(k-G) = 0 \ .} \tag{27}$$

with the notation

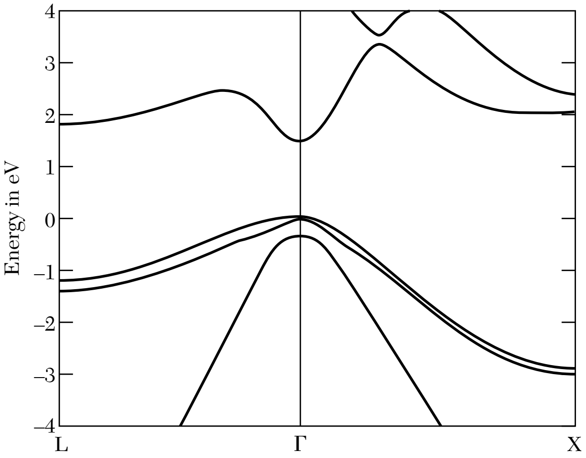
$$\lambda_k = \hbar^2 k^2/2m \ . \tag{28}$$

Equation (27) is a useful form of the wave equation in a periodic lattice, although unfamiliar because a set of algebraic equations has taken the place of the usual differential equation (24). The set appears unpleasant and formidable because there are, in principle, an infinite number of $C(k - G)$ to be determined. In practice a small number will often suffice, perhaps two or four. It takes some experience to appreciate the practical advantages of the algebraic approach.



(a)

Figure 17a Constant energy ellipsoids for electrons in silicon, drawn for $m_l/m_t = 5$.



(b)

Figure 17b Band structure of GaAs, after S. G. Louie.

INTRINSIC CARRIER CONCENTRATION

We want the concentration of intrinsic carriers as a function of temperature, in terms of the band gap. We do the calculation for simple parabolic band edges. We first calculate in terms of the chemical potential μ the number of electrons excited to the conduction band at temperature T. In semiconductor physics μ is called the Fermi level. At the temperatures of interest we may suppose for the conduction band of a semiconductor that $\epsilon - \mu \gg k_BT$, so that the Fermi-Dirac distribution function reduces to

$$f_e \simeq \exp[(\mu - \epsilon)/k_BT] \quad . \tag{35}$$

This is the probability that a conduction electron orbital is occupied, in an approximation valid when $f_e \ll 1$.

The energy of an electron in the conduction band is

$$\epsilon_k = E_c + \hbar^2 k^2/2m_e \quad , \tag{36}$$

where E_c is the energy at the conduction band edge, as in Fig. 18. Here m_e is the effective mass of an electron. Thus from (6.20) the density of states at ϵ is

$$D_e(\epsilon) = \frac{1}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}(\epsilon - E_c)^{1/2} \quad . \tag{37}$$

The concentration of electrons in the conduction band is

$$n = \int_{E_c}^{\infty} D_e(\epsilon) f_e(\epsilon) d\epsilon = \frac{1}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2} \exp(\mu/k_BT) \times \int_{E_c}^{\infty} (\epsilon - E_c)^{1/2} \exp(-\epsilon/k_BT) d\epsilon \ , \tag{38}$$

which integrates to give

$$n = 2\left(\frac{m_e k_B T}{2\pi\hbar^2}\right)^{3/2} \exp[(\mu - E_c)/k_BT] \ . \tag{39}$$

The problem is solved for n when μ is known. It is useful to calculate the equilibrium concentration of holes p. The distribution function f_h for holes is related to the electron distribution function f_e by $f_h = 1 - f_e$, because a hole is the absence of an electron. We have

$$\begin{aligned} f_h &= 1 - \frac{1}{\exp[(\epsilon - \mu)/k_BT] + 1} = \frac{1}{\exp[(\mu - \epsilon)/k_BT] + 1} \\ &\cong \exp[(\epsilon - \mu)/k_BT] \ , \end{aligned} \tag{40}$$

provided $(\mu - \epsilon) \gg k_BT$.

If the holes near the top of the valence band behave as particles with effective mass m_h, the density of hole states is given by

$$D_h(\epsilon) = \frac{1}{2\pi^2}\left(\frac{2m_h}{\hbar^2}\right)^{3/2} (E_v - \epsilon)^{1/2} \ , \tag{41}$$

where E_v is the energy at the valence band edge. Proceeding as in (38) we obtain

$$p = \int_{-\infty}^{E_c} D_h(\epsilon) f_h(\epsilon) d\epsilon = 2\left(\frac{m_h k_B T}{2\pi\hbar^2}\right)^{3/2} \exp[(E_c - \mu)/k_BT] \tag{42}$$

for the concentration p of holes in the valence band.

We multiply together the expressions for n and p to obtain the equilibrium relation, with the energy gap $E_g = E_c - E_v$ as in Fig. 18,

$$np = 4\left(\frac{k_BT}{2\pi\hbar^2}\right)^{3} (m_c m_h)^{3/2} \exp(-E_g/k_BT) \ . \tag{43}$$

This useful result does not involve the Fermi level μ. At 300 K the value of np is 2.10×10^{19} cm^{-6}, 2.89×10^{26} cm^{-6}, and 6.55×10^{12} cm^{-6}, for the actual band structures of Si, Ge, and GaAs, respectively.

We have nowhere assumed in the derivation that the material is intrinsic: the result holds for impurity ionization as well. The only assumption made is that the distance of the Fermi level from the edge of both bands is large in comparison with k_BT.

A simple kinetic argument shows why the product np is constant at a given temperature. Suppose that the equilibrium population of electrons and holes

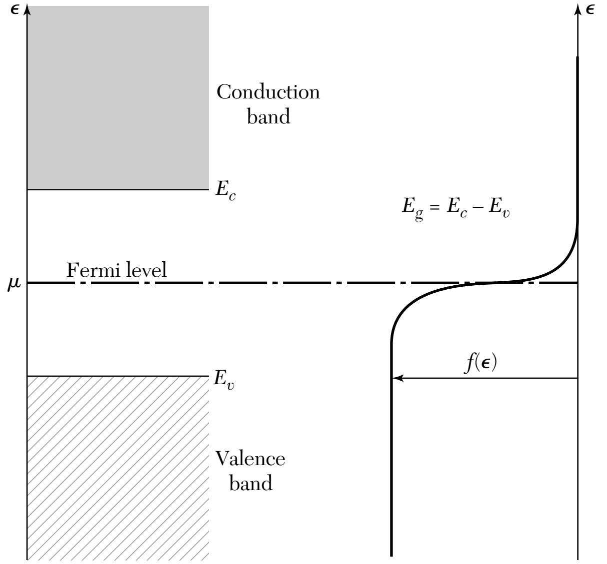


Figure 18 Energy scale for statistical calculations. The Fermi distribution function is shown on the same scale, for a temperature $k_BT \ll E_g$. The Fermi level μ is taken to lie well within the band gap, as for an intrinsic semiconductor. If $\epsilon = \mu$, then $f = \frac{1}{2}$.

is maintained by black-body photon radiation at temperature T. The photons generate electron-hole pairs at a rate $A(T)$, while $B(T)np$ is the rate of the recombination reaction $e + h =$ photon. Then

$$dn/dt = A(T) - B(T)np = dp/dt \cdot \tag{44}$$

In equilibrium $dn/dt = 0$, $dp/dt = 0$, whence $np = A(T)/B(T)$.

Because the product of the electron and hole concentrations is a constant independent of impurity concentration at a given temperature, the introduction of a small proportion of a suitable impurity to increase n, say, must decrease p. This result is important in practice—we can reduce the total carrier concentration $n + p$ in an impure crystal, sometimes enormously, by the controlled introduction of suitable impurities. Such a reduction is called **compensation**.

In an intrinsic semiconductor the number of electrons is equal to the number of holes, because the thermal excitation of an electron leaves behind a hole in the valence band. Thus from (43) we have, letting the subscript i denote intrinsic and $E_g = E_c - E_v$,

$$n_i = p_i = 2\left(\frac{k_BT}{2\pi\hbar^2}\right)^{3/2}(m_em_h)^{3/4}\exp(-E_g/2k_BT) \ . \tag{45}$$

The intrinsic carrier concentration depends exponentially on $E_g/2k_BT$, where E_g is the energy gap. We set (39) equal to (42) to obtain, for the Fermi level as measured from the top of the valence band,

$$\exp(2\mu/k_BT) = (m_h/m_e)^{3/2}\exp(E_g/k_BT) \ ; \tag{46}$$

$$\mu = \tfrac{1}{2}E_g + \tfrac{3}{4}k_BT\ln(m_h/m_e) \ . \tag{47}$$

If $m_h = m_e$, then $\mu = \frac{1}{2}E_g$ and the Fermi level is in the middle of the forbidden gap.

Intrinsic Mobility

The mobility is the magnitude of the drift velocity of a charge carrier per unit electric field:

$$\mu = |v|/E \ . \tag{48}$$

The mobility is defined to be positive for both electrons and holes, although their drift velocities are opposite in a given field. By writing μ_e or μ_h with subscripts for the electron or hole mobility we can avoid any confusion between μ as the chemical potential and as the mobility.

The electrical conductivity is the sum of the electron and hole contributions:

$$\sigma = (ne\mu_e + pe\mu_h) \ , \tag{49}$$

where n and p are the concentrations of electrons and holes. In Chapter 6 the drift velocity of a charge q was found to be $v = q\tau E/m$, whence

$$\mu_e = e\tau_e/m_e \ ; \qquad \mu_h = e\tau_h/m_h \ , \tag{50}$$

where τ is the collision time.

The mobilities depend on temperature as a modest power law. The temperature dependence of the conductivity in the intrinsic region will be dominated by the exponential dependence $\exp(-E_g/2k_BT)$ of the carrier concentration, Eq. (45).

Table 3 gives experimental values of the mobility at room temperature. The mobility in SI units is expressed in m^2/V-s and is 10^{-4} of the mobility in practical units. For most substances the values quoted are limited by the scattering of carriers by thermal phonons. The hole mobilities typically are smaller than the electron mobilities because of the occurrence of band degeneracy at the valence band edge at the zone center, thereby making possible interband scattering processes that reduce the mobility considerably.

Table 3 Carrier mobilities at room temperature, in cm^2/V-s

Crystal	Electrons	Holes	Crystal	Electrons	Holes
Diamond	1800	1200	GaAs	8000	300
Si	1350	480	GaSb	5000	1000
Ge	3600	1800	PbS	550	600
InSb	800	450	PbSe	1020	930
InAs	30000	450	PbTe	2500	1000
InP	4500	100	AgCl	50	—
AlAs	280	—	KBr (100 K)	100	—
AlSb	900	400	SiC	100	10–20

In some crystals, particularly in ionic crystals, the holes are essentially immobile and get about only by thermally-activated hopping from ion to ion. The principal cause of this "self-trapping" is the lattice distortion associated with the Jahn-Teller effect of degenerate states. The orbital degeneracy necessary for self-trapping is much more frequent for holes than for electrons.

There is a tendency for crystals with small energy gaps at direct band edges to have high values of the electron mobility. Small gaps lead to small effective masses, which favor high mobilities. The highest mobility observed in a bulk semiconductor is 5×10^6 cm^2/V-s in PbTe at 4 K, where the gap is 0.19 eV.

IMPURITY CONDUCTIVITY

Certain impurities and imperfections drastically affect the electrical properties of a semiconductor. The addition of boron to silicon in the proportion of 1 boron atom to 10^5 silicon atoms increases the conductivity of pure silicon at room temperature by a factor of 10^3. In a compound semiconductor a stoichiometric deficiency of one constituent will act as an impurity; such semiconductors are known as **deficit semiconductors**. The deliberate addition of impurities to a semiconductor is called **doping**.

We consider the effect of impurities in silicon and germanium. These elements crystallize in the diamond structure. Each atom forms four covalent bonds, one with each of its nearest neighbors, corresponding to the chemical valence four. If an impurity atom of valence five, such as phosphorus, arsenic, or antimony, is substituted in the lattice in place of a normal atom, there will be one valence electron from the impurity atom left over after the four covalent bonds are established with the nearest neighbors, that is, after the impurity atom has been accommodated in the structure with as little disturbance as possible. Impurity atoms that can give up an electron are called **donors**.

Donor States. The structure in Fig. 19 has a positive charge on the impurity atom (which has lost one electron). Lattice constant studies have verified that the pentavalent impurities enter the lattice by substitution for normal atoms, and not in interstitial positions. The crystal as a whole remains neutral because the electron remains in the crystal.

The extra electron moves in the coulomb potential $e/\epsilon r$ of the impurity ion, where ϵ in a covalent crystal is the static dielectric constant of the medium. The factor $1/\epsilon$ takes account of the reduction in the coulomb force between charges caused by the electronic polarization of the medium. This treatment is valid for orbits large in comparison with the distance between atoms, and for slow motions of the electron such that the orbital frequency is low in comparison with the frequency ω_g corresponding to the energy gap. These conditions are satisfied quite well in Ge and Si by the donor electron of P, As, or Sb.