Chapter 8-2

Holes

The properties of vacant orbitals in an otherwise filled band are important in semiconductor physics and in solid state electronics. Vacant orbitals in a band are commonly called holes, and without holes there would be no transistors. A hole acts in applied electric and magnetic fields as if it has a positive charge $+e$. The reason is given in five steps in the boxes that follow.

1. $$\mathbf{k}_h = -\mathbf{k}_e \ . \qquad (17)$$

The total wavevector of the electrons in a filled band is zero: $\Sigma\mathbf{k} = 0$, where the sum is over all states in a Brillouin zone. This result follows from the geometrical symmetry of the Brillouin zone: every fundamental lattice type has symmetry under the inversion operation $\mathbf{r} \rightarrow -\mathbf{r}$ about any lattice point; it follows that the Brillouin zone of the lattice also has inversion symmetry. If the band is filled all pairs of orbitals $\mathbf{k}$ and $-\mathbf{k}$ are filled, and the total wavevector is zero.

If an electron is missing from an orbital of wavevector $\mathbf{k}_e$, the total wavevector of the system is $-\mathbf{k}_e$ and is attributed to the hole. This result is surprising: the electron is missing from $\mathbf{k}_e$ and the position of the hole is usually indicated graphically as situated at $\mathbf{k}_e$, as in Fig. 7. But the true wavevector $\mathbf{k}_h$ of the hole is $-\mathbf{k}_e$, which is the wavevector of the point G if the hole is at E. The wavevector $-\mathbf{k}_e$ enters into selection rules for photon absorption.

The hole is an alternate description of a band with one missing electron, and we either say that the hole has wavevector $-\mathbf{k}_e$ or that the band with one missing electron has total wavevector $-\mathbf{k}_e$.

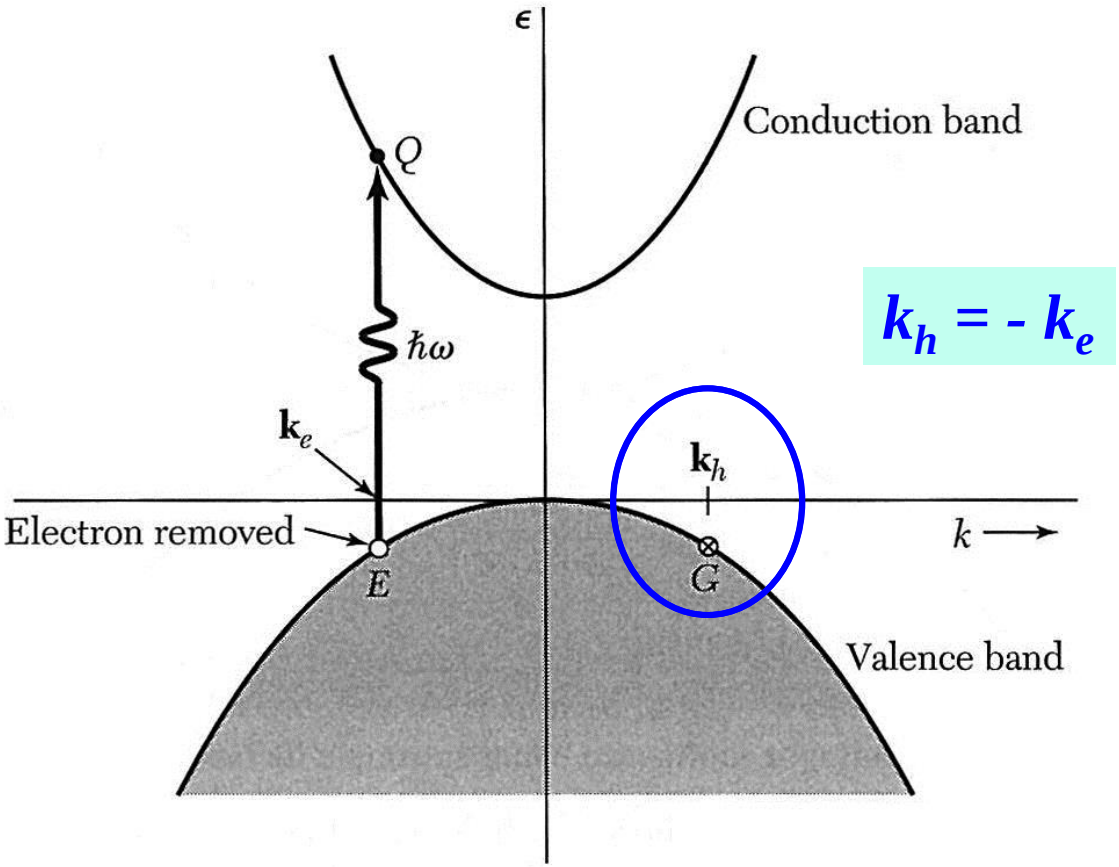


Figure 7 Absorption of a photon of energy $\hbar\omega$ and negligible wavevector takes an electron from E in the filled valence band to Q in the conduction band. If $\mathbf{k}_e$ was the wavevector of the electron at E, it becomes the wavevector of the electron at Q. The total wavevector of the valence band after the absorption is $-\mathbf{k}_e$, and this is the wavevector we must ascribe to the hole if we describe the valence band as occupied by one hole. Thus $\mathbf{k}_h = -\mathbf{k}_e$; the wavevector of the hole is the same as the wavevector of the electron which remains at G. For the entire system the total wavevector after the absorption of the photon is $\mathbf{k}_e + \mathbf{k}_h = 0$, so that the total wavevector is unchanged by the absorption of the photon and the creation of a free electron and free hole.

2. $$\epsilon_h(\mathbf{k}_h) = -\epsilon_e(\mathbf{k}_e) \quad . \tag{18}$$

Here the zero of energy of the valence band is at the top of the band. The lower in the band the missing electron lies, the higher the energy of the system. The energy of the hole is opposite in sign to the energy of the missing electron, because it takes more work to remove an electron from a low orbital than from a high orbital. Thus if the band is symmetric,[1] $\epsilon_e(\mathbf{k}_e) = \epsilon_e(-\mathbf{k}_e) = -\epsilon_h(-\mathbf{k}_e) = -\epsilon_h(\mathbf{k}_h)$. We construct in Fig. 8 a band scheme to represent the properties of a hole. This hole band is a helpful representation because it appears right side up.

3. $$\mathbf{v}_h = \mathbf{v}_e \quad .$$

$dE(k_h)/dk_h = dE(k_e)/dk_e$

The velocity of the hole is equal to the velocity of the missing electron. From Fig. 8 we see that $\nabla\epsilon_h(\boldsymbol{k}_h) = \nabla\epsilon_e(\boldsymbol{k}_e)$, so that $\boldsymbol{v}_h(\boldsymbol{k}_h) = \boldsymbol{v}_e(\boldsymbol{k}_e)$.

[1]Bands are always symmetric under the inversion $\mathbf{k} \rightarrow -\mathbf{k}$ if the spin-orbit interaction is neglected. Even with spin-orbit interaction, bands are always symmetric if the crystal structure permits the inversion operation. Without a center of symmetry, but with spin-orbit interaction, the bands are symmetric if we compare subbands for which the spin direction is reversed: $\epsilon(\boldsymbol{k}, \uparrow) = \epsilon(-\boldsymbol{k}, \downarrow)$. See *QTS*, Chapter 9.

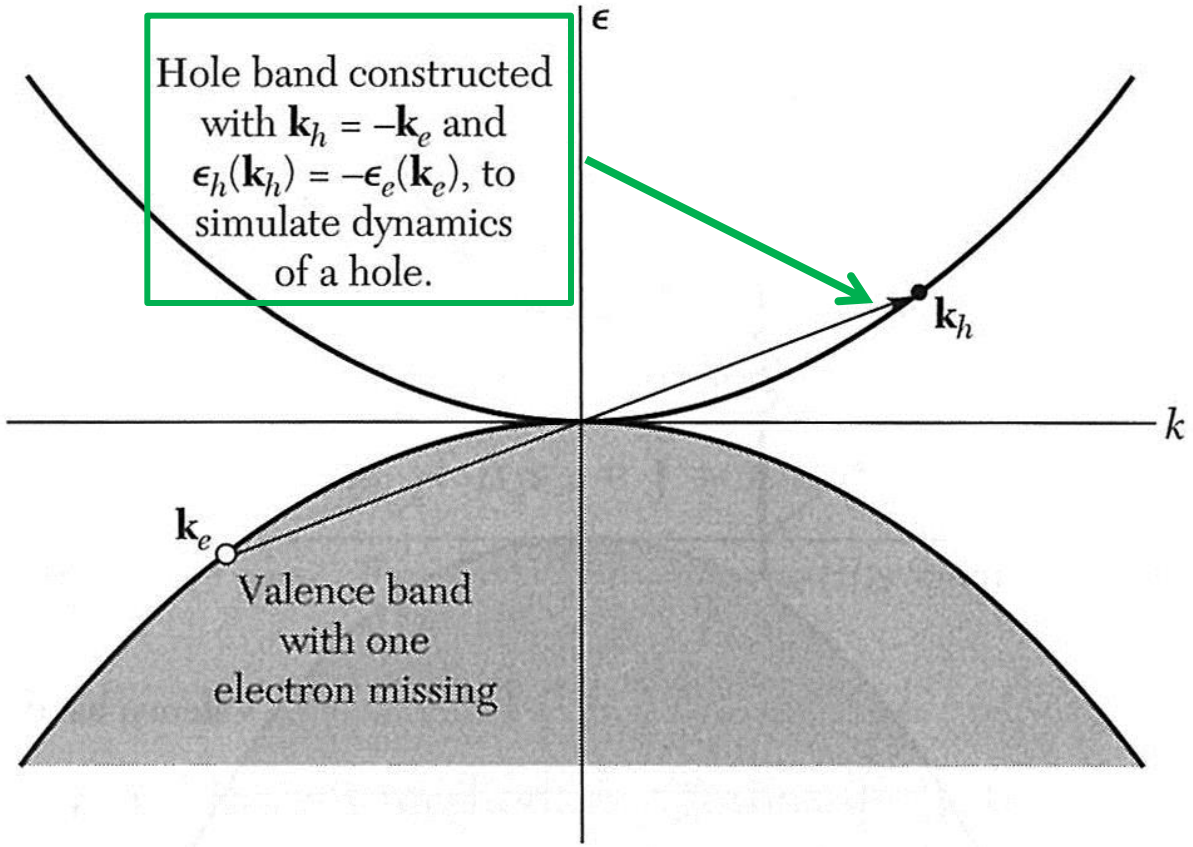


Figure 8 The upper half of the figure shows the hole band that simulates the dynamics of a hole, constructed by inversion of the valence band in the origin. The wavevector and energy of the hole are equal, but opposite in sign, to the wavevector and energy of the empty electron orbital in the valence band. We do not show the disposition of the electron removed from the valence band at $\mathbf{k}_e$.

4. $m_h = -m_e$.

$d^2E(k_h)/dk_h^2 = - dE^2(k_e)/dk_e^2$

We show below that the effective mass is inversely proportional to the curvature $d^2\epsilon/dk^2$, and for the hole band this has the opposite sign to that for an electron in the valence band. Near the top of the valence band m_e is negative, so that m_h is positive.

5. $$\hbar \frac{d\mathbf{k}_h}{dt} = e(\mathbf{E} + \frac{1}{c}\mathbf{v}_h \times \mathbf{B}) . \qquad (21)$$

$V_h = V_e$

This comes from the equation of motion

(CGS) $$\hbar \frac{d\mathbf{k}_e}{dt} = -e(\mathbf{E} + \frac{1}{c}\mathbf{v_e} \times \mathbf{B}) \qquad (22)$$

that applies to the missing electron when we substitute $-\mathbf{k}_h$ for $\mathbf{k}_e$ and $\mathbf{v}_h$ for $\mathbf{v}_e$. **The equation of motion for a hole is that of a particle of positive charge *e*.** The positive charge is consistent with the electric current carried by the valence band of Fig. 9: the current is carried by the unpaired electron in the orbital G:

$$\mathbf{j} = (-e)\mathbf{v}(G) = (-e)[-\mathbf{v}(E)] = e\mathbf{v}(E) , \qquad (23)$$

which is just the current of a positive charge moving with the velocity ascribed to the missing electron at E. The current is shown in Fig. 10.

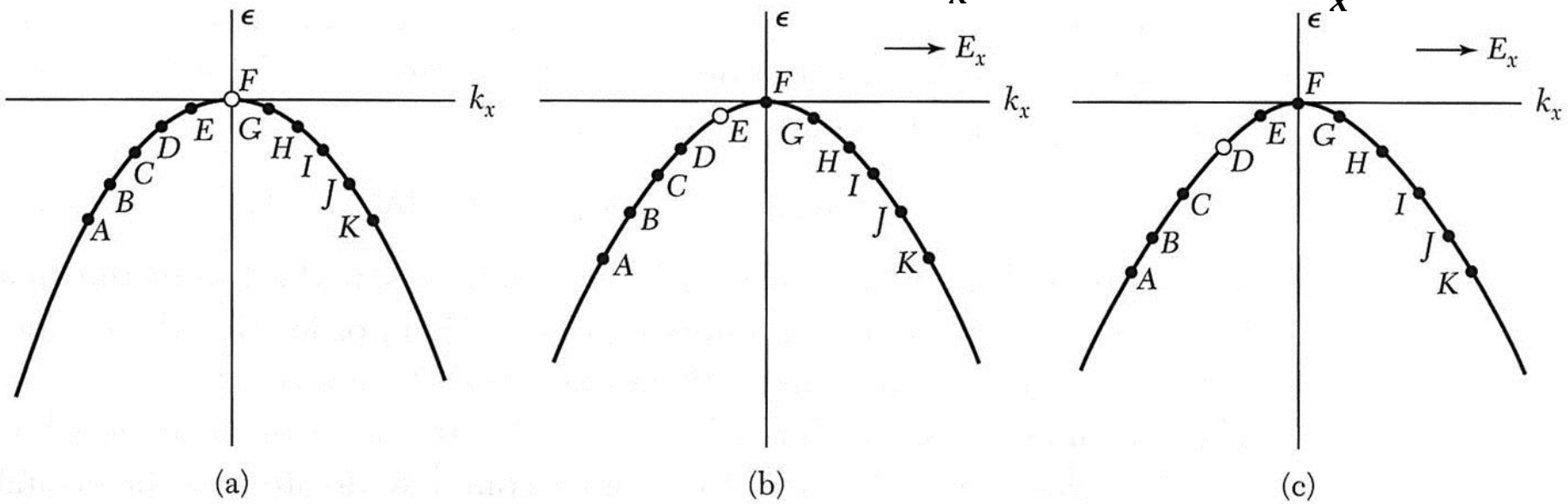
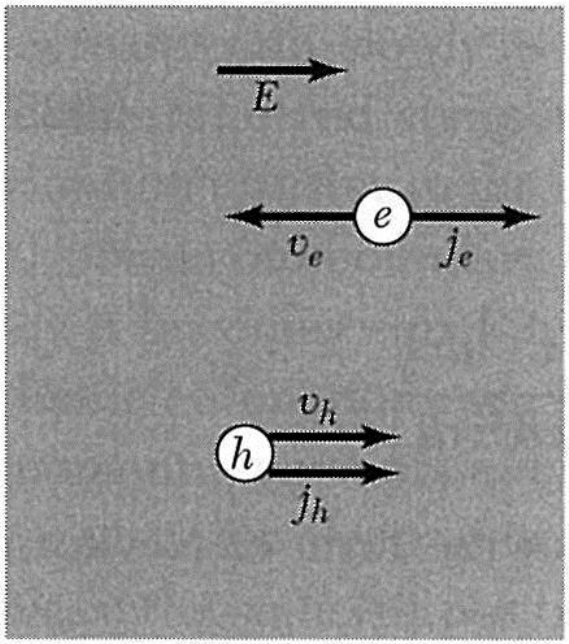


Figure 9 (a) At $t = 0$ all states are filled except F at the top of the band; the velocity v_x is zero at F because $d\epsilon/dk_x = 0$. (b) An electric field E_x is applied in the $+x$ direction. The force on the electrons is in the $-k_x$ direction and all electrons make transitions together in the $-k_x$ direction, moving the hole to the state E. (c) After a further interval the electrons move farther along in k space and the hole is now at D.



From Ch. 6,
drift velocity, $\upsilon = q \tau E/m$

$$\upsilon_e = -\upsilon_h$$

$$J_e = J_h$$

Figure 10 Motion of electrons in the conduction band and holes in the valence band in the electric field E. The hole and electron drift velocities are in opposite directions, but their electric currents are in the same direction, the direction of the electric field.

Effective Mass

When we look at the energy-wavevector relation $\epsilon = (\hbar^2/2m)k^2$ for free electrons, we see that the coefficient of k^2 determines the curvature of ϵ versus k. Turned about, we can say that $1/m$, the reciprocal mass, determines the curvature. For electrons in a band there can be regions of unusually high curvature near the band gap at the zone boundary, as we see from the solutions in Chapter 7 of the wave equation near the zone boundary. If the energy gap is small in comparison with the free electron energy λ at the boundary, the curvature is enhanced by the factor λ/E_g. ~100, Ch. 7, eq. 52

In semiconductors the band width, which is like the free electron energy, is of the order of 20 eV, while the band gap is of the order of 0.2 to 2 eV. Thus the reciprocal mass is enhanced by a factor 10 to 100, and the effective mass is reduced to 0.1–0.01 of the free electron mass. These values apply near the band gap; as we go away from the gap the curvatures and the masses are likely to approach those of free electrons.

$E_g = 2U$

To summarize the solutions of Chapter 7 for U positive, an electron near the lower edge of the second band has an energy that may be written as

$$\epsilon(K) = \epsilon_c + (\hbar^2/2m_e)K^2 \ ; \qquad m_e/m = 1/[(2\lambda/U)-1] \ . \qquad (24)$$

from Ch. 7, eq. 52

$m_e \propto U \propto E_g$

Here K is the wavevector measured from the zone boundary, and m_e denotes the effective mass of the electron near the edge of the second band. An electron near the top of the first band has the energy

$$\epsilon(K) = \epsilon_v - (\hbar^2/2m_h)K^2 \ ; \qquad \boxed{m_h/m = 1/[(2\lambda/U) + 1]} . \tag{25}$$

The curvature and hence the mass will be negative near the top of the first band, but we have introduced a minus sign into (25) in order that the symbol m_h for the hole mass will have a positive value—see (20) above.

The crystal does not weigh any less if the effective mass of a carrier is less than the free electron mass, nor is Newton's second law violated for the crystal *taken as a whole,* ions plus carriers. The important point is that an electron in a periodic potential is accelerated relative to the lattice in an applied electric or magnetic field as if the mass of the electron were equal to an effective mass which we now define.

We differentiate the result (1) for the group velocity to obtain

$$\frac{dv_g}{dt} = \hbar^{-1}\frac{d^2\epsilon}{dk\,dt} = \hbar^{-1}\left(\frac{d^2\epsilon}{dk^2}\frac{dk}{dt}\right) . \tag{26}$$

We know from (5) that $dk/dt = F/\hbar$, whence

$$\frac{dv_g}{dt} = \left(\frac{1}{\hbar^2}\frac{d^2\epsilon}{dk^2}\right)F \; ; \qquad \text{or} \qquad F = \frac{\hbar^2}{d^2\epsilon/dk^2}\frac{dv_g}{dt} . \tag{27}$$

If we identify $\hbar^2/(d^2\epsilon/dk^2)$ as a mass, then (27) assumes the form of Newton's second law. We define the **effective mass** m^* by

$F = ma$

$$\boxed{\frac{1}{m^*} = \frac{1}{\hbar^2}\frac{d^2\epsilon}{dk^2} .} \tag{28}$$

It is easy to generalize this to take account of an anisotropic electron energy surface, as for electrons in Si or Ge. We introduce the components of the reciprocal effective mass tensor

$$\boxed{\left(\frac{1}{m^*}\right)_{\mu\nu} = \frac{1}{\hbar^2}\frac{d^2\,\epsilon_k}{dk_\mu\,dk_\nu} \; ;} \qquad \boxed{\frac{dv_\mu}{dt} = \left(\frac{1}{m^*}\right)_{\mu\nu}F_\nu \; ,} \tag{29}$$

where μ, ν are Cartesian coordinates.

Effective Masses in Semiconductors

In many semiconductors it has been possible to determine by cyclotron resonance the effective masses of carriers in the conduction and valence bands near the band edges. The determination of the energy surface is equivalent to a determination of the effective mass tensor (29). Cyclotron resonance in a semiconductor is carried out with centimeter wave or millimeter wave radiation at low carrier concentration.

The current carriers are accelerated in helical orbits about the axis of a static magnetic field. The angular rotation frequency ω_c is

$$\text{(CGS)} \quad \omega_c = \frac{eB}{m^*c}, \qquad \text{(SI)} \quad \omega_c = \frac{eB}{m^*} \tag{30}$$

where m^* is the appropriate cyclotron effective mass. Resonant absorption of energy from an rf electric field perpendicular to the static magnetic field (Fig. 12) occurs when the rf frequency is equal to the cyclotron frequency. Holes and electrons rotate in opposite senses in a magnetic field.

We consider the experiment for $m^*/m = 0.1$. At $f_c = 24$ GHz, or $\omega_c = 1.5 \times 10^{11}\ s^{-1}$, we have $B = 860$ G at resonance. The line width is determined by the collision relaxation time τ, and to obtain a distinctive resonance it is necessary that $\omega_c\tau \geq 1$. The mean free path must be long enough to permit the average carrier to get one radian around a circle between collisions. The requirements are met with the use of higher frequency radiation and higher magnetic fields, with high purity crystals in liquid helium.

$1/\omega_c << \tau$ by using a larger B, purer crystal

$\omega_c >> \omega_\tau$

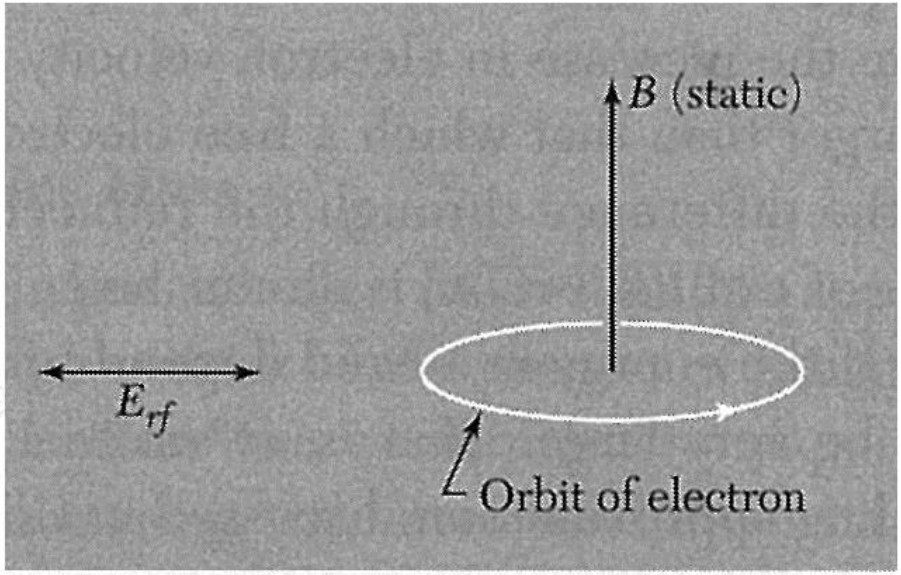


Figure 12 Arrangement of fields in a cyclotron resonance experiment in a semiconductor. The sense of the circulation is opposite for electrons and holes.

CYCLOTRON RESONANCE

Aschroft Mermin, Ch. 28

The effective masses discussed above are measured by the technique of cyclotron resonance. Consider an electron close enough to the bottom of the conduction band (or top of the valence band) for the quadratic expansion (28.2) to be valid. In the presence of a magnetic field $\mathbf{H}$ the semiclassical equations of motion (12.32) and (12.33) imply that the velocity $\mathbf{v}(\mathbf{k})$ obeys the single set of equations

$$\mathbf{M}\frac{d\mathbf{v}}{dt} = \mp\frac{e}{c}\mathbf{v}\times\mathbf{H}. \tag{28.4}$$

In a constant uniform field (taken along the z-axis) it is not difficult to show (Problem 1) that (28.4) has an oscillatory solution

$$\mathbf{v} = \mathrm{Re}\,\mathbf{v}_0 e^{-i\omega t}, \tag{28.5}$$

provided that

$$\omega = \frac{eH}{m^* c}, \tag{28.6}$$

where m^*, the "cyclotron effective mass," is given by

$$m^* = \left(\frac{\det \mathbf{M}}{M_{zz}}\right)^{1/2}. \tag{28.7}$$

This result can also be written in terms of the eigenvalues and principal axes of the mass tensor as (Problem 1):

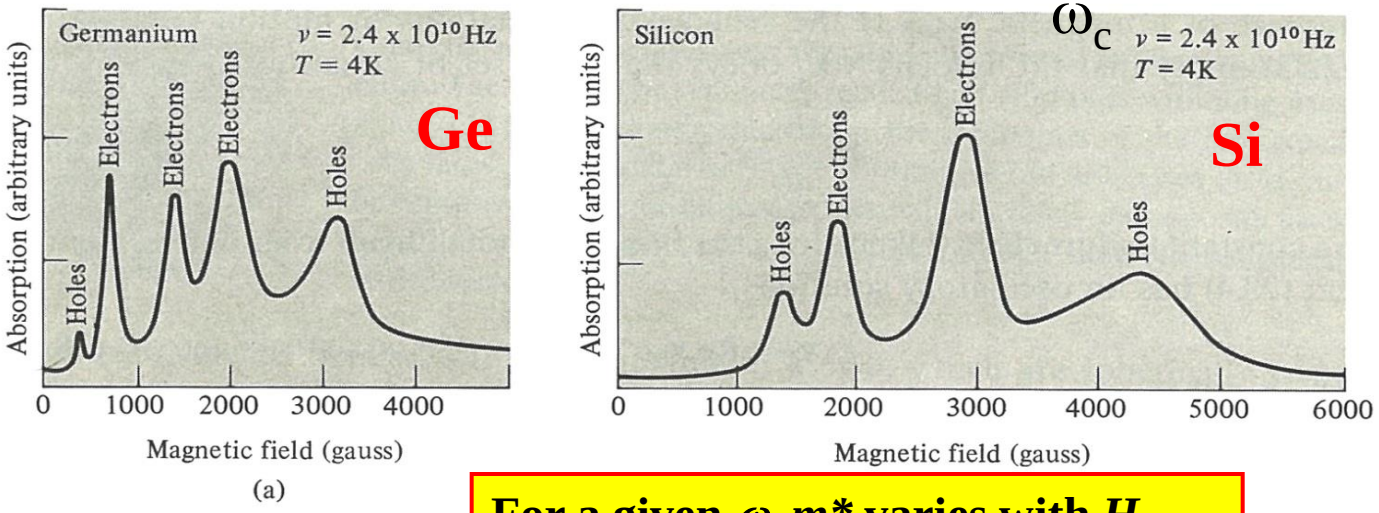
$$m^* = \sqrt{\frac{m_1 m_2 m_3}{\hat{H}_1^{\,2} m_1 + \hat{H}_2^{\,2} m_2 + \hat{H}_3^{\,2} m_3}}, \tag{28.8}$$

where the $\hat{H}_i$ are the components along the three principal axes of a unit vector parallel to the field.

Note that the cyclotron frequency depends, for a given ellipsoid, on the orientation of the magnetic field with respect to that ellipsoid, but not on the initial wave vector or energy of the electron. Thus for a given orientation of the crystal with respect to the field, all electrons in a given ellipsoidal pocket of conduction band electrons (and, by the same token, all holes in a given ellipsoidal pocket of valence band holes) precess at a frequency entirely determined by the effective mass tensor describing that pocket. There will therefore be a small number of distinct cyclotron frequencies. By noting how these resonant frequencies shift as the orientation of the magnetic field is varied, one can extract from (28.8) the kind of information we quoted above.

For a given ω, m^* varies with H

To observe cyclotron resonance it is essential that the cyclotron frequency (28.6) be larger than or comparable to the collision frequency. As in the case of metals, this generally requires working with very pure samples at very low temperatures, to reduce both impurity scattering and phonon scattering to a minimum. Under such conditions the electrical conductivity of a semiconductor will be so small that (in contrast to the case of a metal (page 278)) the driving electromagnetic field can penetrate far enough into the sample to excite the resonance without any difficulties associated with a skin depth. On the other hand, under such conditions of low temperatures and purity the number of carriers available in thermal equilibrium to participate in the resonance may well be so small that carriers will have to be created by other means—such as photoexcitation. Some typical cyclotron resonance data are shown in Figure 28.9.



For a given ω, m^* varies with H

Figure 28.9
Typical cyclotron resonance signals in (a) germanium and (b) silicon. The field lies in a (110) plane and makes an angle with the [001] axis of 60° (Ge) and 30° (Si). (From G. Dresselhaus et al., *Phys. Rev.* **98**, 368 (1955).)

In direct-gap semiconductors with band edges at the center of the Brillouin zone, the bands have the structure shown in Fig. 13. The conduction band edge is spherical with the effective mass m_e

$$\epsilon_c = E_g + \hbar^2 k^2 / 2m_e \ , \tag{31}$$

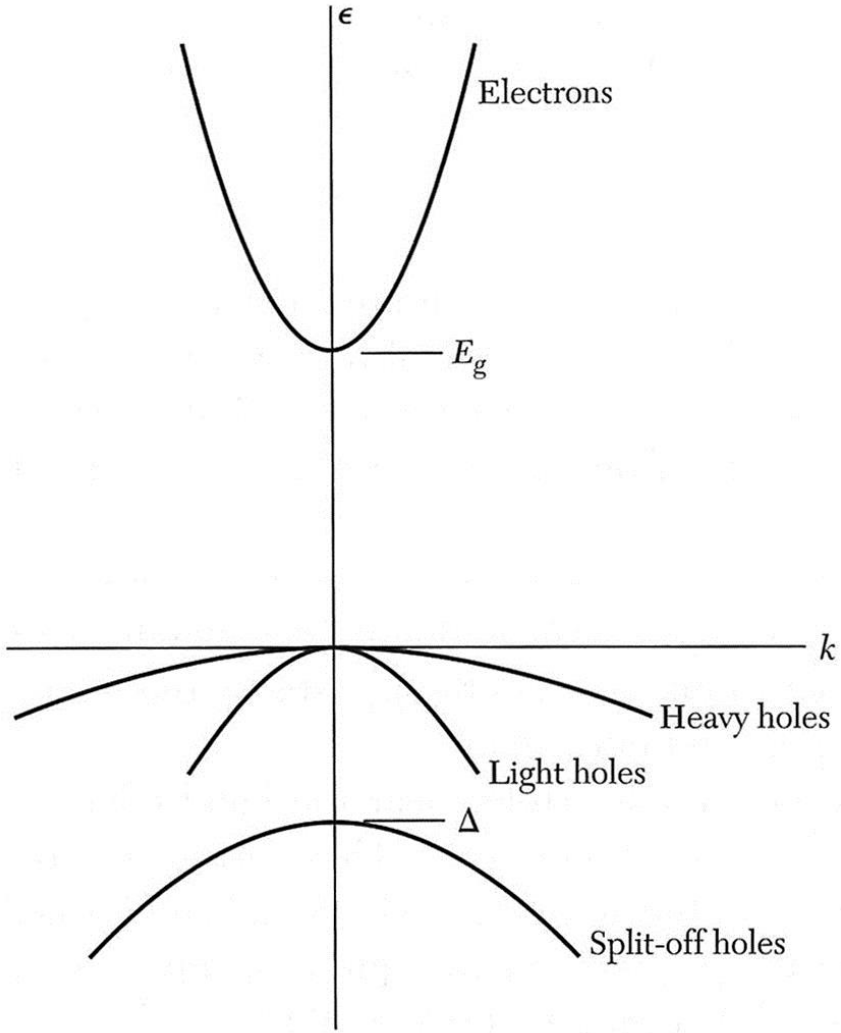
referred to the valence band edge. The valence bands are characteristically threefold near the edge, with the heavy hole *hh* and light hole *lh* bands degenerate at the center, and a band *soh* split off by the spin-orbit splitting Δ:

$$\epsilon_v(hh) \cong -\hbar^2 k^2 / 2m_{hh} \ ; \qquad \epsilon_v(lh) \cong -\hbar^2 k^2 / 2m_{lh} \ ;$$
$$\epsilon_v(soh) \cong -\Delta - \hbar^2 k^2 / 2m_{soh} \ . \tag{32}$$

Values of the mass parameters are given in Table 2. The forms (32) are only approximate, because even close to $k = 0$ the heavy and light hole bands are not spherical—see the discussion below for Ge and Si.

The perturbation theory of band edges (Problem 9.8) suggests that the electron effective mass should be proportional to the band gap, approximately, for a direct gap crystal. We use Tables 1 and 2 to find the fairly constant values $m_e/(mE_g) = 0.063$, 0.060, and 0.051 in $(\text{eV})^{-1}$ for the series InSb, InAs, and InP, in agreement with this suggestion.

$m_e \propto E_g$, *see Ch. 8, eq. 24, 25*



Note in general $m^*_h > m^*_e$

Figure 13 Simplified view of the band edge structure of a direct-gap semiconductor.

Table 2 Effective masses of electrons and holes in direct-gap semiconductors

Crystal	Electron m_e/m	Heavy hole m_{hh}/m	Light hole m_{lh}/m	Split-off hole m_{soh}/m	Spin-orbit Δ, eV
InSb	0.015	0.39	0.021	(0.11)	0.82
InAs	0.026	0.41	0.025	0.08	0.43
InP	0.073	0.4	(0.078)	(0.15)	0.11
GaSb	0.047	0.3	0.06	(0.14)	0.80
GaAs	0.066	0.5	0.082	0.17	0.34
Cu_2O	0.99	—	0.58	0.69	0.13

Z

E_g larger

m* larger

$m^* \propto E_g$

$m^*/(m\,E_g) \propto$ ~0.06

Physical Interpretation of the Effective Mass

How can an electron of mass m when put into a crystal respond to applied fields as if the mass were m^*? It is helpful to think of the process of Bragg reflection of electron waves in a lattice. Consider the weak interaction approximation treated in Chapter 7. Near the bottom of the lower band the orbital is represented quite adequately by a plane wave $\exp(ikx)$ with momentum $\hbar k$; the wave component $\exp[i(k - G)x]$ with momentum $\hbar(k-G)$ is small and increases only slowly as k is increased, and in this region $m^* \simeq m$. An increase in the reflected component $\exp[i(k - G)x]$ as k is increased represents momentum transfer to the electron from the lattice.

Near the boundary the reflected component is quite large; at the boundary it becomes equal in amplitude to the forward component, at which point the eigenfunctions are standing waves, rather than running waves. Here the momentum component $\hbar(-\frac{1}{2}G)$ cancels the momentum component $\hbar(\frac{1}{2}G)$.

A single electron in an energy band may have positive or negative effective mass: the states of positive effective mass occur near the bottom of a band because positive effective mass means that the band has upward curvature ($d^2\epsilon/dk^2$ is positive). States of negative effective mass occur near the top of the band. A negative effective mass means that on going from state k to state $k + \Delta k$, the momentum transfer to the lattice from the electron is larger than the momentum transfer from the applied force to the electron. Although k is increased by Δk by the applied electric field, the approach to Bragg reflection can give an overall decrease in the forward momentum of the electron; when this happens the effective mass is negative (Fig. 11).

As we proceed in the second band away from the boundary, the amplitude of $\exp[i(k - G)x]$ decreases rapidly and m^* assumes a small positive value. Here the increase in electron velocity resulting from a given external impulse is larger than that which a free electron would experience. The lattice makes up the difference through the reduced recoil it experiences when the amplitude of $\exp[i(k - G)x]$ is diminished.

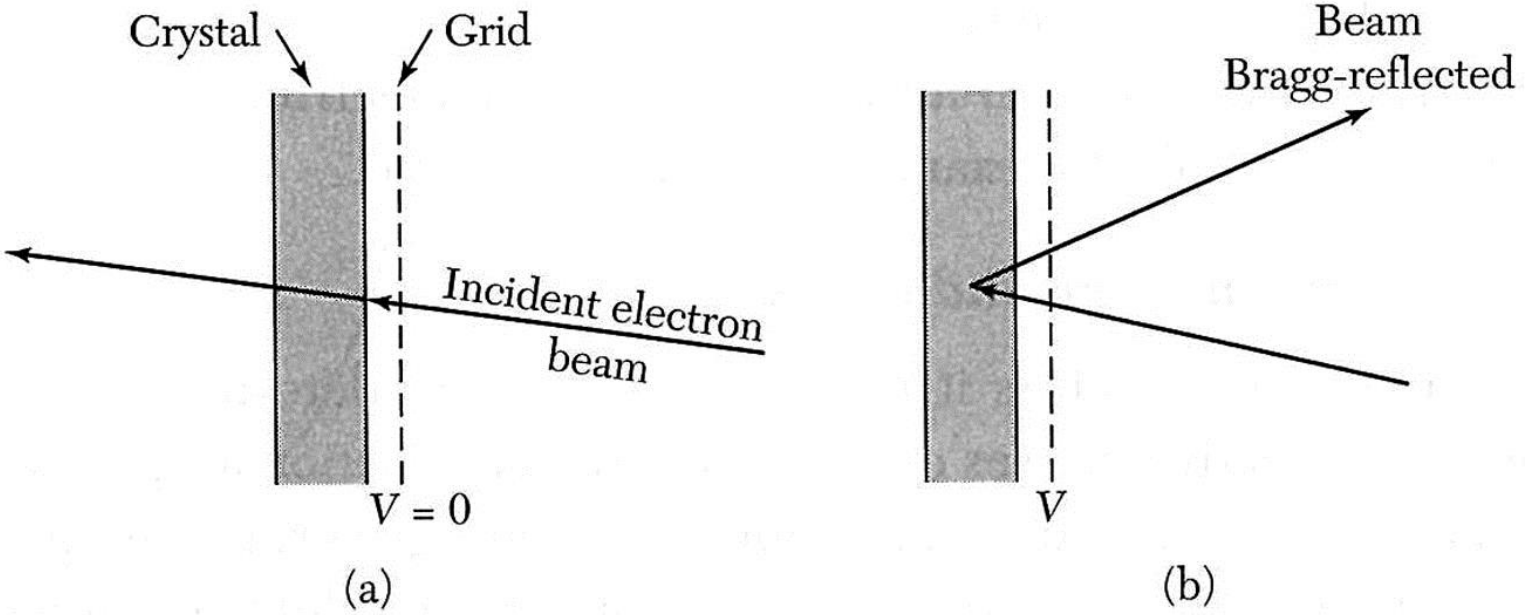


Figure 11 Explanation of negative effective masses which occur near, but below, a Brillouin zone boundary. In (a) the energy of the electron beam incident on a thin crystal is slightly too low to satisfy the condition for Bragg reflection and the beam is transmitted through the crystal. The application of a small voltage across the grid may, as in (b), cause the Bragg condition to be satisfied, and the electron beam will then be reflected from the appropriate set of crystal planes.

If the energy in a band depends only slightly on k, then the effective mass will be very large. That is, $m^*/m \gg 1$ when $d^2\epsilon/dk^2$ is very small. The tight-binding approximation discussed in Chapter 9 gives quick insight into the formation of narrow bands. If the wavefunctions centered on neighboring atoms overlap very little, then the overlap integral is small; the width of the band narrow, and the effective mass large. The overlap of wavefunctions centered on neighboring atoms is small for the inner or core electrons. The $4f$ electrons of the rare earth metals, for example, overlap very little.