IV

ELECTRONS IN A PERFECT LATTICE

4.1. Bloch theorem

IN this and the next few chapters we shall be concerned with the properties of metals. To define a metal, the physicist would refer in the first place to its high electric conductivity, the chemist to its electropositive nature, and the engineer to its ductility and other characteristic mechanical properties. We shall fix our attention on the electric conductivity, and we shall see later that we can understand why substances can be divided into two distinct classes, metals and non-metals, and that the first class has all the characteristic properties that one associates with metals in practice. One can also understand the existence of borderline cases, ranging from 'bad' metals to semi-conductors.

It is well known that the electric conductivity is, in fact, due to the presence of electrons with high mobility, and we are therefore concerned with the problem of an electron moving through a crystal lattice. In this chapter we shall discuss this problem in the case of a perfect lattice, i.e. we shall neglect the lattice vibrations as well as any lattice defects or impurities, and also the effect of any other conduction electrons, except that the space charge due to the other electrons must be allowed for in some average, since we would otherwise find very large and unrealistic electrostatic forces.

Consider, then, the problem of an electron in a periodic field of force, with a potential energy $V(\mathbf{r})$, which must have the translational symmetry of the lattice

$$V(\mathbf{r}+\mathbf{a_n}) = V(\mathbf{r}), \tag{4.1}$$

where $\mathbf{a_n}$ is any of the lattice vectors defined in Chapter I. The Schrödinger equation of the electron is

$$\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r})+\{E-V(\mathbf{r})\}\psi(\mathbf{r}) = 0 \tag{4.2}$$

We again choose a cyclic boundary condition to avoid the complications arising from the external surfaces of the crystal,

$$\psi(x+L_1,y,z) = \psi(x,y+L_2,z) = \psi(x,y,z+L_3) = \psi(x,y,z). \tag{4.3}$$

The translational symmetry allows us to classify the wave function in a similar manner to the classification of normal modes in § 1.5. This result, which was derived by Bloch (1928) amongst others, can be obtained in the following manner:

Change the origin of the coordinates in (4.2) by a_n and use (4.1). It is then evident that $\psi(r+a_n)$ is also a solution to (4.2). It still satisfies the cyclic condition (4.3). We then have two possibilities: either the energy value E is non-degenerate, so that all solutions of (4.2) must be multiples of each other, or there is degeneracy. In the first case, the new solution is obviously a multiple of the original one,

$$\psi(r+a_n) = c\psi(r). \tag{4.4}$$

The integral of the square modulus of the new function over the crystal is evidently the same as for $\psi(r)$, since it differs merely by a change of origin, and therefore

$$|c|^2 = 1. \tag{4.5}$$

In the case of degeneracy, the general solution can be expressed as an arbitrary linear combination of a basic set of solutions, and the choice of this set is to some degree arbitrary. This freedom is just sufficient to allow us to find a basic set such that each function has the property (4.4). Let the basic set at first be $\psi_1, \psi_2, ..., \psi_n$. Then the effect of the displacement on each of them gives again a solution of the equation, which is a linear combination of the basic functions

$$\psi_\mu(r+a_n) = \sum_{\nu=1}^{n} C_{\mu\nu}\psi_\nu(r), \qquad \mu = 1, 2, ..., n.$$

We may assume the original set orthogonal and normalized,

$$\int \psi_\mu^*(r)\psi_\nu(r) \, d^3r = \delta_{\mu\nu},$$

and, since these integrals are not affected by a change of the origin, the same will be true of the new functions. Hence

$$\sum C_{\nu\mu}^* C_{\mu\lambda} = \delta_{\nu\lambda},$$

which shows that the matrix C is unitary. It is well known that any unitary matrix can, by a unitary transformation, be brought to diagonal form. There exists, therefore, a set of linear combinations of the ψ which are again orthogonal and normalized, and which belong to a diagonal matrix C. This is equivalent to (4.4) being valid for each of the functions.

We would like (4.4) to apply for an arbitrary lattice vector a_n (the constant c will of course depend on n) and we must therefore make sure that, having made C diagonal for a particular displacement a_n, we are still free to achieve the same for other displacements without destroying what we have already achieved. This follows from the fact that any two displacements are commuting operations, i.e. the result is the same regardless of the order in which the operations are carried out. This

shows that the matrices C belonging to any two lattice vectors must commute, and it is well known that a set of commuting matrices can simultaneously be brought to diagonal form.

We may therefore assume, from now on, that all eigenfunctions are chosen so as to satisfy the condition (4.4) for any displacement.

The displacement $a_n + a_{n'}$, which is the result of two successive displacements, multiplies the wave function by cc'. Hence $\log c$ is additive, and must therefore depend linearly on the displacement vectors, which shows, in conjunction with (4.5), that

$$c = e^{i\mathbf{k} \cdot \mathbf{a}_n}, \tag{4.6}$$

so that

$$\psi(\mathbf{r} + \mathbf{a}_n) = e^{i\mathbf{k} \cdot \mathbf{a}_n}\psi(\mathbf{r}). \tag{4.7}$$

The vector $\mathbf{k}$ is defined only apart from an arbitrary vector $\mathbf{K}$ of the reciprocal lattice, since, by the definition of the latter,

$$e^{i\mathbf{K} \cdot \mathbf{a}_n} = 1 \tag{4.8}$$

for any $\mathbf{n}$ (1.28). We may use this freedom again to choose $\mathbf{k}$ within the basic cell of the reciprocal lattice, defined in § 6.1, which in this connexion is often called the basic 'Brillouin zone'.

The cyclic condition again restricts the components of $\mathbf{k}$ to be multiples of $2\pi/L_1$, $2\pi/L_2$, and $2\pi/L_3$, respectively. The permitted vectors $\mathbf{k}$ are thus the same as the wave vectors $\mathbf{f}$ of Chapter I, and their number is equal to the number of unit cells in the crystal.

We may therefore label the wave functions by the corresponding vector $\mathbf{k}$. There will, of course, be an infinite series of wave functions belonging to the same $\mathbf{k}$, and we shall call these $\psi_{kl}(\mathbf{r})$ and the corresponding energy eigenvalues $E_l(\mathbf{k})$. The eigenvalues belonging to the same $\mathbf{k}$ are all different, except in very special cases.

The property (4.7) may also, following Bloch, be expressed in a different form. Let

$$\psi_{kl}(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} u_{kl}(\mathbf{r}), \tag{4.9}$$

where u is defined by this equation. Then (4.7) shows that

$$u_{kl}(\mathbf{r} + \mathbf{a}_n) = u_{kl}(\mathbf{r}). \tag{4.10}$$

In other words, $u_{kl}(\mathbf{r})$ has the translational symmetry of the lattice.

This function satisfies the equation

$$\left[\frac{\hbar^2}{2m}(\nabla - i\mathbf{k})^2 + \{E_l(\mathbf{k}) - V(\mathbf{r})\}\right]u_{kl}(\mathbf{r}) = 0, \tag{4.11}$$

which, because of the periodicity of $u_{kl}(\mathbf{r})$, need only be solved for one unit cell, with boundary conditions on opposite faces of the cell, to make the periodic continuation possible.

We see at once that by going to complex conjugate quantities we have the same equation for $-\mathbf{k}$ with u^* as solution. Hence

$$u_{-\mathbf{k}l}(\mathbf{r}) = u^*_{\mathbf{k}l}(\mathbf{r}); \qquad E_l(-\mathbf{k}) = E_l(\mathbf{k}). \qquad (4.12)$$

One also knows that a small change in the coefficients of a differential equation will cause a small change in its solution. If we assign the label correctly, $E_l(\mathbf{k})$ must therefore be a continuous function of $\mathbf{k}$, and have a continuous derivative. (It can, in fact, be shown to be an analytic function of $\mathbf{k}$, but we do not need to use this.)
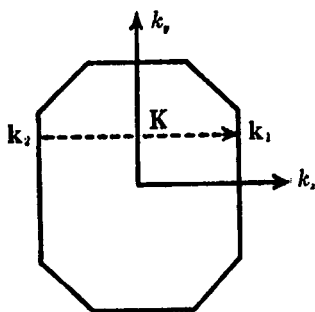

FIG. 7.

This continuity also applies as between points on opposite faces of the basic cell in $\mathbf{k}$ space, if we remember that the location of $\mathbf{k}$ values in this cell was purely a matter of convention. This is illustrated by Fig. 7 for the plane case. The polygon indicates the basic cell, the point $\mathbf{k}_1$ a value on the boundary. Now this differs from $\mathbf{k}_2$ on the opposite boundary just by a reciprocal lattice vector, $\mathbf{K}$. Points to the immediate right of $\mathbf{k}_2$ might just as well have been placed to the immediate right of $\mathbf{k}_1$. This shows that

$$E_l(\mathbf{k}_1) = E_l(\mathbf{k}_2); \quad \operatorname{grad} E_l(\mathbf{k}_1) = \operatorname{grad} E_l(\mathbf{k}_2). \qquad (4.13)$$

Now suppose that the symmetry of the lattice permits a reflection in the plane $x = 0$. From this we may conclude that

$$E_l(\mathbf{k}_1) = E_l(\mathbf{k}_2); \qquad \frac{\partial E_l(\mathbf{k}_1)}{\partial k_y} = \frac{\partial E_l(\mathbf{k}_2)}{\partial k_y};$$

$$\frac{\partial E_l(\mathbf{k}_1)}{\partial k_x} = -\frac{\partial E_l(\mathbf{k}_2)}{\partial k_x}. \qquad (4.14)$$

The first two statements are identical with those derived before from continuity, but the last one differs in sign and hence

$$\frac{\partial E_l(\mathbf{k}_1)}{\partial k_x} = 0. \qquad (4.15)$$

In general, provided the lattice has sufficient symmetry, the derivative of E in a direction normal to a cell boundary vanishes on that boundary. This is clearly also true in three dimensions.

There is a possible exception to this rule when two different functions, say $E_1(\mathbf{k})$ and $E_2(\mathbf{k})$, happen to be equal on the boundary. In that case

(4.13) need not apply; it may instead be that the gradient of E_1 at $\mathbf{k}_1$ equals the gradient of the other function E_2 at $\mathbf{k}_2$. We shall meet an example of this later, but shall also see that it is very exceptional.

Fig. 8 shows, for the case of a linear chain, some typical $E(k)$ curves which are compatible with the results so far obtained.†
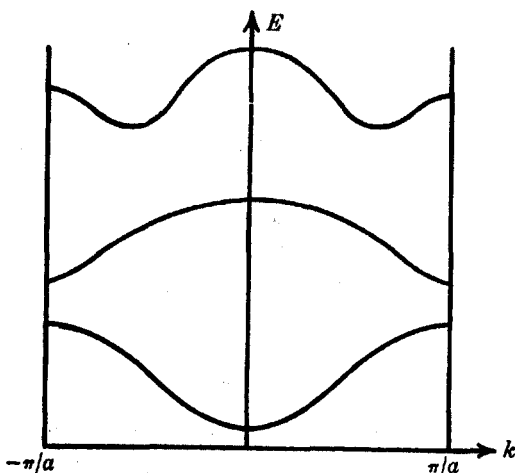

FIG. 8.

4.2. Strong binding

So far we have made no assumption about the potential function $V(\mathbf{r})$ beyond its periodicity, but one may gain further insight into the problem by solving the equations for two limiting cases, that of strongly bound and that of nearly free electrons.

In the limit of strong binding, we assume the lattice constant so large, or the atomic radius so small, that the wave functions of atoms at adjacent lattice points would overlap very little.

To avoid inessential complications, we assume a linear problem, and identical atoms. We may then write the potential in the form

$$V(x) = \sum_n U(x-na), \tag{4.16}$$

where $U(x)$ is the potential energy of the electron near an atom at the origin, and the sum goes over all atoms in the chain. We then expect that the eigenfunction will bear some relation to the eigenfunction $\phi(x)$

---

† Dr. Shockley has pointed out to me that one can prove in the one-dimensional case that the minima or maxima of $E(k)$ may occur only at $k = 0$ or $k = \pm \pi/a$. The upper curve of Fig. 8 cannot therefore be realized in one dimension, but it may occur for a linear chain of three-dimensional atoms.

$h(n-m)$ contains all potentials except that of the mth atom, and since the integrand contains ϕ_m as a factor, the integral is small. $J(n-m)$ is small, unless $n = m$.

The condition that (4.20) be stationary is

$$\sum_m \{H(n-m)-EJ(n-m)\}A_m = 0,$$

or, using (4.22),

$$\sum_m \{h(n-m)-(E-E_0)J(n-m)\}A_m = 0. \tag{4.23}$$

We can now find a solution in the form

$$A_n = Ae^{i\kappa n}. \tag{4.24}$$

Insertion in (4.23) shows that this is satisfied, provided

$$E-E_0 = \frac{\sum\limits_n h(n)e^{i\kappa n}}{\sum\limits_n J(n)e^{i\kappa n}}. \tag{4.25}$$

Now both $h(n)$ and $J(n)$ decrease very rapidly with increasing n, since they contain the product of two atomic wave functions displaced a distance na relative to each other. However, $h(0)$ and $h(1)$ are of comparable magnitude and $J(0)$ is unity if ϕ is normalized. Then, to the leading order,

$$E-E_0 = h(0)+2h(1)\cos\kappa. \tag{4.26}$$

Our 'zero-order' wave function (4.19) with the coefficients (4.24) evidently satisfies (4.7) with

$$\kappa = ka. \tag{4.27}$$

The first-order energy (4.26) is similar to the lower two curves sketched in Fig. 8.

We have not at any point assumed that the eigenvalue E_0 represents the lowest energy of the atom, and the procedure can be carried out for each atomic level. In this way we obtain a set of energy levels $E_l(k)$ and corresponding wave functions, which form a complete orthogonal set. They are therefore suitable for improving the approximation further by perturbation theory of higher order if required.

In this approximation each of the atomic levels is drawn out into a continuous band of width $4h(1)$. This width must be small compared to the spacing of the atomic levels if our approximations are justifiable.

We have here tacitly assumed that the atomic state is non-degenerate, which, in fact, is usually the case in one dimension. If we are dealing with a degenerate level of s components, our wave function will have to be taken as a combination of sN different atomic functions. After splitting

of a single atom. Indeed, any of the N functions $\phi_n(x) = \phi(x-na)$ is very nearly a solution of the wave equation

$$\frac{\hbar^2}{2m} \frac{\partial^2\psi(x)}{\partial x^2} + \left\{E - \sum_n U(x-na)\right\}\psi(x) = 0 \qquad (4.17)$$

with the eigenvalue E_0 of the single atom. This can be seen by inserting ϕ_n in (4.17). Then all products of potential terms with the wave function are small, except for the term with the same n. Neglecting the others, we are left with

$$\frac{\hbar^2}{2m} \frac{d^2\phi(x-na)}{dx^2} + \{E - U(x-na)\}\phi(x-na) = 0, \qquad (4.18)$$

which is the equation for the single atom, which is satisfied when $E = E_0$.

To improve the solution we apply perturbation theory. Since we have N functions which satisfy (4.17) equally well and with the same energy value, we expect the correct solution to be close to some linear combination

$$\sum_n A_n \phi_n. \qquad (4.19)$$

We cannot, however, determine the best choice of the A_n by the usual method of perturbation theory, since we cannot divide the potential into a large part and a small perturbation. Each part of the potential is large for one of the wave functions contained in (4.19).

Instead we select the coefficients of the linear combination (4.19) by using the variation principle; we choose the coefficients A_n in such a way as to make the expectation value of the energy stationary. This expectation value is

$$E = \frac{\sum\limits_{n,m} A_n^* A_m H_{n,m}}{\sum\limits_{n,m} A_n^* A_m J_{n,m}}, \qquad (4.20)$$

where

$$J_{n,m} = \int \phi_n \phi_m \, dx; \qquad H_{n,m} = \int \phi_n\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V\right)\phi_m \, dx. \qquad (4.21)$$

We note, first of all, that $J_{n,m}$ and $H_{n,m}$ depend only on the difference of the suffixes $n-m$, as one can see at once by a change of origin in the integrals (4.21), and that $J_{n,m} = J_{m,n}$, $H_{m,n} = H_{n,m}$. We may therefore denote them by $J(n-m)$ and $H(n-m)$.

Using (4.18), one finds that

$$H(n-m) = E_0 J(n-m) + \int \phi_n\{V(x) - U(x-ma)\}\phi_m \, dx$$

$$= E_0 J(n-m) + h(n-m). \qquad (4.22)$$

off an exponential factor (4.24) we are still left with a set of s-fold simultaneous equations. I shall not cover this case in detail.

These results are immediately generalized to three dimensions. The zero-order wave function takes the form

$$\psi_{k,l}(\mathbf{r}) = \sum_{n,j} A_j e^{i\mathbf{k}.\mathbf{a}_n} \Phi_l(\mathbf{r} - \mathbf{a}_n - \mathbf{d}_j), \qquad (4.28)$$

where $\Phi_l(\mathbf{r})$ is again a single-centre eigenfunction. In a simple lattice, with one atom per unit cell, there is no summation over j, and (4.28) determines the solution completely. In an element with a more complex lattice, A_j has still to be obtained from a set of r simultaneous equations. If the atomic state is degenerate, the number of equations is further increased. For a crystal containing different species of atoms, the sum over j should cover only the atoms of one kind, since the unperturbed energy levels of different atoms are in general different, and the corresponding states not degenerate.

In the simplest case of $r = 1$, and an atomic s-level, (4.25) becomes

$$E - E_0 = \frac{\sum_n h(\mathbf{n}) e^{i\mathbf{k}.\mathbf{a}_n}}{\sum_n J(\mathbf{n}) e^{i\mathbf{k}.\mathbf{a}_n}}, \qquad (4.29)$$

where the meaning of the coefficients is evident by a simple generalization of (4.21) and (4.22).

For example, for a body-centred cubic lattice, for which the reciprocal lattice is face-centred cubic, and has as basic cell a rhombohedron, the leading terms in (4.29), corresponding to the approximation (4.26), give

$$E - E_0 = h_0 + 8h_1 \cos \tfrac{1}{2}k_x a \cos \tfrac{1}{2}k_y a \cos \tfrac{1}{2}k_z a. \qquad (4.30)$$

Fig. 9 shows two sections through the basic reciprocal cell, for $k_z = 0$, and $k_z = \pi/2a$, respectively, with lines of constant energy to the approximation (4.30). Near $\mathbf{k} = 0$, the surfaces of constant energy are spheres and, for negative h_1, this region corresponds to the energy minimum. The highest energy is attained at the corners, for which any one of the components of $\mathbf{k}$ is $2\pi/a$, and the others zero. These points are all equivalent. In their neighbourhood the surfaces of constant energy are parts of spheres, which may be thought of as forming a continuous sphere, cut into octants by our convention about the choice of $\mathbf{k}$.

Between these extremes, the energy surfaces are more complicated in shape. For the special case $E = E_0 + h_0$, the surface is a cube.

From a physical point of view, the approximation of this section may be pictured in the following way: For tight binding an electron is

almost in a stationary state when it describes an orbit near one of the attractive centres. There is, however, a probability of a transition in which it changes to an orbit about the adjacent centre, in spite of the fact that there is a potential barrier between them (tunnel effect). It follows that the electron will progress slowly through the lattice. What is less obvious from this picture is the fact that in a perfect lattice this motion has associated with it the wave vector **k**, which is a constant of the motion and which, as we shall see, results in the electron progressing, on the average, with uniform velocity.

In a real crystal, this approximation would be reasonable for the electrons in the inner shells, whose radius is small compared to the inter-atomic distances, so that the integrals h and J are small. These levels will be drawn out into bands whose width is very small compared to the distances between them. In the outer shells, which are of interest for the conduction electrons, the overlap will be large, and the results of this section have at best qualitative significance.
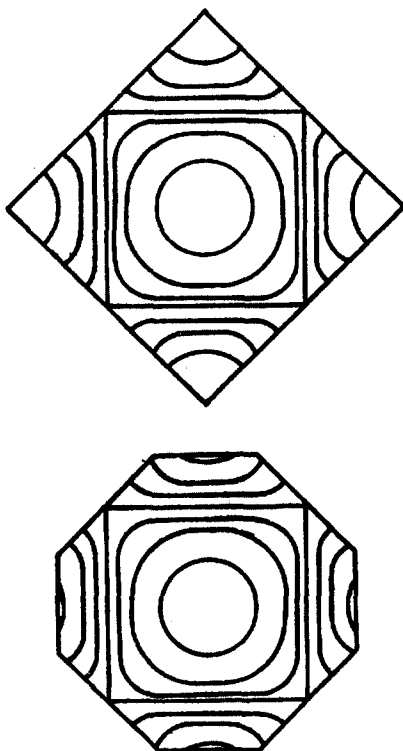

FIG. 9.

4.3. Nearly free electrons

One may also solve the problem in the opposite limit of nearly free electrons. We assume the potential $V(\mathbf{r})$ to be weak, so that we may start from free electrons and then apply a perturbation.

The condition for the validity of perturbation theory is that the matrix elements of the perturbing potential be small compared to the energy differences between the levels to which these matrix elements refer. We again begin, for simplicity, with the one-dimensional case.

The unperturbed wave functions are

$$\psi_p = \frac{1}{\sqrt{L}} e^{ipx}, \tag{4.31}$$

where L is the length of the chain, and p the momentum measured in units of $\hbar$. The energy is

$$E(p) = \frac{\hbar^2}{2m} p^2. \qquad (4.32)$$

The matrix element of the potential between two states p, p' is

$$V(p, p') = (1/L) \int V(x) e^{i(p-p')x} \, dx.$$

Because of the periodicity of V, this integral vanishes unless $p' - p$ is a multiple of $2\pi/a$ and in that case it is

$$V_\kappa = (1/a) \int_{-a/2}^{a/2} V(x) e^{2\pi i \kappa x / a} \, dx,$$

$$\kappa = \frac{a(p-p')}{2\pi} = 0, \pm 1, \pm 2, \dots. \qquad (4.33)$$

The case $\kappa = 0$ represents the diagonal element, and hence the first-order change in the energy. This is equal to the space average of the potential, and it is convenient to choose the zero level of energy in such a way that this average vanishes. Then $V_0 = 0$, and the first-order shift of all energy levels is zero.

Now consider a matrix element with $\kappa \neq 0$. The energy difference between the corresponding states is

$$\frac{\hbar^2}{2m} \left\{ \left(p - \frac{2\pi\kappa}{a} \right)^2 - p^2 \right\} = \frac{2\pi\hbar^2\kappa}{ma} \left(\frac{\pi\kappa}{a} - p \right). \qquad (4.34)$$

This evidently vanishes when $p = \pi\kappa/a$, and in that case $p' = -p$. This is the case of a wave for which Bragg scattering is possible, and this state is obviously connected by the potential with another state of the same energy. In the neighbourhood of such states the condition for perturbation theory is therefore never satisfied. On the other hand, for states well away from these exceptional cases, the quantity (4.34) is finite, and we may postulate that it is larger than V_κ. In that region, the energy shift is of second order in V, precisely

$$\Delta E = \sum_\kappa \frac{|V_\kappa|^2}{E(p) - E\{p - (2\pi\kappa/a)\}} \qquad (4.35)$$

apart from terms of higher order.

In the neighbourhood of the Bragg reflection of order κ, we may assume that all but one of the energy differences appearing in (4.35) are large. We may then apply the perturbation theory for nearly degenerate states by assuming that the energy difference between the two states in question is comparable with the matrix element V. Denoting the

two states for brevity by 1 and 2, and their unperturbed energies by E_1 and E_2, we then try to find a solution of the perturbed equation which lies close to

$$A_1\psi_1 + A_2\psi_2. \tag{4.36}$$

The coefficients A_1, A_2 are to be determined in such a manner as to diagonalize that part of the total energy operator whose matrix elements
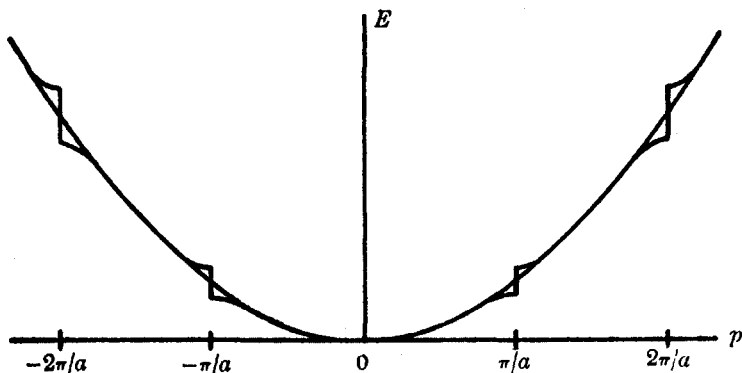

FIG. 10.

do not connect the two chosen states with any others. This leads to the equations

$$\left.\begin{array}{l}(E - E_1)A_1 + V_\kappa A_2 = 0 \\ V_\kappa^* A_1 + (E - E_2)A_2 = 0\end{array}\right\} \tag{4.37}$$

which have a solution if the determinant vanishes. This requires

$$E = \tfrac{1}{2}(E_1 + E_2) \pm \sqrt{\left\{\left(\frac{E_1 - E_2}{2}\right)^2 + |V_\kappa|^2\right\}}. \tag{4.38}$$

If the energy difference is large compared with V_κ, then the two energy values resulting from this lie close to E_1 and E_2 respectively, except for terms of the order $|V_\kappa|^2/(E_1 - E_2)$, which may then be taken together with the terms from transitions to other states. If, on the other hand, $E_1 = E_2$, as for exact Bragg reflection, then the two values of (4.38) differ by $2|V_\kappa|$. For small difference $E_1 - E_2$ the gap between the perturbed energy values differs from $2|V_\kappa|$ only by an amount proportional to the square of $E_1 - E_2$.

Fig. 10 indicates how the energy of free electrons becomes modified by a small disturbance of this kind. The cases $\kappa = \pm 1, \pm 2$ are shown.

It is easy to see from (4.37) that for $E_1 = E_2$, i.e. for points on either side of the discontinuity, $|A_1/A_2| = 1$. The wave functions (4.36) are therefore combinations of two plane waves (4.31) travelling in opposite directions with equal amplitude, and thus are standing waves. The one

with the lower energy has its maximum amplitude at the centres, where the potential is most attractive, and the one with the higher energy has its nodes there.

We must now connect these results with the general theory of § 4.1. We note, for this purpose, that already the unperturbed wave (4.31) satisfies the theorem (4.7), which is not surprising, since the absence of a potential is quite consistent with the periodicity condition (4.1). However, p is not adjusted to meet the convention for k, and to get back to the same variables we must take k as

$$k = p - \frac{2\pi\kappa}{a}, \qquad \left. -\frac{\pi}{a} \right] \leqslant k \leqslant \frac{\pi}{a}. \qquad (4.39)$$

In Fig. 11 the curve of Fig. 10 has been re-plotted in terms of k. The parabola for the free electrons is now broken up into sections, each of which now represents an energy curve $E_l(k)$. At any point $k = 0$, or $k = \pm\pi/a$, two such curves intersect, and hence the reasoning leading to (4.15) fails. But if we apply a periodic potential, however weak, the intersection disappears, and the curve has zero slopes at the zone boundary.
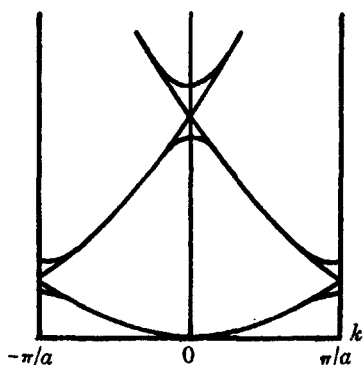
We see therefore that the qualitative appearance of the energy curves is very similar in the two extreme limits of very strong and very weak binding, the principal difference being that in the one case we have narrow bands widely separated from each other, and in the other wide energy bands separated by small gaps. Another important difference is the uneven curvature of the curves of Fig. 11, with its very narrow turns near the boundaries.

FIG. 11.

It is evident that the present method is best at high energies, since we have neglected the matrix elements of V in comparison with the vertical distance of the various curves of Fig. 11, except the nearest one. These vertical distances increase as we go upwards, whereas the matrix elements (4.33) decrease with increasing order.

Our present results also give a convenient method for discussing electron diffraction. If an external beam of electrons is incident on a crystal, and its energy lies in one of the energy gaps appearing in Fig. 10 or 11, it must be totally reflected, since there is no possible motion of an electron inside the crystal with that energy. By the method of Chapter

III we should have calculated an infinitely intense scattered wave for a sharply defined wavelength, and hence for a single value of the energy. Instead, we now find a finite reflected amplitude (equal to the amplitude of the wave incident on the crystal) but over a finite range of energies. This result is equivalent to Ewald's dynamical theory of diffraction, which was developed for the X-ray case. For a quantitative treatment, surface effects must, however, also be taken into consideration.

The generalization of the treatment of nearly free electrons to three dimensions presents little difficulty. In momentum space the condition

$$(\mathbf{p} - \mathbf{K})^2 = \mathbf{p}^2,   \qquad (4.40)$$

which is just a form of the Bragg condition, represents, for each reciprocal lattice vector $\mathbf{K}$, a plane across which a discontinuity will appear. These planes dissect momentum space into polyhedra, of which the one containing the origin is the 'basic Brillouin zone'. Of the others we can always find a set which, after displacements by suitable reciprocal vectors $\mathbf{K}$, will together just cover the basic zone again, and then belong to a continuous energy function. The polyhedra, which together make up such a set, are known as Brillouin zones.

4.4. Velocity and acceleration

Since the wave function $\psi_l(\mathbf{k})$ of a particular electron state is complex, we would in general expect it to have a non-vanishing mean velocity. This can be derived easily from the equation of continuity. Suppose we are not dealing with a strictly stationary state, but with a wave packet made up of a number of waves with adjacent values of $\mathbf{k}$. If the spread of $\mathbf{k}$ values used in building the wave packet is to be small compared to the dimensions of the basic Brillouin zone, the extent of the wave packet in space must be large compared to the size of a lattice cell, but it may still be very small compared to the size of the crystal.

We can then find the displacement in time of the mean position of the packet. It must, by a well-known argument, equal the group velocity

$$\mathbf{v} = \frac{\partial \omega}{\partial \mathbf{k}} = \frac{1}{\hbar} \frac{\partial E_l(\mathbf{k})}{\partial \mathbf{k}}.   \qquad (4.41)$$

(By the derivative with respect to the vector $\mathbf{k}$ we mean that each component of $\mathbf{v}$ equals the derivative with respect to the corresponding component of $\mathbf{k}$.) Now the rate at which the wave packet travels must equal the transport in the wave packet, and, since the states from which it is built are very similar, we may identify the mean transport with the transport for each of them. Hence (4.41) gives the transport, or the

average velocity, for a state **k**. From (4.15) it follows that the normal velocity component vanishes at a zone boundary, in line with the fact that at such a boundary Bragg reflection makes any progress of the electron impossible.

Next consider the action of an applied electric field. This will induce transitions from one stationary state to others. In the weak fields that, in practice, can be applied to metals, we may disregard transitions in which l changes, since this will involve a much greater change of the electron energy than the field can supply with any reasonable probability (this is not the case when one is concerned with the electric breakdown of insulators). We may therefore suppose that the only effect of the field is to alter the value of **k**. Now if **k** is the mean wave vector of a wave packet, the change in the electron energy with time is

$$\frac{\partial E_l}{\partial \mathbf{k}} \cdot \frac{d\mathbf{k}}{dt} \,.$$

On the other hand, the work done by the field per unit time is

$$e\mathbf{v}.\mathbf{F} = \frac{e}{\hbar} \frac{\partial E_l(\mathbf{k})}{\partial \mathbf{k}} . \mathbf{F},$$

where the second expression is taken from (4.41). From energy conservation these two quantities must be equal, and hence

$$\frac{d\mathbf{k}}{dt} = \frac{e}{\hbar}\mathbf{F} \tag{4.42}$$

must give the rate of change of **k** in the field. This result is obvious for free electrons, for which $\hbar\mathbf{k}$ is the momentum, apart from an additive constant.

That it should always hold is at first sight surprising if you look at curves like those of Figs. 8 or 11, which show that there are always regions in which the slope of the energy curve decreases with increasing **k**. Since this slope measures the mean velocity, it follows that it is possible for the field to accelerate the electron in a direction opposite to the usual. This is less strange if we remember the connexion with diffraction. What happens is that the field pushes the electron near to the state in which complete reflection brings it, in the mean, to rest.

By analogy one would expect that the effect of a magnetic field would be to change **k** at the rate

$$\frac{d\mathbf{k}}{dt} = \frac{e}{\hbar c}\mathbf{v} \wedge \mathbf{H}, \tag{4.43}$$

where c is the velocity of light, **v** the velocity (4.41) and the sign $\wedge$ indicates a vector product. This formula is indeed correct for many

purposes. It neglects, however, the fact that, in a magnetic field, the electron describes closed orbits (in the plane at right angles to the field direction) and that therefore the energy levels will be in part discrete. This fact, which is important for the theory of diamagnetism, cannot be expressed in terms of an acceleration; we shall return to this problem at a later stage.

4.5. Many electrons. Statistics

We are now ready to consider the distribution of electrons over the stationary states we have previously described. To do this rigorously one would have to allow for the interaction of the electrons with each other, but this would give rise to formidable mathematical difficulties. The inability to include the electron interaction is the greatest source of uncertainty in the application of the electron theory of metals. In certain problems we shall be able to see what the effect of the electron interaction would be and even to allow for it approximately, but in many cases a large measure of doubt must remain.

For the present we assume, therefore, that the electrons do not interact, except that their mean charge distribution is taken into account in the definition of our potential.

We must, however, take into account Pauli's exclusion principle, according to which any one quantum state can contain at most two electrons of opposite spin. The state of lowest energy of the whole system is therefore obtained by filling the lowest $Nrz/2$ states, where, as before, N is the number of unit cells in the crystal, r the number of atoms in a cell, and z the number of electrons per atom. The energy of the highest state so occupied will be called η_0. Now there are two distinct possibilities. Either η_0 coincides with the upper end of an energy band, so that this band is completely filled and the next higher one empty, or η_0 lies inside a band.

In the first case, the state of lowest energy is not the beginning of a continuous energy spectrum for the whole crystal, but is isolated from all high states by a finite gap. This is evident since no other arrangement of the electrons is possible without moving at least one of them up into the next band, which takes a finite amount of energy. Such a substance could not respond to any weak external disturbance, and could not, for example, redistribute the electrons to cancel an external potential difference. It must therefore be an insulator. On the other hand, if there are free states in the last band, it is possible to shift the electrons to states of slightly different energy so as to set up a net current, or, by the

formation of wave packets, produce a non-uniform charge distribution. This is the case of a metal.

In the case of one atom per unit cell, where the number of states in a band equals the number of atoms, it is not possible to fill a number of bands completely, unless the number of electrons per atom is even. All elements of odd atomic number with a structure for which $r = 1$ should be metals.

At first sight one would expect the converse to be true, that all even elements with $r = 1$ should be non-metals. However, this is not correct, since the bands may not be distinct in energy. It is true that for given $\mathbf{k}$ the different energy values $E_l(\mathbf{k})$ are in general distinct, but the highest value of $E_l(\mathbf{k})$ may well be above the minimum of E_{l+1}. This will always be the case at very high energy, where from the work of § 4.3 we saw that the energy gaps at a zone boundary are very small, whereas the variation of energy along the boundary will be appreciable.

Amongst elements with simple structures the metals are therefore predominant, but in the case of more complex structures the number of states in a band is only $1/r$ times the number of atoms, and it is then more likely that they will be separated in energy. We shall see later that there is, in fact, a tendency to favour a structure for which this happens to be the case.

Bands which are completely filled may in general be left out of consideration completely in the discussion of conduction phenomena and the like. These certainly include the states corresponding to the innermost atomic shells. We shall refer to the incompletely filled band as the conduction band. From the above discussion it is clear that there may be, and in a metal of even atomic number with $r = 1$ there must be, more than one conduction band.

A particularly simple case is that in which a certain band contains a small number of electrons. These will then move in states near the bottom of the band, where the energy may be regarded as a quadratic function of $\mathbf{k}$. If this energy minimum is attained for only a single value of $\mathbf{k}$, then the energy surface in its neighbourhood must have the same symmetry as the structure. If the crystal is cubic, the quadratic function must be isotropic and therefore

$$E(\mathbf{k}) = \text{constant} + b(\mathbf{k} - \mathbf{k}_0)^2 \qquad (4.44)$$

with constant b. Apart from the change in origin, this is similar to the energy function for free electrons, and we may write

$$b = \frac{\hbar^2}{2m^*},$$

where m^* is of the dimension of a mass, and may be called the effective mass. In such states, the behaviour of electrons is in many respects the same as that of free charged particles of mass m^*. It may also be, however, that the energy function, which has the crystal symmetry, reaches its lowest value at a number of points which are symmetrically placed. In that case the function is still quadratic near each of those points, but we can no longer claim that it must be isotropic. The surfaces of constant energy near one minimum may then be ellipsoids, and it merely follows, for a cubic crystal, that a rotation by 90° about a crystal axis must convert this ellipsoid into the corresponding ellipsoid at another minimum.

An equally simple case is that in which the band is almost full, with only a small fraction of the states vacant. In that case it is convenient to specify the vacant states rather than the occupied ones. Since the full band has no conductivity, all conduction processes may be described as being due to the presence of this small number of holes from which electrons have been removed, as compared to a completely filled band. The absence of an electron means the addition of a positive charge. An electric field in the x-direction will then shift all electrons to states of increased k_x, according to (4.42), which, by (4.41), near an energy maximum means decreasing velocity. The vacant states or holes are also now moving with decreasing velocity in the x-direction, as if they were real positive charges. The absence of electrons travelling in the $-x$-direction represents, of course, a current in the positive x-direction, so that the current is in the direction of the field. The energy required to produce a vacancy is, of course, a minimum for the state of highest energy, and if, near a single maximum, the energy is

$$\text{constant} - b(\mathbf{k} - \mathbf{k}_0)^2,$$

the constant b can be linked to an effective mass of the positive holes as before. It is hardly necessary to point out the similarity of this reasoning to Dirac's theory of the positron.

At a finite temperature, when the system is not in the state of lowest energy, the electron distribution is governed by Fermi-Dirac statistics. The average number of electrons in a state of energy E is given by

$$2f(E) = \frac{2}{1 + e^{(E - \eta)/kT}}, \tag{4.45}$$

and the free energy is

$$F = N_e \eta - 2kT \sum_i \log\{1 + e^{-(E_i - \eta)/kT}\}, \tag{4.46}$$

where N_e is the total number of electrons, and i labels the states, E_i being

the corresponding energies. The factor 2 in (4.45) and (4.46) takes account of the two spin directions.

The constant η-has to be determined in such a way as to make the total number of electrons in all states equal to N_c, or

$$N_c = \sum_i 2f(E_i).\tag{4.47}$$

In practice, kT is very small compared to the width of a band, and compared to the width of the filled and the vacant parts of the band, unless the number of conduction electrons or vacancies is exceptionally small. We may then apply the usual approximation for a highly degenerate Fermi gas.

In particular, when we are dealing with an integral over the Fermi function of the form

$$\int_{-\infty}^{\infty} dE \frac{dG}{dE} f(E),\tag{4.48}$$

where G is some function of the energy which is smooth near $E = \eta$, we may integrate by parts to obtain

$$-G(-\infty) - \int_{-\infty}^{\infty} dE\ G(E) \frac{df}{dE}.$$

Now df/dE is greatest near $E = \eta$, and vanishes rapidly if $|E - \eta|$ exceeds kT. It is therefore convenient to expand G in a Taylor series near η, and we obtain

$$-G(-\infty) + G(\eta) + G'(\eta)f_1 + \tfrac{1}{2}G''(\eta)f_2 + \tfrac{1}{6}G'''(\eta)f_3 + \cdots,$$

where the accents denote derivatives, and $f_1, f_2, \ldots$ the moments

$$f_n = -\int_{-\infty}^{\infty} dE(E - \eta)^n \frac{df}{dE}.$$

Now since

$$-\frac{df}{dE} = \frac{1}{kT} \frac{1}{(1 + e^{(E-\eta)/kT})(1 + e^{-(E-\eta)/kT})}\tag{4.49}$$

is evidently an even function of $(E - \eta)$, all odd moments $f_1, f_3, \ldots$ vanish. The second moment f_2 is

$$f_2 = \frac{\pi^2}{3}(kT)^2,$$

so that, apart from terms of order T^4 or higher, the required integral is

$$\int_{-\infty}^{\infty} dE \frac{dG}{dE} f(E) = G(\eta) - G(-\infty) + \frac{\pi^2}{6}(kT)^2 G''(\eta).\tag{4.50}$$

This result may be used to evaluate condition (4.47), which may be written in integral form, if $Z(E)$ denotes the number of energy levels below E,

$$N_e = 2 \int_{-\infty}^{\infty} dE \frac{dZ}{dE} f(E) = 2Z(\eta) + \frac{\pi^2}{3} \left(\frac{d^2Z}{dE^2}\right)_\eta (kT)^2. \qquad (4.51)$$

Hence we can find the temperature dependence of η,

$$\eta - \eta_0 = -\frac{(d^2Z/dE^2)_\eta}{(dZ/dE)_\eta} \frac{\pi^2}{6} (kT)^2. \qquad (4.52)$$

The ratio of the first to the second derivative of Z is the size of the energy interval over which the density of states changes appreciably. As long as this is larger than kT, $\eta - \eta_0$ is small compared to kT, and may therefore for most purposes be neglected.

4.6. Specific heat

The energy content of the electron system is easily calculated from the results of the preceding section.

The total energy is evidently

$$E(T) = \sum_i E_i f(E_i), \qquad (4.53)$$

where i again numbers all possible electron states. Introducing again the number $Z(E)$ of states with energy below E,

$$E(T) = 2 \int_{-\infty}^{\infty} f(E) E \frac{dZ}{dE} \, dE, \qquad (4.54)$$

and, using (4.50), with

$$G(E) = \int_{-\infty}^{E} E \frac{dZ}{dE} \, dE, \qquad (4.55)$$

we find

$$E(T) = 2 \int_{-\infty}^{\eta} E \frac{dZ}{dE} \, dE + \frac{\pi^2}{3} (kT)^2 \left(\frac{dZ}{dE} + E \frac{d^2Z}{dE^2}\right)_\eta.$$

We have already seen that η is very little different from η_0, and therefore we may take the first integral to the limit η_0 instead of η and correct to first order in the difference:

$$E(T) = 2 \int_{-\infty}^{\eta_0} E \frac{dZ}{dE} \, dE + 2 \left(E \frac{dZ}{dE}\right)_{\eta_0} (\eta - \eta_0) + \frac{\pi^2}{3} (kT)^2 \left(\frac{dZ}{dE} + E \frac{d^2Z}{dE^2}\right)_{\eta_0}.$$

Inserting for $\eta - \eta_0$ from (4.52),

$$E(T) = 2 \int_{-\infty}^{\eta_0} E \frac{dZ}{dE} \, dE + \frac{\pi^2}{3} (kT)^2 \left(\frac{dZ}{dE}\right)_{\eta_0}$$
$$= E(0) + \frac{\pi^2}{3} (kT)^2 \left(\frac{dZ}{dE}\right)_{\eta_0}. \tag{4.56}$$

The electronic specific heat is

$$\frac{dE}{dT} = 2\frac{\pi^2}{3} k \cdot kT \left(\frac{dZ}{dE}\right)_{\eta_0}. \tag{4.57}$$

The results (4.56) and (4.57) can easily be interpreted. The constant energy term in (4.56) is just the energy of the lowest state of the system, obtained by filling all states up to the level of the Fermi energy η_0 and leaving the rest vacant. At a finite temperature T the only states whose occupation numbers are different from these at zero temperature are those within an energy region of the order of kT about η_0. The electrons in this region are the only ones which can adjust themselves to the rising temperature, the rest are still 'frozen in'. It is therefore reasonable that the specific heat should be of the order of k for each of those electrons. Their number is the density of states dZ/dE, times the width kT of the energy region.

In a non-metallic crystal η_0 coincides with the upper limit of an energy band, which at this point does not overlap with any other. Hence in that case $(dZ/dE) = 0$, and there is no electronic specific heat, as long as the approximation (4.50) is justified.

In a metal, we may as a crude estimate for orientation expect dZ/dE to be of the order of the number of atoms divided by an energy of the order of the band width. Since this is in general of the order of a few eV, whereas kT at room temperature is of the order of $3 \cdot 10^{-2}$ eV, the electronic specific heat at room temperature is something like 1 per cent. of the Dulong-Petit value.

This result, which was first pointed out by Sommerfeld, removes the discrepancy that existed between the evident fact that metals contained large numbers of apparently free electrons, and the lack of any significant contribution to the specific heat that could be ascribed to them.

However, in an ordinary metal, the electronic specific heat is more easily observable either at rather higher, or at very low temperatures. At higher temperatures, when the vibrational specific heat is approximately constant, a term proportional to the temperature is more easily noticed. However, we have seen in § 2.3 that the crystal lattice itself gives

a correction to the specific heat proportional to T, and a precise knowledge of this effect would be needed to isolate the electron contribution.

It is easier to observe the effect at low temperatures when the vibrational specific heat becomes proportional to T^3, so that, in the limit of very low T, the electronic part will be dominant. Experiments in this region therefore give an estimate of the density of electron levels near the Fermi energy.

Our crude estimate of dZ/dE will be unreasonable in two cases: (a) When the Fermi energy η_0 lies in an exceptionally narrow band. This happens in the case of the transition metals like Pd, in which the free atom contains an unsaturated inner shell, so that some conduction electrons are in rather small orbits, for which the overlap integrals of § 4.2 are small. (b) The other exception arises if we are dealing with a band that is very nearly empty, or very nearly filled. In that case the energy is a quadratic function of k near the limiting value, and in that case we have seen that the distribution of states is nearly the same as for free electrons, except for the value of the mass. In that case one easily verifies that

$$\frac{dZ}{dE} = \frac{V}{4\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{\frac{3}{2}}|E-E_1|^{\frac{1}{2}} = \frac{3m^*}{\hbar^2}\left(\frac{V}{6\pi^2}\right)^{\frac{2}{3}}(Z_1)^{\frac{1}{3}}, \qquad (4.58)$$

where m^* is the effective mass, V the volume, E_1 the energy of the minimum or maximum, and Z_1 the number of electrons in the band or the number of empty places, according to which is small.

In case (a) the density of states is much larger than the crude estimate given previously, and in case (b) it is smaller. In both cases the further derivatives of Z with respect to E will be larger, and the higher terms neglected in (4.50) may no longer be negligible, even at fairly low temperatures. We say in that case that the electron gas is only partly degenerate.

4.7. Surface problems

So far we have taken the crystal as infinitely extended for all practical purposes and used the cyclic boundary condition merely to obtain the correct number of states. In some problems we are, however, concerned with phenomena that specifically involve the surface of the metal, and then a more careful approach is needed. The most important new question is then to place the position of the energy bands absolutely, choosing, for example, the most natural zero of energy, that of an electron at rest outside the metal. At first sight one might think that the potential energy function $V(\mathbf{r})$ could be obtained in principle from a knowledge

of the structure of the atomic cores, and that, for example, in the tight binding approximation of § 4.2 a knowledge of the atomic core potential $U(\mathbf{r})$ would be sufficient, so that the determination of the energy bands is then merely a matter of calculation.

The situation is not so simple, however, because, as was pointed out earlier, the atomic cores carry a positive charge and to get realistic results one must allow for the mean negative charge density due to all the conduction electrons. In the inside of the metal this is bound to be distributed uniformly, so that the total charge in a unit cell just cancels that due to the core, and each cell is neutral. But there is no reason to suppose that this exact cancellation is still valid near the surface. To be sure, we expect the crystal as a whole to have no net charge, so that for each of the surface atoms too there will somewhere be a negative charge to compensate the positive charge of the core. We do not know, however, how this will be distributed in space. It may be, for example, that the average charge distribution of the electrons near the surface layer is unsymmetrical and spreads out into space more than it spreads into the inside of the crystal.

We can imagine this state reached by building the crystal first from electrically neutral cells, and then displacing the charge of the surface atoms a little outwards, without moving the cores. The result is an electric double layer, which will lower the potential inside the metal by a constant amount, and will, of course, shift the position of the energy bands by the same amount.

The problem of actually determining this shift is very involved. It could in principle be obtained by the method of self-consistent fields, since the values of the potential near the surface affect the electronic wave functions, and hence the charge distribution of the conduction electrons. The problem is further complicated by the fact that the exact position of the surface atoms is also not known, since the absence of neighbours on one side will disturb the symmetry, so that the ideal lattice points are no longer equilibrium positions. In general, one has to treat the potential shift near the surface as an empirical parameter.

This shift will not only depend on the physical state of the surface, and be rather sensitive to adsorbed layers of other atoms, but it will also be different according to the crystallographic orientation of the surface. For this reason it is not convenient in practice to count the energy from the external zero. It is better to think of the position of the energy bands inside as given in relation to some arbitrary level and then state the value U of the external zero in relation to the same level. Then the position of

the energy bands is not sensitive to the state of the surface, and the surface dependence is contained in U.

One possible choice is to choose the bottom of the conduction band as reference level. Then U is the energy needed to move an electron which is at the bottom of the conduction band to a state of rest outside.

At low temperatures there are electrons in all states in the metal up to energy η. The least energy required to remove an electron from the metal at zero temperature is therefore

$$W = U - \eta, \qquad (4.59)$$

which is called the 'work function'. The quantities on the right-hand side are dependent on the reference level. W is not, but, like U, it depends on the state of the surface.

Now consider two different metals in contact with each other. The nature of the contact is unimportant as long as electrons are able to move from the one to the other. Then in order to have the whole in statistical equilibrium the electron distribution in both metals must be represented by the Fermi function $f(E)$ with the same value of η. Otherwise electrons are transferred and form a double layer near the contact surface until the condition has been achieved.

Except for the immediate neighbourhood of the surface of contact, each metal still has its normal properties, and conditions at the external surfaces will still be normal. The contact has, however, equalized the Fermi energies η. (4.59) therefore shows that the potentials just outside the two metals differ now by

$$U_1 - U_2 = W_1 - W_2. \qquad (4.60)$$

This potential difference is known as contact potential, and according to (4.60) it equals the difference in the work functions. In this sense surfaces of a metal crystal of different crystallographic orientations also have a contact potential relative to each other.

It follows from our discussion of (4.52), which gives the temperature dependence of η, that for two normal metals we would expect the contact potential to be practically independent of temperature.

Next we consider the escape of electrons from the metal at high temperatures. As a preliminary, consider the surface region in equilibrium, neglecting (in this case unrealistically) the effect of the space charge outside the metal which is produced by the electrons. In that case the Fermi distribution must be valid also outside the metal, and, with the same choice of the zero level of energy as previously, the number

of electrons in a state of kinetic energy E_{kin} outside the metal is

$$2f(U+E_{kin}) = \frac{2}{1+e^{(U+E_{kin}-\eta)/kT}}. \tag{4.61}$$

At all reasonable temperatures the exponent of the exponential is large, and hence the 1 in the denominator may be neglected. Using also the definition (4.59) the expression may be written as

$$2e^{-(W+E_{kin})/kT}. \tag{4.62}$$

Using the result of statistical mechanics that the number of states of a free particle is
$$(2\pi\hbar)^{-3}\, d\Omega, \tag{4.63}$$

where $d\Omega$ is the volume element in phase space, we can evaluate the density of electrons outside the metal and their velocity distribution (which is, of course, the Maxwell distribution, since the dependence of (4.62) on kinetic energy is just the ordinary Boltzmann function).

These results (apart from the neglect of space charge) apply to thermal equilibrium, or in other words to the condition in which as many electrons are incident on the metal from some external source as are leaving it. It may also be used to give us the rate of emission from the surface if we assume the same distribution for electrons travelling away from the metal surface, but remove the electrons travelling towards the metal. This is justified if the electrons just outside the metal have made their last collision inside. Since the rate of emission is extremely small, if measured in terms of the density of electrons inside the metal and their velocities, the effect of the collisions inside the metal is still the same as in equilibrium. The model depends therefore on the assumption that the electrons are not likely to suffer any disturbance on leaving the metal. Since there is a rise of potential near the surface over a distance comparable to the electron wavelength, there is, in fact, a finite probability of the electron being scattered back at the surface, and this will reduce the emission by a numerical factor which is not very important in practice.

We require, therefore, the flux of electrons from (4.62) and (4.63), counting only electrons whose velocity is directed away from the surface, and an elementary integration gives the themionic emission current per unit area as

$$\frac{em}{2\pi^2\hbar^3}(kT)^2e^{-W/kT}. \tag{4.64}$$

The exponential factor is well confirmed by experiment, and for clean surfaces the relation of the work function found in this way with the contact potential is in good agreement with observation.

The absolute magnitude of the other factor and its variation as T^2 is much harder to confirm, since it is very sensitive to small errors in the exponent, and to any slight variation, due to secondary causes, of the work function with temperature. In addition, any small inhomogeneity of the surface, due to impurities, geometrical shape, or crystal structure, will complicate the problem. In a microcrystalline surface, in which not all the grains expose crystallographically equivalent faces, (4.64) would be replaced by a combination of such expressions with slightly different work functions.

So far we have neglected space charge. In fact, the escape of a few electrons produces a space charge which raises the potential at some distance from the metal, and thereby prevents the escape of further electrons. For this reason the current from the surface depends on the presence of an external field, and the 'saturation current' (4.64) is obtained only if the field is strong enough to make the space charge effect negligible.

From what was said before about contact potentials it is evident that, in the case of a metal consisting of several layers of different materials, the work function to be used in (4.64) is that of the actual surface material. Hence the practice of covering cathodes with a substance of low work function, which by itself would not have mechanical or electrical properties suitable for use as cathode material.

Even at low temperatures, electron emission is possible if a strong field is applied in a direction to pull electrons away from the surface. In that case, the potential goes down as we go away from the surface, and, at a certain distance x_1 from the surface, the potential energy will have dropped to the value η. This distance is, in fact,

$$x_1 = W/eF \qquad (4.65)$$

if F is the field intensity. An electron inside the metal near the edge of the Fermi distribution has therefore enough energy also to move outside the metal at distances greater than x_1 from the surface, but it does not have enough energy to pass over the potential 'barrier' which is formed by the intervening region in which the potential is higher.

Now it is well known that in quantum mechanics a particle has a finite probability of making a transition through such a barrier. This probability contains as the dominant factor an exponential with the exponent

$$\frac{2}{\hbar} \int \{2m(V-E)\}^{\frac{1}{2}} \, dx,$$

where E is the energy of the particle, V the potential energy, and the

integral is to extend over the region in which $V > E$. In the present problem $V - E$ is equal to the work function at the surface, and from there decreases linearly to zero at x_1. Hence the exponential factor becomes

$$e^{-\lambda W^{\frac{3}{2}}/F}, \qquad \lambda = \frac{4(2m)^{\frac{1}{2}}}{3e\hbar}. \tag{4.66}$$

As in the previous case, there are factors depending on the field intensity multiplying the exponential, but again they are hard to isolate in practice because of minor variations of the exponent over the surface and in this case also because any geometrical irregularity of the surface may lead to a local increase of the electric field.

In this discussion of surface effects we have taken into account only the influence of the surface on the position of the main energy bands. In addition, the surface will cause changes in the electron states. One very obvious consequence of the presence of a surface is that, instead of progressive waves, we must introduce standing waves which do not carry any current in a direction normal to the surface. The phase of such a standing wave depends on the behaviour of the potential near the surface and, since this phase influences the electron density near the surface, will itself influence the potential.

A less obvious feature is the fact that in addition to wave functions which change periodically with a displacement by successive multiples of a lattice vector (4.7), we cannot now rule out wave functions which decrease exponentially as we pass from the surface into the metal. In other words, the component of $\mathbf{k}$ in a direction at right angles to the surface may now be complex. This gives rise to surface waves. Their number is always small compared to that of the periodic solutions, and they are therefore always negligible in statistical arguments concerning the whole metal. For problems in which the surface is of special importance, they may, however, have to be taken into account.