

CHAPTER 1

PERIODIC STRUCTURES

Again! again! again!

THOMAS CAMPBELL

1.1 Translational symmetry

A theory of the physical properties of solids would be practically impossible if the most stable structure for most solids were not a regular crystal lattice. The N -body problem is reduced to manageable proportions by the existence of *translational symmetry*. This means that *there exist basic vectors, \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 such that the atomic structure remains invariant under translation through any vector which is the sum of integral multiples of these vectors.*

In practice, this is only an ideal. A laboratory specimen is necessarily finite in size, so that we must not carry our structure through the boundary. But the only regions where this matters are the layers of atoms near the surface, and in a block of N atoms these constitute only about $N^{\frac{2}{3}}$ atoms—say 1 atom in 10^8 in a macroscopic specimen. Most crystalline solids are also structurally imperfect, with defects, impurities and dislocations to disturb the regularity of arrangement of the atoms. Such imperfections give rise to many interesting physical phenomena, but we shall ignore them, except incidentally, in the present discussion. We are mainly concerned here with the perfect ideal solid, and with the properties it shows; the phenomena which are associated with the solid as the matrix, vehicle, or background for little bits of dirt, or tiny cracks and structural flaws, belong to a different realm of discourse.

We represent the translational group by a *space lattice* or *Bravais net*. Start from some point and then construct all points reached from it by the basic translations. These are the *lattice sites*, defined by the set

$$\mathbf{l} = l_1 \mathbf{a}_1 + l_2 \mathbf{a}_2 + l_3 \mathbf{a}_3, \quad (1.1)$$

where l_1, l_2, l_3 are integers (Fig. 1).

But a solid is a physical structure—not a set of mathematical points. Suppose that there are some atoms, etc., in the neighbourhood of our origin O . The translational invariance insists that there must be exactly similar atoms, placed similarly, about each lattice site (Fig. 2).

It is obvious that we can define the physical arrangement of the whole crystal if we specify the contents of a single *unit cell*—for example, the parallelepiped subtended by the basic vectors \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 . The whole crystal is made up of endless repetitions of this object stacked like bricks in a wall. But the actual definition of a unit cell is to some extent arbitrary. It is obvious enough that any parallelepiped of the right size, shape and orientation would do—as we see in

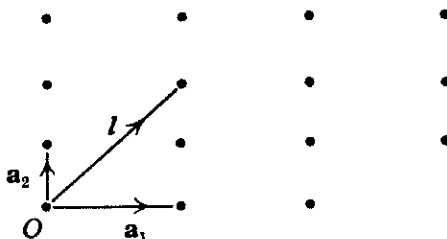


Fig. 1

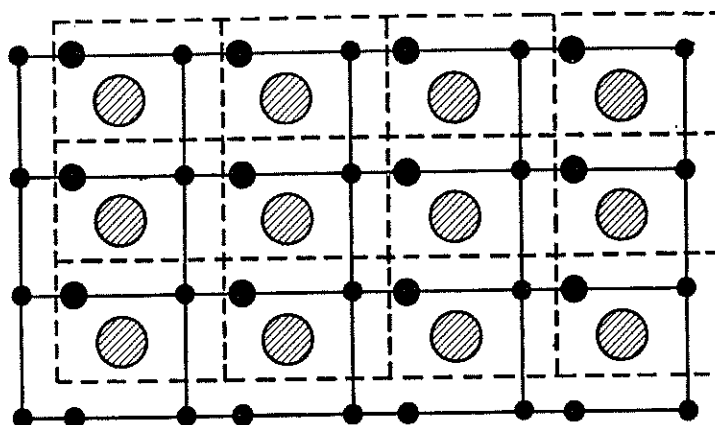


Fig. 2. Alternative unit cells.

Fig. 2. What is, perhaps, not quite so obvious is that the shape can be altered to some extent. Suppose, for example, that there is some central symmetry about some point in the structure (and hence, about all equivalent points). This would be a convenient point to choose as the centre of a cell, itself with central symmetry. One can do this systematically by constructing a *Wigner-Seitz* cell, that is, by drawing the perpendicular bisector planes of the translation vectors from the chosen centre to the nearest equivalent lattice sites. The volume inside all the bisector planes is obviously a unit cell—it is the region whose elements lie nearer to the chosen centre than to any other lattice site.

The unit cell can contain one or more atoms. Naturally, if it contains only one atom, we centre that on the lattice site, and say that we have a *Bravais lattice*. If there are several atoms per unit cell, then we have a *lattice with a basis*. In most of what follows, we shall assume, without special notice, that the structure is a Bravais lattice. This is for simplicity; in reality only a few elementary solids, such as the alkali metals, have this structure.

The science of *crystallography* is concerned with the enumeration and classification of all possible types of crystal structure, and the determination of the actual structure of actual crystalline solids.

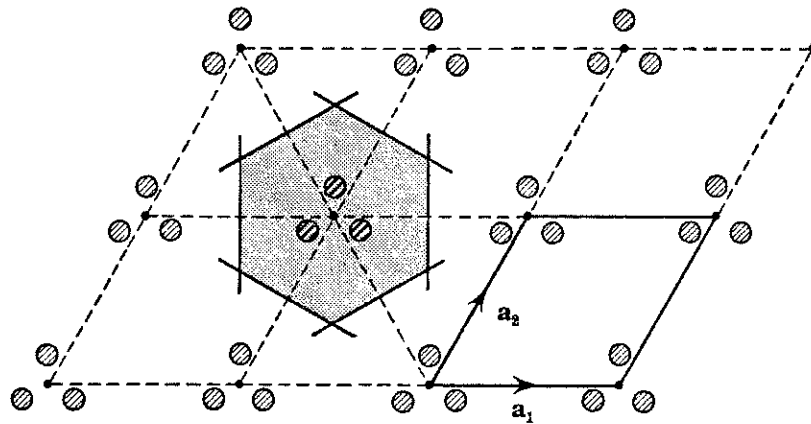


Fig. 3. Wigner-Seitz cell.

Structures are classified according to their symmetry properties, such as invariance under rotation about an axis, reflection in a plane, etc. These symmetries are often of great importance in the simplification of theoretical computations, and can be used with great power in the discussion of the numbers of parameters that are necessary to define the macroscopic properties of solids. However, to take full advantage of this theory, one needs the mathematics of *group theory*, which would take us too far away from our main topic. If we restrict ourselves mainly to very simple solids, then most of the symmetry properties can be discovered by inspection without formal algebraic analysis. In any case, there are many excellent books on crystallography and on group theory and its applications to the theory of solids.

In these books, the various types of Bravais lattice, etc., are set out in detail. We shall consider here only one case, which exemplifies many of the principles of the subject, and which is also of great importance as a structure which is actually assumed by some elements. This is the *body-centred cubic* (B.C.C.) structure illustrated in Fig. 4.

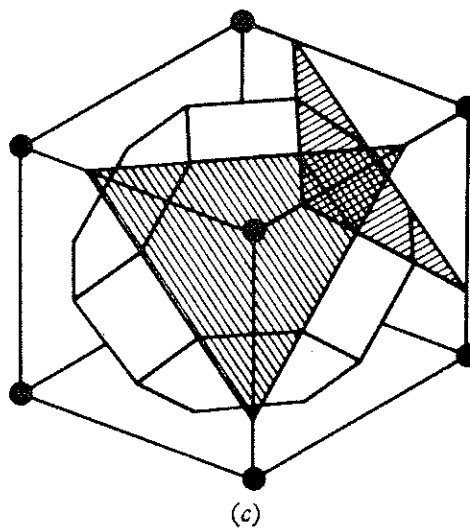
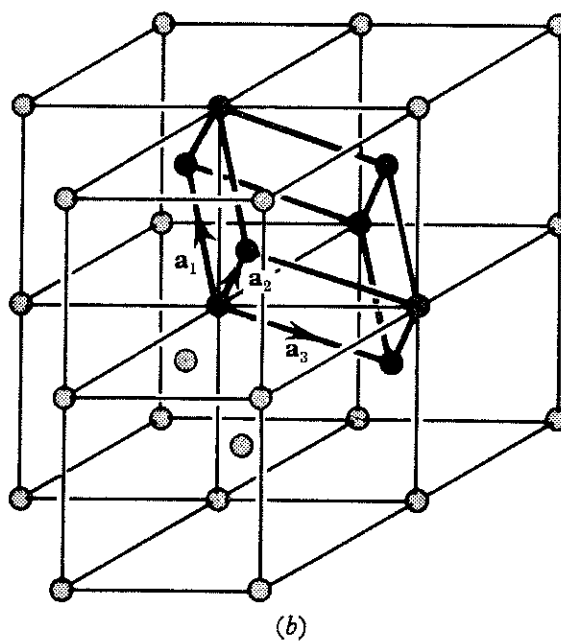
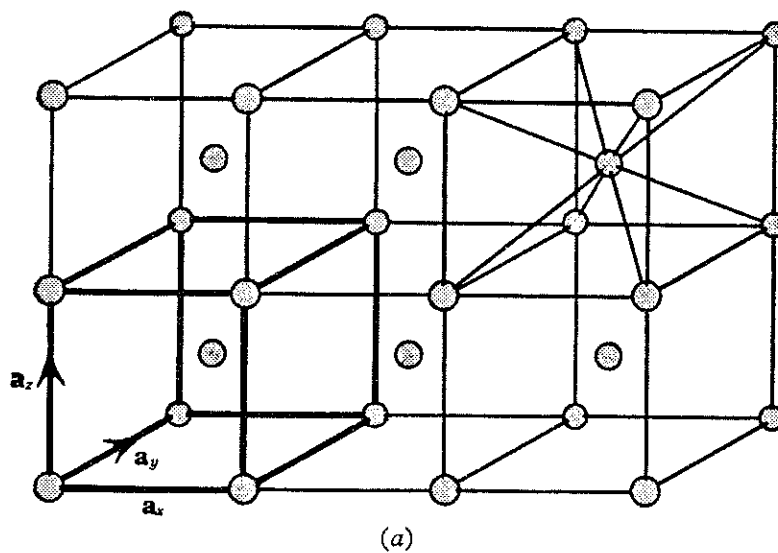


Fig. 4. Body-centred cubic lattice. (a) Cubic unit cell. (b) Generators of the Bravais lattice. (c) Wigner-Seitz cell.

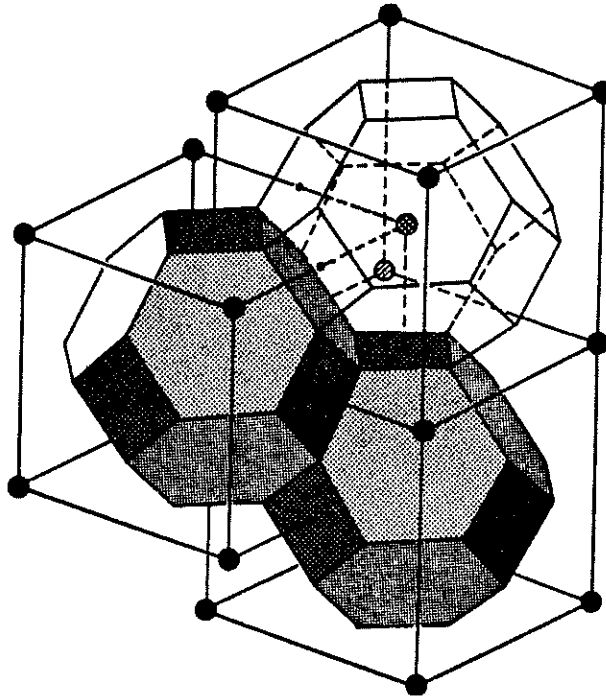


Fig. 5. Stacking Wigner-Seitz cells of the B.C.C. lattice.

At first sight, this is a cubic lattice with two atoms per unit cell, or two interpenetrating simple cubic sublattices defined by

$$\left. \begin{aligned} l &= l_x \mathbf{a}_x + l_y \mathbf{a}_y + l_z \mathbf{a}_z, \\ l' &= (l_x + \frac{1}{2}) \mathbf{a}_x + (l_y + \frac{1}{2}) \mathbf{a}_y + (l_z + \frac{1}{2}) \mathbf{a}_z, \end{aligned} \right\} \quad (1.2)$$

where l_x, l_y, l_z are all integers. But if we write

$$\left. \begin{aligned} \mathbf{a}_1 &= \frac{1}{2}(-\mathbf{a}_x + \mathbf{a}_y + \mathbf{a}_z), \\ \mathbf{a}_2 &= \frac{1}{2}(\mathbf{a}_x - \mathbf{a}_y + \mathbf{a}_z), \\ \mathbf{a}_3 &= \frac{1}{2}(\mathbf{a}_x + \mathbf{a}_y - \mathbf{a}_z), \end{aligned} \right\} \quad (1.3)$$

we can generate all the points of both sublattices by

$$l = l_1 \mathbf{a}_1 + l_2 \mathbf{a}_2 + l_3 \mathbf{a}_3 \quad (1.4)$$

with l_1, l_2, l_3 integers. We shall find ourselves at a cube centre, or corner, according as $(l_1 + l_2 + l_3)$ is odd, or even.

Thus (Fig. 4(b)) this is really a Bravais lattice. Instead of using a cubic unit cell we may use the Wigner-Seitz cell (Fig. 4(c)), which is constructed by chopping off all the corners of a cube half way along a diagonal from the centre to a corner point. This figure obviously has the same symmetry as a cube—for example, the original vectors, $\mathbf{a}_x, \mathbf{a}_y, \mathbf{a}_z$ are axes of four-fold symmetry. It also shows, more clearly

perhaps than the original cube, that the vectors \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 , i.e. the diagonals of the cube, are axes of three-fold rotational symmetry, for they pass through the hexagonal faces of the cell. It is a good exercise to visualize how these polyhedra can be packed to make the original lattice (Fig. 5).

1.2 Periodic functions

To define a physical model of a crystal structure we need to give values to some function $f(\mathbf{r})$ in space—local electron density, electrostatic potential, etc., which can be recognized as an arrangement of atoms (e.g. Fig. 6). Our assertion of translational symmetry means that this must be a *multiply periodic function*

$$f(\mathbf{r} + \mathbf{l}) = f(\mathbf{r}) \quad (1.5)$$

for all points \mathbf{r} in space, and for all lattice translations.

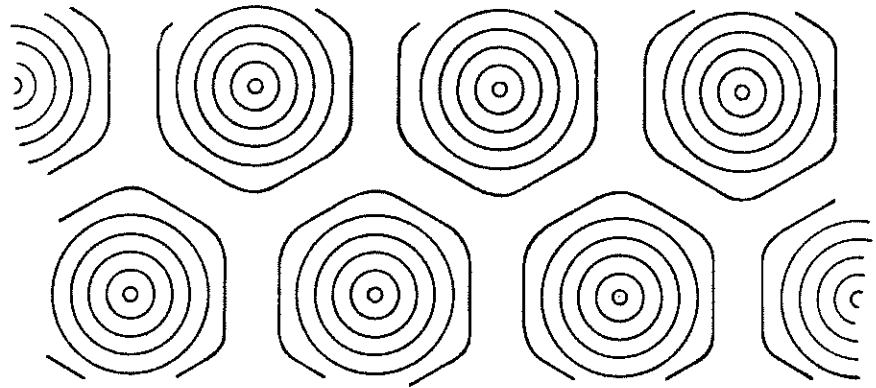


Fig. 6. Multiply periodic function.

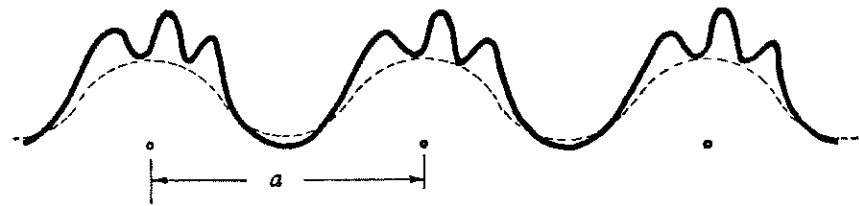


Fig. 7

In one dimension we are perfectly familiar with periodic functions. Thus, as in Fig. 7, we may have

$$f(x + l) = f(x), \quad (1.6)$$

where l is of the form $l_1 a$, with l_1 an integer and a the period of the function. We also know that any such function can be expressed as a *Fourier series*,

$$f(x) = \sum_n A_n e^{2\pi i n x / a}, \quad (1.7)$$

where n is an integer. Let us write this in the form

$$f(x) = \sum_g A_g e^{igx}, \quad (1.8)$$

where the quantities g belong to the set of *reciprocal lattice lengths*,†

$$g_n = n \frac{2\pi}{a}. \quad (1.9)$$

The coefficients in (1.8) are well known to be defined by

$$A_g = \frac{1}{a} \int_{\text{cell}} f(x) e^{-igx} dx, \quad (1.10)$$

where in this case the range of integration is, say $0 < x \leq a$ —it need only be one cell of the lattice. The elementary proof that (1.8) implies (1.6) can be derived from the condition that, for any value of g ,

$$e^{igl} = 1 \quad (1.11)$$

for all translations l . This means no more than that

$$gl = n \frac{2\pi}{a} l_1 a = n l_1 2\pi = \text{integer} \times 2\pi \quad (1.12)$$

when g takes one of its allowed values g_n and $l = l_1 a$. The derivation of (1.10) also follows from this same relation. We are not concerned here with mathematically pathological functions, and may use naïve Fourier theory quite freely.

Extension of this theorem to a structure with three rectangular axes is easy enough. If these axes are $\mathbf{a}_x, \mathbf{a}_y, \mathbf{a}_z$, then, if

$$f(\mathbf{r} + \mathbf{l}) \equiv f(\mathbf{r} + l_x \mathbf{a}_x + l_y \mathbf{a}_y + l_z \mathbf{a}_z) = f(\mathbf{r}), \quad (1.13)$$

we have

$$f(\mathbf{r}) = \sum_{g_x, g_y, g_z} A(g_x, g_y, g_z) \exp\{i(g_x x + g_y y + g_z z)\}, \quad (1.14)$$

where each of g_x , etc., is a reciprocal length out of the set $2\pi n/a_x$, etc. One can prove this by first analysing $f(\mathbf{r})$ as a Fourier series in the x direction, then showing that each coefficient in this series is a periodic function of y , and can be analysed into a further series—and similarly for z .

Let us rewrite (1.14) in the form

$$f(\mathbf{r}) = \sum_{\mathbf{g}} A_{\mathbf{g}} e^{i\mathbf{g} \cdot \mathbf{r}}, \quad (1.15)$$

† It is convenient to include the factor 2π in this definition, although in crystallography the convention is to leave 2π explicitly in the exponent.

where \mathbf{g} is the vector with components (g_x, g_y, g_z) . This vector has the property expressed in (1.11), i.e. for any such vector

$$\begin{aligned}\mathbf{g} \cdot \mathbf{l} &= (g_x l_x a_x + g_y l_y a_y + g_z l_z a_z) \\ &= \frac{2\pi n_x}{a_x} l_x a_x + \frac{2\pi n_y}{a_y} l_y a_y + \frac{2\pi n_z}{a_z} l_z a_z \\ &= 2\pi \times \text{integer},\end{aligned}\tag{1.16}$$

whatever the value of \mathbf{l} . Thus,

$$e^{i\mathbf{g} \cdot \mathbf{l}} = 1\tag{1.17}$$

for all lattice vectors \mathbf{l} , and for all *reciprocal lattice vectors* \mathbf{g} .

It is obvious that this condition is sufficient to make the series (1.15) represent a function like (1.5) with the periodicity of the lattice:

$$\begin{aligned}f(\mathbf{r} + \mathbf{l}) &= \sum_{\mathbf{g}} A_{\mathbf{g}} e^{i\mathbf{g} \cdot (\mathbf{r} + \mathbf{l})} = \sum_{\mathbf{g}} A_{\mathbf{g}} e^{i\mathbf{g} \cdot \mathbf{r}} e^{i\mathbf{g} \cdot \mathbf{l}} \\ &= \sum_{\mathbf{g}} A_{\mathbf{g}} e^{i\mathbf{g} \cdot \mathbf{r}} = f(\mathbf{r}).\end{aligned}\tag{1.18}$$

One can readily devise a proof that it is a necessary condition, i.e. that the sum (1.15) may only contain terms corresponding to values of \mathbf{g} that satisfy the condition (1.17).

It only remains to construct the reciprocal lattice vectors for a non-rectangular lattice. This is easily done as follows: take the *reciprocal triad* of the basic vectors, i.e.

$$\mathbf{b}_1 = \frac{\mathbf{a}_2 \wedge \mathbf{a}_3}{\mathbf{a}_1 \cdot \mathbf{a}_2 \wedge \mathbf{a}_3}, \quad \mathbf{b}_2 = \frac{\mathbf{a}_3 \wedge \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_2 \wedge \mathbf{a}_3}, \quad \mathbf{b}_3 = \frac{\mathbf{a}_1 \wedge \mathbf{a}_2}{\mathbf{a}_1 \cdot \mathbf{a}_2 \wedge \mathbf{a}_3},\tag{1.19}$$

and write

$$\begin{aligned}\mathbf{g} &= g_1 \mathbf{b}_1 + g_2 \mathbf{b}_2 + g_3 \mathbf{b}_3 \\ &= 2\pi n_1 \mathbf{b}_1 + 2\pi n_2 \mathbf{b}_2 + 2\pi n_3 \mathbf{b}_3,\end{aligned}\tag{1.20}$$

where n_1 , etc., are integers.

By elementary vector analysis, we verify that $\mathbf{b}_1 \cdot \mathbf{a}_1 = 1$, etc., and $\mathbf{b}_1 \cdot \mathbf{a}_2 = 0$, etc., so that

$$\begin{aligned}\mathbf{g} \cdot \mathbf{l} &= (g_1 \mathbf{b}_1 + g_2 \mathbf{b}_2 + g_3 \mathbf{b}_3) \cdot (l_1 \mathbf{a}_1 + l_2 \mathbf{a}_2 + l_3 \mathbf{a}_3) \\ &= g_1 l_1 + g_2 l_2 + g_3 l_3 \\ &= 2\pi \times \text{integer}.\end{aligned}\tag{1.21}$$

Thus, every vector of the set (1.20) satisfies the condition (1.17).

We can also generalize the formula (1.10) for each coefficient in the series (1.15). It must be a volume integral, which we write

$$A_{\mathbf{g}} = \frac{1}{v_{\text{cell}}} \int_{\text{cell}} f(\mathbf{r}) e^{-i\mathbf{g} \cdot \mathbf{r}} d\mathbf{r}.\tag{1.22}$$

The proof follows by multiplying the series through by $\exp(-i\mathbf{g}\cdot\mathbf{r})$, and integrating. An integral of $\exp(i\mathbf{g}\cdot\mathbf{r})$ through a unit cell is obviously zero, if \mathbf{g} is of the form (1.20), unless $\mathbf{g} = 0$.

1.3 Properties of the reciprocal lattice

The vectors defined by (1.20), i.e.

$$\mathbf{g} = n_1 \cdot 2\pi\mathbf{b}_1 + n_2 \cdot 2\pi\mathbf{b}_2 + n_3 \cdot 2\pi\mathbf{b}_3, \quad (1.23)$$

generate a lattice with basic cell spanned by the vectors $2\pi\mathbf{b}_1$, $2\pi\mathbf{b}_2$, $2\pi\mathbf{b}_3$. This is the *reciprocal lattice* of our original direct lattice. It is an invariant geometrical object, whose properties are fundamental in the theory of solids. Some of the elementary geometrical properties are easily deduced.

(i) *Each vector of the reciprocal lattice is normal to a set of lattice planes of the direct lattice.*

Choose a particular reciprocal lattice vector \mathbf{g} , and a lattice vector \mathbf{l} and consider the relation (1.21)

$$\begin{aligned} \mathbf{g}\cdot\mathbf{l} &= 2\pi(n_1 l_1 + n_2 l_2 + n_3 l_3) \\ &= 2\pi N, \end{aligned} \quad (1.24)$$

where N is an integer. This tells us that the projection of the vector \mathbf{l} on the direction of \mathbf{g} has the length

$$d = \frac{2\pi N}{|\mathbf{g}|}. \quad (1.25)$$

But there are infinitely many points of the lattice with this property. For example, suppose \mathbf{l}' is a lattice point represented by the integers

$$l'_1 = l_1 - mn_3, \quad l'_2 = l_2 - mn_3, \quad l'_3 = l_3 + m(n_1 + n_2), \quad (1.26)$$

where m is an integer. It is obvious that

$$\mathbf{g}\cdot\mathbf{l}' = \mathbf{g}\cdot\mathbf{l} = 2\pi N. \quad (1.27)$$

Thus \mathbf{l}' has the same projection on \mathbf{g} and must therefore also be on the plane normal to \mathbf{g} , at distance d from the origin. Thus, if there is one lattice point on this plane, there is an infinity of such points; we have constructed one of the *lattice planes*. This relationship between the direct and reciprocal lattices is, of course, a special case of the familiar duality between points and planes in three-dimensional geometry.

(ii) If the components of \mathbf{g} have no common factor, then $|\mathbf{g}|$ is inversely proportional to the spacing of the lattice planes normal to \mathbf{g} .

This follows from (1.25) (see Fig. 8). If (n_1, n_2, n_3) have no common factor, then we can always find a lattice vector \mathbf{l}'' with components such that

$$\mathbf{g} \cdot \mathbf{l}'' = 2\pi(N + 1). \quad (1.28)$$

This means that the lattice plane containing \mathbf{l}'' is at a distance

$$d'' = \frac{2\pi(N + 1)}{|\mathbf{g}|} \quad (1.29)$$

from the origin—i.e. is spaced $2\pi/|\mathbf{g}|$ from the plane containing \mathbf{l} .

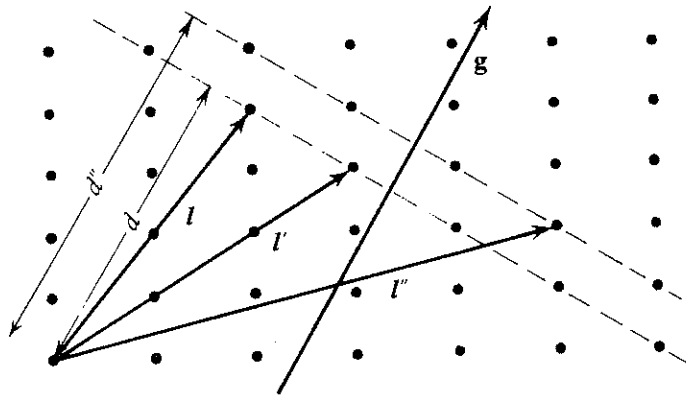


Fig. 8

From these two elementary geometrical results we see that the simplest way of characterizing the planes of a lattice is by their normals, expressed as vectors of the reciprocal lattice. The most prominent planes in the direct lattice are those which are most densely populated with lattice sites. Since the density of lattice sites is constant in space the most prominent planes must be those which are most widely separated, i.e. those with the smallest reciprocal lattice vectors.

The labelling of lattice planes by their corresponding reciprocal lattice vectors is equivalent to the *Miller indices* used in classical crystallography. Suppose that we have a lattice plane, with normal \mathbf{g} , such that (1.24) is satisfied for all points \mathbf{l} upon it. Then if we take a point such that $l_2 = l_3 = 0$, we have $l_1 = N/n_1$, so that the intercept of this plane along the \mathbf{a}_1 axis has length

$$d_1 = \left(\frac{N}{n_1}\right) a_1. \quad (1.30)$$

Similarly this plane will cut the \mathbf{a}_2 axis at a distance

$$d_2 = \left(\frac{N}{n_2}\right) a_2 \quad (1.31)$$

from the origin. The intercepts of this plane along the axes, measured in units of the lengths of the corresponding basic vectors, are inversely proportional to the integers n_1, n_2, n_3 . These integers, after removal of any common factors, are the Miller indices of the plane, expressed in the form (n_1, n_2, n_3) .

It is obvious that densely populated planes—i.e. ones with small Miller indices—are those which are most likely to show up in natural crystals, either in growth or after cleavage. The study of the geometry of such faces was of the essence of classical crystallography, and the

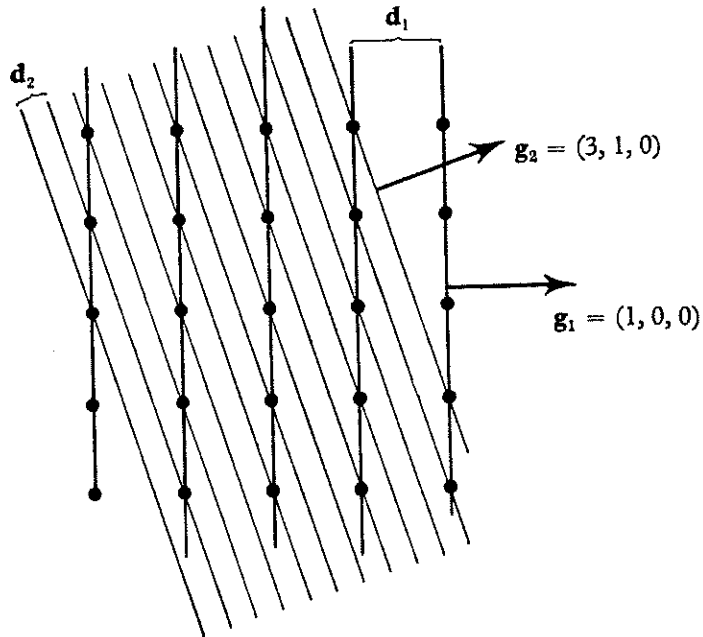


Fig. 9. The spacing of planes with large Miller indices is closer than the spacing of the principal symmetry planes.

key discovery was that a single set of vectors a_1, a_2, a_3 could be found such that all the observed faces of a macroscopic crystal could be represented by small Miller indices.

Two conventions of notation should be mentioned. The symbol $(\bar{1}\bar{1}\bar{1})$ represents the planes $(1, -1, -1)$; the minus sign is put above the number for conciseness. The symbol $\{110\}$, with curly brackets, represents all the different sets of planes that are equivalent by symmetry to the set (110) . Thus, for a cubic structure it might include (101) , (011) , $(\bar{1}\bar{1}0)$, $(1\bar{1}0)$, etc.

(iii) *The volume of a unit cell of the reciprocal lattice is inversely proportional to the volume of a unit cell of the direct lattice.*

This follows by elementary vector analysis. The reciprocal lattice cell is spanned by $2\pi\mathbf{b}_1, 2\pi\mathbf{b}_2, 2\pi\mathbf{b}_3$. Using (1.19), its volume would be

$$\begin{aligned}
 (2\pi)^3 (\mathbf{b}_1 \cdot \mathbf{b}_2 \wedge \mathbf{b}_3) &= (2\pi)^3 \frac{(\mathbf{a}_2 \wedge \mathbf{a}_3) \cdot \{(\mathbf{a}_3 \wedge \mathbf{a}_1) \wedge (\mathbf{a}_1 \wedge \mathbf{a}_2)\}}{(\mathbf{a}_1 \cdot \mathbf{a}_2 \wedge \mathbf{a}_3)^3} \\
 &= (2\pi)^3 \frac{(\mathbf{a}_2 \wedge \mathbf{a}_3) \cdot [\{\mathbf{a}_3 \cdot (\mathbf{a}_1 \wedge \mathbf{a}_2)\} \mathbf{a}_1 - \{\mathbf{a}_1 \cdot (\mathbf{a}_1 \wedge \mathbf{a}_2)\} \mathbf{a}_3]}{(\mathbf{a}_1 \cdot \mathbf{a}_2 \wedge \mathbf{a}_3)^3} \\
 &= \frac{(2\pi)^3}{(\mathbf{a}_1 \cdot \mathbf{a}_2 \wedge \mathbf{a}_3)} \\
 &= \frac{8\pi^3}{v_c}, \tag{1.32}
 \end{aligned}$$

where v_c is the volume of the unit cell spanned by $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$.

The factor $8\pi^3$ comes in here because of the way we have defined the reciprocal lattice. A more common convention is to write

$$e^{i\mathbf{g} \cdot \mathbf{l}} \equiv \exp(2\pi i \mathbf{K}_g \cdot \mathbf{R}_l), \tag{1.33}$$

where \mathbf{R}_l is a lattice vector, and the vectors \mathbf{K}_g are defined as vectors of a reciprocal lattice with unit cell volume $1/v_c$. The disadvantage of this notation is that the factor 2π keeps turning up in the exponent; it is also incommensurable with the conventional symbolism

$$\psi = e^{i\mathbf{k} \cdot \mathbf{r}} \tag{1.34}$$

for a free-electron wave-function.

(iv) *The direct lattice is the reciprocal of its own reciprocal lattice.*

This is implicit in the name, and can be verified by constructing, say, the vector $(\mathbf{b}_1 \wedge \mathbf{b}_2)/(\mathbf{b}_1 \cdot \mathbf{b}_2 \wedge \mathbf{b}_3)$, and showing that it is identical with \mathbf{a}_3 . One can see this by inspection of the derivation of (1.32).

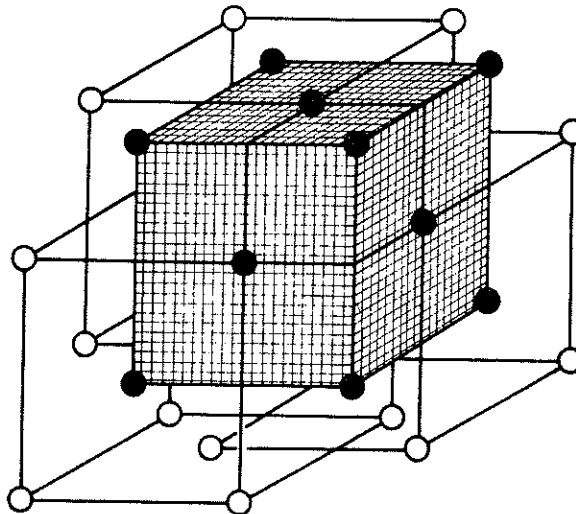
(v) *The unit cell of the reciprocal lattice need not be a parallelepiped.* In fact, we almost always deal with the Wigner-Seitz cell of the reciprocal lattice. This is called a *Brillouin zone*.

As an example of the properties of the reciprocal lattice, consider an important and commonly observed structure—the *face-centred cubic* lattice (F.C.C. structure). This is built up out of four interpenetrating simple cubic lattices, arranged so that if we look at any one of them, we see an atom at the centre of each face of its unit cell, as well as on the sites at the corners of the cube (Fig. 10(a)).

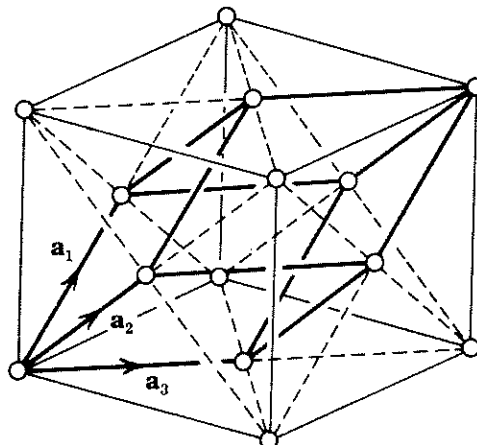
This looks at first, like a lattice with a basis, with four atoms in the cubic unit cell generated by $(\mathbf{a}_x, \mathbf{a}_y, \mathbf{a}_z)$.

But if we choose the half diagonals of the faces of the cube,

$$\left. \begin{aligned} \mathbf{a}_1 &= (0, \frac{1}{2}a, \frac{1}{2}a), & \mathbf{a}_2 &= (\frac{1}{2}a, 0, \frac{1}{2}a), \\ \mathbf{a}_3 &= (\frac{1}{2}a, \frac{1}{2}a, 0), \end{aligned} \right\} \quad (1.35)$$



(a)



(b)

Fig. 10. Face-centred cubic lattice. (a) As four interpenetrating sublattices. (b) As Bravais lattice.

then we can reach any site by a combination of multiples of these (Fig. 10 (b)). Thus, the f.c.c. lattice is truly a Bravais lattice with $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ as generators.

We could now construct the reciprocal lattice by algebra. But it is easy to use the geometrical properties noted above. It is obvious, for example, that there are important lattice planes normal to the edges

of the cube. Relative to ordinary Cartesian axes along $\mathbf{a}_x, \mathbf{a}_y, \mathbf{a}_z$, these planes must have Miller indices (100) (010) (001), etc., i.e. they are the set $\{100\}$. We can see, also, that they are spaced a distance $\frac{1}{2}a$ apart. Thus, they correspond to reciprocal lattice vectors of length

$$|\mathbf{g}| = 2\pi/\frac{1}{2}a$$

along rectangular Cartesian axes in reciprocal space. Our reciprocal lattice must include the whole simple cubic lattice generated by these vectors.

There is also an important lattice plane normal to the diagonal of the cubic unit cell. This plane makes equal intercepts on all three axes, and must therefore have indices in the set $\{111\}$. These planes are

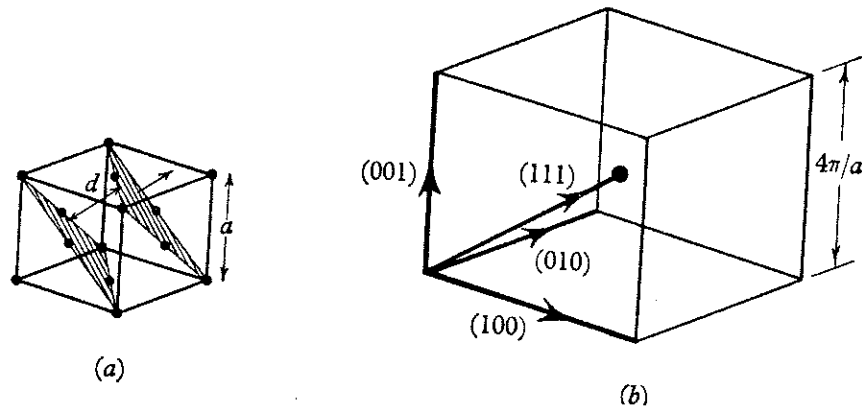


Fig. 11. (a) The $\{111\}$ planes of the f.c.c. lattice. (b) The corresponding reciprocal lattice cell.

spaced apart by a distance equal to one-third of the whole diagonal of the cube, i.e. a distance $a/\sqrt{3}$. Thus, the corresponding reciprocal lattice vector has three equal components and total length $2\pi\sqrt{3}/a$; it must be

$$\mathbf{g}' = \frac{2\pi}{a}(1, 1, 1). \quad (1.36)$$

In reciprocal space, it is obviously directed along the diagonal of the cubic cell—but is only half its length. Thus, it generates the body centres of the simple cubic lattice.

One can now study all other lattice planes—there are, for example, the $\{110\}$ sets, which are normal to the diagonals of the faces of the basic cube—but the reciprocal lattice vectors for these are all contained already in the points generated by the $\{100\}$ and $\{111\}$ planes. Thus, the reciprocal of the f.c.c. lattice is b.c.c. (and, of course, *vice versa*).

The Brillouin zone of the F.C.C. lattice is the Wigner-Seitz cell of this reciprocal lattice. We have already seen this figure (Fig. 4(c))—a polyhedron with square and hexagonal faces. The centres of the square faces correspond to the directions of the normals of the cube planes of the direct lattice. The centres of the hexagonal faces are the normals of the diagonal planes.

It is interesting to note, further, that the F.C.C. structure is 'close-packed'. It can be built up by superposing successive (111) planes, each of which is a hexagonal network, as of rigid spheres. This is one of the reasons why it is often adopted by metals, whose cohesion is not associated with strongly directional bonds but is largely a volume effect.

1.4 Bloch's theorem

We have learnt how to represent functions with the periodicity of the lattice. But these are not enough for a physical theory; we need to consider various excitations of the structure, which will destroy the exact translational symmetry. Of these there are several types, of which the most important are *lattice waves*, i.e. vibrations of the atoms about their equilibrium positions, and *electron states*, i.e. motion of electrons in the field of the static lattice. In Chapter 10 we shall also consider *spin waves*, which are excitations of the spins localized on the atoms of the crystal.

All these excitations are characterized by a dynamical equation, or a Schrödinger equation, or a spin-exchange Hamiltonian, which is *invariant under lattice translation*. Thus, if $\mathcal{V}(\mathbf{r})$ is the potential seen by an electron at \mathbf{r} , then $\mathcal{V}(\mathbf{r} + \mathbf{l}) = \mathcal{V}(\mathbf{r})$ for all \mathbf{l} . The electron wavefunction $\psi(\mathbf{r})$ must satisfy a Schrödinger equation

$$\left(-\frac{\hbar^2}{2m} \nabla^2 + \mathcal{V}(\mathbf{r}) - \mathcal{E} \right) \psi = 0, \quad (1.37)$$

which remains the same after we have substituted $\mathbf{r} + \mathbf{l}$ for \mathbf{r} in the operator that acts on ψ .

In the case of lattice waves and spin waves, the formalism becomes more complicated but the principle is the same. Suppose that

$$(u_1, u_2, \dots, u_{\mathbf{n}}, \dots)$$

are the displacements (or spin displacements, or spin deviation operators, or something) of the atoms at sites $1, 2, \dots, \mathbf{n}, \dots$. The equations of motion (or the Hamiltonian) will contain these variables

(or operators) in such a way that if we make a lattice translation \mathbf{l} —that is, if we put $u_{\mathbf{n}+\mathbf{l}}$ in the formulae wherever $u_{\mathbf{n}}$ occurred previously—we shall not be able to tell the difference. In other words, all cells of the lattice are equivalent, and indistinguishable. It is rather like the ‘cosmological principle’ which says that the universe looks essentially the same from whatever point we view it.

To formalize this property, let us use the following notation. Let $\mathcal{H}(0)$ and $|0\rangle$ represent the Hamiltonian operator and wave-function before a lattice translation; let $\mathcal{H}(\mathbf{l})$ and $|\mathbf{l}\rangle$ represent the same mathematical objects after the translation \mathbf{l} . Thus, for electrons $|0\rangle \equiv \psi(\mathbf{r})$, and $|\mathbf{l}\rangle \equiv \psi(\mathbf{r} + \mathbf{l})$. For lattice waves, if $|0\rangle$ represents the state in which the atom at \mathbf{n} is going through some particular motion, then $|\mathbf{l}\rangle$ represents the state in which the atom at $(\mathbf{n} + \mathbf{l})$ is performing that same motion.

The statement of translational invariance is then

$$\mathcal{H}(\mathbf{l}) = \mathcal{H}(0). \quad (1.38)$$

Now the eigenstates must satisfy equations of the form †

$$\mathcal{H}(0) |0\rangle = \mathcal{E} |0\rangle. \quad (1.39)$$

This equation is identical with

$$\mathcal{H}(\mathbf{l}) |\mathbf{l}\rangle = \mathcal{E} |\mathbf{l}\rangle, \quad (1.40)$$

which comes merely by a formal relabelling of all the variables in both operators and state functions. But, by (1.38), this implies

$$\mathcal{H}(0) |\mathbf{l}\rangle = \mathcal{E} |\mathbf{l}\rangle, \quad (1.41)$$

i.e. the state function $|\mathbf{l}\rangle$ is also a solution of the equation satisfied by $|0\rangle$. Since $|\mathbf{l}\rangle$ is not necessarily identical with $|0\rangle$, it looks as if we can generate an immense number of solutions of the equations of motion just by translations of any one solution that we have happened to find. All these solutions are, of course, degenerate in energy.

This is obviously ridiculous; the new solutions must all be equivalent, in some way, to our original solution. There are two cases to consider:

(i) Suppose $|0\rangle$ is really *non-degenerate* (which is never actually the case, because all lattices have reflection symmetries, as well as translational invariance). Then the only possibility is that each new state

† In the case of the classical equations of motion of lattice waves, the square of the frequency ω plays the role of the energy, and the ‘Hamiltonian’ is the matrix of the linked linear differential equations derived upon the assumption that we are dealing with small deviations from equilibrium. See §2.1.

$|l\rangle$ is a multiple of $|0\rangle$, and hence physically indistinguishable from it. There must be some number λ_1 such that

$$|\mathbf{a}_1\rangle = \lambda_1 |0\rangle \quad (1.42)$$

for the result of a single step in the direction \mathbf{a}_1 . Normalization requires that

$$|\lambda_1|^2 = 1, \quad (1.43)$$

so that we might have $\lambda_1 = e^{ik_1}$, (1.44)

where k_1 is a real number. Similarly, we might have

$$|\mathbf{a}_2\rangle = e^{ik_2} |0\rangle; \quad |\mathbf{a}_3\rangle = e^{ik_3} |0\rangle, \quad (1.45)$$

for unit translations in the other basic directions. Then a general translation must lead to the following relation

$$\begin{aligned} |l\rangle &\equiv |l_1 \mathbf{a}_1 + l_2 \mathbf{a}_2 + l_3 \mathbf{a}_3\rangle \\ &= e^{ik_1} |(l_1 - 1) \mathbf{a}_1 + l_2 \mathbf{a}_2 + l_3 \mathbf{a}_3\rangle \\ &= e^{ik_1 l_1} |0 + l_2 \mathbf{a}_2 + l_3 \mathbf{a}_3\rangle \\ &= e^{i(k_1 l_1 + k_2 l_2 + k_3 l_3)} |0\rangle \end{aligned} \quad (1.46)$$

as we make l_1 steps in direction \mathbf{a}_1 , l_2 steps in direction \mathbf{a}_2 , etc., and multiply by the appropriate factor for each step.

Let us define a vector \mathbf{k} , in the form

$$\mathbf{k} = k_1 \mathbf{b}_1 + k_2 \mathbf{b}_2 + k_3 \mathbf{b}_3, \quad (1.47)$$

where $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ are the reciprocal triad of $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ as in (1.19). The relation (1.46) becomes

$$|l\rangle = e^{i\mathbf{k} \cdot \mathbf{l}} |0\rangle, \quad (1.48)$$

which is the result we are seeking to prove:

For any wave-function/state function that satisfies the Schrödinger equation (or its classical or quantal equivalent) there exists a vector \mathbf{k} such that translation by a lattice vector \mathbf{l} is equivalent to multiplying by the phase factor $\exp(i\mathbf{k} \cdot \mathbf{l})$.

Each different state function may have a different *wave-vector* \mathbf{k} , but each one must satisfy this condition. It is a strong condition imposed on the form of an elementary excitation by the strong initial circumstance that the lattice is translationally invariant.

But we need the proof in a more general case.

(ii) Suppose $|0\rangle$ is degenerate. For simplicity, suppose that it is doubly degenerate, so that there are certainly two distinct states $|0\rangle_1$

and $|0\rangle_2$ of the same energy. Then the lattice translation \mathbf{a}_1 can produce, at most, linear combinations of these same two states. Thus

$$\left. \begin{aligned} |\mathbf{a}_1\rangle_1 &= T_1^{11} |0\rangle_1 + T_1^{12} |0\rangle_2, \\ |\mathbf{a}_1\rangle_2 &= T_1^{21} |0\rangle_1 + T_1^{22} |0\rangle_2. \end{aligned} \right\} \quad (1.49)$$

The numbers T_1^{11} , etc., are written like this because they are the elements of a matrix \mathbf{T}_1 , which must be unitary to preserve normalization. For conciseness, let us write this pair of equations in the form

$$(|\mathbf{a}_1\rangle) = \mathbf{T}_1(|0\rangle). \quad (1.50)$$

But $|0\rangle_1$ and $|0\rangle_2$, being degenerate, are not uniquely defined. We could have started with any two other states which were linear combinations of these. For example, we might have started with

$$\left. \begin{aligned} |0\rangle_1 &= S^{11} |0\rangle_1 + S^{12} |0\rangle_2, \\ |0\rangle_2 &= S^{21} |0\rangle_1 + S^{22} |0\rangle_2, \end{aligned} \right\} \quad (1.51)$$

where the numbers S^{11} , etc., are elements of another (arbitrary) unitary matrix \mathbf{S} .

Now let us choose \mathbf{S} so that the matrix $\mathbf{S}\mathbf{T}_1\mathbf{S}^{-1}$ is a diagonal matrix. By standard algebraic theory, this is always possible. Suppose, for example, that it reduces to the form

$$\mathbf{S}\mathbf{T}_1\mathbf{S}^{-1} = \begin{pmatrix} e^{ik_1} & 0 \\ 0 & e^{ik'_1} \end{pmatrix}. \quad (1.52)$$

It can easily be verified that the states $|0\rangle_1$ and $|0\rangle_2$ will now transform under translation exactly as if they were non-degenerate, i.e. as in (1.42–1.45),

$$|\mathbf{a}_1\rangle_1 = e^{ik_1} |0\rangle_1, \quad |\mathbf{a}_1\rangle_2 = e^{ik'_1} |0\rangle_2. \quad (1.53)$$

The state $|0\rangle_1$ has a wave-vector with component k_1 for the direction \mathbf{a}_1 ; the state $|0\rangle_2$ has wave-vector with component k'_1 (in general, different) for the same translation direction.

But now what about translations in other directions. Consider \mathbf{a}_2 . For the initial states $|0\rangle_1$ and $|0\rangle_2$ there will be another matrix \mathbf{T}_2 defining the result of this translation, i.e. symbolically as in (1.49) and (1.50)

$$(|\mathbf{a}_2\rangle) = \mathbf{T}_2(|0\rangle). \quad (1.54)$$

We could now reduce this matrix \mathbf{T}_2 to diagonal form, just as in (1.52), by choosing a suitable set of starting functions. At first sight,

this does not seem compatible with the set we have used to make \mathbf{T}_1 diagonal. But consider the following alternative reductions:

$$(|\mathbf{a}_1 + \mathbf{a}_2\rangle) = \left\{ \begin{array}{l} \mathbf{T}_2(|\mathbf{a}_1\rangle) = \mathbf{T}_2 \mathbf{T}_1(|0\rangle), \\ \mathbf{T}_1(|\mathbf{a}_2\rangle) = \mathbf{T}_1 \mathbf{T}_2(|0\rangle). \end{array} \right\} \quad (1.55)$$

The matrices \mathbf{T}_1 and \mathbf{T}_2 thus commute with one another. There is a theorem of matrix algebra which tells us that there then exists a unitary matrix \mathbf{S} such that *both* \mathbf{T}_1 and \mathbf{T}_2 can be reduced *simultaneously* to diagonal form. Thus, the states $|0\rangle_1$ and $|0\rangle_2$ will not be mixed by the translation \mathbf{a}_2 , but will simply be multiplied by phase factors

$$|\mathbf{a}_2\rangle_1 = e^{ik_2} |0\rangle_1, \quad |\mathbf{a}_2\rangle_2 = e^{ik'_2} |0\rangle_2. \quad (1.56)$$

A similar result holds for translation by \mathbf{a}_3 , so that we are back at essentially the same as (1.45). For each of $|0\rangle_1$ and $|0\rangle_2$ there exists a wave-vector such that

$$|l\rangle = e^{i\mathbf{k}\cdot l} |0\rangle. \quad (1.57)$$

The extension to more than two states degenerate in energy is obviously trivial.

Thus, the theorem is proved in general, in the sense that any degenerate solution of the wave-function can be represented as a linear combination of solutions of the same energy each of which must satisfy a condition of this sort, though not all with the same value of \mathbf{k} .

This is *Bloch's theorem*. Actually, there is a general group-theoretical proof, which follows as a corollary from the theorem that 'in the field of complex numbers any representation of an Abelian group can be reduced to a sum of one-dimensional representations'. The group of translations of the crystal is Abelian—that is to say, all the elements of the group commute with one another. This follows from the obvious circumstance that the translation $(\mathbf{a}_1 + \mathbf{a}_2)$, say, is identical with $(\mathbf{a}_2 + \mathbf{a}_1)$ —we can get to a lattice point by many different paths, but the result is just the same. The matrices \mathbf{T}_1 , \mathbf{T}_2 , etc., are representations of these translations, and so can be reduced to diagonal form.

1.5 Reduction to a Brillouin zone

Bloch's theorem is of such generality that it is hard to grasp it at this stage. In the case of electron waves, it means that we can label every wave-function by its wave-vector \mathbf{k} , and write

$$\psi_{\mathbf{k}}(\mathbf{r} + \mathbf{l}) = e^{i\mathbf{k}\cdot l} \psi_{\mathbf{k}}(\mathbf{r}). \quad (1.58)$$

We notice that an ordinary free-electron wave satisfies this condition;

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}. \quad (1.59)$$

This is to be expected, for we are still dealing with the solution of the Schrödinger equation in a periodic potential; but this potential happens to be zero everywhere. The *empty-lattice test* is often very valuable in the theory of solids.

It is sometimes convenient to try to make an electron wavefunction of given value of \mathbf{k} look as much as possible like a free-electron wave. We put

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) \quad (1.60)$$

and hope that $u_{\mathbf{k}}$ will be nearly constant. In fact, as may be proved at once from (1.58), the function $u_{\mathbf{k}}(\mathbf{r})$ must be periodic, i.e.

$$u_{\mathbf{k}}(\mathbf{r} + \mathbf{l}) = u_{\mathbf{k}}(\mathbf{r}). \quad (1.61)$$

Bloch's theorem is sometimes stated in this form.

The factor $\exp(i\mathbf{k}\cdot\mathbf{l})$ that appears in the theorem is similar to the factor $\exp(i\mathbf{g}\cdot\mathbf{l})$ that appeared when we studied periodic functions. A wave-vector \mathbf{k} evidently has the same dimensions as a reciprocal lattice vector \mathbf{g} ; it belongs in reciprocal space. If some state happened to have wave-vector \mathbf{g} , then it would be a periodic function:

$$\begin{aligned} \psi_{\mathbf{g}}(\mathbf{r} + \mathbf{l}) &= e^{i\mathbf{g}\cdot\mathbf{l}} \psi_{\mathbf{g}}(\mathbf{r}) \\ &= \psi_{\mathbf{g}}(\mathbf{r}), \end{aligned} \quad (1.62)$$

because
for all \mathbf{l} .

$$e^{i\mathbf{g}\cdot\mathbf{l}} = 1 \quad (1.63)$$

Again, suppose the state $\psi_{\mathbf{k}}$ has wave-vector \mathbf{k} such that

$$\mathbf{k} = \mathbf{g} + \mathbf{k}', \quad (1.64)$$

where \mathbf{g} is some reciprocal-lattice vector, and \mathbf{k}' is another vector. Then by (1.58) and (1.63)

$$\begin{aligned} \psi_{\mathbf{k}}(\mathbf{r} + \mathbf{l}) &= e^{i(\mathbf{g} + \mathbf{k}')\cdot\mathbf{l}} \psi_{\mathbf{k}}(\mathbf{r}) \\ &= e^{i\mathbf{g}\cdot\mathbf{l}} e^{i\mathbf{k}'\cdot\mathbf{l}} \psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}'\cdot\mathbf{l}} \psi_{\mathbf{k}}(\mathbf{r}). \end{aligned} \quad (1.65)$$

That is to say, the state $\psi_{\mathbf{k}}$ satisfies Bloch's theorem as if it had the wave-vector \mathbf{k}' . The original label \mathbf{k} is not unique; every state has a whole host of possible wave-vectors, differing from one another by the vectors of the reciprocal lattice. This does not, of course, contradict the theorem, which merely asserted that each state must have at least *one* wave-vector.

We are thus faced with a problem; how are we to define uniquely the wave-vector of a given state? One could use (1.60) as a guide, and try to choose \mathbf{k} so that $u_{\mathbf{k}}(\mathbf{r})$ is as constant as possible—but this is arbitrary, and, as we shall see later, even misleading. The proper procedure is as follows:

Consider what happens in one dimension (cf. 1.8–1.12): the analogue of the reciprocal lattice is the set of reciprocal lattice lengths

$$g_n = n \frac{2\pi}{a}. \quad (1.66)$$

A state may be assigned any wave-number in the set

$$k = n \frac{2\pi}{a} + k', \quad (1.67)$$

that is to say, \mathbf{k} is only defined *modulo* $(2\pi/a)$. All the points \mathbf{k} in Fig. 12 are equivalent.

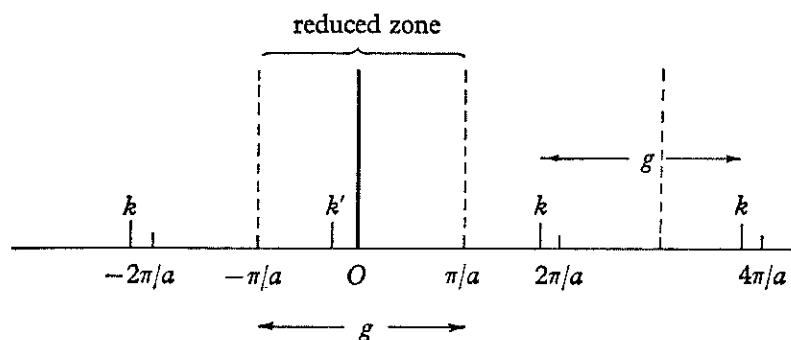


Fig. 12. The points k all reduce to k' in the one-dimensional reciprocal lattice.

It is natural to take k' as the representative of all of them, with $|k'|$ as small as possible. In other words, we always choose for a wave-number the value that lies in the range

$$-\frac{\pi}{a} < k \leq \frac{\pi}{a}. \quad (1.68)$$

It is evident that this range is equivalent to a Brillouin zone for our one-dimensional system; it is a unit cell of the reciprocal lattice. In three dimensions we do the same thing—we choose our wave-vector to lie in the first Brillouin zone in reciprocal space. That this is always possible follows by elementary geometry. We choose the value of \mathbf{g} in (1.64) to make $|\mathbf{k}'|$ as small as possible—that is, to lie as near to the origin of the reciprocal lattice as it can. This means that $|\mathbf{k}'|$ is to lie nearer to the origin than to any other sites of the reciprocal lattice—which amounts to saying that it lies in the Wigner–Seitz cell of that

lattice, i.e. in the Brillouin zone. It is evident that we can *reduce* any point \mathbf{k} in reciprocal space to a point in the Brillouin zone, so that any state can be given a label in the *reduced zone scheme*.

Any state may thus be characterized by its *reduced wave-vector*. But there may be many states with the same reduced wave-vector, and different energies. The use of the reduced zone scheme thus prevents us from assigning a distinct value of \mathbf{k} to each state.

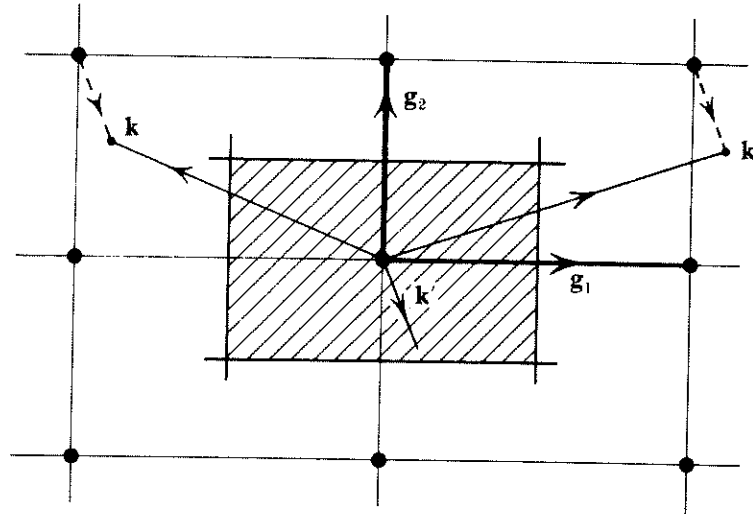


Fig. 13. The wave-vectors \mathbf{k} all reduce to \mathbf{k}' , which lies in the Brillouin zone.

It is interesting to see what happens to free-electron states in an empty lattice. Suppose

$$\begin{aligned}\psi &= e^{i\mathbf{k}\cdot\mathbf{r}} \\ &= e^{i(\mathbf{k}-\mathbf{g})\cdot\mathbf{r}} e^{i\mathbf{g}\cdot\mathbf{r}} \\ &= e^{i\mathbf{k}'\cdot\mathbf{r}} (e^{i\mathbf{g}\cdot\mathbf{r}}),\end{aligned}\tag{1.69}$$

where \mathbf{k}' is the reduced value of the actual wave-vector, \mathbf{k} . This is in the form (1.60), since $\exp(i\mathbf{g}\cdot\mathbf{r})$ is a periodic function in the lattice: but it is an extremely artificial representation of a plane wave. If the energy is given, in the one-dimensional case, by

$$\mathcal{E}(\mathbf{k}) = \frac{\hbar^2 k^2}{2m},\tag{1.70}$$

then it will appear as a multivalued function of \mathbf{k}' in the reduced zone. In Fig. 14, the section AB of the parabola is moved to $A'B'$ by a translation of a reciprocal lattice vector and so on. It is obvious that the *extended zone scheme*, in which each state is represented by its 'true' wave-vector, \mathbf{k} , would be more natural in this case.

1.6 Boundary conditions: counting states

All this assumes complete translational symmetry, which would require an infinite lattice. This is tiresome mathematically, because there would be an infinite number of atoms, and of states to deal with. We can only deal with these by treating a system of a finite number of atoms, and then going to the limit as this number tends to infinity. We might introduce boundaries as if they were real surfaces, at which, say, the wave-functions must vanish. But this also is inconvenient, because the exact stationary states of the system would then have to take account of all the reflections from the boundaries, and would turn out to be standing waves. To use such functions to represent the states of the conduction electrons, which are usually scattered incoherently by impurities or thermal vibrations within a very short distance, is mathematically clumsy and essentially unphysical.

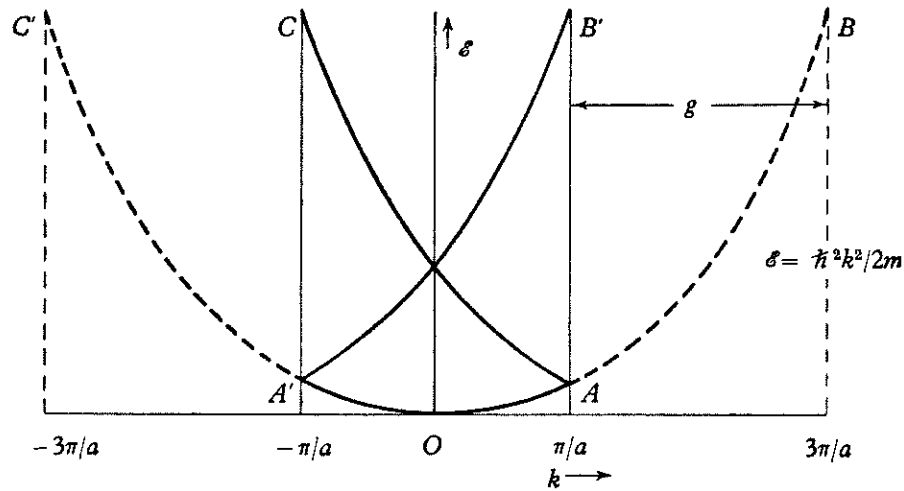


Fig. 14. The energy of free electrons in the reduced zone.

There is a mathematical device which deals satisfactorily with the problem of counting states, yet which does not introduce any direct physical effects from the boundaries. This is to use *cyclic*, or *Born-von Kármán* boundary conditions.

In one dimension, let the 'crystal' have L cells, which are joined in a circle. The consequence of this for an electron wave-function might be expressed by the condition

$$\psi(x + La) = \psi(x), \quad (1.71)$$

ensuring that there is no discontinuity at the junction point.

But our Bloch condition in one dimension says

$$\psi_k(x + La) = e^{ikLa} \psi_k(x), \quad (1.72)$$

so that we must have $e^{ikLa} = 1$, (1.73)

$$\text{i.e.} \quad k = \frac{2\pi m}{La}, \quad (1.74)$$

where m is an integer.

In a reduced zone in one dimension we have $-\pi/a < k \leq \pi/a$, so that the integers in the range

$$-\frac{1}{2}L < m \leq \frac{1}{2}L \quad (1.75)$$

will give all essentially different values of reduced wave-number. There are exactly L such values (when one has taken care of the different cases where L is odd and L is even—a trivial point) and they are spaced $(1/L)(2\pi/a)$ apart. Since L is supposed to be very large, this distribution is effectively continuous, and of constant density in reciprocal space.

In three dimensions we extend this argument by asserting that our macroscopic system is cyclic in three dimensions. We take it to be a crystal of dimensions $L_1 \mathbf{a}_1$, $L_2 \mathbf{a}_2$, $L_3 \mathbf{a}_3$ in the three basic directions of the lattice, and we write

$$\psi(\mathbf{r} + L_1 \mathbf{a}_1) = \psi(\mathbf{r}), \quad \psi(\mathbf{r} + L_2 \mathbf{a}_2) = \psi(\mathbf{r}), \quad \psi(\mathbf{r} + L_3 \mathbf{a}_3) = \psi(\mathbf{r}), \quad (1.76)$$

just as in (1.71).

For a Bloch state of wave-vector \mathbf{k} , these conditions imply that

$$e^{i\mathbf{k} \cdot (L_1 \mathbf{a}_1)} = e^{i\mathbf{k} \cdot (L_2 \mathbf{a}_2)} = e^{i\mathbf{k} \cdot (L_3 \mathbf{a}_3)} = 1, \quad (1.77)$$

which can only be achieved by having \mathbf{k} in the form

$$\begin{aligned} \mathbf{k} &= k_1 \mathbf{b}_1 + k_2 \mathbf{b}_2 + k_3 \mathbf{b}_3 \\ &= \frac{2\pi m_1}{L_1} \mathbf{b}_1 + \frac{2\pi m_2}{L_2} \mathbf{b}_2 + \frac{2\pi m_3}{L_3} \mathbf{b}_3, \end{aligned} \quad (1.78)$$

where m_1 , m_2 , m_3 are integers, and the vectors \mathbf{b}_1 , \mathbf{b}_2 , \mathbf{b}_3 , are, of course, the reciprocal triad of the set $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ as in (1.19).

But this result is obviously comparable to the definition of a reciprocal lattice vector

$$\mathbf{g} = n_1 \cdot 2\pi \mathbf{b}_1 + n_2 \cdot 2\pi \mathbf{b}_2 + n_3 \cdot 2\pi \mathbf{b}_3, \quad (1.79)$$

where n_1, n_2, n_3 are integers. The *allowed values* of \mathbf{k} in (1.78) are obtained by dividing the generators of the cell of the reciprocal lattice into L_1 parts in direction \mathbf{b}_1 , L_2 parts in direction \mathbf{b}_2 and L_3 parts in direction \mathbf{b}_3 . Thus, a fine rash of points is distributed evenly through reciprocal space, as in Fig. 15.

To calculate the density of these points, it is only necessary to

notice that we can cover a unit cell of the reciprocal lattice by allowing the integers m_1, m_2, m_3 to run through the values

$$0 \leq m_1 < L_1, \quad 0 \leq m_2 < L_2, \quad 0 \leq m_3 < L_3. \quad (1.80)$$

But this is not the cell we should naturally use as a basic zone for \mathbf{k} -vectors. We would do better, perhaps, with the range

$$-\frac{1}{2}L_1 < m_1 < \frac{1}{2}L_1, \quad -\frac{1}{2}L_2 < m_2 < \frac{1}{2}L_2, \quad -\frac{1}{2}L_3 < m_3 < \frac{1}{2}L_3, \quad (1.81)$$

by analogy with (1.75). This would give a parallelepiped unit cell centred on the origin. The best choice of unit cell, however, is the Wigner-Seitz cell of the reciprocal lattice, i.e. our old friend the Brillouin zone.

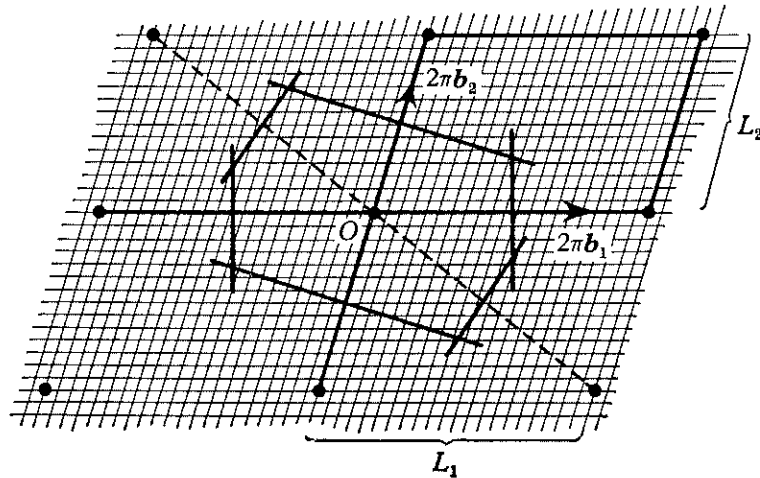


Fig. 15. The Brillouin zone covers the same area as the parallelepiped unit cell of the reciprocal lattice, and contains just N 'allowed \mathbf{k} -vectors'.

Now the volume of the Brillouin zone is the same as the parallelepiped unit cell, so that it must contain exactly the same number of 'allowed values' of \mathbf{k} . From (1.80) or (1.81) this number is obviously $L_1 \times L_2 \times L_3 = N$ which is exactly the number of unit cells in our whole macroscopic crystal. This is a most important theorem: *there are exactly as many allowed wave-vectors in a Brillouin zone as there are unit cells in the block of crystal.*

We have already shown (in (1.32)) that the volume of a zone is $8\pi^3/v_c$. If there are N unit cells in the volume V of crystal, then

$$v_c = V/N. \quad (1.82)$$

The volume in \mathbf{k} -space per allowed \mathbf{k} -vector is thus

$$\frac{1}{N} \frac{8\pi^3}{v_c} = \frac{8\pi^3}{V}. \quad (1.83)$$

Thus, there are $V/8\pi^3$ allowed \mathbf{k} -vectors per unit volume of reciprocal space.

In practice, N is very large, so that this distribution is treated as continuous. We often express a sum over \mathbf{k} -vectors as an integral

$$\sum_{\mathbf{k}} \rightarrow \int d\mathbf{k} \equiv \frac{V}{8\pi^3} \iiint d^3k, \quad (1.84)$$

using the single integral as a concise notation for the limit of the sum. It is important to remember, however, that when we actually come to a definite quadrature in \mathbf{k} -space we must include the weight factor $V/8\pi^3$. We shall usually assume, for simplicity, that $V = 1$, so that N is the number of cells per unit volume of crystal and $1/N$ is the volume of a unit cell.

The result (1.83) is actually independent of any assumed zone structure; it is well known in the theory of radiation, and in the case of free electrons. But for many purposes the fact that the Brillouin zone contains exactly N allowed points is more useful. It shows that the zone is essentially invariant, and depends only on the crystal structure; increasing the size of the whole crystal merely increases the density of states in \mathbf{k} -space.

Actually, it must be confessed that the Born-von Kármán conditions cannot be achieved physically. In two dimensions a network of cells drawn on the surface of a torus is cyclic in both directions, but for a three-dimensional lattice all three conditions cannot be satisfied simultaneously by any topological contortions. However, it works admirably as a mathematical device, and it can be justified by some messy mathematics. The fact is that the density of states in the asymptotic limit of large quantum numbers is very insensitive to the precise form of the boundary conditions. So long as we do not propose to discuss boundary effects themselves, the cyclic conditions are far and away the most convenient.

CHAPTER 8

OPTICAL PROPERTIES

But, soft! what light through yonder window breaks. . . .
Romeo and Juliet

8.1 Macroscopic theory

In this chapter we discuss the propagation of electromagnetic waves into and through solids. It is well known that some solids are *transparent* and others are *opaque*, that some solid surfaces are strongly *reflecting*, whilst others tend to *absorb* the radiation that falls on them. These effects depend on the frequency of the radiation; we therefore include the whole electromagnetic spectrum, from long-wave radio waves to soft X-rays.

Electromagnetic waves are solutions of Maxwell's equations. We shall write these as follows

$$\left. \begin{aligned} \nabla \wedge \mathbf{H} &= \frac{\epsilon}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi\sigma}{c} \mathbf{E}, & \nabla \cdot \mathbf{E} &= 0, \\ \nabla \wedge \mathbf{E} &= -\frac{\mu}{c} \frac{\partial \mathbf{H}}{\partial t}, & \nabla \cdot \mathbf{H} &= 0. \end{aligned} \right\} \quad (8.1)$$

In this chapter we shall not consider any magnetic effects: we take $\mu = 1$.

The difference between propagation in free space, and propagation in the solid is expressed by the two coefficients—the *dielectric constant* ϵ and the *conductivity* σ . The former defines the magnitude of displacement currents due to time variation of \mathbf{E} : the latter is a measure of the real currents created by the electric field.

The magnetic field may be eliminated between these equations

$$\nabla^2 \mathbf{E} = \frac{\epsilon}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} + \frac{4\pi\sigma}{c^2} \frac{\partial \mathbf{E}}{\partial t}. \quad (8.2)$$

This represents a wave propagated with dissipation. If we choose the frequency, ω , and write

$$\mathbf{E} = \mathbf{E}_0 \exp \{i(\mathbf{K} \cdot \mathbf{r} - \omega t)\}, \quad (8.3)$$

then our wave equation requires

$$-\mathbf{K}^2 = -\epsilon \frac{\omega^2}{c^2} - \frac{4\pi\sigma i\omega}{c^2}, \quad (8.4)$$

i.e.

$$K = \frac{\omega}{c} \left(\epsilon + \frac{4\pi\sigma i}{\omega} \right)^{\frac{1}{2}}. \quad (8.5)$$

Thus, in general the propagation constant K comes out to be a complex number. Now in free space we should have, simply

$$K = \frac{\omega}{c}, \quad (8.6)$$

as for a wave travelling with the velocity of light. In the medium the velocity is modified: we say that the phase velocity is divided by a *complex refractive index*

$$N = \left(\epsilon + \frac{4\pi\sigma i}{\omega} \right)^{\frac{1}{2}}. \quad (8.7)$$

The whole theory of the optical properties, as observed macroscopically, can be expressed in terms of N . For example, suppose we have a disturbance of frequency ω trying to propagate as a plane wave in the z -direction, and suppose we write

$$N = n + ik \quad (8.8)$$

for the real and imaginary parts of N . The propagation constant becomes

$$K = \frac{n\omega}{c} + \frac{ik\omega}{c}, \quad (8.9)$$

so that the wave (8.3) becomes

$$\mathbf{E} = \mathbf{E}_0 \exp \left\{ i\omega \left(\frac{nz}{c} - t \right) \right\} \exp \left(-\frac{k\omega z}{c} \right), \quad (8.10)$$

the velocity is reduced to c/n , and the wave is *damped*, as it progresses, by a fraction $\exp(-2\pi k/n)$ per wavelength.

The damping of the wave is, of course, associated with the absorption of electromagnetic energy. To calculate this we should use Maxwell's equations to find the current associated with (8.10). This is the right-hand side of the first equation of (8.1)

$$\begin{aligned} \mathbf{J} &= \left(-\frac{i\omega\epsilon}{c} + \frac{4\pi\sigma}{c} \right) \mathbf{E} \\ &= -\frac{i\omega}{c} N^2 \mathbf{E} \end{aligned} \quad (8.11)$$

by (8.7). The rate of production of Joule heat is the real part of

$$\mathbf{J} \cdot \mathbf{E} = -\frac{i\omega}{c} \mathbf{N}^2 E^2. \quad (8.12)$$

Thus the *absorption coefficient*—the fraction of energy absorbed in passing through unit thickness of the material—is given by

$$\eta = \frac{\text{Re}(\mathbf{J} \cdot \mathbf{E})}{n|\mathbf{E}|^2} = \frac{2k\omega}{c}. \quad (8.13)$$

The other situation which is often studied is where radiation is incident upon a plane surface of the material. In general this is a complicated problem in optics, but we need consider only the case of

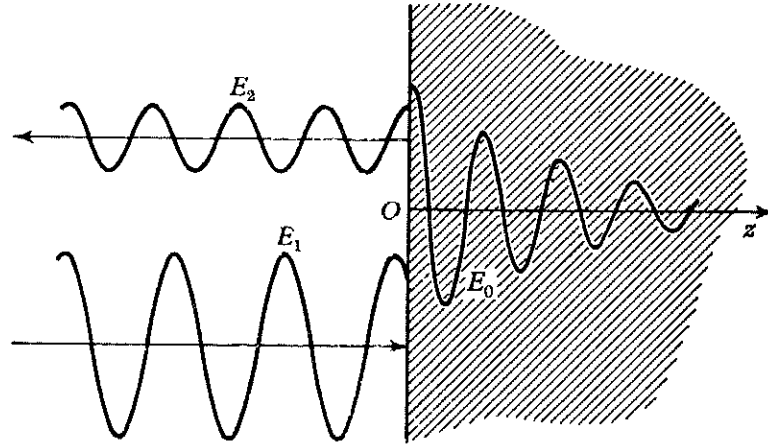


Fig. 136

normal incidence. We want to construct a solution of Maxwell's equations having the form (8.10) inside the medium, but matched to an incident and a reflected wave outside. Thus, for $z > 0$ we write

$$E_x = E_0 \exp\{i\omega(Nz/c - t)\}, \quad (8.14)$$

whilst for $z < 0$, in the free space outside, we write

$$E_x = E_1 \exp\{i\omega(z/c - t)\} + E_2 \exp\{-i\omega(z/c + t)\}, \quad (8.15)$$

corresponding to a wave of amplitude E_1 , travelling to the right and a wave of amplitude E_2 travelling to the left. Matching these on the boundary

$$E_0 = E_1 + E_2. \quad (8.16)$$

But there are also magnetic fields associated with these waves, with magnetic vector in the y -direction, say. Using Maxwell's equations to calculate H_y we have

$$-NE_0 = E_2 - E_1 \quad (8.17)$$

when matched to the wave outside. Thus, the ratio of the complex amplitudes of the reflected and incident waves is

$$\frac{E_2}{E_1} = \frac{1 - N}{1 + N}, \quad (8.18)$$

which corresponds to a real *reflection coefficient*,

$$R = \left| \frac{1 - N}{1 + N} \right|^2 = \frac{(n - 1)^2 + k^2}{(n + 1)^2 + k^2} \quad (8.19)$$

in terms of the real and imaginary parts (8.8) of the complex refractive index N .

It is evident that independent measurements of the reflection coefficient and of the absorption coefficient are sufficient to fix the values of n and of k . These are, therefore, the optical constants which one usually quotes, and whose theory we shall consider in this chapter. In practice the experiments may be much more complicated—e.g. reflection at an angle to the surface, or transmission through a thin film—but the results ought, in principle, all to be described by these same two coefficients.

However, the coefficients n and k are not quite independent of one another. They are linked by *dispersion relations*. A quantity such as N^2 in (8.11) is an example of a *generalized susceptibility*, $\alpha(\omega)$, say, in a relation

$$D(\omega) = \alpha(\omega) F(\omega) \quad (8.20)$$

between the Fourier components of a generalized ‘displacement’, D , and a ‘force’ F , at some frequency ω . As in elementary a.c. circuit theory, this would have to be a complex quantity

$$\alpha(\omega) = \alpha'(\omega) + i\alpha''(\omega) \quad (8.21)$$

to describe phase differences between D and F .

But (8.20) is merely the Fourier transform of

$$D(t) = \int_{-\infty}^{\infty} A(t - t') F(t') dt', \quad (8.22)$$

in which the displacement at time t is the resultant *linear response* of the system to forces acting at all other times t' . But there can be no displacement until *after* the application of a force: the response function is subject to the rigid condition of *causality*:

$$A(t - t') \equiv 0 \quad \text{for } t' > t. \quad (8.23)$$

In other words, the Fourier integral for the susceptibility is fully

defined by

$$\alpha(\omega) = \int_{-\infty}^{\infty} A(t) e^{i\omega t} dt \equiv \int_0^{\infty} A(t) e^{i\omega t} dt. \quad (8.24)$$

without any contributions from negative values of t .

To give meaning to such an integral, it is usual to treat $\alpha(\omega)$ as a function of the *complex* variable ω , with, say, an infinitesimal positive imaginary part $i\epsilon$. The form of (8.24) implies, moreover, that this function has no singularities above the real axis in the complex ω plane, and tends to zero as $|\omega| \rightarrow \infty$ in all directions in that region. A little bit of functional analysis, in which Cauchy's theorem is applied to a contour running along the real axis and returning in an infinite semicircle, yields the formula

$$\alpha(\Omega) = \frac{1}{i\pi} \int_{-\infty}^{\infty} \frac{\alpha(\omega)}{\omega - \Omega} d\omega \quad (8.25)$$

where the integral takes its principal value and the frequency variables are real.

The definition of the Fourier transforms for 'negative frequencies' implies that $\alpha(-\omega)$ is the complex conjugate of $\alpha(\omega)$. From (8.25) we deduce the *Kramers-Kronig relations* between the real and imaginary parts of the susceptibility (8.21):

$$\alpha'(\Omega) = \frac{2}{\pi} \int_0^{\infty} \frac{\omega \alpha''(\omega) d\omega}{\omega^2 - \Omega^2} + \text{const.}, \quad (8.26)$$

and

$$\alpha''(\Omega) = -\frac{2\Omega}{\pi} \int_0^{\infty} \frac{\alpha'(\omega) d\omega}{\omega^2 - \Omega^2}. \quad (8.27)$$

In our case, the real and imaginary parts of N^2 are

$$N^2 = (n^2 - k^2) + 2nki, \quad (8.28)$$

so that we have the conditions

$$n^2(\omega) - k^2(\omega) = \frac{2}{\pi} \int_0^{\infty} \frac{\omega' 2n(\omega') k(\omega') d\omega'}{\omega'^2 - \omega^2} + \text{const.}, \quad (8.29)$$

$$\text{and} \quad 2n(\omega) k(\omega) = -\frac{2\omega}{\pi} \int_0^{\infty} \frac{\{n^2(\omega') - k^2(\omega')\} d\omega'}{\omega'^2 - \omega^2}. \quad (8.30)$$

From these relations, if we know, say, the absorption coefficient as a function of frequency for all frequencies, then we can evaluate both $n(\omega)$ and $k(\omega)$ separately. They are not accidental consequences of some particular model; the proof, which depends solely on the causality condition (8.23), is perfectly general and applies to any linear response function such as the dielectric function (5.16) or the surface impedance (8.106).

8.2 Dispersion and absorption

The simplest model of a solid is a fixed assembly of independent neutral atoms. What is the effect of an electromagnetic wave on such a system? In case this theory is not well known, we sketch it out for a simple case, in which each atom contains only one electron, in a ground-state orbital $\phi_0(\mathbf{r})$, which can be excited to higher orbitals $\phi_j(\mathbf{r})$.

Suppose that the electric field in the wave, near the atom, is given by

$$\mathbf{E}(t) = E_x(e^{i\omega t} + e^{-i\omega t}), \quad (8.31)$$

ignoring the variation with distance as in (8.3). We follow time-dependent perturbation theory, and suppose that the electron wavefunction at any particular moment can be written

$$\psi(\mathbf{r}, t) = \phi_0 \exp(-i\mathcal{E}_0 t/\hbar) + \sum_j c_j(t) \phi_j \exp(-i\mathcal{E}_j t/\hbar). \quad (8.32)$$

We treat $e\mathbf{E} \cdot \mathbf{r}$ as the perturbing potential acting on the electron. It is easy to prove, by putting (8.32) into the time-dependent Schrödinger equation, multiplying through by ϕ_0 and integrating, that the coefficient of the higher state must approximately satisfy the differential equation

$$i\hbar \frac{dc_j}{dt} = \int \phi_j^* e\mathbf{E}(t) \cdot \mathbf{r} \phi_0 \mathbf{dr} e^{i(\mathcal{E}_j - \mathcal{E}_0)t/\hbar} \quad (8.33)$$

which has the solution

$$\begin{aligned} c_j(t) &= \frac{1}{i\hbar} \int_0^t eE_x x_{j0} (e^{i\omega t} + e^{-i\omega t}) e^{i(\mathcal{E}_j - \mathcal{E}_0)t/\hbar} \\ &= eE_x x_{j0} \left\{ \frac{1 - e^{i(\hbar\omega + \mathcal{E}_j - \mathcal{E}_0)t/\hbar}}{\hbar\omega + (\mathcal{E}_j - \mathcal{E}_0)} - \frac{1 - e^{i(-\hbar\omega + \mathcal{E}_j - \mathcal{E}_0)t/\hbar}}{\hbar\omega - (\mathcal{E}_j - \mathcal{E}_0)} \right\}, \end{aligned} \quad (8.34)$$

where

$$ex_{j0} \equiv \int \phi_j^* ex \phi_0 \mathbf{dr}, \quad (8.35)$$

i.e. it is the matrix element of the dipole moment of the electron in the direction of the electric vector of the field, between states ϕ_0 and ϕ_j .

The usual procedure now, in books of quantum theory, is to study $|c_j(t)|^2$, and to show that it oscillates strongly except for such values of ω as satisfy the condition

$$\hbar\omega = \hbar\omega_j \equiv (\mathcal{E}_j - \mathcal{E}_0). \quad (8.36)$$

One then proves that there will be a quantum jump from state ϕ_0 to state ϕ_j , with probability determined by the square of the matrix element (8.35).

But we now follow another line, and calculate the value of the dipole moment of the atom in this time-dependent state. From (8.32) and (8.34) we have the following

$$\begin{aligned}\langle ex(t) \rangle &= \int \psi^*(\mathbf{r}, t) ex \psi(\mathbf{r}, t) d\mathbf{r} \\ &= \sum_j \{ ex_{0j} c_j(t) e^{-i\omega_j t} + ex_{j0} c_j^*(t) e^{i\omega_j t} \} \\ &= E_x \sum_j e^2 |x_{0j}|^2 \frac{1}{\hbar} \left\{ \frac{1}{\omega_j - \omega} + \frac{1}{\omega_j + \omega} \right\} (e^{i\omega t} + e^{-i\omega t}) \quad (8.37)\end{aligned}$$

together with terms that oscillate rapidly and out of phase with the applied field.

Thus, the dipole moment of the atom is proportional to the field. We say that the *atomic polarizability* is given by

$$\alpha(\omega) = \sum_j \frac{e^2 |x_{0j}|^2}{\hbar} \frac{2\omega_j}{\omega_j^2 - \omega^2}. \quad (8.38)$$

To discuss the order of magnitude of such an expression, we may use the *Thomas-Reiche-Kuhn* sum rule—a simpler case of (5.48). This reads

$$\frac{2m}{\hbar^2} \sum_j \hbar \omega_j |x_{0j}|^2 = 1; \quad (8.39)$$

it is convenient to define the *oscillator strength* of the j th transition

$$f_j = \frac{2m}{\hbar^2} \hbar \omega_j |x_{0j}|^2, \quad (8.40)$$

and write (8.38) in the form

$$\alpha(\omega) = \frac{e^2}{m} \sum_j \frac{f_j}{\omega_j^2 - \omega^2}, \quad (8.41)$$

where the numbers f_j add to unity over all transitions.

If now we bring together N such atoms per unit volume, we have a medium whose dielectric constant has a real part

$$\begin{aligned}\epsilon(\omega) &= 1 + 4\pi N \alpha(\omega) \\ &= 1 + \frac{4\pi N e^2}{m} \sum_j \frac{f_j}{\omega_j^2 - \omega^2}.\end{aligned} \quad (8.42)$$

Strictly speaking, we should distinguish between this *transverse* dielectric constant, and the dielectric constant for *longitudinally* polarized waves of the electrostatic field discussed in Chapter 5. In the limit of long wavelengths, however—i.e. for normal ‘optical’ phenomena—these quantities are the same, since the screening or atomic polarization effects are relatively localized and depend mainly

on the strength of the electric field in the immediate vicinity of each ion or atom.

This is the prototype of a dispersion formula. It is interesting to note that it reduces to the case already discussed in §5.7. If the electrons are 'free', in the sense that all the ω_j are zero and the sum of the f_j are unity, we may write, as in (5.53)

$$\omega_p^2 = \frac{4\pi N e^2}{m} \quad (8.43)$$

to define a plasma frequency for our N electrons per unit volume; (8.42) then reproduces (5.67), the formula leading to the phenomenon of ultra-violet transparency. Notice, however, that there is practically no direct excitation of longitudinal plasmons by transverse photons. To observe *plasmon lines* we need to excite the electron gas with high energy electrons, or look through thin films where *surface plasmon modes* may be observed.

But in general the lowest value of ω_j would correspond to an optical or infra-red frequency, and the variation of ϵ with ω is more complicated. At low frequencies we might find the static dielectric constant

$$\epsilon(0) = 1 + \sum_j f_j \frac{\omega_p^2}{\omega_j^2}, \quad (8.44)$$

which is essentially larger than unity. But $\epsilon(0)$ would not be very large, because ω_p for most solids is comparable with an ordinary optical frequency for the free atom.

Then, as ω increases, $\epsilon(\omega)$ increases until there is a singularity at $\omega = \omega_j$. Thereafter, $\epsilon(\omega)$ becomes negative for a certain range of frequencies before becoming positive again near the next resonance frequency. Eventually, however, $\epsilon(\omega)$ will tend to the formula (5.67);

$$\epsilon(\omega) \rightarrow 1 - \frac{\omega_p^2}{\omega^2}, \quad (8.45)$$

for $\omega >$ all ω_j .

Now in the neighbourhood of each ω_j we have a region of *anomalous dispersion*. In particular, if ϵ is negative, then by (8.7) the refractive index N becomes pure imaginary; thus

$$n = 0, \quad k = |\epsilon|^{\frac{1}{2}}, \quad (8.46)$$

so that by (8.19) there would be *total reflection* of light from the surface of the solid. In effect, the crystal would seem to remain transparent, although with a very high refractive index, up to $\omega = \omega_j$, would then

suddenly become opaque and perfectly reflecting, and would then become transparent again at a higher frequency.

But this would not be consistent with our dispersion relations (8.29) and (8.30). There must also be some absorption, as measured by the product $2nk\omega$. This, indeed, is the aspect of the time-dependent perturbation that we set aside—the absorption associated with the atomic transition (8.36). This comes as a sharp line at the frequency ω_j —presumably a delta function. In order to satisfy the dispersion relation (8.29) and give the correct dispersion terms in (8.42), we must have

$$2\omega n(\omega) k(\omega) = \frac{\pi}{2} \frac{4\pi N e^2}{m} \sum_j f_j \delta(\omega - \omega_j); \quad (8.47)$$

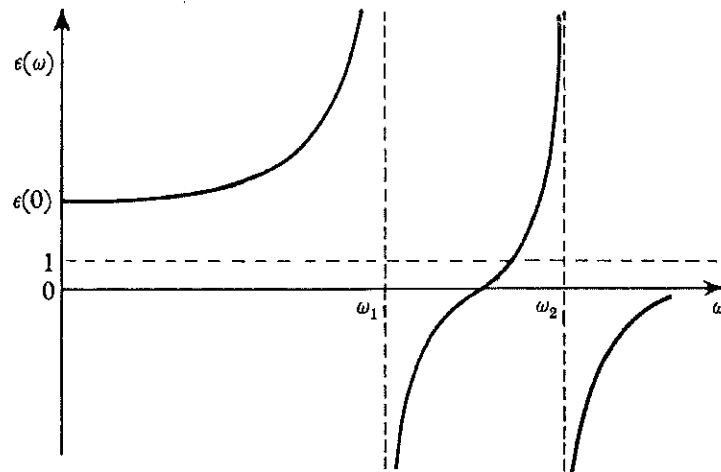


Fig. 137. Dispersion of light.

this must be the imaginary part of the complex dielectric constant which takes the form

$$\epsilon(\omega) = 1 + \frac{4\pi N e^2}{m} \sum_j f_j \left\{ \frac{1}{\omega_j^2 - \omega^2} + \frac{i\pi}{2\omega} \delta(\omega^2 - \omega_j^2) \right\}. \quad (8.48)$$

The large refractive index near ω_j turns into a sharp absorption line at this frequency.

In practice we never see an infinitely sharp line. There is always some broadening, from impurities, etc., or just from the natural radiative relaxation of the levels. As is shown in standard quantum theory, such effects can be incorporated phenomenologically in the analysis by inserting in (8.34), say, a decay factor $\exp(-\frac{1}{2}\Gamma t)$, corresponding to a decay time (in the *square* of the amplitude) of the order of $1/\Gamma$ seconds.

It is easy to follow such a factor through the algebra as the addition

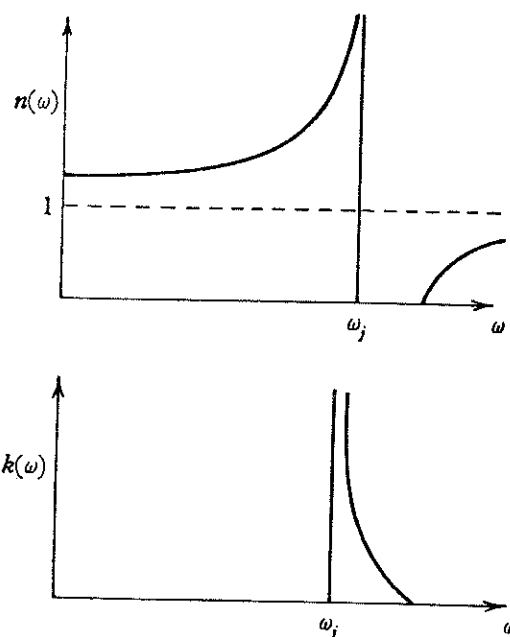


Fig. 138. Real and imaginary parts of the refractive index.

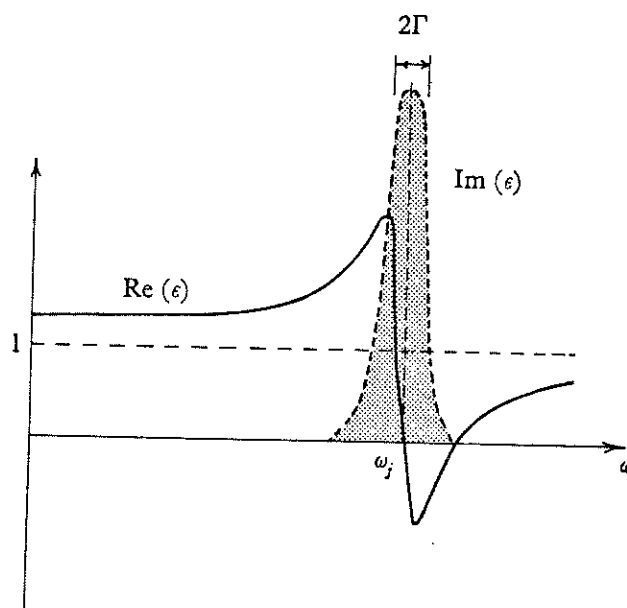


Fig. 139. Broadening of dispersion curve.

of $\frac{1}{2}i\Gamma$ to ω_j , in (8.37) or (8.38). If we ignore Γ^2 relative to ω_j^2 , then we get for the real part of the polarizability, and hence in the real part of $\epsilon(\omega)$, terms like

$$f_j \frac{\omega_j^2 - \omega^2}{(\omega_j^2 - \omega^2)^2 + \omega^2 \Gamma^2} \quad (8.49)$$

in place of $f_j/(\omega_j^2 - \omega^2)$. This has the effect of removing the singularity

in n and spreading the dispersion function over a range of frequency of the order of 2Γ in width.

As is well known, the effect of relaxation on the absorption line itself is to spread out the delta function into a finite function of the form

$$\frac{\Gamma/2\pi}{(\omega_j - \omega)^2 + \frac{1}{4}\Gamma^2} \approx \frac{2\Gamma\omega^2/\pi}{(\omega_j^2 - \omega^2)^2 + \Gamma^2\omega^2} \quad (8.50)$$

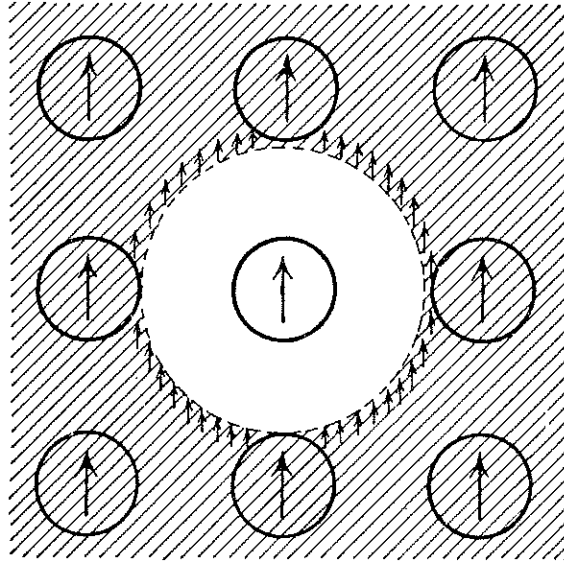


Fig. 140. The Lorentz correction replaces the local field due to the dipoles on neighbouring atoms by the field of a spherical cavity in a uniformly polarized medium.

in the neighbourhood of $\omega = \omega_j$. In fact, (8.49) and (8.50) can be combined into a formula like (8.48)

$$\begin{aligned} \epsilon(\omega) &= 1 + \frac{4\pi N e^2}{m} \sum_j f_j \left\{ \frac{\omega_j^2 - \omega^2}{(\omega_j^2 - \omega^2)^2 + \omega^2 \Gamma_j^2} + \frac{i\Gamma_j \omega}{(\omega_j^2 - \omega^2)^2 + \omega^2 \Gamma_j^2} \right\} \\ &= 1 + \frac{4\pi N e^2}{m} \sum_j \frac{f_j}{(\omega_j^2 - \omega^2) - i\omega \Gamma_j}, \end{aligned} \quad (8.51)$$

which is more or less the most general type of dispersion formula. Although calculated for a model of independent atoms, it provides a phenomenological expression for any system whose absorption spectrum is a series of discrete lines.

There is one further correction that should be made in this analysis. The relation (8.42) implies that the *local* field polarizing each atom is the same as the *macroscopic* field applied to the crystal. This is not strictly true; the atom is not situated in an absolutely continuous

medium, but occupies a site in a lattice where it is surrounded, but not permeated, by 'other' atoms, similarly polarized (see § 2.3). It is well known, in elementary electrostatics and magnetostatics, that there must be a correction to the local field; if the 'cavity' surrounding the atom is more or less spherical, then we have the *Lorentz correction* to the local field

$$\mathbf{E}_{\text{local}} = \mathbf{E}_{\text{macroscopic}} + \frac{4}{3}\pi\mathbf{P}, \quad (8.52)$$

where \mathbf{P} is the average polarization. This leads to the *Clausius-Mosotti relation*

$$\left(\frac{\epsilon - 1}{\epsilon + 2}\right) = \frac{4}{3}\pi N\alpha \quad (8.53)$$

in place of (8.42). The result is that the observed dielectric constant and refractive index will follow a more complicated formula than (8.51); for example, the régime of 'total' reflection may be avoided.

When the dielectric polarization is due to the lining up of *permanent* dipole moments on the molecules of a liquid or solid, we must introduce further corrections, such as the *Onsager local field* model, which takes account of fluctuations in the orientation of neighbouring atomic moments.

8.3 Optical modes in ionic crystals

Let us return to the equations of motion of the lattice, as in § 2.1. It was there shown that the displacement of the s th ion in the l th cell satisfies (2.7), i.e.

$$M_s \ddot{\mathbf{u}}_{sl} = - \sum_{s'\mathbf{h}} \mathbf{G}_{ss'}(\mathbf{h}) \cdot \mathbf{u}_{s',l+\mathbf{h}}. \quad (8.54)$$

Now suppose we drive this system with an electromagnetic wave of the form of (8.3). We must add to the right-hand side of (8.54) a force

$$e_s \mathbf{E}_0 \exp\{i(\mathbf{K} \cdot \mathbf{l} - \omega t)\}, \quad (8.55)$$

where e_s is the charge of the s th ion in the cell.

We only need a particular integral of the differential equations (8.54) to take account of such an inhomogeneous term. It is obvious that the assumption of (2.8), with $\mathbf{q} = \mathbf{K}$, will remove the space-varying factor; we get an equation for $\mathbf{U}_{s\mathbf{K}}$, the amplitude of the mode of wave-vector \mathbf{K} ;

$$M_s \ddot{\mathbf{U}}_{s\mathbf{K}} = - \sum_{s'\mathbf{K}} \mathbf{G}_{ss'}(\mathbf{K}) \cdot \mathbf{U}_{s'\mathbf{K}} + e_s \mathbf{E}_0 e^{-i\omega t}. \quad (8.56)$$

To see the form of the solution of these equations, let us consider the simple model of the diatomic linear chain as in (2.22). In practice we are thinking of electromagnetic waves of very great wavelength on the

atomic scale, so we may take $\cos Ka \approx 1$. If the two atoms are of opposite charge, as in a typical ionic crystal, then we have

$$\left. \begin{aligned} M_1 \ddot{U}_1 &= 2\alpha(U_2 - U_1) + eE_0 e^{-i\omega t}, \\ M_2 \ddot{U}_2 &= 2\alpha(U_1 - U_2) - eE_0 e^{-i\omega t}. \end{aligned} \right\} \quad (8.57)$$

These have the solution

$$(U_1 - U_2) = \frac{eE_0(1/M_1 + 1/M_2)}{\omega_T^2 - \omega^2} e^{-i\omega t}, \quad (8.58)$$

where ω_T is the natural frequency of the system at $q = 0$, as given by (2.25). We may calculate the dipole moment associated with this motion and express it as a polarizability of the lattice

$$4\pi N\alpha(\omega) = \frac{4\pi N e^2(1/M_1 + 1/M_2)}{\omega_T^2 - \omega^2}. \quad (8.59)$$

We are thus led, by a purely classical argument, to a dispersion formula similar to (8.42). The numerator of (8.59) may be written $\Omega_p'^2$, indicating that it is very like the plasma frequency (6.81) for the assembly of ions. Since the ions are much heavier than electrons, this must be much smaller than ω_p^2 . On the other hand, ω_T is just the frequency of a lattice vibration of long wavelength, which is much smaller than the energy of any electronic transition in the atoms or ions. Over a wide frequency range, therefore, we may write

$$\epsilon(\omega) \approx \epsilon(\infty) + \frac{\Omega_p'^2}{\omega_T^2 - \omega^2}, \quad (8.60)$$

where the error arises from local field effects such as (8.53). In this formula, $\epsilon(\infty)$ represents the dielectric constant deduced from observations of the refractive index for $\omega \gg \omega_T$, before electronic transitions are excited. From this we may deduce the *Born equation* for the static dielectric constant

$$\epsilon(0) \approx \epsilon(\infty) + \Omega_p'^2/\omega_T^2. \quad (8.61)$$

In a three-dimensional ionic crystal, ω_T would be the frequency of a *transversely* polarized optical mode, whose dipole moment interacts strongly with a transverse electromagnetic wave. We have here, in fact, a *polariton*—a mixed mode of light and crystal polarization which is the true mode of propagation in the material. When ω passes through ω_T , the dielectric constant becomes negative: just as in (5.67) and (8.46), the radiation would be totally reflected from the crystal surface. The *Reststrahlen effect*—a forbidden region for electromagnetic propa-

gation in the crystal—persists up to the frequency ω_L , say, where (8.60) again becomes positive, i.e.

$$\omega_L^2 = \omega_T^2 + \Omega_p'^2/\epsilon(\infty). \quad (8.62)$$

But the definition of a dielectric constant implies (cf. (5.59)) that a very large electric field variation may then exist within the crystal at this frequency without external excitation. This is the analogue of a plasma oscillation in the ionic lattice, with contributions from both local interatomic forces and long range electrostatic fields: ω_L is the long wave limit of the *longitudinal* optical branch of the lattice spectrum. From (8.61) and (8.62) we derive a companion to (8.60)—the *Lyddane–Sachs–Teller relation*

$$\omega_L^2/\omega_T^2 = \epsilon(0)/\epsilon(\infty) \quad (8.63)$$

which is quite general for any polar material.

In practice, (8.59) and (8.60) should contain an imaginary term in the denominator (cf. (8.51)) corresponding to dissipative effects. For example, energy may be absorbed from the electromagnetic field by the creation of phonons, which would be linked to the optical modes by anharmonicity of the force constants (§§ 2.11, 7.10, 8.4). These effects would be temperature-dependent and would also be observable as a variation of the effective force constant α in the harmonic approximation (2.25) or (8.53). It can even happen that the balance of interatomic forces for some set of lattice displacements is so delicate that this temperature dependence takes ω_T^2 to *negative* values below some critical temperature T_C —an event that we should recognize macroscopically as a transition to a *ferro-electric* phase (§10.5) with a permanent built-in dielectric polarization of the lattice structure.

The relatively low frequency of the longitudinal optical modes is important in the theory of the *polaron*. An electron in an ionic crystal carries an electrostatic field that polarizes the lattice around it. For a stationary electron we could use the dielectric constant $\epsilon(0)$ but the response of the lattice to a moving electron is limited dynamically by the frequency ω_L , which describes the natural oscillations of dielectric polarization of the medium.

The situation is very much the same as in the theory of plasmons (§5.8). If the energy of the electron exceeds $\hbar\omega_L$ an optical phonon will be produced, and the electron will be decelerated and strongly scattered. Even when the electron is moving slowly, we must take account of the virtual excitation and re-absorption of quanta in the

optical modes; in other words, the electron behaves like a particle surrounded by a cloud of virtual phonons. The self-energy of this compound object is a function of the velocity, so that the effective 'mass' of the carrier is considerably increased.

The idealized Hamiltonian of the *large polaron* has been extensively studied as a model of a strongly-interacting field-theoretical system, but it is not an easy entity to observe in practice. The model itself is only valid when the polarization region is somewhat larger than a unit cell of the lattice; otherwise the self-trapped *small polaron* model is more appropriate (§ 6.7).

8.4 Photon-phonon transitions

In the language of field theory, infra-red absorption by a crystal is described as the interaction of a *photon* with one or more phonons. The selection rules for all such processes have already been deduced in §§ 2.7, 2.8, under the guise of *inelastic diffraction*. In the simplest case, for example, the *change* in wave-vector of the diffracted beam had to be equal to the wave-vector of the phonon emitted, i.e.

$$\mathbf{k} - \mathbf{k}' = \mathbf{q} \quad (8.64)$$

and also the change in frequency of the radiation had to be equal to the frequency of the phonon,

$$\omega - \omega' = \nu \quad (8.65)$$

as in (2.101) and (2.102). In special cases it is possible to satisfy these conditions without any resultant diffracted wave; the whole energy and momentum of the photon is transferred to the crystal.

The main point to notice here is that we are dealing with light in the visible and infra-red region, whose wavelength is much larger than the lattice constant of the crystal. The magnitude of $|\mathbf{k} - \mathbf{k}'|$ must therefore be extremely small on the scale of the Brillouin zone. For all ordinary optical phenomena we can take $\mathbf{k} - \mathbf{k}' \approx 0$, and concentrate our attention on the possible changes of energy allowed in the zone scheme; this aspect is complementary to the 'diffraction' point of view appropriate for X-rays and neutrons, where the change in *momentum* is large but the energy differences are not so easy to detect.

The process which we are here considering, the absorption or emission of one optical phonon, is therefore indicated by a nearly vertical line at $\mathbf{q} = 0$, in the reduced zone scheme. If the optical spectrum had several branches, then several different lines ought in principle to be observed, as in Fig. 141(b). But the magnitude of such processes

would depend upon the matrix element for the coupling of the electromagnetic wave to the crystal. As we have seen in § 8.3, this coupling is so strong for the transverse optical modes of most polar crystals that a description in terms of weakly interacting photons and phonons is inadequate and we get the Reststrahlen effect. But optical modes of very weak dipole moment are found in some compound semiconductors, where *one-phonon absorption* may be observed in light transmitted through thin films.

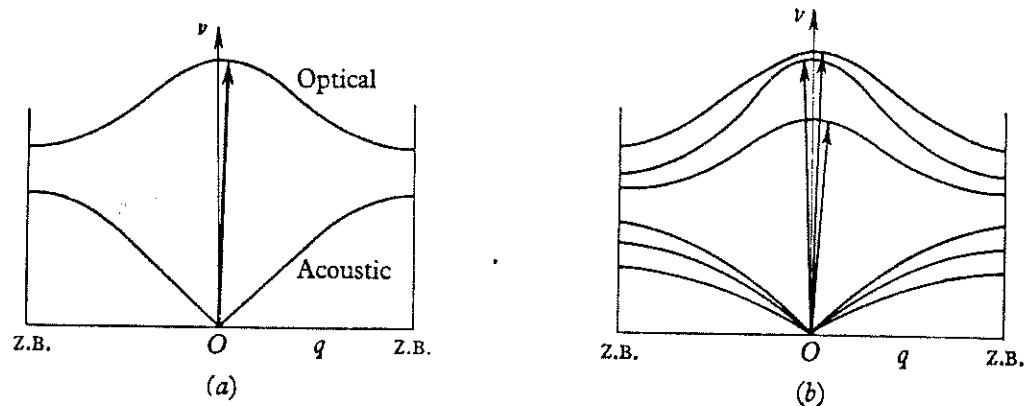


Fig. 141. Absorption by optical modes: (a) one-dimensional scheme; (b) three-dimensional scheme.

The observation that photon absorption (and emission) processes are 'vertical' in the Brillouin zone suggests that we may have other 'vertical' transitions, as indicated in Fig. 142. This is a transition involving an acoustic mode and an optical mode. The energy of the photon absorbed is evidently made up as follows

$$\hbar\omega = \hbar\nu_{\text{optical}} - \hbar\nu_{\text{acoustic}} \quad (8.66)$$

In effect, a photon, and an acoustic phonon have disappeared; an optical phonon has been created. The momentum conditions are also satisfied. We may ignore the momentum of the photon, and simply assert that

$$\hbar\mathbf{q}_{\text{optical}} = \hbar\mathbf{q}_{\text{acoustic}} \quad (8.67)$$

which is implied by the statement that the transition must be vertical.

Such processes are allowed by the selection rules, and therefore may be observed. They give rise to whole bands of absorption in the far infra-red. The width and general structure of such bands obviously

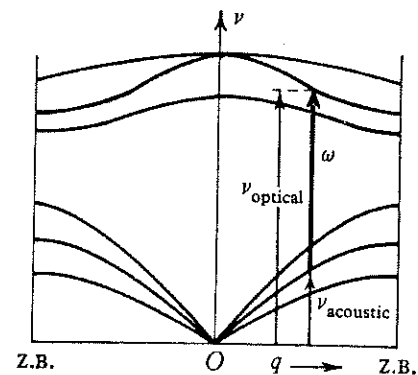


Fig. 142. Multiphonon absorption.

depends on the variety of possible transitions from acoustic to optical modes and to their spacing. The complete discussion of such processes is obviously rather complicated, since we are dealing with the difference of two functions, each of which is a function of \mathbf{q} in the Brillouin zone.

This does not exhaust all the possible processes. The above transitions we may describe as ‘*difference*’ processes. It is also possible to have *summation* bands, in which *two* phonons are created out of the photon energy, but of equal and opposite momenta. In that case

$$\hbar\omega = \hbar\nu_{\text{optical}} + \hbar\nu_{\text{acoustic}}, \quad (8.68)$$

with

$$\hbar\mathbf{q}_{\text{optical}} = -\hbar\mathbf{q}_{\text{acoustic}}.$$

To actually calculate the transition probability we need to know the non-linear term in the dipole moment as a function of lattice displacement. We need, for example, the tensor $\mathbf{B}_{ss'}$ in an expression like

$$\mathbf{M}_l = \sum_s \mathbf{A}_s \cdot \mathbf{u}_{sl} + \sum_{ss'} \mathbf{u}_{sl} \cdot \mathbf{B}_{ss'} \cdot \mathbf{u}_{s'l} \quad (8.69)$$

for the dipole moment in the l th cell, due to the displacements \mathbf{u}_{sl} . We then use (2.8), (2.10) and (2.129) to express each of the displacements in terms of the annihilation and creation operators for phonons in the various branches of the spectrum. The term in $\mathbf{B}_{ss'}$, corresponding to the product of two such operators, gives rise to the two-phonon processes for which we are looking. The actual probability of a transition will contain the squares of two matrix elements, of the form

$$\left. \begin{aligned} \langle n_{\mathbf{q}} + 1 | a_{\mathbf{q}}^* | n_{\mathbf{q}} \rangle &= (n_{\mathbf{q}} + 1)^{\frac{1}{2}}, \\ \langle n_{\mathbf{q}} - 1 | a_{\mathbf{q}} | n_{\mathbf{q}} \rangle &= n_{\mathbf{q}}^{\frac{1}{2}}, \end{aligned} \right\} \quad (8.70)$$

or

corresponding to the increase or decrease of the occupation number of the phonon states by one quantum. Since the average value of $n_{\mathbf{q}}$ is a function of the temperature, as in (2.46), these processes will be temperature dependent. In general, however, the non-linear coefficient $\mathbf{B}_{ss'}$ is small, so that these bands are weak by comparison with the one-phonon processes.

It is worth remarking that the function $\nu_{\mathbf{q}}$, for any branch, is a continuous periodic function of \mathbf{q} with the period of the reciprocal lattice (cf. § 2.2). It follows that the sum or difference of any pair of such functions has the same properties. Thus, van Hove’s theorem (§ 2.5) applies to the distribution of frequencies; only the four types of singularity indicated in Fig. 27 can occur in the observed spectrum (unless the matrix elements introduce extra singularities).

In (8.64), (8.65) we wrote down selection rules for the inelastic diffraction of light—i.e. for a photon $|\mathbf{k}, \omega\rangle$ to be transformed into another photon $|\mathbf{k}', \omega'\rangle$ with the emission of a phonon $|\mathbf{q}, \nu\rangle$. This would be observed as a shift in the frequency of some of the light scattered by the crystal. This first-order *Raman effect* is evidently equivalent to one-phonon absorption. The second-order Raman spectrum of a crystal contains bands from two-phonon processes—and so on. Where acoustic modes of finite wave-vector are involved, we refer to *Brillouin scattering*.

The selection rules for momentum (8.64), (8.67), (8.68) depend upon Bloch's theorem (see § 2.7). In the absence of perfect lattice translational symmetry, transitions would be allowed to any point in the Brillouin zone. An impurity or other imperfection of the lattice thus induces infra-red absorption and similar optical phenomena which would otherwise be forbidden. The detailed calculation of such effects is obviously very complicated, but sharp lines from *localized modes* and peaks from resonances (§ 2.12) can thus be observed directly.

8.5 Interband transitions

When the electrons are in Bloch states, forming broad energy bands, the electronic transitions from full to empty states produce a broad absorption band. Again, for optical frequencies, \mathbf{K} is small on the scale of the Brillouin zone, so that one can ignore the wave-vector of the photon, and assume that all transitions are 'vertical', as indicated in Fig. 143.

We have already written down, in (5.45), the formula for the dielectric constant of a semiconductor. In that case we were more interested in the static behaviour at all wavelengths; we are now interested in the frequency dependence of $\epsilon(\mathbf{q}, \omega)$ at long wavelengths. We have from (5.57)

$$\begin{aligned} \epsilon(\mathbf{q}, \omega) &= 1 + \frac{4\pi e^2}{q^2} \sum_{\mathbf{k}} \frac{|\langle \mathbf{k} | e^{i\mathbf{q}\cdot\mathbf{r}} | \mathbf{k} + \mathbf{q} + \mathbf{g} \rangle|^2 2\{\mathcal{E}(\mathbf{k}) - \mathcal{E}(\mathbf{k} + \mathbf{q} + \mathbf{g})\}}{(\hbar\omega)^2 - \{\mathcal{E}(\mathbf{k}) - \mathcal{E}(\mathbf{k} + \mathbf{q} + \mathbf{g})\}^2} \\ &\approx 1 + \frac{4\pi e^2}{m} \sum_{\mathbf{k}} \frac{f_{\mathbf{k}}}{\omega_{\mathbf{k}}^2 - \omega^2}, \end{aligned} \quad (8.71)$$

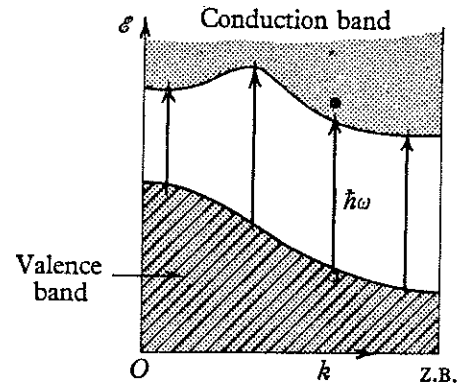


Fig. 143. Vertical interband transitions in a semiconductor.

where $f_{\mathbf{k}}$ is an oscillator strength, as in (8.40), for the transition between $|\mathbf{k}\rangle$ in the valence band and $|\mathbf{k} + \mathbf{g}\rangle$ (directly above it in the reduced zone scheme) in the conduction band, these states differing in energy by the amount $\hbar\omega_{\mathbf{k}}$. But since \mathbf{k} is continuous, the sum becomes an integral, of the form

$$\epsilon(0, \omega) \approx 1 + \frac{4\pi e^2}{m} \int \frac{f(\omega') \mathcal{N}_d(\omega') d\omega'}{\omega'^2 - \omega^2}, \quad (8.72)$$

where $\mathcal{N}_d(\omega') d\omega'$ is the number of levels having a vertical energy difference $\hbar\omega'$, in the range $d\omega'$, and $f(\omega')$ is an oscillator strength—i.e. a number of the order of unity—for transitions in this range.

This formula for the real part of the dielectric constant is not so interesting as the imaginary part. Appeal to the dispersion relations (8.29) and (8.30) shows that the absorption coefficient must be of the form

$$\begin{aligned} 2\omega n(\omega) k(\omega) &= \frac{\pi}{2} \frac{4\pi e^2}{m} \int f(\omega') \mathcal{N}_d(\omega') \delta(\omega - \omega') d\omega' \\ &= \frac{2\pi^2 e^2}{m} f(\omega) \mathcal{N}_d(\omega). \end{aligned} \quad (8.73)$$

We could have arrived at the same result by noting that $\epsilon(\mathbf{q}, \omega)$ in (5.45) was provided with an imaginary part, governed by the parameter α , which was introduced in (5.1) as the decay constant of the perturbation being screened by the electron system. It is evident that this parameter plays the same role as the line-width parameter Γ_j in (8.51), which is, indeed, essentially of the form of (5.45) and (5.57). Thus, we return to (8.73) by retracing the steps from (8.51) to (8.47)—effectively, by letting the line-width for each transition tend to zero.

The function $\mathcal{N}_d(\omega)$ in (8.73), being the spectrum of the difference of the energies of the valence and conduction bands, each of which is a continuous periodic function in the reciprocal lattice, has only the van Hove singularities noted in § 2.5. The most important of these is the *absorption band edge*, corresponding to the minimum vertical energy difference $\hbar\omega_0$ between the bands. As shown in (2.72), the behaviour of the spectrum in that neighbourhood must be of the form

$$\mathcal{N}_d(\omega) \propto (\omega - \omega_0)^{\frac{1}{2}}. \quad (8.74)$$

It should be noted, however, that $\hbar\omega_0$ is not necessarily the same as the 'energy gap', \mathcal{E}_{gap} , between the top of the valence band and the bottom of the conduction band. These points may not be vertically above one another in \mathbf{k} -space: the smallest *vertical* separation of the bands may be somewhat larger.

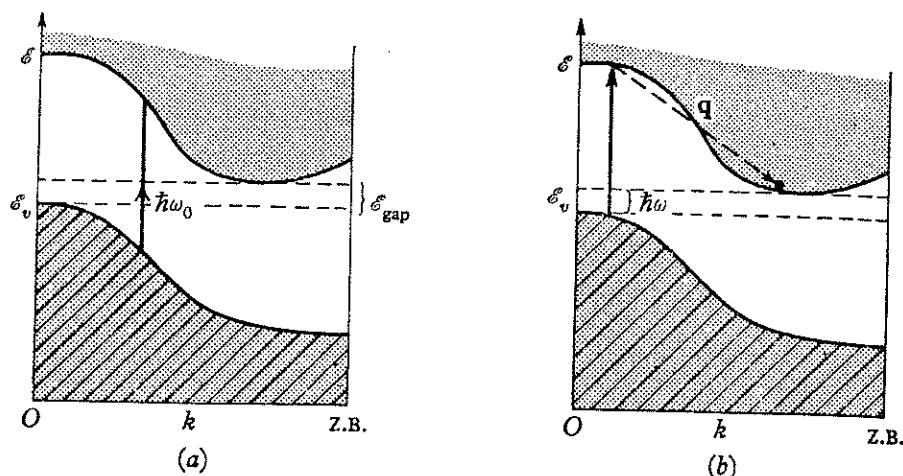


Fig. 144. (a) Direct transitions do not necessarily give energy gap.
(b) Phonon-assisted transition.

However, it is possible to observe optical transitions corresponding to $\hbar\omega \approx \mathcal{E}_{\text{gap}}$, if one allows the simultaneous emission or absorption of a phonon. Such *indirect* or *phonon-assisted* transitions are easily derived by second-order perturbation theory. We have two steps; for example, upwards, vertically by the absorption of a photon, and then to the appropriate minimum in the conduction band by the emission or absorption of a phonon of large enough wave-vector, \mathbf{q} . We need not worry about energy conservation in the intermediate virtual state; the over-all condition will be

$$\hbar\omega = \mathcal{E}_{\text{gap}} \pm \hbar\nu_{\mathbf{q}}, \quad (8.75)$$

where $\nu_{\mathbf{q}}$ is the frequency of the phonon. Since $\hbar\nu_{\mathbf{q}}$ is a small energy (for example, 0.03 eV.) on the scale of the energy gap, we find that the absorption band seems to start at about \mathcal{E}_{gap} . But the probability of indirect transitions is much smaller than that of direct transitions, and depends on the temperature through the occupation number of the phonon states, as indicated in (8.70).

This discussion of electronic transitions in perfect crystals assumes a simple one-electron model. In reality, however, the final state for a transition such as Fig. 144 contains a hole in the valence band as well as an electron in the conduction band. If these two particles do not immediately escape from one another, they may interact, as in § 6.7, to form bound states—Wannier exciton levels—whose total energy is less than the band gap from which they were created. In other words, the optical spectrum shows *exciton lines* below the fundamental absorption edge. This phenomenon is particularly noticeable in ionic and crystals, where the Frenkel excitons, although ‘small’ and practically immobile, have large oscillator strengths corresponding to the

atomic or molecular excited states from which they derive. In such materials the coupling between the electromagnetic field and the excitonic amplitude may be strong enough to produce polariton modes of propagation (cf. § 8.3).

The optical properties of semiconductors and ionic crystals are also greatly complicated and enriched by the effects of impurities and imperfections. Thus, one can observe a variety of absorption lines associated with transitions between the hydrogen-like levels of charged impurities (§ 6.4).

The theory of optical transitions in colour centres is made somewhat more complicated by the lattice distortion about the imperfection. Since this distortion is affected by electrostatic interaction with the optically active electron, it must depend on the electronic state of the system. The eigenstates of the imperfection must therefore be described by both electronic and atomic co-ordinates. But an optical transition takes place in the periodic time of the frequency absorbed or emitted, whereas the time needed for the lattice to distort into the configuration required by the new electronic state would be the periodic time of a dynamical mode, which is very much longer. The transition must therefore be treated as *non-adiabatic* or *sudden* from the point of view of the lattice (cf. § 6.11).

This is the basis of the *Franck-Condon* principle, illustrated in Fig. 145. Suppose that the distortion of the lattice about the imperfection can be measured by a single co-ordinate u , such as the amplitude of a 'breathing' displacement of the ions, inwards or outwards. When the electron is in state $|a\rangle$, say, the potential energy of the lattice has a minimum at some value u_a , which is not the same as the equilibrium displacement, u_b for a different electron state $|b\rangle$. Optical absorption from $|a\rangle$ to $|b\rangle$ will proceed without change of lattice distortion—i.e. 'vertically' along the line $u = u_a$. On the other hand, emission must occur along the u_b , corresponding to the equilibrium displacement in the initial electron state $|b\rangle$. Evidently, $\hbar\omega_{ab} \neq \hbar\omega_{ba}$. In practice, however, we must allow for thermal excitation of the lattice modes near each potential minimum. Phonon emission will broaden the lines and eventually, at high temperatures, smear out the difference between the two processes.

Interband optical transitions are, of course, also observable in metals—being modified to allow for the fact that the electrons may not completely fill a band, or may fall into several different zones. The theory of such transitions follows, in principle, the argument of this section—but of course the phenomenon may be somewhat masked by

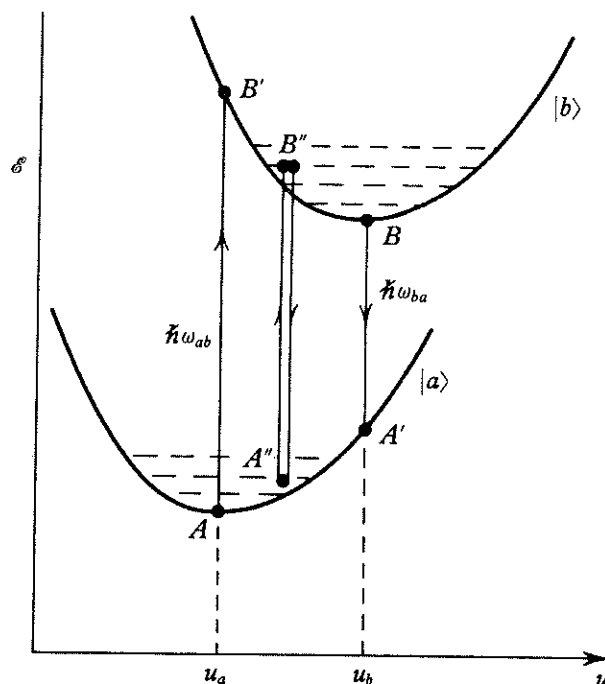


Fig. 145. Optical transitions at F -centre, showing effect of lattice distortions: absorption, AB' requires higher energy than emission, BA' . At high temperatures, phonon levels combine with electronic transitions in both directions, $A''B''$.

the high reflectivity associated with the conductivity of the electron system (see § 8.6).

Another phenomenon which is essentially equivalent to an interband transition is the *emission of soft X-rays*. When an electron is removed from one of the deepest levels in an atom of the solid, X-rays are emitted by the transition of an electron down from various other levels. We are concerned with transitions from the valence bands of a metal—it is intuitively obvious that the spectrum of the emitted radiation will show a band of frequencies reflecting the band of states occupied by the conduction or valence electrons.

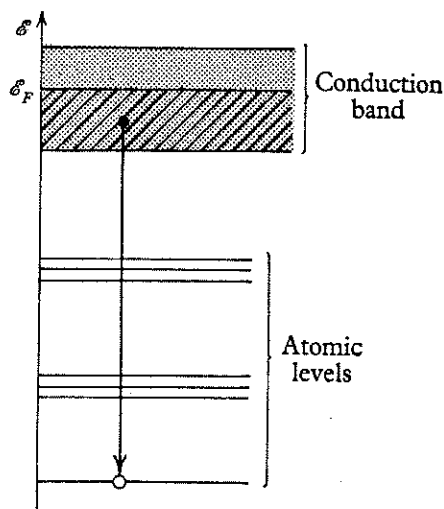


Fig. 146. X-ray emission.

The actual shape of this spectrum will depend on the product of the density of states in the band with the square of the appropriate matrix element for the transition. It is well known that the interaction between an electron and the electromagnetic field proceeds via the momentum operator for the electron (i.e. by the current produced by

its motion) so that one must study matrix elements like

$$\int \psi_{\mathbf{k}}^*(\mathbf{r}) e^{i\mathbf{K}\cdot\mathbf{r}} \text{grad } \phi_a(\mathbf{r}) \, d\mathbf{r}, \quad (8.76)$$

where ϕ_a is the atomic orbital from which the electron is missing and \mathbf{K} is the wave-vector of the X-rays. The wavelengths associated with such transitions are still long on the scale of atomic distances, so this wavy factor can be ignored—and, since $\phi_a(\mathbf{r})$ is a localized function there is no selection rule for momentum.

However, this matrix element may vary considerably through the band, according as the level ϕ_a is an s -state or a p -state, and according as $\psi_{\mathbf{k}}(\mathbf{r})$ is more or less s -like or p -like in the interior of the atom. There are also effects due to electron–electron interaction. The exact shape of the emission spectrum is not, therefore, necessarily a direct measure of the density of states, although it should reflect some of the features of that function, especially the sharp cut-off at the Fermi level.

Optical or X-ray absorption by a crystal is accompanied by the excitation of electrons into higher levels. In a semiconductor or insulator these excited electrons are normally mobile, hence giving rise to *photoconductivity*. But carriers produced in this way are very susceptible to *trapping* by impurities and imperfections, as witnessed by a multitude of complex experimental phenomena.

Given sufficient energy, the electron may be *photo-emitted* through the crystal surface. The elementary Einstein formula for the *photoelectric effect* merely tells us that its energy outside the surface cannot exceed

$$\mathcal{E}_{\text{max}} = \hbar\omega - \phi_W \quad (8.77)$$

since it must have surmounted the work-function barrier, ϕ_W (§ 6.9). But photo-electrons of lower energy must have come from levels within the crystal below the Fermi energy. Direct measurements of (\mathcal{E}, ω) —the energy distribution of emitted electrons for given photon frequency—should give a great deal of information about the probabilities of transitions from occupied to empty band states, similar to those responsible for the absorption of light.

To interpret these observations, we naturally construct an energy level diagram (Fig. 147) and count vertical or *direct* transitions in the manner of Fig. 143. In principle, with control over both \mathcal{E} and ω we should thus obtain somewhat more information about the band structure than is given by, say, the interband spectral density of (8.73). In practice, the photo-emission spectrum for many crystalline solids does not conform to this pattern, but seems to behave like

$$n(\mathcal{E}, \omega) \sim \rho_f(\mathcal{E}) \rho_i(\mathcal{E} - \hbar\omega), \quad (8.78)$$

as if 'non-direct' transitions were allowed indiscriminately between initial states of energy density $\rho_i(\mathcal{E} - \hbar\omega)$ and final states of energy density $\rho_f(\mathcal{E})$. The experiments are technically difficult, but unequivocal: it is not yet clear whether this discrepancy with simple theory is due simply to the breakdown of crystal momentum selection rules near the emitting surface as in L.E.E.D. (§ 6.9), or whether many-electron effects are responsible.

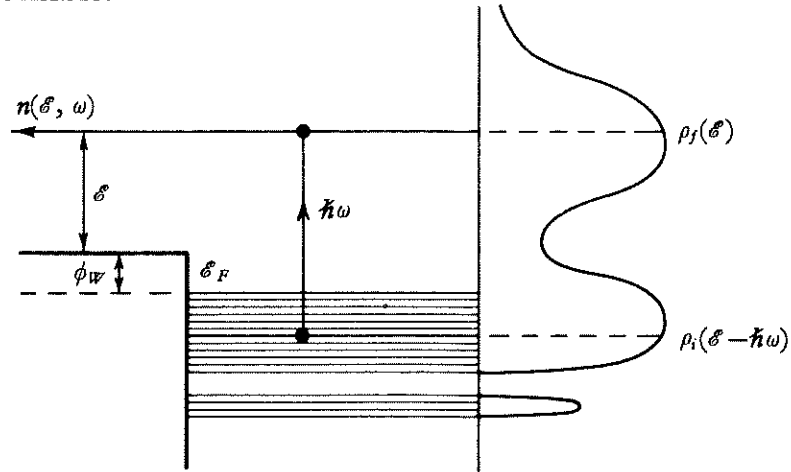


Fig. 147. Photo-emission.

8.6 Interaction with conduction electrons

What happens to the optical properties when the solid is a relatively good conductor? Suppose, for example, that we ignore ϵ in the formula (8.7) for the complex refractive index—an assumption that is valid for metals.

Then we have

$$N^2 = \frac{4\pi\sigma}{\omega} i, \quad (8.79)$$

and the real and imaginary parts are equal in magnitude;

$$n + ik = \left(\frac{2\pi\sigma}{\omega}\right)^{\frac{1}{2}} (1 + i). \quad (8.80)$$

The most obvious consequence of this is that the reflecting power of the solid becomes very high. From (8.19),

$$R \approx 1 - 2\left(\frac{\omega}{2\pi\sigma}\right)^{\frac{1}{2}}, \quad (8.81)$$

which is known as the *Hagen-Rubens relation*. The deviation from perfect reflectivity is proportional to

$$\left(\frac{\omega}{2\pi\sigma}\right)^{\frac{1}{2}} \sim \left(\frac{2\omega}{\omega_p^2 \tau}\right)^{\frac{1}{2}}, \quad (8.82)$$

where τ is the relaxation time in the classical formula (7.33) for the conductivity, and ω_p is the ubiquitous plasma frequency of the electron gas. We know that this ratio will be quite small even when ω approaches infra-red frequencies.

But this elementary macroscopic theory assumes that σ is independent of the frequency. This will not be true when the field oscillates so rapidly that the electrons are not given time to make collisions, i.e. when

$$\omega\tau > 1. \quad (8.83)$$

To investigate this region we must return to our transport equation (7.14).

For completeness, and later reference, let us write down the Boltzmann equation for a distribution that may vary in space and time:

$$e\mathbf{E} \cdot \mathbf{v}_k \left(-\frac{\partial f^0}{\partial \mathcal{E}} \right) = \frac{g_k}{\tau} + \mathbf{v}_k \cdot \frac{\partial g_k}{\partial \mathbf{r}} + \frac{\partial g_k}{\partial t}. \quad (8.84)$$

We assume a relaxation time as in (7.17). The variation of the 'out of balance' distribution g_k with time may be included explicitly; as we see from (7.7), this is what is left over when all the other contributions to $\partial f_k / \partial t$ have been accounted for.

Now suppose

$$\mathbf{E} = \mathbf{E}_0 \exp \{i(\mathbf{K} \cdot \mathbf{r} - \omega t)\}, \quad (8.85)$$

as in (8.3), and that g_k varies, in space and time, in the same manner, by 'sympathy'. Thus, let us write

$$g_k = \left(-\frac{\partial f^0}{\partial \mathcal{E}} \right) \Phi(\mathbf{k}) \exp \{i(\mathbf{K} \cdot \mathbf{r} - \omega t)\}. \quad (8.86)$$

Substituting in (8.84), we have

$$e\mathbf{E}_0 \cdot \mathbf{v}_k = \frac{\Phi(\mathbf{k})}{\tau} + i\mathbf{K} \cdot \mathbf{v}_k \Phi(\mathbf{k}) - i\omega \Phi(\mathbf{k}), \quad (8.87)$$

which has the solution

$$\Phi(\mathbf{k}) = \frac{e\tau \mathbf{v}_k \cdot \mathbf{E}_0}{1 - i\omega\tau + i\tau \mathbf{K} \cdot \mathbf{v}_k}. \quad (8.88)$$

The linearity of our Boltzmann equation makes this elementary solution possible.

For the conductivity we have, as in (7.20)–(7.24),

$$\begin{aligned} \boldsymbol{\sigma} \cdot \mathbf{E}_0 &= \frac{e}{4\pi^3} \int \Phi(\mathbf{k}) \mathbf{v}_k \frac{dS_F}{\hbar v_k} \\ &= \frac{e^2}{4\pi^3} \int \frac{\tau \mathbf{v}_k \mathbf{v}_k}{1 - i\tau(\omega - \mathbf{K} \cdot \mathbf{v}_k)} \frac{dS_F}{\hbar v_k} \cdot \mathbf{E}_0. \end{aligned} \quad (8.89)$$

Thus the conductivity itself has real and imaginary parts; the real part of σ will contribute to the imaginary part of N^2 , whilst the imaginary part of σ will look like part of the real dielectric constant.

For an ordinary electromagnetic wave, in which the phase velocity is greater than the Fermi velocity of an electron, we can drop the term $\mathbf{K} \cdot \mathbf{v}_k$, which is much smaller than ω . For formal simplicity assuming that the metal has cubic symmetry, we get from (8.89)

$$\begin{aligned}\sigma(\omega) &= \frac{e^2}{12\pi^3} \int \frac{\tau v(1+i\omega\tau)}{(1+\omega^2\tau^2)} dS_F \\ &= \sigma(0) \frac{1+i\omega\tau}{1+\omega^2\tau^2},\end{aligned}\tag{8.90}$$

where $\sigma(0)$ is the ordinary static conductivity of the metal.

This formula should be put into (8.7) and (8.19). If we take $\epsilon = 1$, i.e. if we ignore any internal polarizability of the ions, then we have, from (8.7), (8.8), (7.33), and (5.53), for the real and imaginary parts of N^2

$$\begin{aligned}n^2 - k^2 &= 1 - \frac{4\pi\sigma(0)\omega\tau}{\omega(1+\omega^2\tau^2)} \\ &= 1 - \frac{\omega_p^2\tau^2}{1+\omega^2\tau^2}\end{aligned}\tag{8.91}$$

and

$$\begin{aligned}2nk &= \frac{4\pi\sigma(0)}{\omega(1+\omega^2\tau^2)} \\ &= \frac{\omega_p^2\tau}{\omega(1+\omega^2\tau^2)},\end{aligned}\tag{8.92}$$

where, once more, the plasma frequency of the electron gas plays a most important role.

There are three different frequency regions covered by this *Drude theory*.

(i) $\omega \ll 1/\tau$. This is the ordinary low-frequency region. The imaginary part of N^2 is much larger than the real part, so that the metal is strongly reflecting, and the *Hagen-Rubens* relation (8.81) is valid. The absorption coefficient is more or less independent of ω , and is proportional to the conductivity. The real part of N^2 is negative, and much larger than unity in magnitude.

(ii) $1/\tau \ll \omega \ll \omega_p$. This is the *relaxation region* where $\omega^2\tau^2$ takes over in the denominators of (8.91) and (8.92). The absorption coefficient falls rapidly, proportionately to $1/\omega^2$ —and, strangely, is

inversely proportional to the conductivity. The imaginary part of N^2 becomes less than the real part—but this is still large and negative, being now of the form

$$n^2 - k^2 = 1 - \frac{\omega_p^2}{\omega^2} \quad (8.93)$$

as in (5.67) and (8.45). Thus, the metal is still strongly reflecting, with

$$R \approx 1 - \frac{1}{\sqrt{\pi\sigma_0\tau}} \approx 1 - \frac{2}{\omega_p\tau}. \quad (8.94)$$

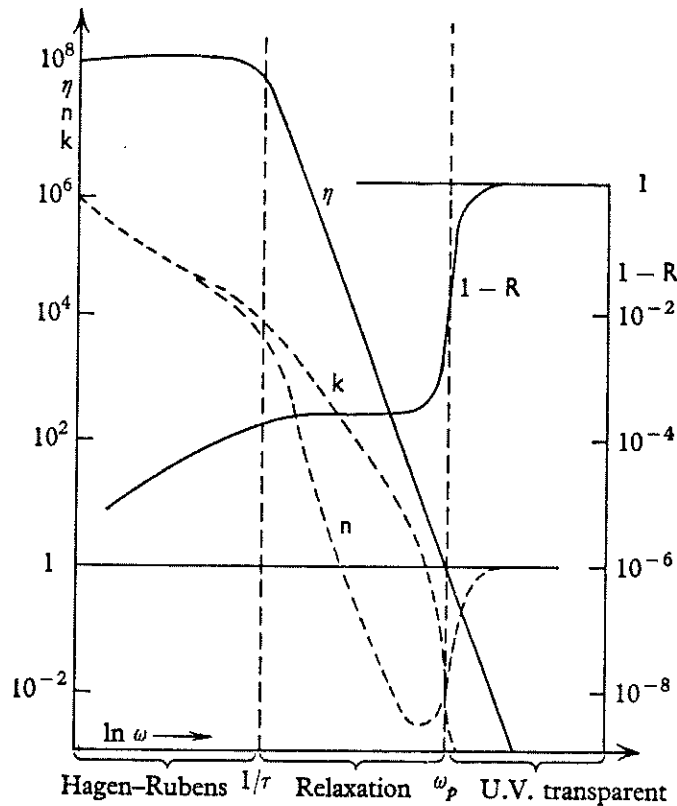


Fig. 148. Schematic behaviour of optical properties of metals, showing real and imaginary parts of dielectric constant, reflection coefficient and absorption coefficient. Note logarithmic scale of frequency.

(iii) $\omega_p \ll \omega$. The real part of N^2 becomes positive, and the reflecting power falls to zero. The metal should now appear more or less transparent, with an absorption coefficient

$$\frac{2\omega n(\omega) k(\omega)}{c} \approx \frac{\omega_p^2}{\omega^2 \tau c}. \quad (8.95)$$

The formulae (8.91) and (8.92) imply certain relations between n and k , as functions of ω , which are not, in fact, always obeyed. But

this can, to some extent, be understood if one supposes that the relaxation time, τ , in the integral (8.89), is not a constant, but varies over the Fermi surface. Thus, we can, in principle, distinguish three different integrals

$$\int \tau v dS_F, \quad \int \tau^2 v dS_F, \quad \int \frac{v}{\tau} dS_F, \quad (8.96)$$

corresponding to three different types of average of $\tau(\mathbf{k})$ over the Fermi surface. There is no reason why these different averages should all be equal.

We also notice that the real part of N^2 at high frequencies depends upon the integral

$$\int v dS_F, \quad (8.97)$$

which is, for free electrons, proportional to the number of electrons and inversely proportional to their mass. In the density of states, on the other hand, there appears a different type of average of the velocity—a harmonic mean

$$\int \frac{dS_F}{v} \quad (8.98)$$

as in (4.6). The difference between these two types of average can give information about the anisotropy of the electron velocity over the Fermi surface. It is not wise to 'correct' observed values of (8.97) by a factor associated with an 'optical mass', which then turns out to be very different from the 'thermal mass' that would be required to make (8.98) fit the free-electron formula for the electronic specific heat. The implicit assumption that all Fermi surfaces are spheres often turns out to be unjustified even for monovalent metals.

8.7 The anomalous skin effect

The Hagen-Rubens relation (8.81) may break down in another way. The absorption coefficient given by (8.79) is very high and the electromagnetic wave is very rapidly attenuated as it enters the metal. The damping distance, as derived from (8.10) and (8.80), is of the order

$$\begin{aligned} \delta &= \frac{c}{k\omega} \\ &= \frac{c}{(2\pi\sigma\omega)^{\frac{1}{2}}}. \end{aligned} \quad (8.99)$$

This phenomenon is well known, classically, as the *skin effect*. We refer to δ , as given by (8.99), as the *classical skin depth*, δ_{cl} .

At high frequencies this may become very small. If we have a metal of great purity at low temperatures, we find that the skin depth can

become much less than the electron mean free path, Λ . The ordinary theory of the electrical conductivity is no longer valid; the effective field acting on a carrier is varying rapidly over the distance the carrier moves between collisions.

To deal with this case we must try to solve the Boltzmann equation with $g_{\mathbf{k}}$ varying in space. If we ignore the time variation of \mathbf{E} and $g_{\mathbf{k}}$ (and this phenomenon can easily be observed at frequencies well below the onset of the relaxation régime discussed above), then we have, as in (8.84),

$$e\mathbf{E}(\mathbf{r}) \cdot \mathbf{v}_{\mathbf{k}} \left(-\frac{\partial f^0}{\partial \mathcal{E}} \right) = \frac{g_{\mathbf{k}}}{\tau} + \mathbf{v}_{\mathbf{k}} \cdot \frac{\partial g_{\mathbf{k}}}{\partial \mathbf{r}} \quad (8.100)$$

as the differential equation to be solved for $g_{\mathbf{k}}$.

The most direct procedure for the solution of (8.100)—a procedure that is applicable to a variety of problems involving surfaces, thin films, wires, etc.—is to use the *Chambers formula*,

$$g(\mathbf{v}, \mathbf{r}) = \frac{e}{v} \left(-\frac{\partial f^0}{\partial \mathcal{E}} \right) \int^{\mathbf{r}} \mathbf{v} \cdot \mathbf{E}(\mathbf{r}') e^{-s'/\tau v} ds'. \quad (8.101)$$

It is not very difficult to prove, quite formally, that this is a solution of the differential equation; it is simply a generalization of the usual formula for the solution of an equation of the form $dy/dx + \alpha y = f(x)$.

What it means, physically, is that the electron current at the point \mathbf{r} , in the direction \mathbf{v} , depends upon the previous history of the electrons that contribute to that current. The integral is not over all space, but simply over the distance s' back along the trajectory passing through \mathbf{r} in the direction \mathbf{v} . Thus, at the point \mathbf{r}' the electrons are accelerated by the force $e\mathbf{E}(\mathbf{r}')$. But this contribution decays with time, because of relaxation processes; the exponential factor measures this effect as a function of distance along the trajectory. This semi-classical formula is closely related in principle to the Kubo formula (7.15) where we similarly integrate over the previous history of the system.

From (8.101), one can calculate the total electric current by the usual formula (7.20). But we notice that this current will be a function of \mathbf{r} —and also of the boundary conditions on the function $g(\mathbf{v}, \mathbf{r})$. Thus, there is an integral equation defining $\mathbf{J}(\mathbf{r})$ as a functional of the distribution of electric field $\mathbf{E}(\mathbf{r})$. On the other hand, these two quantities are linked by Maxwell's equations in a form more general than (8.2), i.e.

$$\begin{aligned} \nabla^2 \mathbf{E} &= \frac{4\pi}{c^2} \frac{d\mathbf{J}}{dt} \\ &= \frac{4\pi i\omega}{c^2} \mathbf{J}(\mathbf{r}) \end{aligned} \quad (8.102)$$

(ignoring the ordinary displacement current). Thus, one can solve the integro-differential equation, derived from (8.101) and (8.102), and obtain the distribution of $\mathbf{J}(\mathbf{r})$ and $\mathbf{E}(\mathbf{r})$ in the neighbourhood of the surface.

The exact solution of this equation is rather elaborate, and depends upon the boundary conditions—it depends on whether the electrons are specularly reflected at the surface, or whether they are randomly scattered there. The actual formulae are not so important as their physical interpretation. The fact is that not all the electrons are participating in the absorption and reflection of the electromagnetic wave. Only those that are running inside the skin depth for most of a mean free path Λ are capable of picking up much energy from the

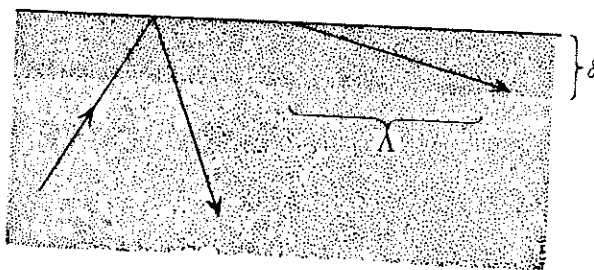


Fig. 149. Only electrons in the skin depth are 'effective'.

electric field. If, for example, the skin depth is δ' , then one may suppose that only a fraction δ'/Λ of the electrons are *effective* in the conductivity. We may write

$$\sigma' = \frac{3}{2}\beta \frac{\delta'}{\Lambda} \sigma_0 \quad (8.103)$$

for the apparent conductivity—the number β being just a 'fudge factor'.

But then, a surface with this conductivity would, according to (8.99), itself have a skin depth

$$\begin{aligned} \delta' &= \frac{c}{(2\pi\sigma'\omega)^{\frac{1}{2}}} \\ &= \frac{c}{\left(2\pi \frac{3}{2}\beta \frac{\delta'}{\Lambda} \sigma_0 \omega\right)^{\frac{1}{2}}}, \end{aligned} \quad (8.104)$$

which can be solved for δ' , and for σ' . Thus

$$\sigma' = \left(\frac{9\beta^2}{8\pi}\right)^{\frac{1}{2}} \frac{1}{\omega^{\frac{1}{2}}} \left(\frac{c\sigma_0}{\Lambda}\right)^{\frac{3}{2}}; \quad (8.105)$$

the effective conductivity will appear to behave as $\omega^{-\frac{1}{2}}$, but will be independent of the electron mean free path, which cancels with the ordinary static conductivity σ_0 . We are then in the conditions of the *anomalous skin effect*. What one does, in practice, is to make the surface of the metal part of a resonant cavity, and then observe the effect upon the resonance properties of the cavity. For examples, one measures the *surface impedance*, which is just

$$Z(\omega) = -4\pi i\omega \left. \frac{E(z)}{(\partial E/\partial z)} \right]_{z=0} = \frac{4\pi c}{N} \quad (8.106)$$

in the notation of § 8.1.

There are various other changes in the theory of the apparent optical properties of metals in the anomalous limit which can be worked out, especially when one gets into the relaxation region of frequencies. For example, there is an additional absorption of power from a reflected wave due to the actual scattering of electrons from the surface, if this is 'rough'—a contribution that is as large as the effects calculated in the previous section where such scattering was ignored.

The most interesting property of the anomalous skin effect is, however, its dependence on the geometry of the Fermi surface. As we saw above, only the electrons travelling nearly parallel to the surface are 'effective' in the conductivity. If our specimen is a single crystal, then those electrons derive from a narrow belt running round the Fermi surface. If the Fermi surface is highly anisotropic, we should see different values of the surface impedance according as we measure on different cuts of the metal crystal, and at different orientations of the field.

The geometrical property that is being measured can be derived by an extension of the *ineffectiveness concept*, (8.103). We assume that we need only consider electrons whose velocity vectors lie within an angle $\pm\beta\delta'/\Lambda$ of parallelism with the (x, y) -plane, which defines the metal boundary (we take the electric field in the x -direction).

At some general point, P , on the belt of effective electrons, the velocity \mathbf{v} makes an angle θ with \mathbf{E} . The width of the belt there will be $2\beta\delta' |\rho|/\Lambda$, where $|\rho|$ is the radius of curvature in the plane of \mathbf{v} and of the z -axis. An element of the circumference of the belt, of length ds , will thus correspond to an area of Fermi surface

$$dS = \frac{2\beta\delta' |\rho|}{\Lambda} ds, \quad (8.107)$$

and will evidently contribute to the conductivity in the x -direction, by the usual formula (7.23),

$$\begin{aligned} d\sigma'_{xx} &= \frac{e^2}{4\pi^3\hbar} \tau v_x dS_x \\ &= \frac{e^2}{4\pi^3\hbar} \Lambda \cos\theta \frac{2\beta\delta' |\rho|}{\Lambda} \cos\theta ds. \end{aligned} \quad (8.108)$$

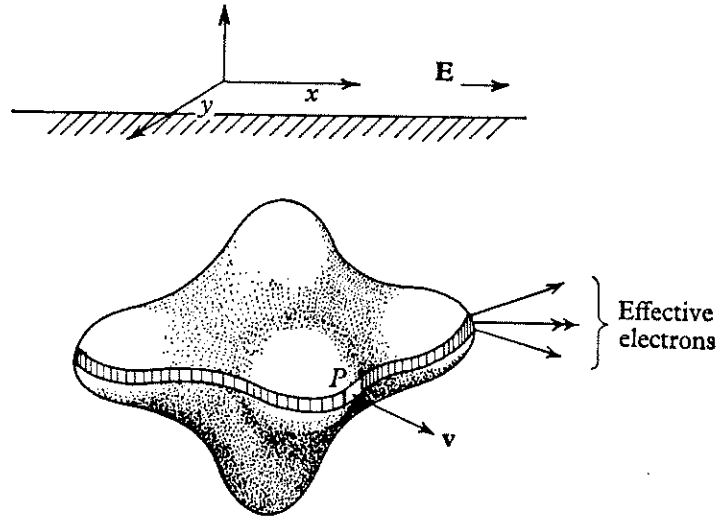


Fig. 150. Belt of effective electrons on a Fermi surface.

Thus, the total apparent conductivity of the metal will be

$$\begin{aligned} \sigma' &= \frac{e^2\beta\delta'}{2\pi^3\hbar} \int |\rho| \cos^2\theta ds \\ &= \frac{e^2\beta\delta'}{2\pi^3\hbar} \oint |\rho_y| dk_y, \end{aligned} \quad (8.109)$$

where ρ_y means the radius of curvature, on the belt, of a section of the Fermi surface parallel to the (x, z) -plane, and we integrate over the range of the momentum component k_y around the belt. This is the quantity we put into (8.104), and then solve for δ' and σ' separately. A more complete analysis shows that

$$\beta = \frac{8\pi}{3\sqrt{3}} \quad (8.110)$$

is the proper value for this arbitrary 'fudge factor'.

The beauty of this result is that the surface impedance does not depend upon the mean free path, or velocity, of the electrons, but only on the local curvature integrated round a well-defined belt. The

integral (8.109) is especially sensitive to regions of large radius of curvature—that is, to relatively flat portions of the Fermi surface. It thus provides information that is to some extent complementary to other ‘effects’, which often depend on small details and little bits of disconnected Fermi surface. It is not, however, easy to invert the relation between a given surface and the values of a set of integrals like (8.109) for belts drawn in different orientations; a tedious procedure of trial and error is usually necessary.

8.8 Ultrasonic attenuation

Although not strictly an ‘optical’ property, the theory of the attenuation of high-frequency elastic waves by interaction with the conduction electrons in a metal comes most naturally in this chapter. A sound wave in a solid gives rise to electric fields which accelerate electrons in much the same way as an electromagnetic wave. But the velocity of sound is so much less than the velocity of light—less even than the Fermi velocity of the electrons—that certain special effects are observed.

In the ordinary low-frequency region, the attenuation can be calculated by appeal to elementary kinetic theory. A gas of electrons, of mass m , number density n , average velocity v_F and mean free path Λ would have a viscosity

$$\eta = \frac{1}{3}nm\Lambda v_F. \quad (8.111)$$

This viscosity can be included as a term in the ordinary classical equation for the forces acting on an element of the medium. The attenuation constant for longitudinal elastic waves of frequency ω comes out as

$$\alpha = \frac{4}{5} \frac{\omega^2}{Ds^3} \eta, \quad (8.112)$$

where D is the mass density and s is the velocity of sound. These may be put together into various forms, for example

$$\alpha = \frac{4}{15} \frac{\omega^2 m v_F^2}{e^2 D s^3} \sigma, \quad (8.113)$$

in terms of the conductivity (7.33).

If the metal is very pure, and the measurement is made at low temperatures, the relaxation time of the electrons is greatly increased. Yet it is still difficult, with ultrasonic waves, to get into the relaxation region where $\omega\tau > 1$. However, because the velocity of sound is so

much less than the Fermi velocity, it is possible to get the mean free path longer than the wavelength of the sound, i.e. to have

$$q\Lambda > 1. \quad (8.114)$$

To discuss this case, we return to (8.89), where we constructed a formula for the solution of the Boltzmann equation in an electric field varying sinusoidally in space and time. In effect, we found a generalized conductivity

$$\sigma(\mathbf{q}, \omega) = \frac{e^2}{4\pi^3} \int \frac{\tau \mathbf{v} \mathbf{v}}{1 - i\tau(\omega - \mathbf{q} \cdot \mathbf{v})} \frac{dS_F}{\hbar v}. \quad (8.115)$$

When discussing this formula in § 8.6, we took $q = 0$, on the grounds that the velocity of an electromagnetic wave is so large that its wave-number may be taken to be nearly zero. But when the electric field is

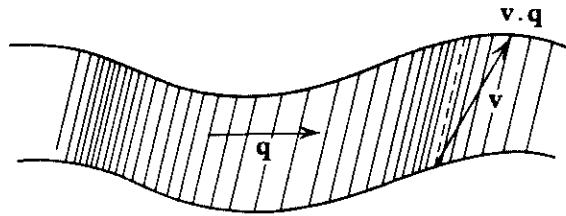


Fig. 151. Surf-riding resonance.

generated by an acoustic wave, which travels so much more slowly than the electrons, we must retain this term. The real part of the conductivity then comes predominantly from regions of the Fermi surface where $\omega \sim \mathbf{q} \cdot \mathbf{v}$. This is another way of saying that the component of the electron velocity in the direction of propagation of the wave is equal to the velocity of sound.

This is known as the *surf-riding resonance*. An electron that happens to be on the crest of the wave can continue to draw energy from the field by travelling in a direction nearly normal to the propagation vector of the wave. It is evident that only a few electrons on the Fermi surface can satisfy this condition. If we neglect ω completely, it defines a belt round the Fermi surface where the electron velocity is normal to \mathbf{q} —a belt of exactly the same form as the one defined in the theory of the anomalous skin effect (§ 8.7).

It is not difficult to evaluate (8.115) in the limit $q\Lambda \gg 1$. The effective width of the belt depends upon the rate at which $\tau \mathbf{q} \cdot \mathbf{v}$ increases as we go away from the exact resonance position. The larger the value of τ , the narrower does it become. Just as in (8.107), the element of 'effective' area is proportional to the local radius of curva-

ture of the surface, and inversely proportional to Λ . For example, one can prove by elementary differential geometry the exact formula for the limit $q\Lambda \gg 1$

$$\sigma_{xx}(\mathbf{q}, 0) = \frac{e^2}{4\pi^3\hbar q} \oint |\rho_y| dk_y, \quad (8.116)$$

where the directions x and y are normal to the direction of propagation, \mathbf{q} , and the other symbols are exactly as in (8.109).

The interesting feature of this formula is that it is independent of Λ . It is essential that there should be some mechanism by which the electrons are finally scattered, but in this limit the strength of the

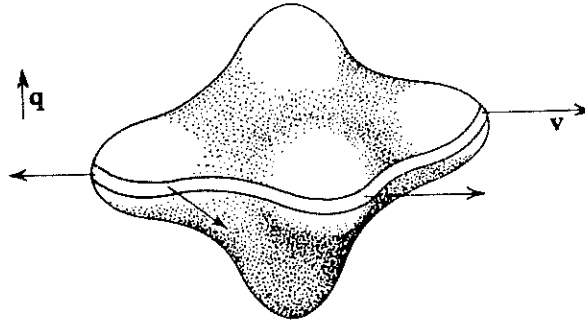


Fig. 152. Belt of electrons interacting with phonon \mathbf{q} .

scattering does not matter. This is the situation we have already encountered in § 8.7. Indeed, one can use (8.116) as the starting point for a complete derivation of the formula (8.109) for the anomalous skin effect, including the value of the fudge factor β . It is a matter, essentially, of expressing the relation (8.102), derived from Maxwell's equations, in Fourier components, and also putting in boundary conditions that allow for the behaviour of the electrons at the metal surface.

It would be very agreeable if we could use (8.116) by itself in a theory of ultrasonic attenuation. The experiment of measuring the attenuation of waves propagating in various directions through a single crystal is obviously much easier, technically, than cutting crystal surfaces in various orientations, and keeping them clean and perfect whilst measuring the surface impedance.

Unfortunately, this will not give direct information about the shape of the Fermi surface. The attenuation does, indeed, tend to a value independent of the mean free path of the electrons—(8.113) is no longer valid—but the actual magnitude depends upon the details of the coupling of the electrons to the lattice wave. In other words, the electric field seen by an electron as a result of lattice displacements or lattice strains cannot be calculated directly, and varies over the Fermi

surface. One may think of having to insert, in the integrand of (8.116) a *deformation potential* of the sort discussed in § 6.14—but much more complicated, much less easily expressed by a few parameters, on a multiply connected Fermi surface, say, than in a simple free-electron metal.

Ultrasonic attenuation by free carriers can also occur in a semiconductor. An ionic material without a centre of symmetry—e.g. a III-V or II-VI compound (§ 4.2)—is usually *piezo-electric*. A local strain $W_{ij}(\mathbf{r})$ gives rise to a local electric field

$$\delta E_k(\mathbf{r}) = \sum_{ij} \beta_{ijk} W_{ij}(\mathbf{r}) \quad (8.117)$$

in addition to the deformation potential (6.95). This travelling wave of electric field carried by a macroscopic sound wave affects the carriers, and is the dominant mechanism of interaction until one reaches relatively high microwave frequencies. The details of the phenomenon would then depend upon the components of the piezo-electric tensor β_{ijk} projected on the polarization vector of the sound wave.

In a medium of relatively low conductivity, the *acousto-electric field* produced by the beam of ultrasonic phonons transferring momentum to the carriers may be directly observed. This is obviously the inverse of the phonon drag effect (§ 7.11). A closely related phenomenon is *acousto-electric amplification* of sound by the application of a D.C. electric field along the ultrasonic beam. This may be regarded as a form of maser action: energy from the electric field unbalances the carrier distribution, which is then stimulated into the coherent emission of phonons. We may note the analogy with Cerenkov radiation: the electric field gives the carrier a drift velocity (7.31) that exceeds the velocity of sound, hence permitting the direct creation of phonons with conservation of energy and momentum.

We have based the above derivation upon a quasi-classical treatment of the Boltzmann equation. But the same results may be obtained, some people might say more rigorously, from quantum theory. For example, consider the energy condition (2.101) for the absorption of a phonon of frequency ω in the transition from state \mathbf{k} to state $\mathbf{k} + \mathbf{q}$. This gives us

$$\begin{aligned} \hbar\omega &= \mathcal{E}(\mathbf{k} + \mathbf{q}) - \mathcal{E}(\mathbf{k}) \\ &\approx \mathbf{q} \cdot \frac{\partial \mathcal{E}(\mathbf{k})}{\partial \mathbf{k}} \\ &= \hbar \mathbf{q} \cdot \mathbf{v}_{\mathbf{k}}, \end{aligned} \quad (8.118)$$

essentially the same as (8.115). In other words, the 'surf-riding resonance' is no more than an elementary process of the electron-phonon interaction, as discussed in §§ 2.8, 6.13.

Again (8.115) can be derived from quite a different source. Let us write the denominator in the form

$$1 - i\tau(\omega - \mathbf{q} \cdot \mathbf{v}) = \frac{i\tau}{\hbar} \{ \mathcal{E}(\mathbf{k} + \mathbf{q}) - \mathcal{E}(\mathbf{k}) - \hbar\omega - i\hbar\alpha \}, \quad (8.119)$$

where $\alpha = -1/\tau$. We recognize the denominator of the sum occurring in (5.16) and elsewhere in Chapter 5. With a little juggling, and use of (8.118), we find that we may write

$$\epsilon(\mathbf{q}, \omega) \approx 1 + \frac{4\pi i \sigma_l(\mathbf{q}, \omega)}{\omega} \quad (8.120)$$

when q and ω are small. Here σ_l means the longitudinal conductivity—the component of $\boldsymbol{\sigma}$ in the direction of propagation of the phonon.

In other words, we have rederived (8.7). The complex dielectric constant of the electron gas, as calculated in § 5.1, already contains the conductivity as its imaginary part. In (5.1) we took $-\alpha$ to be an arbitrary decay constant for the perturbing field—but the identification with $1/\tau$ is obviously appropriate. There is an extraordinary amount of physics packed up in (5.16).