AD No. 25 862
ASTIA FILE COPY

ELECTRONIC STRUCTURE OF SOLIDS II: THE PERTURBED PERIODIC LATTICE

JOHN C. SLATER

TECHNICAL REPORT NO. 5

DECEMBER 15, 1953

SOLID-STATE AND MOLECULAR THEORY GROUP
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

Technical Report No. 5

ELECTRONIC STRUCTURE OF SOLIDS II: THE PERTURBED PERIODIC LATTICE

John C. Slater

SOLID-STATE AND MOLECULAR THEORY GROUP
Massachusetts Institute of Technology
Cambridge, Massachusetts

O. N. R. Contract N5ori-07856

December 15, 1953

CONTENTS

Preface - iii

THE PERTURBED PERIODIC LATTICE

Introduction - 1
1. The Wannier Function Method for Perturbed Lattices - - - - - - - - - - - - - 1
2. The One-Dimensional Case - 7
3. The Two- and Three-Dimensional Cases - - - - - - - - - - - - - - - - - - - 20
4. Wannier Functions for Overlapping Bands - - - - - - - - - - - - - - - - - - 35
5. Impurity Levels with the Tight Binding Approximation - - - - - - - - - - - 42
6. General Method for Solving the Difference Equations - - - - - - - - - - - - 45
7. Additional Types of Perturbative Potentials - - - - - - - - - - - - - - - - 54
8. Periodic Perturbative Potentials - 58
9. A Constant Accelerating Field - 65
Bibliography - 69

PREFACE

This Technical Report forms a continuation of the series started with Technical Report No. 3, Electronic Structure of Atoms and Molecules, and Technical Report No. 4, Electronic Structure of Solids I: The Energy Band Method. As was mentioned in the prefaces of those reports, it forms part of a series of advanced lectures on the theory of molecules and solids, in which I am endeavoring to cover the main features of the electronic structure of matter. It has been decided that it is more practical to issue this material a chapter at a time; hence there are still further reports to come in the series, dealing with configuration interaction in solids, magnetic problems, and various related topics which depend on the many-electron aspect of the structure of solids. This report and the preceding one, in contrast, deal with the one-electron problem.

It will be recognized that a great deal of the material in the present report is new, particularly the methods for handling perturbation problems, in terms of Wannier functions, using the technique of difference equations. This material is also described very briefly in the Quarterly Progress Reports of October 15, 1953, and January 15, 1954. As soon as possible, it will be written up for publication, so that it will be convenient to make reference to it.

John C. Slater

Cambridge, Mass.
December, 1953

CHAPTER 2

THE PERTURBED PERIODIC LATTICE

In Chapter 1* we have treated the problem of an electron moving in a periodic potential such as we find in a crystal, when we solve the problem by the method of self-consistent field. Later we shall go on to the problem of setting up determinantal wave functions made from such one-electron orbitals, and of the configuration interaction between such determinantal wave functions. Before we do so, however, we shall take up the one-electron problem in a lattice which is no longer perfectly periodic, but has suffered some sort of perturbation. This perturbation may be in the form of an impurity atom which is substituted for one of the atoms of the lattice; it may be displacements of atomic positions, either on account of dislocations or other lattice defects, or regular displacements such as would be produced by a sound wave traversing the crystal. Other perturbations are present at the boundaries of real crystals, where we may not use the periodic boundary conditions so useful in the interior of a crystal. All of these questions are of obvious practical importance, in such problems as the impurity atoms in semiconductors, the behavior of alloys, the effect of imperfections or thermal agitation on the scattering of electrons, surface states on the surface of a crystal, the effect of missing or interstitial atoms, and many other problems. The method used for attacking all these types of problems can be worked out in a general way, and we shall do so in this chapter, later extending it to specific applications.

1. The Wannier Function Method for Perturbed Lattices

The most generally useful method for handling the perturbed lattice depends on the use of Wannier functions. We assume that we start with an unperturbed periodic problem, with Hamiltonian H_o, consisting of the kinetic energy operator, and a periodic potential, and that we have solved this problem exactly, so that we know its energy levels and wave functions as functions of the propagation constant $\vec{k}$. Our problem has a Hamiltonian H, equal to $H_o + H_1$, where H_1 is a perturbative potential. In the case of an impurity atom, for instance, H_1 would represent the difference between the potential around the impurity atom and the periodic potential which we should see if the impurity atom were replaced by one of the regular atoms of the lattice, so that it would be vanishingly small except in the immediate neighborhood of the impurity atom. We shall now expand the solutions of the perturbed problem in terms of the Wannier functions of the unperturbed problem. The reason why the Wannier functions are so convenient for this purpose is that they are rather localized, so that only those located near the impurity atom are appreciably affected by the perturbation.

We saw in Section 3, Chapter 1, that corresponding to each energy band we can

---

*References to Chapter 1 refer to Technical Report No. 4, Solid-State and Molecular Theory Group, M. I. T., entitled Electronic Theory of Solids: The Energy Band Method (Chapter 1), July 15, 1953.

set up Wannier functions, one centered on each unit cell, such that the Wannier functions associated with different unit cells are orthogonal. It is also easy to prove, though we did not do it, that the Wannier functions associated with different energy bands are orthogonal. If $a_n(\vec{r} - \vec{R}_j)$ represents the Wannier function corresponding to the nth band, localized in the unit cell whose vector position is $\vec{R}_j$, then as in Eq. (1.32) of Chapter 1 we can show that

$$H_o a_n(\vec{r} - \vec{R}_j) = \sum(\vec{R}_i)\ \mathscr{E}_n(\vec{R}_i - \vec{R}_j)\ a_n(\vec{r} - \vec{R}_i). \tag{2.1}$$

That is, $\mathscr{E}_n(\vec{R}_i - \vec{R}_j)$ is the matrix component of H_o between the Wannier functions at $\vec{R}_j$ and $\vec{R}_i$. We can express this otherwise by multiplying Eq. (2.1) by $a_n^*(\vec{r} - \vec{R}_i)$, and integrating over the periodic volume of N unit cells, over which the a's are normalized and orthogonal. Then we have at once

$$\int a_n^*(\vec{r} - \vec{R}_i)\ H_o a_n(\vec{r} - \vec{R}_j)\ dv = \mathscr{E}_n(\vec{R}_i - \vec{R}_j). \tag{2.2}$$

We see from Eq. (2.1) that H_o has no matrix components between Wannier functions associated with different energy bands.

We can now set up a Schrödinger equation for the unperturbed problem. We express the wave function in the form $\sum(n, j)\ U_n(\vec{R}_j)\ a_n(\vec{r} - \vec{R}_j)$, where the U's are coefficients. We know that we can make such an expansion, for we mentioned in Chapter 1 that the a's form a complete orthogonal set of functions, and any function can be expanded in terms of them. We set up the unperturbed Schrödinger equation $H_o u = Eu$, where u is the wave function. We substitute the sum of Wannier functions in this equation, multiply by the conjugate of one of the Wannier functions, and integrate over the volume. Using (2.1) or (2.2), we obtain

$$\sum(\vec{R}_j)\ \mathscr{E}_n(\vec{R}_i - \vec{R}_j)\ U_n(\vec{R}_j) = EU_n(\vec{R}_i). \tag{2.3}$$

This set of linear equations would, of course, lead to a secular equation of the ordinary sort for the energy E. We can solve it more conveniently, however, by assuming that $U_n(\vec{R}_i) = \exp(i\vec{k} \cdot \vec{R}_i)$. This assumption is equivalent to building up a Bloch sum of the Wannier functions in the nth band, just as in Eq. (1.28), Chapter 1. When we make this substitution in Eq. (2.3), we find that all our equations are simultaneously satisfied, provided

$$E = \sum(\vec{R}_s)\ \mathscr{E}_n(\vec{R}_s) \exp(-i\vec{k} \cdot \vec{R}_s). \tag{2.4}$$

Eq. (2.4) is identical with Eq. (1.33) of Chapter 1, though our method of derivation is a little different. Since the energy E of Eq. (2.4) depends on $\vec{k}$, and on the energy band we are dealing with, we should have a notation indicating these two quantities; we may denote the energy by the symbol $E_n(\vec{k})$.

We now ask what effect on these results is produced by the perturbative Hamiltonian H_1. Let the matrix component of H_1 between two Wannier functions, in the nth and m^{th} bands, and in the unit cells at $\vec{R}_i$ and $\vec{R}_j$, be

$$\int a_n^*(\vec{r} - \vec{R}_i)\, H_1 a_m(\vec{r} - \vec{R}_j)\, dv = V_{nm}(\vec{R}_i, \vec{R}_j).$$

Then in place of (2.3), Schrödinger's equation becomes

$$\sum(m, \vec{R}_j)\left[\mathscr{E}_n(\vec{R}_i - \vec{R}_j)\,\delta_{nm} + V_{nm}(\vec{R}_i, \vec{R}_j)\right] U_m(\vec{R}_j) = EU_n(\vec{R}_i). \qquad (2.5)$$

This set of equations provides an exact solution of our problem, if we can solve them. In some important cases, as we shall show later, an exact solution is possible. More often we can only approximate a solution. Our first aim will be to look into the qualitative nature of the problem, so as to understand the type of solution to be expected. Specially simple cases result when the components V_{nm} for $n \neq m$ and for $\vec{R}_i \neq \vec{R}_j$ can be neglected; that is, when the perturbations do not result in mixing of the Wannier functions of different bands, or different unit cells. Unfortunately, we shall see later that in most important cases this simplification does not hold, but there are some cases where it does. In case it does, Eq. (2.5) simplifies in that we can omit the summation over m, replacing m by n on the left side. We are dealing, then, only with the quantities $U_n(\vec{R}_i)$ relating to a single band.

The most important feature of our problem is the fact that $U_n(\vec{R}_i)$, though really a set of coefficients, nevertheless has much of the character of a continuous function $U_n(\vec{r})$, whose values at the discrete points $\vec{r} = \vec{R}_i$ give the $U_n(\vec{R}_i)$'s. For the unperturbed problem, where V is zero, we have seen that a solution of Schrödinger's equation is given by assuming that $U_n(\vec{R}_i) = \exp(i\vec{k} \cdot \vec{R}_i)$, which obviously means that our continuous function is the plane wave $\exp(i\vec{k} \cdot \vec{r})$, whose values for $\vec{r} = \vec{R}_i$ give the U's. But such a plane wave is the solution of a Schrödinger problem for a free particle. The remarkable fact that we find is that in many cases, in the presence of the perturbation, the continuous function $U_n(\vec{r})$ has the character of the solution of the Schrödinger equation for a particle in the presence only of the perturbative potential H_1. In other words, by using the Wannier functions, we can often take account almost automatically of the periodic potential and the band nature of the solutions, and have a problem almost as simple as a free electron in a perturbing field. Let us examine the extent to which this is true.

The possibility of making this identification arises when the function $U(\vec{r})$ (we shall drop the index n, for simplicity, since here we are dealing only with a single band) varies so slowly from one unit cell to the next that we can replace the differences in Eq. (2.5) by derivatives. Then we can convert Eq. (2.5) from a difference equation into a differential equation, which can often take a character similar to Schrödinger's equation for a free particle under the influence of the perturbative potential H_1. An elegant mathematical method for proving this was given by Wannier,[1] providing the main foundation for the treatment of perturbed lattices which goes by his name. The proof starts by considering the energy $E_n(\vec{k})$, or $E(\vec{k})$, of Eq. (2.4). The propagation vector $\vec{k}$, in this expression, can be related to a momentum-like quantity $\vec{p}$, by the equation $\vec{k} = 2\pi\vec{p}/h$. That is, $\vec{p}$ is the momentum of a particle whose de Broglie wave length would be such that the propagation constant would be $\vec{k}$. Sometimes $\vec{p}$ is called the pseudomomentum, since it is not the real momentum of the electron. We now take $E(\vec{k})$, express $\vec{k}$ in terms of $\vec{p}$, and then convert this into a differential operator, by applying to $\vec{p}$ the usual rule used in setting up a Schrödinger equation. That is, we replace $\vec{p}$ by $-i\hbar\nabla$, or replace $\vec{k}$ by $-i\nabla$. We then allow this to operate on the function $U(\vec{r})$, and shall find that the result of this operation helps us in solving Eq. (2.5).

We can set up this operator in either of two ways. In the first place, we can use Eq. (2.4), and replace $\vec{k}$ where it appears in the exponent by $-i\nabla$. We then can use a simple property of such an operator. Let us consider the quantity $\exp(-ik_x R_{sx})$. We can expand this by power series, to give $1 - ik_x R_{sx} - \frac{1}{2!} k_x^2 R_{sx}^2 \ldots$. If we replace k_x by $-i\partial/\partial x$, according to our prescription, and allow the result to operate on the function $f(x)$, the result is $f(x) - R_{sx}\,\partial f/\partial x + \frac{1}{2!} R_{sx}^2\,\partial^2 f/\partial x^2 \ldots = f(x - R_{sx})$ by Taylor's theorem. We thus find that $\exp(-i\vec{k}\cdot\vec{R}_s)\,U(\vec{r})$ becomes equivalent to $U(\vec{r} - \vec{R}_s)$, when the exponential is converted into an operator. Using (2.4), then, we see that the quantity $E(\vec{k})\,U(\vec{r})$ becomes converted into $\sum(\vec{R}_s)\ \mathscr{E}_n(\vec{R}_s)\,U(\vec{r} - \vec{R}_s)$, when $\vec{k}$ is converted into an operator. But we may now rewrite the first term of Eq. (2.5) in the form $\sum(\vec{R}_s)\ \mathscr{E}_n(\vec{R}_s)\,U(\vec{R}_i - \vec{R}_s)$, by letting $\vec{R}_s$ be equivalent to $\vec{R}_i - \vec{R}_j$. Thus we see that we can write the first term of (2.5) in terms of our operator.

Our simplification of the method holds only in case $V_{nm}(\vec{R}_i, \vec{R}_j)$ is a completely diagonal matrix: that is, not only do the components vanish if $n \neq m$, but also they vanish if $\vec{R}_i \neq \vec{R}_j$. If we assume that this is the case, then we can define a continuous perturbation function $V(\vec{r})$, whose value equals $V(\vec{R}_i, \vec{R}_i)$ when $\vec{r} = \vec{R}_i$. In that case, (2.5) can be rewritten in the form

$$E(\vec{k})\,U(\vec{r}) + V(\vec{r})\,U(\vec{r}) = EU(\vec{r}), \tag{2.6}$$

---

[1] G. H. Wannier, Phys. Rev. 52, 191 (1937); see also J. C. Slater, Phys. Rev. 76, 1592 (1949).

where $E(\vec{k})$ is to be interpreted as the operator which we have described. But now we can approach the operator $E(\vec{k})$ in a different way, by expanding E in power series, instead of in Fourier series as in Eq. (2.4). Let us expand about a point where E, regarded as a function of $\vec{k}$, has a minimum or maximum value. Around such a point we have

$$E(\vec{k}) = E|_o + \frac{1}{2!}\left[\left.\frac{\partial^2 E}{\partial k_x^2}\right|_o (k_x - k_{xo})^2 + 2\left.\frac{\partial^2 E}{\partial k_x \partial k_y}\right|_o (k_x - k_{xo})(k_y - k_{yo}) + \ldots\right] \tag{2.7}$$

+ terms in third and higher powers.

In making such an expansion, we are assuming that we do not have the situation described in Eq. (43), Chapter 1, in which we had degenerate bands, and could not carry out a power series expansion in $\vec{k}$. We can now replace k_x by $-i\partial/\partial x$, etc., to convert (2.7) into an operator. We find that $U(\vec{r})$ does not itself vary slowly with $\vec{r}$, but that rather we can set up an expression

$$U(\vec{r}) = \exp(i\vec{k}_o \cdot \vec{r})\, W(\vec{r}), \tag{2.8}$$

where $W(\vec{r})$ will prove to vary slowly. We find immediately that for instance the operator formed from $k_x - k_{xo}$, operating on $\exp(i\vec{k}_o \cdot \vec{r})\, W(\vec{r})$, equals $\exp(i\vec{k}_o \cdot \vec{r})(-i\partial/\partial x)\, W(\vec{r})$, and similarly with any powers or products of the operators like $k_y - k_{xo}$. Hence, substituting (2.8) into Eq. (2.6), we can cancel the exponential, and are left with

$$E|_o\, W(\vec{r}) - \frac{1}{2!}\left[\left.\frac{\partial^2 E}{\partial k_x^2}\right|_o \frac{\partial^2 W(\vec{r})}{\partial x^2} + \left.\frac{2\partial^2 E}{\partial k_x \partial k_y}\right|_o \frac{\partial^2 W(\vec{r})}{\partial x \partial y} + \ldots\right]$$

$$+ \text{terms in third and higher derivatives of } W(\vec{r}) \tag{2.9}$$

$$+ V(\vec{r})\, W(\vec{r}) = E W(\vec{r}).$$

In some cases $W(\vec{r})$ will vary slowly enough with position so that the terms in the third and higher derivatives can be neglected in comparison with the second derivative terms which we have retained. As far as the second derivative terms are concerned, we can perform a rotation of axes in ordinary space, in such a way as to remove the cross derivatives; we may as well assume that this has already been done, so that $\partial^2 E/\partial k_x \partial k_y$, etc., are zero. Then we can rewrite (2.9) in the form

$$\frac{-h^2}{8\pi^2}\left(\frac{1}{m_x}\frac{\partial^2 W}{\partial x^2}+\frac{1}{m_y}\frac{\partial^2 W}{\partial y^2}+\frac{1}{m_z}\frac{\partial^2 W}{\partial z^2}\right)+\left[E_o+V(\vec{r})\right]W=EW, \tag{2.10}$$

where

$$m_x=\hbar^2/(\partial^2 E/\partial k_x^2),\ \text{etc.} \tag{2.11}$$

In other words, in case all our assumptions are justified, we have converted our difference equations (2.5) into a form which is like Schrödinger's equation for W, only with different effective masses for motion in the three coordinate directions. If we are dealing with a minimum of $E(\vec{k})$, so that the second derivatives are positive, the effective mass is positive, while if we are at a maximum of $E(\vec{k})$ the effective mass is negative; in such a case it is more convenient to change the sign of Eq. (2.10), treat the effective mass as being defined by the negative of (2.11), and make a compensating change in the sign of the potential energy, so that it is as if we were dealing with a positively charged particle instead of a negatively charged electron, which then would be subjected to a force, and potential energy, which would be the negative of those acting on an electron.

The simplest case which we can have is that in which m_x, m_y, m_z are all equal, in which case (2.10) becomes just an ordinary Schrödinger equation for a particle of mass m in a potential V. A good deal of discussion of impurity levels in semiconductors has been based on this assumption, plus the further assumption that V(r) is a Coulomb potential, in a material of dielectric constant equal to that of the semiconductor. Then the problem becomes mathematically like that of hydrogen, and we can get solutions for bound states just like those in hydrogen. We easily find that the linear extension of the wave functions in space is greater than in hydrogen, by a factor equal to the dielectric constant of the material, divided by the ratio of the effective mass of the particle to the true mass of the electron, and the binding energy is smaller than in hydrogen, by a factor equal to the square of the dielectric constant, divided by the ratio of masses. In the case of germanium, for instance, with a dielectric constant in the neighborhood of 16, and effective masses of the order of half the electronic masses, this leads to wave functions extended over a good many Angstroms, but with binding energies which are a very small fraction of a volt.

These widely extended wave functions, which we can often get by solution of (2.10), allow us to examine the justification for the assumption that third and higher derivatives can be neglected in the expansion (2.9). The energy, as we know from (2.4), can be expanded in a Fourier series, and ordinarily the first few terms of this series are the important ones. Thus we see that the successive derivatives of E with

respect to k will be in the ratio of powers of R_s, the lattice spacing. If the function W is falling down like an exponential, e^{-ar}, as we should have for a hydrogen-like case, successive derivatives will be in the ratio of powers of a, inversely proportional to the linear dimensions of the wave function. In other words, we may expect the terms in (2. 9) in successively higher derivatives of W to fall off in the ratio of the lattice spacing to the linear dimensions of the wave function. If the wave function extends over many lattice spacings, as in the case of germanium, we may expect successive terms to get rapidly smaller, so that the reduction of the equation to the Schrödinger form (2. 10) is legitimate. On the other hand, if it does not extend out very far, as is the more usual case in a material of smaller dielectric constant, it is not legitimate to neglect the higher derivatives. In such a case we can try to retain additional terms in (2. 9), but it is much better and more straightforward to return to the difference equation (2. 5), and try to solve this directly. The writer suspects that the number of actual cases where the approach by means of the differential equation is justified is very limited, and that in most cases we are on much firmer ground to use the difference equation. However, the differential equation is often useful in pointing out the general nature of the solution to be expected.

2. The One-Dimensional Case

Actual crystals of course are three-dimensional; but in a corresponding one-dimensional problem it is easier to solve our difference equations exactly, and the insight gained in this way can help us to understand the more difficult three-dimensional case. Accordingly we shall consider in this section the special case of Eq. (2. 5) which arises when we have a set of equally spaced atoms along a line. We shall limit ourselves for the present to the case where only one band need be considered, though later we shall take up the interaction between different bands. We can then simplify the notation of (2. 5), in an obvious way, and write it

$$\sum(q)\left[\mathscr{E}(p-q) + V(p,q)\right] U(q) = EU(p), \tag{2.12}$$

where p is the index indicating the pth atom along the line. To show the general nature of the problem, we shall start with the simplest example. This is the one in which there is only a nearest-neighbor interaction between atoms, and where the perturbation is found only on a single atom. By nearest-neighbor interaction, we mean that the matrix component of H_o between two Wannier functions, as given in (2. 2), vanishes if the Wannier functions are located on two atoms more distant than nearest neighbors. That is, $\mathscr{E}(s)$ is zero if s is greater than 1. It is at once obvious from (2. 2) that $\mathscr{E}(-s) = \mathscr{E}(s)$, so that $\mathscr{E}(-1) = \mathscr{E}(1)$, and we have only two non-vanishing coefficients, $\mathscr{E}(0)$ and $\mathscr{E}(1)$. By stating that the perturbation is located on only a single atom, we

mean that $V(p, q) = 0$ unless p, q are each equal to the index of the impurity atom, which we may as well take to be zero.

With these simple assumptions, Eq. (2.12) reduces to

$$\begin{aligned} &\left[\mathcal{E}(0) - E\right] U(p) + \mathcal{E}(1)\left[U(p+1) + U(p-1)\right] = 0 \text{ if } p \neq 0 \\ &\left[\mathcal{E}(0) + V(0) - E\right] U(0) + \mathcal{E}(1)\left[U(1) + U(-1)\right] = 0. \end{aligned} \tag{2.13}$$

These equations can easily be solved exactly, for the type of boundary conditions we are interested in. We try a solution $\exp(ikRp)$ for $U(p)$, where k is a propagation constant as before, R is the interatomic distance, so that Rp is the distance to an atom p units away. Then the general equation, written in the first line of (2.13), will automatically be satisfied, as in Section 1, provided

$$E = \mathcal{E}(0) + 2\mathcal{E}(1)\cos kR. \tag{2.14}$$

The second equation of (2.13), however, is a special case, and will not be satisfied by our exponential solution. We can combine exponentials with equal positive and negative k's, however, or what is the same thing, use a cosine function with a phase to be determined. That is, we can let $U(p) = \cos(kRp - a)$, where a is a phase. This assumption will satisfy the first equation of (2.13) as well as the exponential.

Before considering the second equation of (2.13), we note one property of the solution, determined by symmetry. The potential is symmetric on reflection in the impurity atom. Therefore the solutions must be symmetric or antisymmetric, or must be even or odd functions of p. If they are odd functions of p, $U(0) = 0$, and $U(-1) = -U(1)$, so that the second equation of (2.13) is automatically satisfied. In other words, an antisymmetric solution can be written in the form $\sin kRp$, and as far as such solutions are concerned, the perturbation has no effect at all. We need then consider only the symmetric solutions. We might at first think that these should be written simply $\cos kRp$, but this is wrong, for we already know that this function will not satisfy the second equation of (2.13). Instead, we must let $U(p) = \cos(kRp - a)$ when $0 \leq p$, but define $U(-p) = U(p)$. This will give a function with discontinuous slope at $p = 0$; but this does not trouble us, for we are interested in $U(p)$ only for integral values of p, and the continuity of the function does not concern us.

Let us then assume $U(p)$ is as given above, and substitute in the second equation of (2.13). We have

$$\left[\mathcal{E}(0) + V(0) - \mathcal{E}(0) - 2\mathcal{E}(1)\cos kR\right]\cos a + 2\mathcal{E}(1)\cos(kR - a) = 0. \tag{2.15}$$

We expand cos (kR - α), cancel terms, and find

$$V(0)\cos\alpha + 2\mathscr{E}(1)\sin kR\sin\alpha = 0,$$
$$\tan\alpha = -V(0)/(2\mathscr{E}(1)\sin kR). \tag{2.16}$$

In (2.16) we find how to choose α so that our assumed values of U(p) will form a complete solution of Eq. (2.13). We thus have a general solution of the problem, for arbitrary energy values. The form of (2.14), however, leads to a real k only if the energy E is between $\mathscr{E}(0) \pm 2\mathscr{E}(1)$, for only in this case is |cos kR| less than 1. That is, for any energy within the single energy band described by our single Wannier function, we can set up a sinusoidal solution of this type. If we have to satisfy a boundary condition, such as that the function vanish at a fixed boundary, this will lead to quantum conditions determining discrete energy levels; we shall take up such boundary conditions later.

The only effect of the perturbation on these energy levels within the band, as we see, is to determine the phase of the wave function. It is interesting to note that if V(0) is very large numerically, so that tan α in Eq. (2.16) has large magnitude, then α approaches $\pi/2$, so that U(p) approaches sin kRp, which goes to zero at p = 0. In other words, the effect of a very large perturbation is to make the wave functions of the states within the energy band avoid the perturbing atom. On the other hand, we shall now show that the perturbation introduces another bound state, which tends to be concentrated on the perturbing atom, falling off exponentially as we go away from it. To describe such a case, clearly our exponential type of solution is not appropriate. This bound state, we shall find, lies outside the limits of the band, so that cos kR in Eq. (2.14) would be greater numerically than unity, and k imaginary. It is better, then, to start from the beginning by assuming a pure exponential function for U(p). Let us then assume that U(p) = exp (- γRp) for positive p, and again, for the symmetric state which alone comes into consideration, that U(- p) = U(p). We could, of course, also have a solution increasing exponentially with p, but this obviously would be inadmissible.

If we substitute our solution U(p) = exp(- γRp) in the first equation of (2.13), we find at once that we have a solution if the energy is given by

$$E = \mathscr{E}(0) + 2\mathscr{E}(1)\cosh\gamma R, \tag{2.17}$$

the analogue of (2.14). For γ = 0, the case where the wave function falls off infinitely slowly with r, the energy is $\mathscr{E}(0) + 2\mathscr{E}(1)$, lying at the top of the band if $\mathscr{E}(1)$ is positive, at the bottom if it is negative. We shall find that the bound state we are interested in becomes detached from the band if the perturbative energy V(0) has the correct

sign to push the energy of this state away from the band: that is, if it is positive if the state is at the top of the band, or negative if at the bottom. The case we are interested in is then that where $V(0)$ and $\mathscr{E}(1)$ have the same sign. We may now substitute our expression for $U(p)$, and (2.17) for the energy, in the second equation of (2.13), and we find, remembering that $U(-i) = U(i)$,

$$\sinh \gamma R = V_o / 2\mathscr{E}(1). \tag{2.18}$$

This equation determines the value of γ, showing that it is zero for $V_o = 0$, but becomes greater without limit as V_o increases. We may substitute (2.18) in (2.17), to obtain an explicit formula for the energy of the bound state:

$$E = \mathscr{E}(0) + 2\mathscr{E}(1)\sqrt{1 + (V_o/2\mathscr{E}(1))^2} \tag{2.19}$$

We see that, for small values of $V_o/2\mathscr{E}(1)$, the energy departs from the value at the band edge by a small quantity of the second order, while the value of γ is small of the first order.

We have just seen that a single energy level pushes away from the energy band under the action of the perturbation V_o, and that its wave function falls off exponentially as we go away from the perturbing atom. As V_o becomes very large compared to the band width, the exponential fall-off becomes very rapid, though it is small for small V_o. At this point we can give an elementary check of our procedure of the preceding section, in which we replaced the difference equation by a differential equation of the Schrödinger type. We are directed to start with the energy expression (2.14), and replace k in it by the operator $-i\,d/dx$. We then use this operator to set up a Schrödinger equation of form (2.9). In our case, this leads to

$$\mathscr{E}(0)\,W(x) + 2\mathscr{E}(1)\left[W + \frac{1}{2!}R^2\frac{d^2W}{dx^2} + \frac{1}{4!}R^4\frac{d^4W}{dx^4} + \ldots\right] + V(x)\,W(x) = EW(x), \tag{2.20}$$

where $V(x)$ is the perturbative potential, which is zero except when $x = 0$, so that except at $x = 0$ we can omit the term in V. The simplified Schrödinger equation corresponding to (2.11) is obtained by omitting everything beyond the second derivative in (2.20). Now we can, in this simple case, solve either the approximate or the exact problem by assuming $W = \exp(-\gamma x)$, the only difference being in the relation which we find between γ and E. If we substitute this expression in (2.20), we find a solution of the exact equation if $E = \mathscr{E}(0) + 2\mathscr{E}(1)\left[1 + (\gamma R)^2/2! + (\gamma R)^4/4! + \ldots\right] = \mathscr{E}(0) + 2\mathscr{E}(1)\cosh \gamma R$, in agreement with (2.17). If on the contrary we neglect fourth and higher derivatives, we have $E = \mathscr{E}(0) + 2\mathscr{E}(1)\left[1 + (\gamma R)^2/2!\right]$, which is a satisfactory

approximation only if $(\gamma R)^2$ is small in comparison to unity. This is the case in which the wave function $\exp(-\gamma x)$ falls off by only a small amount in the distance R, which agrees with the condition for the applicability of the second-order differential equation in Section 1. If the perturbation is greater, so that the wave function falls off more rapidly, and the extra level is further from the band, the second-order differential equation is inapplicable, and we must use the difference equation, or the differential equation (2.20) of infinite order, which of course is rigorously equivalent to it.

We can gain additional insight into our problem if we impose definite boundary conditions, and see what happens to each of the energy levels and wave functions as the perturbation V_0 increases from zero. Thus, let us impose the periodic boundary condition that $U(p + 2N) = U(p)$, where N is a definite integer. We must evidently have perturbing atoms not only at the origin, but at $p = \pm 2N, \pm 4N$, etc. For the antisymmetric solutions we easily see that our functions must be antisymmetric with respect to the point $p = N$, and the symmetric solutions are symmetric about this point. For the antisymmetric solutions of (2.13), we may set $U(p) = \sin kRp$, and the boundary conditions will be satisfied if $kRN = n\pi$, where n is an integer. We thus have solutions for $\kappa R = n\pi/N$, where in the usual way we shall find independent solutions only for the $N - 1$ values $n = 1, 2, \ldots N - 1$; the values $n = 0, N$ automatically give nothing, and any integral values of n outside this range repeat values of U(p) for each atom as found for one of the values $n = 1, \ldots N - 1$.

For the symmetric solutions, we can most conveniently proceed, not as in our earlier discussion, but by setting $U(p) = \cos kR(p - N)$, which will automatically satisfy the boundary condition at $p = N$. The energy is given by (2.14), as before, but now in place of (2.17) and (2.18) we find easily

$$- \tan kRN \sin kR = V(0)/2\mathscr{E}(1). \tag{2.21}$$

We can solve this equation graphically, by plotting the left side as a function of k, and finding the k's for which this equals $V(0)/2\mathscr{E}(1)$. A plot of this function, for $N = 6$, as a function of kR, is given in Fig. 1. We see that the axis of abscissas cuts this curve in seven places, $kR = 0, \pi/N, 2\pi/N, \ldots \pi$. For other horizontal lines, where there are only six intersections, it is easy to show that there is also one intersection corresponding to an imaginary value of k, corresponding to the bound state which we have been discussing in the preceding paragraphs. The generalization of this is that there are $N + 1$ symmetric states, which together with the $N - 1$ antisymmetric states give 2N in all. This is the number which we should expect on account of our periodicity.

It is very informing, using this graphical solution, to find the values of the $N + 1$ energy values of the symmetric states, as functions of the quantity $V(0)/2\mathscr{E}(1)$.

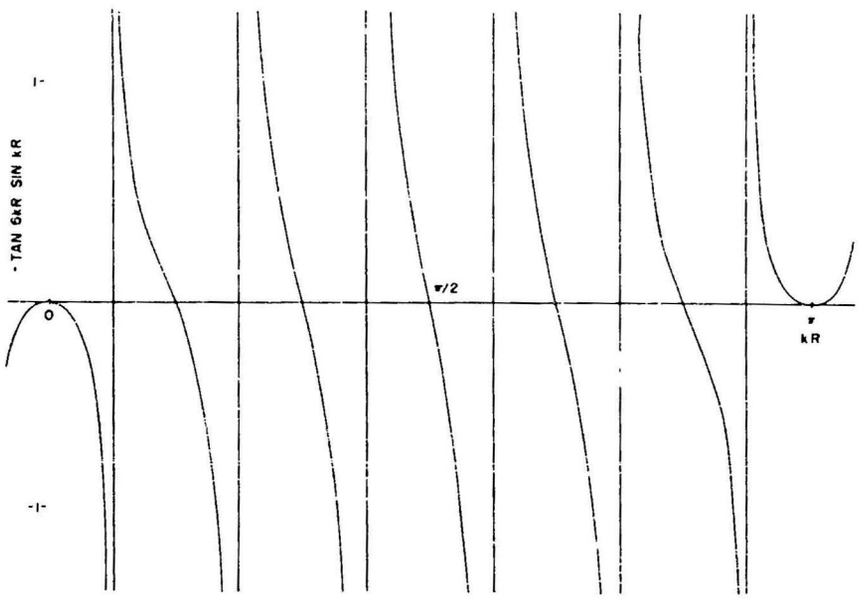


Fig. 1
Function - tan kRN sin kR, for N = 6.

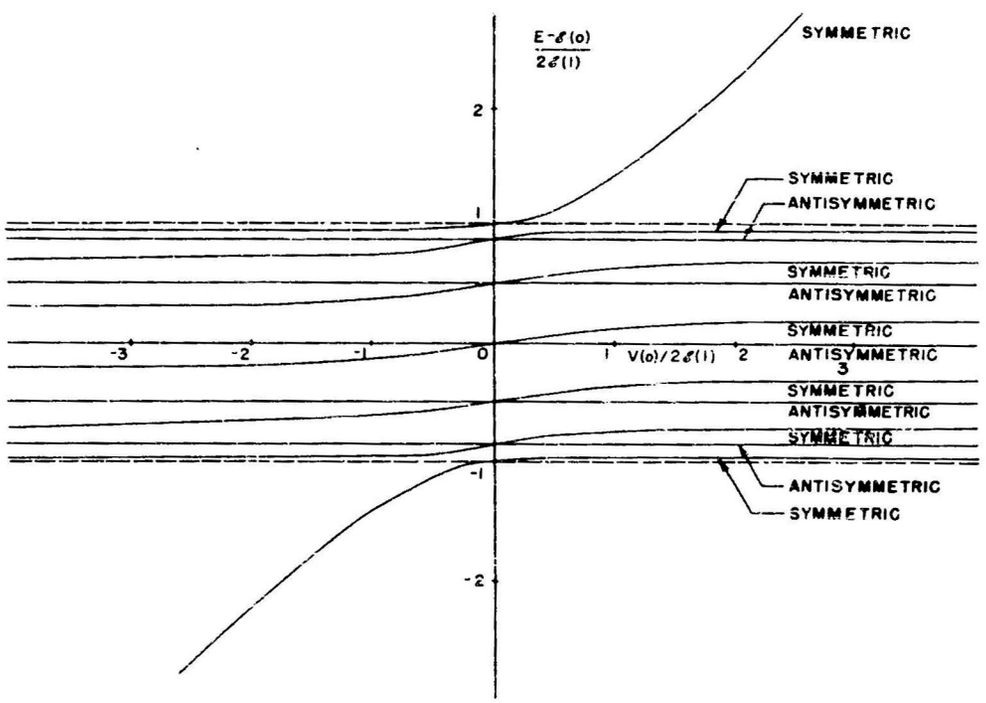


Fig. 2
Energy of perturbed energy levels, as function of perturbative energy.

These values are plotted in Fig. 2; we show also, for comparison, the values of the antisymmetric energies, which are not affected by the perturbation. The interesting thing to notice is the way in which each of the energy levels within the band is displaced, by the perturbation, only far enough to approach the midpoint between neighboring antisymmetric states, with the exception of the limiting state, which becomes detached from the band, and forms the separate impurity state which we have discussed earlier. The case which we have shown is that of positive $\mathscr{E}(1)$, in which case it is the top of the band where the separate level appears for positive V(0). We have shown the case for a small N value; but if N is very large, it is clear that the displacements of all levels but the one which becomes detached from the band will be negligible.

We have discussed the case in which the discrete level becomes detached from the band at k = 0. That is, if $\mathscr{E}(1)$ is positive, the maximum energy of the band comes at k = 0, and a positive value of V(0) results in a level rising from this maximum point, while if $\mathscr{E}(1)$ and V(0) are negative a level is pushed out of the band at the energy minimum, which again comes at k = 0. We can also have the case where the discrete level emerges from the band at the extreme value of k, equal to $k = \pi/R$. Such a case arises if $\mathscr{E}(1)$ is positive and V(0) negative, or vice versa. This follows formally from Eq. (2.21), just as does the case we have considered, and the curves for energy form the part of Fig. 2 for negative values of $V(0)/2\mathscr{E}(1)$. The quantities U(p) have a different form in this case, however, as we should deduce from Eq. (2.8) of the preceding section. This case corresponds to $U(r) = \exp(ik_o r)\, W(r)$ of Eq. (2.8), where the expression $\exp(ik_o r)$ will change by a factor -1 when r increases by R. In other words, the solution corresponding to a bound state, in the absence of the boundary condition that U(p) goes to zero when p = N, will here have the form $U(p) = (-1)^p \exp(-\gamma Rp)$, instead of the value $\exp(-\gamma Rp)$ previously encountered. When we substitute this value in Eq. (2.13), we find that we have a consistent solution for our present case in which $V(0)/2\mathscr{E}(1)$ is negative; there are appropriate changes of sign in the equations which we obtain as substitutes for (2.18) and (2.19).

We have noted earlier that the wave functions of the states in the band become small at the position of the perturbing atom, as the perturbation becomes large. We can see that this must be so in a very fundamental way. Let us consider the case of periodic boundary conditions. We are building up wave functions as linear combinations of the 2N Wannier functions on the atoms of a fundamental period. These Wannier functions are an orthonormal set of functions, as are the final linear combinations, so that the U(p)'s, which are the transformation coefficients from one set of functions to the other, form a unitary matrix. This is a matrix with 2N rows and columns: we have 2N energy levels, and 2N p's. If we now form the sum of the quantities $U^*(p)\,U(p)$, for a given p value, summed over the 2N energy levels, we know on account of the unitary property that this sum must be unity. Let us carry out this sum

for $p = 0$, corresponding to the impurity atom. For the separated state, with large perturbative potential $V(0)$, we know that the wave function will fall off as $\exp(-\gamma Rp)$, where γ is very large, so that the coefficients $U(p)$ will be very small for $p \neq 0$. Since the wave functions must be normalized, we must have the sum of $U^*(p)\, U(p)$ over p, for a given energy level, equal to unity, which means that if the contributions to the sum for $p \neq 0$ are very small, the term for $p = 0$ must be very nearly equal to unity. We now return to our statement that the sum of $U^*(p)\, U(p)$ over all levels, for a fixed p, must equal unity. For $p = 0$, we have just seen that the contribution from the separated state is almost unity. Therefore the contributions from all other levels must be very small, showing that with a large perturbation, the wave functions of the states in the band avoid the perturbing atom almost completely. This general type of proof is important, for it can be used in more complicated cases, where we cannot get explicit solutions.

We have now proceeded about as far as possible with our very simple one-dimensional problem. We can introduce some further complications, however, and still carry through an exact solution, and in this way we can get some further insight into more difficult problems. In the first place, we can include second-nearest neighbor interactions, without seriously increasing the difficulty of the problem. That is, in Eq. (2.12), we can include a term $\mathscr{E}(2)$ as well as $\mathscr{E}(1)$, so that the energy expression, similar to (2.14), becomes

$$E = \mathscr{E}(0) + 2\mathscr{E}(1) \cos kR + 2\mathscr{E}(2) \cos 2kR. \qquad (2.22)$$

With suitable choice of the constants, this allows us to have the energy as a function

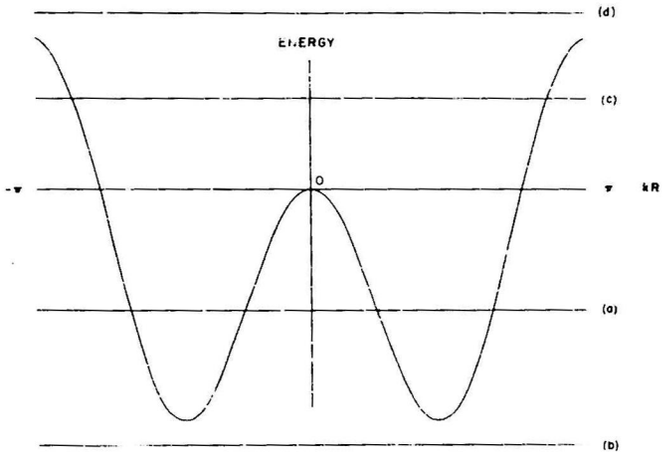


Fig. 3
Energy vs. kR, for second-nearest neighbor interaction.

of kR given by a curve of the type of Fig. 3, in which we can have a minimum or maximum energy coming elsewhere than at $k = 0$ or π/R. This corresponds to a minimum or maximum energy coming somewhere within a Brillouin zone, and not at a symmetry point, in a three-dimensional problem, and can be of practical importance, for there is some reason for thinking that this is the situation actually found in germanium in the conduction band. Let us consider what happens in this case if we introduce a perturbation V(0), like that of Eq. (2.13).

In the first place, we shall still have the same situation as before, that the solutions are divided into antisymmetric and symmetric functions, and that the energy of the antisymmetric states will not be altered by the perturbation. We also have the possibility of setting up solutions of a typical equation of type (2.13) in the form exp(ikRp), which leads at once to the energy expression (2.22). We see now, however, that we have four solutions for k, rather than the two ($\pm$ a single value) which we had in the earlier case. We can see this most simply from Fig. 3. If we cut this curve with a horizontal line, the intersections will correspond to the values of k for which we get solutions of our difference equation, for a given energy. It is clear that there are ranges of energy for which there are four intersections, which we may call $\pm k_1$, $\pm k_2$. In other ranges of energy some of these intersections become imaginary, but it can still be shown that Eq. (2.22) has four different solutions for k for a given energy, though sometimes the k's are real, sometimes complex.

Let us first consider the case where the k's are real. The equations which we must satisfy, similar to (2.13), include the typical equation

$$\left[\mathscr{E}(0) - E\right] U(p) + \mathscr{E}(1)\left[U(p+1) + U(p-1)\right] + \mathscr{E}(2)\left[U(p+2) + U(p-2)\right] = 0 \quad (2.23)$$

which is satisfied by $U(p) = \exp(ikRp)$, provided k satisfies (2.22). This equation holds for $p \geq 2$. The additional equations are different depending on whether we are dealing with antisymmetric or symmetric solutions. For the symmetric case, we have

$$\begin{aligned} &\left[\mathscr{E}(0) + V(0) - E\right] U(0) + 2\mathscr{E}(1)\, U(1) + 2\mathscr{E}(2)\, U(2) = 0 \\ &\left[\mathscr{E}(0) - E\right] U(1) + \mathscr{E}(1)\left[U(2) + U(0)\right] + \mathscr{E}(2)\left[U(3) + U(1)\right] = 0. \end{aligned} \quad (2.24)$$

For the antisymmetric case, the first equation of the form of (2.24) is automatically satisfied, and the second reduces to

$$\left[\mathscr{E}(0) - E\right] U(1) + \mathscr{E}(1)\, U(2) + \mathscr{E}(2)\left[U(3) - U(1)\right] = 0. \quad (2.25)$$

THE PERTURBED PERIODIC LATTICE

It is instructive to examine the antisymmetric case, though we know in advance that its solution is as in the unperturbed case; that is, we know that there are two independent solutions, $U(p) = \sin k_1 Rp$ or $\sin k_2 Rp$, where k_1, k_2 are the two independent solutions of (2.22). Since we have degeneracy, with two independent solutions for the same energy, any linear combination of these two functions is a solution of (2.25), as we readily verify. We can see that there must be two independent solutions directly from the difference equations. From (2.25), we can solve for U(3) in terms of U(1) and U(2) (remembering of course that U(0) = 0 for the antisymmetric case). From (2.23) we can then solve for U(4) in terms of U(3), U(2), and U(1), which means that we can find U(4) in terms of U(1) and U(2); and we can repeat this process step by step, finding each further value of U in terms of U(1) and U(2), writing them in fact as one quantity times U(1) plus another quantity times U(2). We can get a solution, then, for arbitrary values of U(1) and U(2), showing that there must be two arbitrary constants in the solution. This is merely an alternative way of writing the solution as a linear combination of $\sin k_1 Rp$ and $\sin k_2 Rp$.

We can now apply the same sort of argument to the symmetric case, to find what must be the nature of the solution. From the first equation of (2.24), we can find U(2) as a linear combination of U(0) and U(1), for a given energy. From the second equation of (2.24), we can find U(3) in terms of U(0), U(1), and U(2), which then leads to U(3) as a linear combination of U(0) and U(1). By using (2.23) we can extend this process, writing each of the U's as a linear combination of U(0) and U(1), so that again we have two arbitrary constants in the solution. It is no longer trivial, however, to find what these solutions are. The complete solution, as a matter of fact, has not been worked out, the algebra being involved, but we can state the general nature of the solution. We can write U(p) in the symmetric case as a linear combination $A_1 \cos (k_1 Rp - \alpha_1) + A_2 \cos (k_2 Rp - \alpha_2)$, where A_1, A_2 are amplitudes, α_1 and α_2 phases. We can substitute this solution in Eqs. (2.24), which then must be sufficient, since there are two equations, to determine two of the four arbitrary constants, which we may take to be α_1 and α_2. Within limits, we can find a solution for any pair of amplitudes A_1, A_2. It is clear, however, that this cannot be done for completely arbitrary amplitudes, for then we should be able to choose one of the amplitudes to be zero, which would imply that a single function $\cos (k_1 Rp - \alpha_1)$ or $\cos (k_2 Rp - \alpha_2)$ could furnish a solution. This is clearly impossible, for such a function is a solution of the unperturbed problem, for V(0) = 0, and hence cannot satisfy the perturbed problem. There must be, then, only a limited range of the ratio A_2/A_1, not extending either to zero or infinity, over which our conditions can be satisfied with real values of α; outside this range α_1 or α_2 or both will have to become complex. This range will become larger as V(0) approaches zero.

We can then take two independent solutions, say those for which A_2/A_1 has its

maximum and minimum allowable values, and use these as two independent functions, building up all others as linear combinations of these two. One of them will consist largely of the function cos $(k_1Rp - a_1)$, with a minimum contribution of cos $(k_2Rp - a_2)$, and the other will have a maximum contribution of cos $(k_2Rp - a_2)$ and a minimum amount of cos $(k_1Rp - a_1)$. The effect of the perturbation, in other words, is to mix the two waves corresponding to the two different k values together, the mixing increasing as the size of the perturbation increases. It is as if the wave of one k value, when scattered by the perturbing atom, scatters contributions of waves of both k values corresponding to the same energy.

We shall see later that there is a bound state near the perturbing atom, being concentrated more and more on this atom as the perturbation becomes larger. Thus we can conclude, just as in the case of nearest-neighbor interaction, that the wave functions of the other states must become vanishingly small at the perturbing atom, or U(0) must approach zero, as the perturbation becomes infinite. We can also, as in the case of the nearest neighbor interaction, impose boundary conditions, such as making U(p + 2N) = U(p). In this case, as before, we shall find N - 1 antisymmetric solutions. These will come when $kR = n\pi/N$, just as in the earlier case. For such a k value, in general the other k value corresponding to the same energy will not satisfy the periodic boundary conditions, so that the wave function will be just U(p) = sin kRp, with the (N - 1) values of k given above. For the symmetric solutions, the situation is more complicated. But we expect to find, as before, that there are N + 1 symmetric solutions, of which one becomes detached from the band as a discrete state under the action of the perturbation.

Our main interest is in this discrete state, and we can discuss it in a somewhat different way, by analogy to the method used for the case of nearest-neighbor interaction. Let us consider what happens when the energy E in Eq. (2.22), and in Fig. 3, decreases. In the case shown by (a), Fig. 3, a line of constant energy intersects the curve in four places; this is the case which we have been discussing. As E decreases, however, we come to a point where k_1 and k_2 become equal to each other, and then there is no further real intersection. Examination of the form of Eq. (2.22) shows that at this point the values of k for the intersections become complex, the four solutions of (2.22) having the form $\pm k_o \pm i\gamma$, where k_o and γ are real, γ increasing as the energy E decreases below the lowest point on the curve of Fig. 3, or below the bottom of the energy band. The quantities U(p) then can be written in the form $\begin{matrix}\sin\\ \cos\end{matrix} k_oRp \exp(\pm \gamma Rp)$. We are interested only in a symmetric solution; there are no antisymmetric solutions corresponding to bound states. Obviously we wish only a solution which vanishes exponentially as p becomes large. Thus for positive p we may take U(p) = cos $(k_oRp - a) \exp(-\gamma Rp)$. This form of expression reminds us of Eq. (2.8), where we had an energy minimum at a point corresponding to a propagation

vector k_o, and where the solution was $\exp(ik_o r)$ multiplied by a slowly varying function, which in our case corresponds to the exponential function. If now we substitute our solution in Eq. (2.24), we find that by rather tedious algebra we can solve for k_o, and the energy, as functions of γ. That is, we find a bound state of a given energy, and its qualitative behavior is as in the case of nearest-neighbor interaction: as the perturbative potential V(0) is increased numerically (it must of course be negative to make the discrete level fall below the bottom of the band), the discrete level falls in energy, and its γ increases, so that the wave functions falls off more rapidly as we go away from the perturbing atom. In other words, as far as the qualitative situation is concerned, our second nearest-neighbor interaction has not changed the properties of the discrete state. Now, however, it detaches itself from the minimum of the energy curve, which no longer lies at the center or boundary of the Brillouin zone, and as a result of this the wave function U(p) has the modulation factor $\cos(k_o Rp - \alpha)$, as well as falling off exponentially with p.

We have now considered two of the possible four cases shown in Fig. 3: the case where the energy is such that there are four real values of k, and that where the energy falls below the bottom of the band, and there are four complex values of k, as shown in (a) and (b) in Fig. 3. The case of (c), Fig. 3, is one where there are two real k values, and two pure imaginary values: as the energy rises from a value like (a) to that like (c), two solutions coalesce at k = 0, at the energy maximum at the center, and there they become imaginary. No discrete level appears at this point, however, even with a positive value of V(0), though there would be a discrete level if an energy like (c) lay entirely outside the energy band. Instead, if we call the real value of k by the symbol k_1, the imaginary one $i\gamma$, then we build up a solution of the form of a linear combination of $\cos(k_1 Rp - \alpha)$ and of $\exp(-\gamma Rp)$, for positive p's. Both types of terms are required to satisfy Eqs. (2.24), and we shall find solutions of this general type, corresponding to an enhancement of the wave function near the perturbing atom, but merging into an ordinary cosine-like wave function at a distance from this atom. Finally, in the case (d), where the energy is entirely above the band, we find two different values of γ. We build up combinations of the two appropriate exponentials, falling off as we depart from the perturbing atom, and again find, for positive V(0), an impurity level detaching itself from the top of the energy band, becoming further separated as V(0) increases, with the wave function falling off more and more rapidly as the perturbation becomes larger and larger.

This sketch of the case where second-nearest-neighbor interactions are considered shows that the general situation is similar to that with only nearest-neighbor interactions. We shall now consider another form of generalization of our simple nearest-neighbor case. This comes when the perturbation V occurs on more than one atom. As an example, we may suppose that it occurs on three adjacent atoms, in a

symmetrical form. That is, we may suppose that V(1) is different from zero, as well as V(0), but that V(- 1) = V(1). This would allow us to simulate a case in which, for instance, there was a screened Coulomb interaction between the impurity atom and its neighbors, falling off rapidly with distance (so that V(1) would be numerically smaller than V(0)), but appreciable at the nearest neighbor of the perturbing atom. We find by good fortune that Eqs. (2.24) are not appreciably more difficult to solve in this case than when V(1) is absent. Thus if we have second-nearest-neighbor interactions, and also perturbations on the atoms next to the central atom, we can still get analytic solutions of the problem. We find that the situation is altered in an interesting way: we now find three energy levels splitting off from the band under the action of the perturbation, rather than one only. As the perturbation becomes large, we may describe these states as follows: one of them has U(1), U(0), and U(- 1), all of the same sign, and is a symmetric solution: another is a symmetric solution in which U(1) and U(- 1), which are equal to each other, have the opposite sign to U(0); and the third is antisymmetric, with U(0) = 0, U(- 1) = - U(1). These have the same arrangement formally, then, as the three modes of a string carrying three weights. We may generalize this result: if large energy perturbations are applied to n adjacent atomic sites, we shall have n discrete levels splitting off from the band, as the perturbations increase from zero. It is only in the special case where the perturbations apply to only one atom that there is only one discrete level.

If there are n discrete levels, located on n adjacent atoms, widely separated from the rest of the band for large perturbative potentials, then an argument based on the unitary nature of our transformation will tell us, just as in the simpler cases, that the wave functions of the band will avoid these n atoms for large perturbation. In the limit, we can obviously have a splitting of the band into two sub-bands. For instance, if n adjacent atoms have a perturbative term V(0) in their potential, equal on all of them, and then the next n atoms do not have such a term, and after that we alternate between perturbed and unperturbed atoms, we shall split half the levels away from the other half, and shall have effectively a band split in the middle. Situations such as this are met in molecular crystals.

Sometimes, instead of having a situation like that just described, where all the perturbed atoms have potentials differing by a large amount from the unperturbed ones, we have quite a different situation, of which the Coulomb potential is typical. This of course is a three-dimensional problem, in which the perturbing potential is a Coulomb potential centered on a perturbing atom, so that the perturbative potential V, though it falls off fairly rapidly with distance from the perturbing atom, still extends out with appreciable amplitude out to distant atoms. We know that the corresponding hydrogen problem has an infinite number of discrete energy levels, below the continuum. A one-dimensional version of this problem, in a similar way, will not show a finite num-

ber of discrete levels definitely separated from a largely unperturbed band. Rather, all the levels will be perturbed, the lowest ones most (assuming that the perturbation is of the type to depress the discrete levels), but others more or less according to a Rydberg law. We might think, in such a case, that since all the states have a certain resemblance to bound states, the charge distribution produced when all states were occupied would tend to concentrate around the impurity atom. That this is not the case is seen from our argument in terms of the unitary transformations: if all states in the perturbed band are occupied, then the electrons will be uniformly distributed around all atoms, the rather concentrated discrete state wave functions being balanced by higher energy levels, in which the wave functions tend to avoid the perturbed atom.

3. The Two- and Three-Dimensional Cases

The two- and three-dimensional problems are very much more difficult than the one-dimensional case which we have just been taking up. The difficulty arises from various sources. In the first place, we have the same sort of problems which arise in differential equations in two and three dimensions: we are dealing with the analogues of partial differential equations, which do not have simple solutions except in cases of separability, and most of our cases do not correspond to separable equations. Furthermore, we have the symmetry properties which become so important in two and three dimensions. Not only must our solutions of the problem in terms of Wannier functions show proper group-theoretical behavior; we must consider also the symmetry properties of the Wannier functions themselves. This introduces us to a question which was not emphasized in Chapter 1: the nature of the Wannier functions in case we have degenerate bands, such as the p or d bands. We shall have to go into such questions in detail, before we can understand the three-dimensional case. In our first discussion, however, we shall simplify things by not considering degenerate bands; we shall deal with s-like bands, and Wannier functions with s-like symmetry, amd shall continue, as in the preceding section, to assume that we can get a good representation of the perturbed problem by using Wannier functions from only one band. We shall remove this restriction in a later section, where we shall take up the effect of overlapping bands on the nature of impurity levels. Furthermore, in the present section, we shall deal mainly with the case where the perturbative potential is located only on one perturbing atom, as in the larger part of the discussion of the preceding section. In this way, we get as close an approach as we can to a problem of spherical symmetry, and we can use some of the results of the separation of variables in spherical coordinates which we should meet in the corresponding problem of a free particle.

The discussion of Section 1 was of course general, applying to the two- and three-dimensional cases. Hence we can conclude at once that when the wave function

varies slowly with position, we can replace our difference equation by a second-order differential equation, and can use (2.10) to solve our problem. Furthermore, with an energy band of s-like symmetry, the effective masses m_x, m_y, m_z will all be equal. Hence if the potential V has spherical symmetry, our Schrödinger problem (2.10) will be separable in spherical coordinates, and we can really solve our problem. Let us then consider this soluble case in some detail, as a guide to what to expect in more complicated cases where an exact solution is impossible. We consider first the case in which, as in the one-dimensional problem, V has a diagonal matrix, having no components between Wannier functions belonging to different bands or on different atoms, and where furthermore the diagonal component is zero except on one perturbing atom, which we may take to be at the origin.

In this case, the approximate Schrödinger equation (2.10) has $V = 0$ everywhere except at the origin, and hence is the equation for a free particle. The solutions which we wish, then, must be spherical harmonics of the angle, times spherical Bessel functions of the radial distance out from the perturbing atom. We must choose such a solution as to have a suitable singularity at the origin, so that if we take the values of this solution at the various atoms, treat these values as the quantities $U(\vec{R}_i)$, and substitute in the difference equation (2.5), we shall have a solution of this difference equation even at the origin. From our knowledge of the solutions of the wave equation in spherical coordinates, we know that the wave functions of all these cases except for $\ell = 0$, where ℓ is the azimuthal quantum number, will vanish at the origin. Since the perturbation is located only at the origin, this means that only the states for $\ell = 0$, or the s states, will be perturbed; this is analogous to the situation in one dimension, where the symmetric wave functions were perturbed, the antisymmetric ones were not. We may limit our discussion, then, to s states, independent of angle, and in which the corresponding spherical Bessel function, expressing the dependence of the wave function on r, has the form of $\frac{\sin kr}{r}$ or $\frac{\cos kr}{r}$ or $\frac{\exp(\pm \gamma r)}{r}$, depending on circumstances. We note that all these functions, except the particular combinations $\frac{\sin kr}{r}$ or $\frac{\sinh \gamma r}{r}$, become infinite at the origin; there is no one of these functions, in the range where they vary exponentially with distance, which simultaneously goes to zero at infinity, and is finite at the origin. This as once shows us that the behavior of our function at the origin is more complicated than in one dimension, and that we must give special attention to the problem.

We can understand what is going on by considering the Schrödinger problem of a free particle in a spherical square-well potential. That is, we let the particle move in a potential which is zero except within a small sphere of radius R_o, where it equals $-V_o$. If we choose this sphere to be of atomic dimensions, then this problem may serve as a sort of analogue to our actual case, the potential well at the origin corresponding to the perturbation V(0) found at the central atom. In a similar way, in the

one-dimensional case, we could have taken a free particle moving in a field-free space, except for a square well where the potential is $-V_o$ located in the immediate neighborhood of the origin. This problem shows strong analogies to the one-dimensional case which we have taken up in the preceding section. For any value of V_o, even a very small one, a bound state is set up. Its energy, as a function of V_o, shows a behavior similar to that of Eq. (2.19): it is proportional to V_o^2 for small V_o, to V_o for large V_o, and the wave function falls off exponentially as we go away from the perturbation, with a value of γ similar to that of Eq. (2.18). In our present case, however, the situation is quite different, as is known from elementary quantum mechanics. We find that if the potential well is not deep or wide enough, there is no bound state at all.[1] This can have quite profound effects on the nature of the impurity levels.

This difference between the one- and three-dimensional cases is so important that we must examine the reasons for it in detail. We remember that Schrödinger's equation, in spherical coordinates, for the case where the wave function is independent of angle, can be written in the form

$$-\frac{h^2}{8\pi^2 m}\frac{d^2}{dr^2}(ru) + V(ru) = E(ru), \tag{2.26}$$

where u is the Schrödinger wave function. That is, if V is zero, as we have outside our potential well, the quantity ru satisfies the differential equation for a free particle, so that it can be written in the form $\exp(ikr)$, where $k = \frac{2\pi}{h}(2mE)^{1/2}$. Inside the potential well, in a similar way, where the potential energy is assumed to be $-V_o$, we have a solution of the form $\exp(ik_o r)$, where $k_o = \frac{2\pi}{h}(2m(E + V_o)^{1/2})$. For a bound state, we must assume that $E + V_o$ is positive, but that E, the kinetic energy outside the potential well, is negative. Inside, then, we must use a solution $ru = \sin k_o r$; for only by using such a solution will the actual wave function u be finite at the origin. We must use outside a solution $\exp(-\gamma r)$, where $\gamma = \frac{2\pi}{h}(-2mE)^{1/2}$, the energy being negative. These two functions must join smoothly at the boundary of the well. Now the exponential must be sloping down at the boundary of the well, or in the limit where γ is very small it will have a horizontal tangent. Thus we cannot join the functions smoothly unless the sine function has gone through at least a quarter wave length within the well, so as to rise to its first maximum at the boundary of the well. In this limit, then, where E is zero, and the wave length within the well is $h(2mV_o)^{-1/2}$, the condition for the existence of a bound state is that $4R_o$ should be at least a wave length, or that V_o must be at least $h^2/(32mR_o^2)$.

---

[1] For a discussion, see for instance L. Schiff, Quantum Mechanics, (McGraw-Hill Book Company, Inc., New York) 1949, p. 77.

Of course, we can have smaller values of V_o than this; and we naturally ask what happens to the wave function in such a case. The answer is clear from the discussion we have just given: the wave function inside the well joins at the boundary of the well onto a sinusoidal function outside, corresponding to positive energy. If we consider a finite volume within which the electrons move, freely except for the one potential well within the volume, then if the depth of the well is zero, we shall have discrete energy levels, just as in the free-electron model of a conductor. The lowest one of these energy levels will have a very small but finite energy: the wave length corresponding to its wave must be equal to twice the dimension of the cavity, or some such thing, depending on the details of the boundary conditions. Now let the depth of the potential well increase from zero. All the wave functions and energy levels will be slowly modified, and the energy of the lowest level will be gradually reduced to zero. It will not reach zero, however, until the well has reached its critical size. At this point, the wave function will be like $\frac{\sin k_o r}{r}$ within the well, the sine having almost exactly zero slope at the boundary of the well; outside the well, the wave function will be almost exactly $1/r$, fitting smoothly onto the function within the well, and being very slightly modified for very large values of r to fit whatever boundary conditions we impose around the boundary of the cavity. As the well becomes still deeper, this lowest level will fall below the value of zero corresponding to the bottom of the band. Closer examination shows that for small increases of V_o over the critical value, the energy will be proportional to $(V_o - V_{crit})^2$, but at sufficiently great V_o, the energy will be proportional to V_o. The wave function will become more and more concentrated in the potential well as V_o increases, falling off more and more rapidly in the exponential region outside the well, where it has the form $\exp(-\gamma r)/r$. Finally, for an extremely deep well, the wave function will be negligible outside the well, and will become identical with that of a particle in a well with infinitely high barriers at the boundary.

This sketch of the situation shows the general nature of our solution for the bound state. There is now one very important feature connected with it: in any case where a bound state exists, the wave function changes so rapidly with position that we are not justified in replacing the difference equation by a differential equation. We have just seen that the minimum depth of well gives a wave function proportional to $1/r$ outside the well, and for deeper wells it falls off more rapidly, as $\exp(-\gamma_1 r)/r$. But even the function $1/r$ varies so rapidly with position that we are not at all justified in replacing differences by derivatives, or, what amounts to the same thing, disregarding fourth and higher derivatives. The wave function decreases by a factor of 2 in going from an atom which is a nearest neighbor of that at the origin, to an atom twice as far away. Such a variation cannot be regarded as small, by any stretch of the word. We thus conclude that it is never legitimate, in a three-dimensional problem, to re-

place the difference equation by a differential equation, in the problem where there is a perturbation at only one point. We may still expect the discussion which we have just given to hold in a qualitative manner, but cannot expect it to be quantitatively correct. We shall then pass on to the difference equation, and shall show that in fact the qualitative behavior is just as we have described, but that fortunately we can handle the problem quantitatively as well as qualitatively.

Let us take the case of a simple cubic lattice in three dimensions, with only nearest-neighbor interactions. Taking only a non-degenerate band, with s-like symmetry, and considering only the Wannier functions of this band, our difference equations (2.5) then become

$$\begin{aligned}&\left[\mathscr{E}(0) - E\right] U(p, q, r) + \mathscr{E}(1)\left[U(p+1, q, r) + U(p-1, q, r) + U(p, q+1, r) + U(p, q-1, r)\right. \\ &\qquad \left. + U(p, q, r+1) + U(p, q, r-1)\right] = 0, \text{ unless } p = q = r = 0 \\ &\left[\mathscr{E}(0) + V(0) - E\right] U(0, 0, 0) + \mathscr{E}(1)\left[U(1, 0, 0) + U(-1, 0, 0) + U(0, 1, 0) + U(0, -1, 0)\right. \\ &\qquad \left. + U(0, 0, 1) + U(0, 0, -1)\right] = 0\end{aligned} \tag{2.27}$$

Here we have numbered the atoms, by the three integral indices p, q, r. These difference equations are exactly equivalent to Eq. (2.13) for the one-dimensional case. Our problem has the symmetry properties of the cube, and our solution must then form a basis function for an irreducible representation of the group. We know that we are looking for a solution which is an even function of p, of q, and of r. This is the same type of symmetry shown by the s function in the free electron problem; only here, since we do not have spherical symmetry, we must not expect that the continuous function U(p, q, r), whose values for integral values of p, q, r we desire, will show shperical symmetry.

Our difference equations (2.27) fortunately can be solved, though the solution does not seem to appear in the literature. The writer is indebted to Dr. H. C. Schweinler for pointing out that an analogous problem has been solved by McCrae and Whipple,[1] and to Dr. G. F. Koster for showing that the same method of solution can be used for the present problem. The problem is much more involved than in one dimension. We cannot make any direct use of the spherical symmetry of the differential equation which approximately represents the problem, for the real solution is not spherically symmetrical, and there is in fact no way to separate variables. We can, however, of course set up solutions, in this case an infinite number of them, for the

---

[1] W. H. McCrea and F. J. W. Whipple, Proc. Roy. Soc. (Edinburgh) 60, 281 (1940).

general equation of (2.27), and the procedure used is to write our solution as a sum of this infinite number of terms, and then apply suitable limiting conditions to take care of the special equation corresponding to $p = q = r = 0$.

The method of solution introduces an apparent lack of symmetry between the three directions in space: one axis, which we shall take to be the r axis, must be treated differently from the others. Though the solution appears unsymmetrical, this is only in the way of writing it, not in the solution itself. Let us take a function $U(p, q, r) = \cos k_1 Rp \cos k_2 Rq \exp(-\gamma Rr)$. If we substitute this function in the first equation of (2.27), we find that we have a solution provided

$$E = \mathscr{E}(0) + 2\mathscr{E}(1)\left[\cos k_1 R + \cos k_2 R + \cosh \gamma R\right]. \tag{2.28}$$

Since k_1 and k_2 can take on arbitrary values, we have a double infinity of solutions of this type, for a given energy value; the quantity γ is determined from (2.28) in terms of k_1, k_2, and E. Now let us make up a solution as a sum of an infinite number of such terms, all corresponding to the same E; it will clearly be a solution of the first equation of (2.27), and we have merely to satisfy the second equation. Before doing this, however, we must look a little more closely at boundary conditions at large distances.

We are looking specifically for the bound state. Thus for positive values of r we wish to use the exponential $\exp(-\gamma Rr)$, which will go to zero at infinite values of r. However, for negative r we must have a solution which is symmetric with this, so that we must use a similar sum of terms with $\exp(\gamma Rr)$. We shall, then, have a condition of joining of these two solutions along the plane $r = 0$, as well as the condition at $p = q = r = 0$. We also wish the function to go to zero at infinite positive and negative values of p and q. To accomplish this, we shall arrange to have the solution go to zero on the planes $p = \pm N$, $q = \pm N$, where N is a large integer, and eventually we shall let N become infinite. We can accomplish this if we limit $k_1 R$, and $k_2 R$, to the N values $\pi/2N$, $3\pi/2N$, ... $(2N - 1)\pi/2N$; for then each of the cosine functions will become zero at the limits. We can then write our solution in the form

$$U(p, q, r) = \sum(k_1, k_2)\, c(k_1, k_2) \cos k_1 Rp \cos k_2 Rq \exp(-\gamma Rr), \tag{2.29}$$

where the k's are to go over the values just stated, where the c's are coefficients to be determined, and where the solution written above applies only for positive r: for negative r, we use the exponential $\exp(\gamma Rr)$. The quantity γ is to be determined in terms of k_1 and k_2 by Eq. (2.28). We shall now apply the condition of continuity over the plane $r = 0$, but for the case where p and q are not both equal to zero.

To do this, we substitute (2.29) in the first equation of (2.27), for the special

case where r = 0, remembering the situation for negative r's. We find at once that the result, when we make use of (2.28), reduces to the statement that the expression (2.29), computed for r = -1, must equal the same expression computed for r = 1. In other words, though the expansion (2.29) is correct only for positive r's, this shows that it must still converge for r = -1 (so long as we do not have p = q = 0), and must give the correct value of the function at such a point. We can then write this condition in the form

$$\sum(k_1, k_2)\, c(k_1, k_2) \cos k_1 Rp \cos k_2 Rq \sinh \gamma R = 0. \tag{2.30}$$

We shall now show that all such equations, provided p and q are not both zero, can be satisfied by the simple assumption

$$c(k_1, k_2) = 1/\sinh \gamma R, \tag{2.31}$$

where γ is defined in terms of k_1 and k_2 by (2.28).

To do this, we note that in case (2.31) holds, the summation in (2.30) becomes $\sum(k_1, k_2) \cos k_1 Rp \cos k_2 Rq$. Such a sum is immediately shown to be zero on account of the familiar theorems on the orthogonality of the cosine functions. We may do it directly from the summation, or we may remember that we are going to pass to the limit of infinite N, and convert the sum into an integration. If we introduce two continuous variables, α_1 and α_2, such that $\alpha_1 = k_1 Rp$, $\alpha_2 = k_2 Rp$, then each of the α's is to range from 0 to π, in extremely small steps. Thus the summation over k_1 reduces to N times $\int_0^\pi \cos \alpha_1 \, d\alpha_1 = 0$, and similarly for the summation over k_2. We thus see that our expression (2.29), in which the c's are given by (2.31), forms an exact solution of our difference equations (2.27) everywhere, except that we have not yet considered the point p = q = r = 0.

If we substitute our solution in the second equation of (2.27), we find

$$\sum(k_1, k_2)(1/\sinh \gamma R)\left\{\mathscr{E}(0) + V(0) - E + 2\mathscr{E}(1)\left[\cos k_1 R + \cos k_2 R + \exp(-\gamma R)\right]\right\} = 0 \tag{2.32}$$

If we substitute for E from (2.28), we have

$$\sum(k_1, k_2)(1/\sinh \gamma R)\left[V(0) - 2\mathscr{E}(1) \sinh \gamma R\right] = 0 \tag{2.33}$$

We replace the summations over k_1 and k_2 by integrations over α_1 and α_2, as before, and write sinh γR in terms of E, α_1, α_2, using (2.28) and the equation $\cosh^2 x = 1 + \sinh^2 x$. Then (2.33) becomes

$$\int_0^\pi da_1 \int_0^\pi da_2 \left\{ V(0) \left[\left(\frac{E - \mathscr{E}(0)}{2\mathscr{E}(1)} - \cos a_1 - \cos a_2 \right)^2 - 1 \right]^{-1/2} - 2\mathscr{E}(1) \right\} = 0. \tag{2.34}$$

This can be put in the form

$$V(0)\, I = 2\pi^2 \mathscr{E}(1), \tag{2.35}$$

where

$$I = \int_0^\pi \int_0^\pi \left[\left(\frac{E - \mathscr{E}(0)}{2\mathscr{E}(1)} - \cos a_1 - \cos a_2 \right)^2 - 1 \right]^{-1/2} da_1\, da_2. \tag{2.36}$$

In (2.35) we have an equation determining the energy E as a function of the perturbative potential V(0), for the bound state whose wave function is given in (2.29).

The integral (2.36) has not yet been rigorously evaluated, and probably cannot be carried through completely in terms of well known functions. However, it is not hard to give a sufficiently accurate approximation to it for our present purposes. In the first place, we note that unless $\frac{E - \mathscr{E}(0)}{2\mathscr{E}(1)}$ is greater than 3, the integral will diverge, and we have no solution to our problem. The case where it is equal to 3 is easy to interpret. This corresponds to the case where $E = \mathscr{E}(0) + 6\mathscr{E}(1)$, which is the value of the energy at the center of the Brillouin zone, from (2.28). If $\mathscr{E}(1)$ is positive, this is the top of the energy band, and in this case we find a discrete state only when E is above the top of the band; when $\mathscr{E}(1)$ is negative, it is the bottom of the band, and we have a discrete state only below the bottom of the band. This of course is as it should be. Since the integral I by its definition is positive, this means that from (2.35) we have a solution when V(0) has the same sign as $\mathscr{E}(1)$, just as in the one-dimensional case. We shall not take up the opposite case, where the discrete level appears at the value of energy corresponding to the corner of the Brillouin zone rather than the center, but it can be handled as easily as in one dimension, merely by putting in a change of sign on the U's associated with adjacent lattice points.

Let the value of the integral I when $E = \mathscr{E}(0) + 6\mathscr{E}(1)$ be called I_0. The writer has made a rough numerical integration for this quantity, and it comes out so close to $\pi^2/2$ that one suspects that perhaps this is its exact value. (McCrea and Whipple, by a different numerical method, find the value 5.02, as compared with $\pi^2/2 = 4.94$.) For values of $\left[E - \mathscr{E}(0)\right]/2\mathscr{E}(1)$ greater than 3, the value leading to I_0, it is clear that the denominator of (2.36) is increased, and I decreased. Thus in such a case, to satisfy (2.35), V(0) must be increased. In other words, in order to have a discrete state, V(0) must be at least equal to $2\pi^2 \mathscr{E}(1)/I_0$, which is approximately $4\mathscr{E}(1)$, using

our approximate value of I_o. That is, from the difference equation as well as from the differential equation we have shown that there is no discrete state unless the perturbative potential is greater than a critical value. Furthermore, we have been able to evaluate this critical value, for the simple cubic lattice with nearest-neighbor interaction. The total width of the energy band in this case is $12\mathscr{E}(1)$ (since each of the cosines in an energy expression $E = \mathscr{E}(0) + 2\mathscr{E}(1)\left[\cos k_1 R + \cos k_2 R + \cos k_3 R\right]$ of the form of Eq. (2.4) can go from - 1 to 1). In other words, the quantity V(0) must be equal to at least a third of the width of the band, in order that there be a discrete state.

It is interesting, though probably not very significant, to compare the condition for the critical value of V(0) which we find here with the corresponding value for the square potential well and Schrödinger's equation. We have seen that in that case we have a discrete state provided V_o, the depth of the potential well, is greater than $h^2/(32mR_o^2)$, where R_o is the radius of the well. To compare these, we must substitute for the effective mass from Eq. (2.11): if the energy as a function of k is as given above, then (2.11) tells us that the effective mass is $-\hbar^2/2\mathscr{E}(1)R^2$ (the negative sign because it is positive at the bottom of a band, which comes at k = 0 if $\mathscr{E}(1)$ is negative). If we had a Schrödinger equation with this effective mass, then, we find that the critical value of V_o is $(\pi^2/4)(R^2/R_o^2)\,\mathscr{E}(1)$. We may ask what value R_o, the radius of the square well, must be, in order that this critical value of V_o should agree with $4\mathscr{E}(1)$, the critical value of V(0). Clearly $R_o = (\pi/4)R$, so that R_o is of atomic dimensions, R being the interatomic distance in the simple cubic lattice. The volume of the sphere within which the potential is V_o is $(4\pi/3)R_o^3$, which equals $(\pi^4/48)R^3$, or approximately $2R^3$, or twice the volume per atom in the simple cubic lattice. It seems likely that the solution of Schrödinger's equation for such a potential well would give a fairly good approximation to the solution of the difference equation.

Let us now proceed further with our discussion of the energy as a function of V(0). If $\frac{E - \mathscr{E}(0)}{2\mathscr{E}(1)}$ is only slightly greater than 3, so that the energy of the discrete level is only slightly outside the band, we can make an expansion of I. Let $\frac{E - \mathscr{E}(0)}{2\mathscr{E}(1)} = 3 + a$, where a will be a small quantity. Now for a = 0, the integrand of (2.36) has an integrable singularity at $\alpha_1 = \alpha_2 = 0$, and this singularity is removed if a becomes greater than zero; it is only in this part of the range of integration that a small value of a will have a significant effect on the integral. In this region, however, we can expand the integrand in powers of the α's, and convert it into a form which we can handle analytically. If we regard a and the α^2's as small quantities whose squares and products can be neglected, the integrand of (2.36) reduces to $(2a + \alpha_1^2 + \alpha_2^2)^{-1/2}$. For small α's, we can then introduce polar coordinates in an α_1--α_2 space, the radius being r, so that $r^2 = \alpha_1^2 + \alpha_2^2$. The integral can then be expressed in polar coordinates, where we only integrate over the first quadrant; that is, we replace $d\alpha_1\, d\alpha_2$ by

$(\pi/2)\, r\, dr$, and our integral becomes $\int_0 \frac{(\pi/2)\, r\, dr}{(2a + r^2)^{1/2}}$, where we have not written the upper limit, since when r becomes large our approximation is no longer valid. On the other hand, when r becomes large the effect of a becomes negligible, so that we need merely carry out the integral above to some arbitrary upper limit, with and without the term in 2a, to get the difference between I and I_o. We can carry out the integration at once by introducing $x = 2a + r^2$, $dx = 2r\, dr$, and find that

$$I - I_o = (\pi/4)\int_{2a}^{\infty} \frac{dx}{x^{1/2}}, \quad I = I_o - \frac{\pi}{2}(2a)^{1/2}, \tag{2.37}$$

which holds for small values of a. If we substitute in (3.35), we can easily transform to the form

$$\begin{aligned} E &= \mathscr{E}(0) + 6\mathscr{E}(1) + \frac{I_o^4}{\pi^6\, \mathscr{E}(1)} \left[(V(0) - \frac{2\pi^2}{I_o}\, \mathscr{E}(1)\right]^2 \\ &= \mathscr{E}(0) + 6\mathscr{E}(1) + \frac{\pi^2}{16\, \mathscr{E}(1)} \left[V(0) - 4\mathscr{E}(1)\right]^2 \end{aligned} \tag{2.38}$$

if we use the approximate value $I_o = \pi^2/2$. Thus we see that, just as in the one-dimensional case, the energy goes proportionally to the square of the perturbative energy for small displacements of the discrete level from the continuum; only here it is the increase of the perturbative energy over its critical value which is raised to the second power, rather than the perturbative energy itself.

We are also interested in the behavior for large values of V(0). For sufficiently large values of $\frac{E - \mathscr{E}(0)}{\mathscr{E}(1)}$, it is clear that the integrand of (2.36) approaches $\frac{2\mathscr{E}(1)}{E - \mathscr{E}(0)}$, independent of a_1 and a_2, so that I approaches π^2 times this quantity. As a matter of fact, if we expand the integrand of (2.36) in power series, we find that the term of next smaller order of magnitude than the leading one involves the integral of $\cos a_1 + \cos a_2$, which integrates to zero, so that this value for I is quite good. Thus we find that for large perturbations, $E = \mathscr{E}(0) + V(0)$, to quite a good approximation, showing the linear dependence of energy on the perturbation.

It is now instructive to draw the energy as a function of the perturbation V(0), using these two approximations. In Fig. 4, we show in this way E as a function of V(0). Clearly it is easy to interpolate between the two limiting curves, and get a very good estimate as to the behavior of the actual energy level. We have drawn the curve for a positive value of $\mathscr{E}(1)$, so as to indicate a discrete level above the bottom of a band, produced by a repulsive perturbation, if V(0) is positive. Thus this agrees with the case shown in Fig. 2 for the one-dimensional case. The difference between the

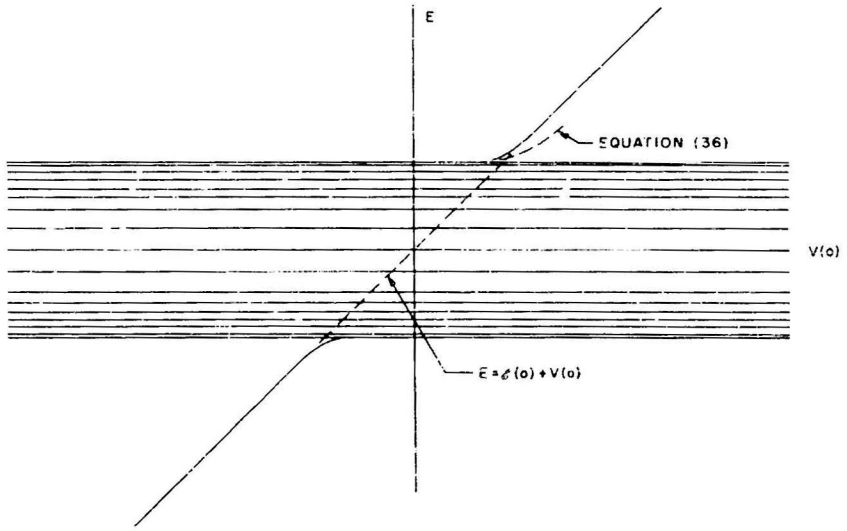


Fig. 4

Energy E as function of perturbative energy V(0), for discrete level in a three-dimensional simple cubic lattice.

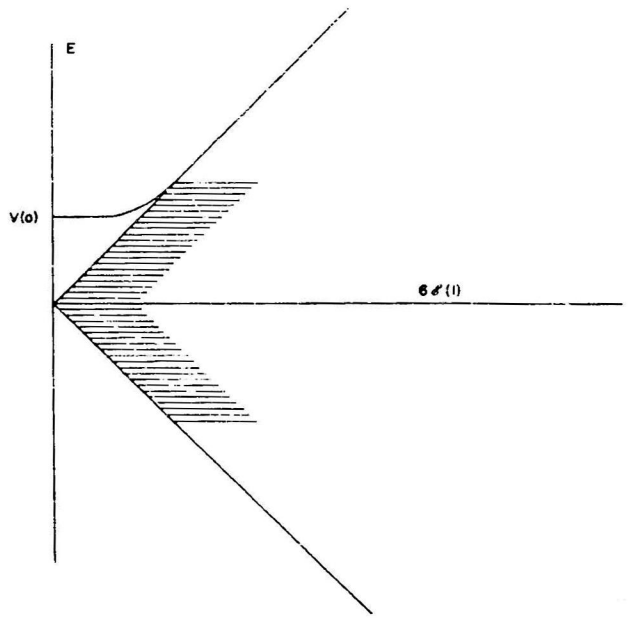


Fig. 5

Energy as function of band width, for constant perturbative potential.

two cases is clear: the discrete state in Fig. 4 does not begin to emerge from the continuum until V(0) reaches its critical value, whereas in Fig. 2 it emerges for infinitesimal V(0).

Another interesting way to show the results is given in Fig. 5. Here we show energy as a function of the band width $\mathscr{E}(1)$, keeping the perturbative potential V(0) constant. This shows what would happen if we started with a lattice of separated atoms, so widely separated that the energy levels were not broadened into a band. At zero band width, the one level of the impurity atom would lie above that of the rest of the atoms (if V(0) is positive). As the atoms are pushed closer together, and the band broadens, we see that it does not begin to perturb the discrete level very much until the band is so broad that it almost touches the discrete level. Then it repels the discrete level upward, so that the discrete level remains above the band but comes closer and closer to it, until finally the discrete level becomes incorporated into the band, at the point where the total band width is about three times V(0).

With this discussion, we have a fairly complete understanding of the problem of a single perturbing atom in a simple cubic lattice; since the results here are so close to those for the continuum theory, we may assume that the same thing will hold without qualitative change in other types of lattices. Now let us consider the change in the situation if the perturbative energy is felt over a number of atoms, instead of just on one atom. This problem is too hard to attempt to solve analytically, but it is not hard to make a good guess as to what will happen. We may expect, as in one dimension, that there will be a number of discrete levels, for sufficiently great perturbative potentials. As a general rule, the discussion which we have just given makes it fairly plain that as the perturbative potential increases, these discrete levels may split off from the continuum successively, rather than all at once. The various wave functions can have different types of symmetry. These functions, as we have mentioned earlier, must belong to the various irreducible representations of the point symmetry group of the crystal with respect to the center of the perturbation. Thus, if the perturbative potential and the crystal both have cubic symmetry, we shall have the various possible symmetry types of the cubic group, one being s-like, a threefold degenerate p-like type, and so on.

The particularly interesting problem in three dimensions is the Coulomb potential. The reason for this comes from its occurrence in semiconductors. We are not going at the moment to examine the nature of the perturbing field to be expected in various types of problems; this is a problem in self-consistent fields, rather than of solving a periodic potential problem, and we wish to separate these two parts of the question, first understanding the nature of the solutions of the one-electron Schrödinger equation before we build up determinantal many-electron wave functions and apply a condition of self-consistency. However, we may anticipate enough to state the general

situation encountered when an impurity atom is introduced substitutionally into a crystal. The result is quite different, depending on whether we are dealing with a good conductor like a metal, or a poor conductor like a semiconductor. Let us assume that the impurity atom has a nuclear charge greater or less than that of the crystal by a few units. If there were no shielding, this would introduce a perturbation of the Coulomb type, which outside the impurity atom itself would be proportional to the difference between the nuclear charges of the impurity atom and the atoms of the crystal. In an actual case, most of this field will be shielded by a rearrangement of electronic charge. There will usually be enough discrete levels with wave functions concentrated close to the impurity atom so that these levels can be occupied by electrons, or emptied, as the case may be, to make the impurity atom approximately uncharged. In a good conductor, the conduction electrons will rearrange their charge distributions so as almost completely to shield this Coulomb potential, just as any electric field in a metal is almost completely neutralized by the conduction electrons. Ordinarily in a metal, then, the perturbative potential arising from the self-consistent problem is confined almost entirely to the perturbing atom. In other words, we have almost precisely the case which we have been taking up so far in this section, leading to discrete levels if the perturbative term $V(0)$ is great enough, but to no new energy levels if it is below a critical value. We shall have to come back to this case, then, and its implications for the determination of the self-consistent field, when we discuss alloys, the effect of an atom of one type substituted in a crystal of another type.

An impurity atom in a metal, as we have just stated, has its field almost completely neutralized even at the position of its nearest neighbors by the conductivity, just as we should conclude from looking at the electrical properties of the metal from a macroscopic point of view. In a semiconductor, on the contrary, the conductivity is not great enough for this to happen, and the problem is much more like the electrostatic one of a charge in a dielectric. The dielectric constants of some of the important semiconductors are very high, that of germanium being of the order of magnitude of 16. The electric field of a point charge in such a dielectric will be less than in empty space, being inversely proportional to the dielectric constant. This dielectric effect of course is itself a manifestation of the self-consistent field. If we look at it microscopically, we see that the charge polarizes all the surrounding atoms, displacing their electrons with respect to their nuclei so as to produce dipoles on them, and it is well known that it is the superposition of the fields of these dipoles which cancels most of the field of the polarizing charge. We should get at this effect, in a completely logical treatment of our problem, by solving for the wave functions of the polarized atoms in the external polarizing field, which would be part of the self-consistent field, and making the problem self-consistent, which would imply that each atom had just such a dipole moment that the sums of these dipole fields produced

the dielectric effect by simple electrostatics. Ordinarily, however, we are willing not to look into the problem in such a fundamental way, but to accept the value of the macroscopic dielectric constant as being given, and to assume that it correctly describes the self-consistent effect of the polarizing charge in the dielectric.

We conclude, then, that if a dielectric contains an impurity atom which normally has an excess or deficiency of charge, the potential produced by this atom at a distance will be that produced by its net charge, computed for the actual dielectric constant. That is, we shall have a Coulomb potential, but ordinarily a good deal less in magnitude than in free space, on account of the large dielectric constant. We thus assume that if it is legitimate to replace the difference equation by the differential equation, we shall have Schrödinger's equation for a hydrogen-like problem, but with a potential decreased in proportion to the dielectric constant, and an effective mass instead of the real mass. The solutions will then be the ordinary hydrogenic solutions, but as we have stated earlier, the wave functions will be much more spread out, and the energies much less. In this case we have an infinite number of stationary states, instead of a finite number as with the perturbation located only on a finite number of atoms. Furthermore, we do not have any critical lower limit for the magnitude of the perturbative potential, below which we do not find discrete states; that is a characteristic of the problem where the perturbations are confined to a finite number of atoms.

We also find something in the Coulomb case which we do not in the case of a single perturbing atom: the differential equation forms a good approximation to the solution of the difference equation. The reason is that in this case the solution of the difference equation really varies slowly from atom to atom, and we have seen that this is the condition for the applicability of the differential equation method. Thus, for the ground state of hydrogen, the wave function is $\exp(-\gamma r)$, which remains finite at $r = 0$, and varies slowly from atom to atom, if γR is small compared to unity, where R is the interatomic distance. We have this situation in the cases we are interested in. This is in striking contrast to the case of a single perturbing atom, where we have seen that the wave function is of the form $\frac{\exp(-\gamma r)}{r}$. We can thus feel that the conventional derivation of the impurity levels in a semiconductor on the basis of hydrogenic wave functions is generally legitimate.

There is, however, a feature of the situation which has generally been overlooked, and which can be very important. The dielectric effect shields the field of the perturbing atom at the positions of all neighboring atoms, but it does not affect the perturbative potential on the perturbing atom itself. In other words, the perturbation $V(0)$, the average of the perturbative potential over the Wannier function located on the impurity atom, can be just as great in a semiconductor as in a metal. This is larger than the value which would be consistent with the hydrogenic problem, roughly in the ratio of the dielectric constant. In other words, the Schrodinger problem which is

really appropriate in this case is one with a Coulomb potential at all neighboring atoms, with a dielectric term to make it small, but in addition a potential well at the central atom. We may now legitimately combine the types of arguments which we have used in discussing the potential well problem, and the Coulomb problem. We may assume that unless the potential well becomes too deep, exceeding a critical value, the wave function and energy level will not be appreciably affected by it. In other words, it is likely that the hydrogenic wave function and energy level corresponding to the Coulomb potential can, so to speak, resist the effect of a perturbation on the central atom, provided this perturbation is not too large. The wave function will be modified in the immediate neighborhood of the central atom, but it extends over so many atoms that this modification is not very important, and will not change the energy very much. If the perturbation of the central atom becomes too great, however, the problem will change completely. The energy level will fall far below the hydrogenic value (we are assuming that the perturbing potential energy is negative, as it would be if the perturbing ion were positively charged, or as in the case of a donor atom below the bottom of a conduction band). At the same time, the wave function will become much more concentrated around the perturbing atom, as in the discrete levels in the problem of a single perturbing atom.

This situation can very likely occur in practice. We understand, in the first place, on the basis of this argument, why in so many cases we seem to get good agreement, quantitative as well as qualitative, between the simple hydrogenic theory and observed impurity levels in semiconductors. These are the cases in which the perturbative energy on the perturbing atom is less than the critical value. Different perturbative atoms, under these circumstances, would have almost, though not quite, the same wave functions and energy levels, the discrepancies coming only in the immediate neighborhood of the perturbing atom. On the other hand, we should also expect cases of greater perturbations, in which the whole character of the wave functions and energy level changes, and the energy level is much further below the conduction band (or much further above the valence band, if we are dealing with acceptor impurities). This may well be the explanation of the deep traps, which seem to occur particularly in silicon. Such trapping levels seem to explain the long time constants observed in some experiments with silicon,[1] and similar levels have been postulated in silicon after neutron bombardment;[2] these, of course, may well come from interstitial rather than substutional impurity atoms, and this is a problem slightly different from those we have discussed so far, but the principle is not different.

- - - - - - - - - - - - - - -

[1] J. R. Haynes and J. A. Hornbeck, Phys. Rev. 90, 152 (L), (1953).

[2] W. E. Johnson and K. Lark-Horovitz, Phys. Rev. 76, 442 (1949).

Finally, we should mention that if the dielectric constant of a semiconductor or insulator is not very great, the Coulomb type of problem will lead to a wave function which is no longer extended over many atoms, but is more concentrated. In such a case it will no longer be justified to replace the difference equation by the differential equation, and we must antitipate quantitative errors in the use of the hydrogenic solution. We hardly expect, however, any very striking qualitative change in the situation.

4. Wannier Functions for Overlapping Bands

In our discussion of the preceding section, we have assumed that we were dealing with a non-degenerate band of s-like character, and that the effective masses in the x, y, and z directions were identical. In many actual cases, however, notably in germanium and silicon, these assumptions are not correct. We have degenerate or overlapping bands, with their various Wannier functions. In our discussions of Chapter 1, we did not take up the nature of the Wannier functions in such a case, since we were not going to encounter these problems in that chapter. Now, however, these questions are essential. We meet a problem having two aspects. First, what are the Wannier functions like in cases of degenerate bands? Secondly, it seems almost obvious in such a case that we must use the Wannier functions of the various overlapping bands in order to get a proper representation of the impurity wave function. We must then look into the question, which we have avoided so far, of building up a solution from the Wannier functions of several bands, rather than of one band only. The way in which this will enter our problem, of course, will be that the perturbative potential will have non-diagonal matrix components between Wannier functions of different bands. We shall have to look into the whole question of the non-diagonal matrix components of the perturbative energy, a question with far-reaching consequences.

Before going into the question of the Wannier functions for overlapping bands, we must remind ourselves of the nature of overlapping bands themselves, which we have discussed in Chapter 1. The method which we found most informing in a qualitative way for discussing this problem was the simplified tight-binding approximation, which we took up in Sections 9 - 12 of Chapter 1. Let us use this method in a first survey of the situation. We remember that the general method starts by setting up Bloch sums of a number of atomic orbitals, which can be of s, p, d, etc., type. We then calculate the matrix components of the one-electron Hamiltonian between these Bloch sums, making many simplifications such as disregarding overlap effects, three-center integrals, and related quantities. We solve the secular equation between the various unperturbed wave functions, to get the energy levels, and the appropriate combinations of unperturbed wave functions giving the best approximations to solutions

of Schrödinger's equation.

Let us assume that our original atomic orbitals are $\phi_n(\vec{r})$, so that an atomic orbital located on an atom in the cell at displacement $\vec{R}_i$ from the origin is $\phi_n(\vec{r} - \vec{R}_i)$. Let us assume that we have solved the secular equation just described, so that the wave function corresponding to the propagation vector $\vec{k}$ can be written

$$B_{m,\vec{k}}(\vec{r}) = \sum(n, \vec{R}_i)\, f_{mn}(\vec{k}) \exp(i\vec{k} \cdot \vec{R}_i)\, \phi_n(\vec{r} - \vec{R}_i). \tag{2.39}$$

This, as we see, is a linear combination of Bloch sums formed from the various atomic orbitals ϕ_n; the coefficients, $f_{mn}(\vec{k})$, depend on the propagation vector $\vec{k}$, as we indicate explicitly. We have spoken as if this determination of the wave functions depended on our simplified tight-binding approximation, but we note that the exact solution of a tight-binding method, taking account of all refinements, would still have the form (2.39), though the coefficients $f_{mn}(\vec{k})$ would have somewhat different numerical values.

A question can well arise concerning the assignment of indices m to the various energy bands. If we have a certain number of atomic orbitals, then the secular equation will yield an equal number of functions $B_{m,\vec{k}}(\vec{r})$. In general, however, there will be no one-to-one correspondence between one of the final energy bands, and one type of atomic orbital. The only thing which we can do to name the bands, or to give them the indices m appearing in the notation $B_{m,\vec{k}}(\vec{r})$ for the wave functions, is to number them up from the lowest energy band. We cannot even identify a single band as an s, or a p, or a d band: in general, we shall have mixtures of all types of atomic orbitals, in each band, though in some cases, as in the tightly bound states, we shall have a band whose wave functions are composed almost entirely of one type. The Wannier functions associated with one band, then, will be combinations of all the different types of atomic orbitals. Sometimes we have a single band, widely separated from others, so that the wave functions of that band are composed almost entirely of one type of atomic orbital. Then the Wannier function will also be composed almost entirely of that type of orbital; but with overlapping bands, like the p and d bands, this will no longer be the case, and in a p band each Wannier function will be made up in a complicated way from the p_x, p_y, p_z atomic orbitals.

There is one very important feature of the bands, and their relation to the Wannier functions: two bands, as we have defined them, can never be degenerate with each other. This is a loose way of describing the following situation. Except at certain symmetry points in the Brillouin zone, all the energy levels of the system will be separated from each other; there will be no degenerate states for an arbitrary $\vec{k}$ value. If then we associate all the lowest energy levels, for all $\vec{k}$ values, to form the lowest energy band, all the next lowest levels to form the second band, and so on,

these bands will all be distinct from each other. The average energy of each will lie higher than the average energy of the next band below it. It is in this sense that there is no degeneracy. But there is an immediate consequence of this: there cannot be any degeneracy in the Wannier functions, in the sense that the Wannier function of one band cannot transform into the Wannier function of another band by one of the symmetry operations of the crystal. We cannot have Wannier functions, for instance, which have the symmetries of p_x, p_y, and p_z atomic functions, for in such a case one of them would transform into another by one of the symmetry operations of the crystal. Rather, in the p case, we have three quite distinct Wannier functions, one corresponding to the lowest p band, the second to the next lowest, the third to the highest. The diagonal matrix components of energy of these three Wannier functions (the quantities which we have been denoting as $\mathscr{E}(0)$) will thus be different for the three, and so will the other matrix components. And each of these three Wannier functions must be built up in some sort of symmetrical way out of p_x, p_y, and p_z atomic orbitals.

Crystal symmetry gives us quite definite information about the Wannier functions, even though it does not lead to degeneracy. It is of course clear that a Wannier function must be transformed into other Wannier functions by any operation of the group of crystal symmetry operations; but we have just seen that there can be no degeneracy, or that really a Wannier function cannot be transformed into a distinct Wannier function by one of these operations. The only alternative is that it should be transformed into a multiple of itself. Thus we arrive at the important fact that a Wannier function must belong to a one-dimensional irreducible representation of the symmetry group of the crystal. For instance, in the cubic group, we remember that there are four one-dimensional representations, which we can describe easily as follows: the first transforms into itself under all operations of the group; the second changes sign on reflection in the planes $x = 0$, $y = 0$, and $z = 0$, but not on reflection in the planes $x = \pm y$, $y = \pm z$, $z = \pm x$; the third changes sign on reflection in the planes $x = \pm y$, etc., but not in $x = 0$, etc; and the fourth changes sign on reflection in each set of planes. There are correspondingly four one-dimensional representations in the group of the square. We now see that each Wannier function, no matter whether made up of s, p, d, or what type of atomic orbitals, must transform according to one of these very limited types of symmetry.

It is somewhat puzzling at first sight to see how the atomic orbitals are combined to give Wannier functions of quite different symmetry type. We can understand it clearly, however, if we examine the detailed way in which the Wannier functions are set up. In Eq. (1.25), Chapter 1, we find that the Wannier function $a_m(\vec{r} - \vec{R}_j)$ can be written in terms of the energy band wave function, which we are here writing $B_{m,\vec{k}}(\vec{r})$, by the equation

$$a_m(\vec{r} - \vec{R}_j) = N^{-1/2} \sum(\vec{k}) B_{m,\vec{k}}(\vec{r}) \exp(-i\vec{k} \cdot \vec{R}_j). \quad (2.40)$$

Here the summation over $\vec{k}$ is to be carried over the N distinct $\vec{k}$ values in the unit cell in $\vec{k}$ space. We can then substitute for $B_{m,\vec{k}}(\vec{r})$ from (2.39). In doing this, it is well to take advantage of the fact that the coefficient $f_{mn}(\vec{k})$ must be a periodic function in $\vec{k}$ space; we know this, since both the exponential functions $\exp(\vec{k} \cdot \vec{R}_i)$ and the Bloch functions $B_{m,\vec{k}}(\vec{r})$ are periodic functions. Thus we can expand the f's in the form

$$f_{mn}(\vec{k}) = \sum(\vec{R}_\ell) F_{mn}(\vec{R}_\ell) \exp(i\vec{k} \cdot \vec{R}_\ell). \quad (2.41)$$

If we now substitute (2.39) and (2.41) in (2.40), we have

$$a_m(\vec{r} - \vec{R}_j) = N^{-1/2} \sum(\vec{k}, n, \vec{R}_i, \vec{R}_\ell) F_{mn}(\vec{R}_\ell) \exp(i\vec{k} \cdot (\vec{R}_i - \vec{R}_j + \vec{R}_\ell)) \phi_n(\vec{r} - \vec{R}_i). \quad (2.42)$$

We may now consider separately the sum over $\vec{k}$ in (2.42). We remember that such a sum is zero, unless the quantity $\vec{R}_i - \vec{R}_j + \vec{R}_\ell = 0$. In that case, since we are summing over N lattice points, it is N. Thus we can reduce the expression (2.42) to

$$a_m(\vec{r} - \vec{R}_j) = N^{1/2} \sum(n, \vec{R}_\ell) F_{mn}(\vec{R}_\ell) \phi_n(\vec{r} - \vec{R}_j + \vec{R}_\ell). \quad (2.43)$$

Eq. (2.43) is a very important result showing how the Wannier function $a_m(\vec{r} - \vec{R}_j)$, for the mth band and the unit cell at $\vec{R}_j$, is written in terms of atomic orbitals $\phi_n(\vec{r} - \vec{R}_j + \vec{R}_\ell)$, of the nth type, located in cells displaced by $\vec{R}_\ell$ from the cell where we are finding the Wannier function. It shows us at once that the way to build up the Wannier function of a band arising from degenerate atomic orbitals is to superpose atomic orbitals, not only on the atom where we are finding the Wannier function, but on neighboring atoms as well; and it shows us that the way to find the coefficients is to get the Fourier coefficients $F_{mn}(\vec{R}_\ell)$ of the transformation coefficients $f_{mn}(\vec{k})$ transforming the original atomic orbitals to the final Bloch functions.

We can now easily work out special cases, to see the sort of superposition of atomic orbitals actually arising in cases met in practice. Thus the two-dimensional p degeneracy in the square lattice has been discussed in Chapter 1, pp. 46-47. We can divide the square Brillouin zone into four regions, by the diagonal lines $k_y = \pm k_x$. In one of these regions, k_x is positive and greater numerically than k_y; in the second, k_y is positive and greater numerically than k_x; in the third, k_x is negative and greater numerically k_y; and in the fourth k_y is negative and greater numerically than k_x. In the first and third of these regions the lower band corresponds to a linear combination

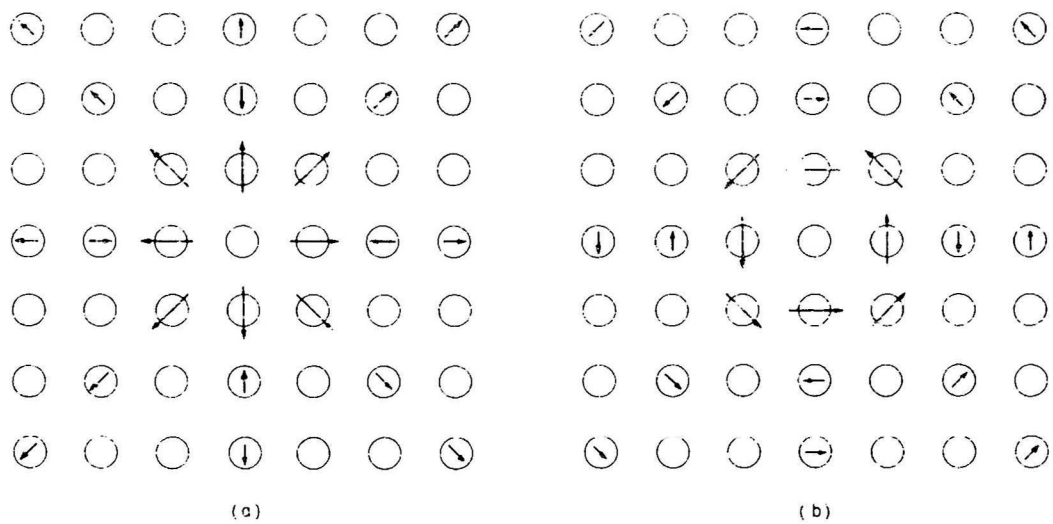


Fig. 6

Wannier function for lower (a) and upper (b) bands arising from degeneracy of two p states in two-dimensional square lattice.

which is composed almost entirely of p_x, hardly at all of p_y, while in the second and fourth the lower band is almost entirely p_y, hardly at all p_x. Thus the function f for the coefficient of the p_x function in the expansion of the lower energy band is almost unity (or a constant, correct to produce normalization) in the first and third regions, almost zero in the second and fourth. Examination of the symmetry of the Wannier function shows that to obtain a function whose symmetry belongs to one of the irreducible representations of the square group, the function f must as a matter of fact have opposite signs in the first and third regions mentioned above, and to get a real Wannier function it must be pure imaginary. We can then approximate the problem by setting up this f equal to i in region 1, - i in region 3, and zero in regions 2 and 4. Similar situations are encountered for the other f's.

The writer[1] has shown how to expand these approximate f's in Fourier series of the form (2.41), and the corresponding expansions of the Wannier functions, of the form (2.43), were arrived at. The results, for the lower and upper p bands, were presented graphically, in the form shown in Fig. 6. Here the circles represent atoms; the Wannier function is being computed for the central atom in each figure. They consist of p-like unperturbed atomic functions on adjacent atoms, pointing in the direction of the arrows, and with amplitudes indicated schematically by the lengths of the arrows. There is no contribution in either case on the central atom itself. We can now

---

[1] J. C. Slater, Phys. Rev. 87, 807 (1952); see particularly pp. 828-833.

see clearly from Fig. 6 how it can be that the p-like atomic orbitals can combine to give a Wannier function whose symmetry properties are quite different from those of the individual orbitals. In the case (a), it is clear that the Wannier function belongs to the representation of the square group which transforms into itself under each operation of the group. In case (b), on the contrary, the Wannier function belongs to the representation which changes sign on reflection in the lines $x = 0$ and $y = 0$, and also on reflection in the lines $x = \pm y$. These, we remember, are two out of the four possible symmetry types for Wannier functions in the square lattice.

One interesting feature of the Wannier functions is observed in Fig. 6: there is no contribution to the Wannier function from atomic orbitals on the central atom, where the Wannier function is centered. This must clearly always be the case if the symmetry of the Wannier function is different from the symmetry type of the atomic orbitals of which it is composed. The case of an s-like atomic orbital is one of the few cases in which an atomic orbital possesses the same symmetry type as the Wannier function; in p and d atomic orbitals each of the orbitals belongs to a two- or three-dimensional representation of the cubic point group, and since the Wannier functions must belong to one-dimensional representations, we cannot have any contributions on the central atom when we are dealing with Wannier functions derived from p or d orbitals. This has the result, of course, that the Wannier functions are concentrated on atoms distant from the central one on which they are formally located, so that they really are much more extended in space than has been generally assumed. The reason for this general misconception is that almost all the discussions of Wannier functions have disregarded orbital degeneracy, and have been carried out as if we were dealing with s orbitals.

The extended nature of the Wannier functions means that there are matrix components of energy between much more distant neighbors than we should have if they were more like single atomic orbitals. We can see this also from Eqs. (2.2) and (2.4). These show us that the matrix component of energy between two Wannier functions displaced by a vector $\vec{R}_i$ with respect to each other equals the corresponding Fourier component in the Fourier expansion of the energy $E(\vec{k})$ of the energy band, as a function of $\vec{k}$. Now this energy $E(\vec{k})$ is found as the eigenvalue of the secular problem of making linear combinations of atomic orbitals to set up the Bloch functions. In such a case as that of the two-dimensional square problem of p degeneracy, which we have mentioned earlier, the energy as a function of $\vec{k}$ can be quite complicated, consisting approximately of the energy of the Bloch function formed from p_x orbitals over part of the Brillouin zone, approximately the energy of the function formed from p_y over the rest. At the center of the Brillouin zone, as we saw in Section 7, Chapter 1, the energy has a singular point, and cannot be expanded in power series; its second derivative is discontinuous at that point. We must expect, then, that the Fourier

expansion of the energy will converge very slowly. This means that there will be components $\mathscr{E}(\vec{R}_i)$ connected with quite large values of $\vec{R}_i$; in other words, the interactions with quite distant neighbors are important. This situation, then, is quite different from that with s-like Wannier functions, where it can be quite a good approximation to consider only interactions with perhaps nearest- and second-nearest neighbors. This can lead to considerable increase in the difficulty of handling the difference equation solution for impurity levels; it can also make it much less legitimate than in the simpler cases to replace the difference equation by a differential equation.

We may well be disconcerted to find that we do not have Wannier functions transforming according to the same symmetry behavior as the atomic orbitals. It is, of course, perfectly possible to have impurity wave functions transforming with the same symmetry properties as atomic orbitals, and we might well wish to have Wannier functions transforming in the same way. The reason why we do not have this is the convention which we have made regarding the definition of the bands. Koster[1] has shown that it is possible to define the bands differently, in such a way as to obtain Wannier functions having the same transformation properties as the atomic orbitals. Thus, for instance, in the two-dimensional square lattice case of the p bands, which we have been using as an illustration, we can proceed as follows. We remember our four regions in the Brillouin zone, 1, 2, 3, and 4, such that in region 1, k_x is positive and its magnitude is greater than that of k_y, and so on. Let us assume that we know the energy of each of the two bands precisely in our Brillouin zone. Now let us define a new band as follows: in region 1, it will consist of the lower of the two energy levels; in 2, of the higher; in 3, of the lower; and in 4, of the higher. Along the lines $k_x = \pm k_y$, the energy will be discontinuous. From this band, let us define a Wannier function in the usual way: it will then prove to have the symmetry behavior of a p_x orbital. In fact, if the non-diagonal matrix component of energy between the Bloch functions formed from p_x and p_y orbitals is vanishingly small, this Wannier function will be precisely the p_x function. Similarly we can define a second band, consisting of the higher of the two energy bands in regions 1 and 3, the lower in 2 and 4; the Wannier function defined from it will behave like a p_y orbital. Under the action of the symmetry operations of the group, these two Wannier functions will transform like a two-dimensional irreducible representation of the transformation group of the square.

This shows us, then, that we can introduce Wannier functions in various different ways, though only the way we first described is consistent with defining energy bands in which the energy is a continuous function of $\vec{k}$. The new Wannier functions must, of course, be linear combinations of the old ones. We can if we choose build

---

[1] G. F. Koster, Phys. Rev. 89, 67 (1953).

up the solution of the problem of the energy level around an impurity atom as a linear combination of these new Wannier functions, rather than of the original ones. But once we realize this, we see that we can really accomplish the same thing, in the tight binding approximation, in a much more straightforward way. Our Wannier functions of the original sort are linear combinations of the atomic orbitals. By making linear combinations of these Wannier functions, we secure other Wannier functions having the same symmetry as the atomic orbitals. But now we see that we could have saved all this work. Any linear combination of all the Wannier functions formed from a set of atomic orbitals ϕ_n can equally well be written directly as a linear combination of the original ϕ_n's themselves. As far as the simplified tight binding approximation is concerned, we can treat the ϕ_n's themselves as the starting point of the calculation of impurity levels, and as we shall see immediately, the same matrix components of energy between these atomic orbitals which we assume to get a correct representation of the energy bands can also be used to discuss impurity levels. This, then, forms a very straightforward method of discussing such levels, which we shall consider next.

5. Impurity Levels with the Tight Binding Approximation

In the preceding section we have shown that if we start with atomic orbitals, apply the simplified tight binding approximation, build up energy bands and Bloch wave functions in this way, solving the secular equation so as to diagonalize the energy, and then derive Wannier functions from the resulting energy band functions, the final Wannier functions are linear combinations of the original atomic orbitals, located on many different atoms. Our next step in building up a solution of the perturbed periodic potential problem would be to expand the solution as a linear combination of the Wannier functions on the various atoms. We are now concerned with the case where a number of bands are close enough together so that the Wannier functions of all these bands must be used to get a good representation of the final wave function. In this case, as we have just pointed out at the close of the preceding section, it is an entirely unnecessary complication to introduce the Wannier functions at all. We can get just the same approximation by expanding our solution directly in terms of the original atomic orbitals. We shall now show that this procedure can be formulated very conveniently. It seems likely that, in the present state of development of the theory of energy bands, it is the most usable method for handling the energy levels of impurity atoms with overlapping bands.

In Eq. (59), Chapter I, we set up an equation for the matrix components of the periodic potential function between two atomic orbitals. If V_o is the periodic potential, the matrix component between orbitals $\phi_n(\vec{r})$ and $\phi_m(\vec{r} - \vec{R}_i)$, where n and m refer to two different types of orbitals, such as an s and a p_x orbital, and where

$\vec{R}_i$ is a lattice displacement, was written

$$\int \phi_n^*(\vec{r})\, V_o(\vec{r})\, \phi_m(\vec{r} - \vec{R}_i)\, dv = \mathscr{H}_{nm}(\vec{R}_i). \tag{2.44}$$

We called the diagonal component of the Hamiltonian, when $n = m$, $\vec{R}_i = 0$, by the symbol E_m, but there is no reason why we cannot use the symbol $\mathscr{H}_{nm}(0)$ for this quantity. We disregarded all lack of orthogonality between the ϕ's on different atoms. We set up in Table 1, Chapter 1, methods of reducing these quantities $\mathscr{H}_{nm}(\vec{R}_i)$ to simpler expressions in terms of two-center integrals. Finally, we used these expressions as disposable parameters, to be determined in such a way that the energy bands derived from the method should fit the results of more accurate calculation, as for instance by the orthogonalized plane wave method.

We were interested in Chapter 1 in solving the periodic potential problem, without perturbations. We did this, though we did not write the steps down explicitly, in the following way. We assumed that the wave function could be written in the form of a sum

$$\psi = \sum(n, \vec{R}_i)\, U_n(\vec{R}_i)\, \phi_n(\vec{r} - \vec{R}_i), \tag{2.45}$$

where the U_n's are coefficients to be determined. Then we wrote the Schrödinger equation for this wave function ψ, and set up the corresponding secular equation for the coefficients. This secular equation, making the assumptions (2.44) regarding the matrix components of energy, was

$$\sum(m, \vec{R})\, \mathscr{H}_{nm}(\vec{R}_i - \vec{R}_j)\, U_m(\vec{R}_j) = E U_n(\vec{R}_i). \tag{2.46}$$

We then assumed that the coefficients $U_n(\vec{R}_i)$ could be written in the Bloch form $c_n \exp(i\vec{k} \cdot \vec{R}_i)$. If we make this substitution in (2.45), we have a wave function of the form already written in Eq. (2.39). The secular equation (2.46) becomes simplified, in that we find that our assumption automatically satisfies the equation for all values of $\vec{R}_i$, provided only we satisfy the secular equation

$$\sum(m)\, c_m \left[\sum(\vec{R}_j)\, \mathscr{H}_{nm}(\vec{R}_j) \exp(-i\vec{k} \cdot \vec{R}_j)\right] = E\, c_n. \tag{2.47}$$

In other words, the c_n's are the amplitudes of various Bloch waves, and the matrix components of energy between these Bloch waves are as given in the square bracket of (2.47), agreeing with Eq. (60) of Chapter 1. We saw in Chapter 1 how often a relatively small secular equation of the form of the form of (2.47) could give satisfactory results. Thus, in the diamond structure, we used eight atomic orbitals, and s and

a p_x, p_y, p_z on each of the two types of atoms in the unit cell, so that we had an eight-by-eight secular equation, whose eigenvalues E gave the energy bands described in Chapter 1.

We notice that Eq. (2.46) is very much like Eq. (2.5), in which we set up the difference equations for the Wannier functions. The only difference (aside from the fact that we have not yet introduced a perturbative potential like the V_{nm} of Eq. (2.5)) is that we now have non-diagonal matrix components $\mathscr{H}_{nm}(R_i - R_j)$ between ϕ's with different subscripts n and m, whereas in (2.5) there are no matrix components between Wannier functions in different bands. This distinction is fundamental: it is by introducing such non-diagonal matrix components between different atomic orbitals that we were able in Chapter 1 to set up the secular equation which gave the whole behavior of the energy bands. We can get rid of these non-diagonal matrix components only by solving (2.47) for the energy band wave functions, defining Wannier functions from these, and setting up difference equations in terms of the Wannier functions. We gain one thing by doing this: we have no non-diagonal matrix components of energy between Wannier functions in different bands. But we lose something: we have the extended type of Wannier function discussed in the preceding section, with non-diagonal matrix components of energy between Wannier functions in the same band on quite distant atoms. In the present section we are exploring the advantages of using the original atomic orbitals, in which we must consider several different energy bands, but in which we can assume that the non-diagonal matrix components $\mathscr{H}_{nm}(\vec{R}_i - \vec{R}_j)$ fall off quite rapidly as the atoms get further apart. For most purposes, it seems that the advantages of this method outweigh the disadvantages.

Let us now supplement Eq. (2.46) by inserting perturbation terms, like those in Eq. (2.5). We can write them in the same form as there. That is, we have

$$\sum(m, \vec{R}_j)\left[\mathscr{H}_{nm}(\vec{R}_i - \vec{R}_j) + V_{nm}(\vec{R}_i, \vec{R}_j)\right] U_m(\vec{R}_j) = EU_n(\vec{R}_i). \qquad (2.48)$$

Except for the one fact that our quantity $\mathscr{H}_{nm}(\vec{R}_i - \vec{R}_j)$ has components if $n \neq m$, this equation is identical with (2.5). We must remember, of course, that the rigorous tight binding calculation would be much more complicated than this. We should have to include terms coming from lack of orthogonality of atomic orbitals on different atoms, from three-center integrals, and so on. But we have already seen the usefulness of the present approximations in the case of finding energy bands of unperturbed lattices. The advantage of the present scheme is that we can now use the same $\mathscr{H}_{nm}$'s for both the unperturbed band and the problem of the perturbed lattice, so that the perturbed energy levels will be consistent with a correct solution, to the same accuracy to which we have been able to fit the energy bands to a correct solution. The V_{nm}'s, of course, represent matrix components of a perturbative potential,

such as that produced by an impurity atom, between different atomic orbitals; and just as with the $\mathscr{H}_{nm}$'s, we might well want to treat these as adjustable parameters, to approximate more accurate solutions determined by other methods, rather than computing them from actual atomic orbitals.

Now that we have set up Eq. (2.48), and understand its implications, we can at once answer the question, how many atomic orbitals, and therefore how many energy bands, must we take into account in setting up the energy band problem properly. Thus, for instance, in diamond or silicon or germanium, if we have perturbed energy levels near the conduction or valence bands, we shall want to use the atomic s and p orbitals on atoms of both types in the unit cell, or eight orbitals, just as in discussing the energy bands; and we cannot expect to obtain a correct picture of the perturbed levels in any simpler way. Similarly, in a simple cubic crystal in which the p levels are well separated from s and d bands, we should have to use the p_x, p_y, p_z orbitals as a minimum. Ordinarily, in practice, the s levels lie so close to the p that we cannot do with less than an s orbital as well as the p's.

6. General Method for Solving the Difference Equations

In the preceding sections, we have seen that our general problem of finding the energy levels in a perturbed periodic lattice can be reduced to the solution of certain difference equations: Eq. (2.5) if we are formulating the problem in terms of Wannier functions, (2.48) if we are using the tight binding approximation. We have solved these difference equations in a number of special cases: the one-dimensional chain with nearest and second-nearest neighbor interactions in Section 2, the three-dimensional simple cubic lattice with nearest-neighbor interactions in Section 3. Our methods have been rather complicated, and applicable only to special cases. Fortunately, however, a far more general method of solving these difference equations exists, reducing to the values previously obtained, but capable of being applied to a far wider range of problems. This method is now being worked out by the writer and Dr. G. F. Koster, and was not available at the time the earlier sections of this chapter were written. Its general outline will be presented here, though its applications have not yet been carried further than the examples already discussed.

This method is based on a suggestion of Dr. Koster, to expand the perturbed function, not in terms of Wannier functions directly, but in terms of Bloch functions. We shall formulate it first for the simplest case, of a single non-overlapping band with a perturbation only at the atom at the origin, but then shall extend it to the general case. Let us use periodic boundary conditions, with N atoms in the repeating region. We have Wannier functions $a(\vec{r} - \vec{R}_i)$, for the band in question. Then the corresponding Bloch functions are

$$B_{\vec{k}}(\vec{r}) = N^{-1/2} \sum(\vec{R}_i)\, e^{i\vec{k}\cdot\vec{R}_i}\, a(\vec{r} - \vec{R}_i). \tag{2.49}$$

These functions are normalized and orthogonal, and there are N of them, corresponding to N $\vec{k}$ values, as shown in Chapter 1. The matrix component of the perturbative energy between two of these Bloch functions will be

$$N^{-1} \sum(\vec{R}_i, \vec{R}_j)\, e^{i(-\vec{k}\cdot\vec{R}_i + \vec{k}'\cdot\vec{R}_j)}\, V(\vec{R}_i, \vec{R}_j), \tag{2.50}$$

where $V(\vec{R}_i, \vec{R}_j)$ is the matrix component of the perturbation between the corresponding Wannier functions. The special case which we shall take up first is that in which $V(\vec{R}_i, \vec{R}_j)$ is zero unless both $\vec{R}_i$ and $\vec{R}_j$ are zero, in which case it is V_o, as in previous applications. Then the matrix component (2.50), between the Bloch functions corresponding to propagation constants $\vec{k}$ and $\vec{k}'$, is merely V_o/N, the same for all pairs of propagation constants. It is this very simple form of the matrix component which leads to a very simple solution for this problem.

Let us now set up our solution as a series of Bloch functions, $\sum(k)\, F_k B_k(r)$. Then, remembering that the periodic potential problem has a diagonal matrix with respect to the Bloch functions, with matrix components E_k, the equations for the F_k's immediately become

$$E_k F_k + \sum(k')(V_o/N)\, F_{k'} = EF_k. \tag{2.51}$$

We can immediately solve these equation, in the form

$$F_k = \frac{(V_o/N) \sum(k')\, F_{k'}}{E - E_k} \tag{2.52}$$

Since $\sum(k')\, F_{k'}$ is a constant independent of k, this at once tells us that F_k is a constant divided by $E - E_k$, which gives a complete solution of the problem of determining the wave function; the constant can be determined by the normalization condition. We can at once find an equation for the energy. Let $(V_o/N) \sum(k')\, F_{k'} = A$. Then from (2.52), $F_k = A/(E - E_k)$. We substitute this in the definition of A above, and find that

$$(V_o/N) \sum(k) \frac{1}{E - E_k} = 1 \tag{2.53}$$

Here V_o, we remember, is the perturbative potential on the atom at the origin, N the number of atoms in the repeating region, the E_k's are the energies of the N energy

levels in the unperturbed energy band. Eq. (2.53) then is an equation for E, determining the energies of the perturbed states as a function of the perturbative potential V_o.

This almost ridiculously simple result is the complete solution for the energy levels of the perturbed problem, when our conditions are satisfied, of a single non-degenerate band, and a perturbing potential only on a single atom. It includes the cases of interaction with neighbors at any distance, including the two- and three-dimensional problems. We shall show in a moment that it reduces to our earlier solutions in the cases taken up before. Before doing this, however, let us consider the general nature of Eq. (2.53), and make some deductions about the energy levels. The function $\sum(k)\, 1/(E - E_k)$ becomes infinite when E equals any one of the E_k's. Its reciprocal, which according to (2.53) equals V_o/N, becomes zero at each of the E_k's, so that if V_o is zero, the solutions are just the E_k's. Between each pair of zeros of this reciprocal function, there will be a simple pole. Thus as V_o goes from zero to positive or negative infinity, each of the energy levels will be displaced, but not further than the next adjacent E_k. The exceptional cases come if we are looking for a solution outside the unperturbed energy band. As the energy gets well outside the unperturbed band, we may approximate each term $E - E_k$ by $E - \mathscr{E}(0)$, where $\mathscr{E}(0)$ is the average energy in the band. Thus the summation will be approximately $N/(E - \mathscr{E}(0))$, so that (2.53) will become approximately $E = \mathscr{E}(0) + V_o$. But this behavior is just what we found in the specific cases we worked out earlier, the discrete state being pushed out of the band, the remaining states of the band being only slightly perturbed. We now see that this behavior is perfectly general. Eq. (2.53), giving V_o as a function of E, or E as a function of V_o, is thus the general equation of curves like those of Figs. 2 and 4.

To show the power of our method, we shall use it to discuss the one-dimensional chain with nearest-neighbor interaction, and then the three-dimensional simple cubic case, which we have taken up earlier. For the one-dimensional case, we know that $E_k = \mathscr{E}(0) + 2\mathscr{E}(1) \cos kR$. We must then evaluate the sum $\sum(k)\, 1/(E - E_k)$ in (2.53). In case we are dealing with the discrete state, so that E lies outside the unperturbed band, we can convert this sum into an integral without trouble. We find that kR goes by equal steps from $-\pi$ to π, or from 0 to 2π, there being N intervals. That is, we can replace $\sum(k)$ by $(N/2\pi) \int_0^{2\pi} d(kR)$. Thus (2.53) reduces to

$$(V_o/N)(N/2\pi) \int_0^{2\pi} (E - \mathscr{E}(0) - 2\mathscr{E}(1) \cos \alpha)^{-1}\, d\alpha = 1. \tag{2.54}$$

This can be integrated by elementary means, and gives

$$V_o / \sqrt{(E - \mathscr{E}(0))^2 - (2\mathscr{E}(1))^2} = 1. \tag{2.55}$$

When we solve (2.55) for the energy E, the result is identical with Eq. (2.19), which we obtained for the discrete state by our straightforward solution of the problem.

Similarly we can take the three-dimensional simple cubic lattice. Here $E_k = \mathscr{E}(0) + 2\mathscr{E}(1)(\cos k_1 R + \cos k_2 R + \cos k_3 R)$, as we found in Section 3, where k_1, k_2, k_3 are the three components of the vector k. Just as before, we can convert the sum into an integral, replacing $\sum(k) b_j (N/(2\pi)^3) \int_0^{2\pi} d(k_1 R) \int_0^{2\pi} d(k_2 R) \int_0^{2\pi} d(k_3 R)$. When we substitute in (2.53), we then have

$$V_0/(2\pi)^3 \int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi} \left[E - \mathscr{E}(0) - 2\mathscr{E}(1)(\cos a_1 + \cos a_2 + \cos a_3)\right]^{-1} da_1\, da_2\, da_3 = 1, \tag{2.56}$$

just analogous to (2.54). We can integrate once, over a_3, without trouble, obtaining

$$V_0/(2\pi)^2 \int_0^{2\pi}\int_0^{2\pi} 1/\sqrt{\left[E - \mathscr{E}(0) - 2\mathscr{E}(1)(\cos a_1 + \cos a_2)\right]^2 - (2\mathscr{E}(1))^2}\; da_1\, da_2 = 1. \tag{2.57}$$

But this equation can be at once converted to Eq. (2.35), which we obtained by our earlier method, so that we are led to exactly the same situation which we have already discussed. The power of the present method is shown by the direct way in which we have been able to get at this result for the energy; and also in the fact that the form (2.56) for the condition exhibits directly the symmetry between x, y, and z which was lacking in our earlier treatment, and is lacking in (2.57), in which we have already integrated once.

Even in cases where the integration cannot conveniently be carried out, Eq. (2.53) can still give useful results, by using it directly as a sum. Thus, for instance, let us consider the one-dimensional chain with second-nearest-neighbor interactions, as in Eq. (2.22). We do not get a convenient result if we convert (2.53) into an integral. However, we may apply periodic boundary conditions with a fairly small value of N, the number of atoms in the repeating region, just as we did in constructing Figs. 1 and 2, where we took N = 6 for the case of nearest-neighbor interactions. Then the N values E_k can be at once computed, and (2.53) immediately gives us V_0 as a function of E, in an algebraic form which can be conveniently handled. In a similar way, even in three dimensions, we could get numerical values by using periodic boundary conditions with small N, dividing our Brillouin zone into a finite network of points, and

finding the E_k's for these points. In this way we get an exact solution for the problem of small N, and we may reasonably assume that if N is chosen fairly large the results will approximate those for a very large N, such as we should use for really accurate results.

Our methods allow us to approximate the wave functions, as well as the energy. We know that the wave function can be written as $\sum(k)\, F_k B_k(r)$, where $F_k = A/(E - E_k)$, B_k is a Bloch function, and A is a constant, which can be determined by normalizing the wave function. We are more interested in finding, not the coefficients of the Bloch functions, but the coefficients of the Wannier functions, since we know from our earlier treatment in this chapter that these coefficients will have the general nature of a continuous function of atomic position. Let us then substitute for the Bloch functions in terms of the Wannier functions, from Eq. (2.49). When we do this, we find that the wave function equals

$$AN^{-1/2} \sum(\vec{R}_i) \left\{ \sum(\vec{k}) \frac{e^{i\vec{k} \cdot \vec{R}_i}}{E - E_k} \right\} a(\vec{r} - \vec{R}_i). \tag{2.58}$$

In other words, the coefficient of the Wannier function on the atom at $\vec{R}_i$, the quantity called $U(\vec{R}_i)$ in our earlier work, is

$$U(\vec{R}_i) = AN^{-1/2} \sum(\vec{k}) \frac{e^{i\vec{k} \cdot \vec{R}_i}}{E - E_k} \tag{2.59}$$

This simple result is the complete solution of the problem of finding the wave function in the presence of an impurity producing a perturbation only on one atom. It is to be understood, of course, that the E which appears in (2.59) is that which results from solution of Eq. (2.53), which is essentially the secular equation. It is interesting to note that since E_k is a periodic function of position in k space, we can make a Fourier expansion of the function $1/(E - E_k)$, the coefficients being associated with the various points R_i of an ordinary space. It is clear from (2.59) that the $U(\vec{R}_i)$'s are, except for a constant factor, just these Fourier coefficients. We can now show by direct integration in simple cases, such as the linear chain with nearest-neighbor interactions, that (2.59) leads to the same results already found for the wave function.

In simple cases, we can use Eq. (2.59) to obtain results about the wave function in an analytic way. Thus, let us consider the problem of a three-dimensional lattice, in which the energy E_k can be expanded about a minimum or maximum energy, which we take to come at $k = 0$, in a power series starting with the terms $a(k_x^2 + k_y^2 + k_z^2)$, where $a = h^2/8\pi^2 m$, m being an effective mass. Let us have a perturbation at the atom at the origin, such that the discrete level lies at an energy E, below

the bottom of the band (and hence negative, if the bottom of the band is at zero energy). Then $E - E_k$ can be approximated by $E - a(k_x^2 + k_y^2 + k_z^2)$, which is numerically small near $k = 0$, but increases rapidly as we go away from this point. The significant contributions to the sum or integral in (2.59) will then come from small values of k, and we shall not make a serious error if we integrate, not merely over the unit cell in k space, but out to infinity. The quantity $1/(E - E_k)$ will depend only on the magnitude of $\vec{k}$, to the approximation we are using, so that in carrying out the integration over $\vec{k}$, we may first average over angles, in which case the exponential $e^{i\vec{k} \cdot R_i}$ can be replaced by the spherical Bessel function $\sin kR_i/(kR_i)$, where here k and R_i stand for the magnitudes of the vectors. Thus the sum in (2.59) becomes proportional to

$$\int_0^\infty \frac{2\pi k^2 \sin kR_i}{(E - ak^2)\, kR_i}\, dk = -\frac{\pi^2}{aR_i} e^{-\gamma R_i}, \tag{2.60}$$

where $\gamma = (-E/a)^{1/2} = (2\pi/h)(-2mE)^{1/2}$. But this simply tells us that the quantity $U(R_i)$ is proportional to $e^{-\gamma R_i}/R_i$, as we deduced for this case in Section 3, and the value of γ we have found here is the same one found there.

We can, in other words, reproduce the results of the free electron approximation for the wave function very easily; but at the same time we have the machinery for easily improving the approximation. We merely have to investigate the change in the sum or integral in (2.59) when we take a better approximation to the energy E_k. In the limit as we have a very large perturbation, so that the energy E departs widely from the energy band, the constant term in the Fourier representation will outweigh any other terms; this means that $U(R_i)$ is much larger for $R_i = 0$, or on the perturbing atom, than on any other lattice site, as we know should be the case from our earlier qualitative discussion. But it would be an easy thing, with any given function E_k, to obtain numerical values for the other Fourier components, or other $U(R_i)$'s, to any desired degree of approximation. By Eqs. (2.53) and (2.59), we have reduced the problem of any lattice with a single perturbing atom to quadratures, and as more examples are worked out, the wave functions and energy levels of the discrete states in such problems can be completely investigated.

Next we shall consider the more complicated case where the perturbative potential is felt over a number of atoms, and where a number of overlapping bands, with their corresponding Wannier functions, are concerned. Before carrying through the analytical treatment of this case, we shall state the results in words. We find that the solution can be written as a linear combination of functions of the type we have just been considering, arising from perturbations on single atoms. One such function is to be centered on each atom on which there is a perturbative potential, and one is to be

constructed from the Wannier function of each band concerned. The number of coefficients of these functions, then, will equal the number of atoms on which there is perturbative potential, multiplied by the number of Wannier functions concerned on each atom. These coefficients are determined by a set of simultaneous homogeneous linear equations, which are incompatible unless the determinant of coefficients vanishes. This furnishes a secular equation for the energy, though it is not of the usual type of secular equation. All the various functions which are superposed to find the final solution are to be computed for the energy determined from this secular equation. Consideration of symmetry properties of the resulting solution will ordinarily lead to relations between the coefficients, reducing greatly the order of the secular equation to be solved; in some cases it can remove entirely the requirement of solving a secular equation, and lead directly to the final wave function.

The statements which we have already made about the wave function can lead to useful information. Thus, for the discrete state, the problem with a single atom on which there is perturbation leads to a wave function falling off exponentially with the distance, or perhaps as $e^{-\gamma r}/r$. If we superpose such functions on a number of neighboring atoms, then no matter what may be the coefficients, we shall still find that at a distance from these perturbed atoms, the total wave function will fall off in much the same way as if there were perturbation on only one atom. Superposition of such functions to give a resulting wave function having the symmetry behavior suggested by various spherical harmonics of angle will then give functions behaving in general like spherical harmonics of angle times functions of angle which, like $e^{-\gamma r}/r$, can be written in terms of spherical Bessel functions of imaginary argument. Considerations like this can lead to a good deal of information about the wave functions of the discrete states, even without the need of solving the problem completely.

Let us now see how to verify the statements which we have made. In the first place, we generalize our earlier work by including several bands. We give the Wannier functions $a_n(\vec{r} - \vec{R}_i)$ a subscript n referring to the band we are dealing with, and similarly for the Bloch function $B_{n,\vec{k}}(\vec{r})$, which is still defined by (2.49). The matrix component of perturbative energy between Bloch functions $B_{n,\vec{k}}(\vec{r})$ and $B_{n',\vec{k}}(\vec{r})$ is

$$N^{-1} \sum(\vec{R}_i, \vec{R}_j)\, e^{i(-\vec{k}\cdot\vec{R}_i + \vec{k}'\cdot\vec{R}_j)}\, V_{n,n'}(\vec{R}_i, \vec{R}_j), \tag{2.61}$$

where $V_{n,n'}(\vec{R}_i, \vec{R}_j)$ is the matrix component of perturbative energy between the Wannier function of the nth band on atomic site at $\vec{R}_i$, and that of the band n' at $\vec{R}_j$. Then as before we set up a solution consisting of a sum of Bloch function, $\sum(n, \vec{k})\, F_{n,\vec{k}}\, B_{n,\vec{k}}(\vec{r})$. The equations for the F's are

$$E_{n,\vec{k}} F_{n,\vec{k}} + \sum(n', \vec{k}', \vec{R}_i, \vec{R}_j)\, N^{-1}\, e^{i(-\vec{k}\cdot\vec{R}_i + \vec{k}'\cdot\vec{R}_j)}\, V_{n,n'}(\vec{R}_i, \vec{R}_j)\, F_{n',\vec{k}'} = E F_{n,\vec{k}}. \tag{2.62}$$

Now let

$$\sum(n', \vec{k}', \vec{R}_j)\, N^{-1} e^{i(\vec{k}' \cdot \vec{R}_j)} V_{n, n'}(\vec{R}_i, \vec{R}_j)\, F_{n', \vec{k}'} = A_{n, \vec{R}_i} . \tag{2.63}$$

Then we can rewrite Eq. (2.62) in the form

$$(E - E_{n, \vec{k}})\, F_{n, \vec{k}} = \sum(\vec{R}_i)\, e^{-i\vec{k} \cdot \vec{R}_i} A_{n, \vec{R}_i}, \tag{2.64}$$

whose solution at once is

$$F_{n, \vec{k}} = \sum(\vec{R}_i)\, A_{n, \vec{R}_i} \frac{e^{-i\vec{k} \cdot \vec{R}_i}}{E - E_{n, \vec{k}}} . \tag{2.65}$$

We can now substitute Eq. (2.65) in the left side of (2.63), obtaining

$$\sum(n', \vec{R}_\ell)\, A_{n', \vec{R}_\ell} \left\{ \sum(\vec{R}_j)\, V_{n, n'}(\vec{R}_i, \vec{R}_j) \sum(\vec{k}')\, N^{-1} \frac{e^{i\vec{k}' \cdot (\vec{R}_j - \vec{R}_\ell)}}{E - E_{n, \vec{k}'}} - \delta_{n, n'} \delta_{\vec{R}_i, \vec{R}_\ell} \right\} = 0. \tag{2.66}$$

From Eq. (2.63), we can see at once that the only non-vanishing $A_{n, \vec{R}_i}$'s are those for which one or more of the $V_{n, n'}(\vec{R}_i, \vec{R}_j)$'s are different from zero; that is, those referring to atomic sites, and Wannier functions, for which there is a diagonal or non-diagonal matrix component of the perturbative energy to some other atomic site and Wannier function. In our earlier case, there is only one such site and Wannier function, so that we met only one coefficient A, relating to the atomic site at the origin. If only a finite number of atomic sites and energy bands are concerned, we shall have a corresponding finite number of non-vanishing $A_{n, \vec{R}_i}$'s. Then in Eq. (2.66) we have a set of simultaneous linear homogeneous equations for these A's, and as usual we cannot get solutions unless the determinant of coefficients vanishes. This furnishes us with a secular equation for the energy E, though it is not of the usual form met in quantum mechanics, since the energy E appears in the denominators rather than in the usual way. Once this secular equation is solved, so that we know the energy, we can then solve Eqs. (2.66) to find the A's, and then substitute in (2.65) to get the wave functions.

As in our earlier simpler case, it is easier to interpret the result if we consider, not the coefficients $F_{n, \vec{k}}$ of the Bloch functions in the expansion of the final wave function, but the corresponding coefficients $U_n(\vec{R}_i)$ in the expansion of the wave function in terms of Wannier functions. If we proceed as before, using Eq. (2.49), we find at once that

6. GENERAL METHOD FOR SOLVING THE DIFFERENCE EQUATIONS

$$U_n(\vec{R}_i) = \sum(\vec{k}, \vec{R}_j) A_{n, \vec{R}_j} N^{-1/2} \frac{e^{i\vec{k} \cdot (\vec{R}_i - \vec{R}_j)}}{E - E_{n, \vec{k}}} \tag{2.67}$$

Now $\sum(\vec{k}) N^{-1/2} \frac{e^{i\vec{k} \cdot (\vec{R}_i - \vec{R}_j)}}{E - E_{n, \vec{k}}}$ is proportional to the solution of the problem of a perturbed lattice with only a perturbative potential at the lattice site $\vec{R}_j$, as we see from Eq. (2.59). Thus we verify our earlier statement that the whole solution of our present problem is a linear combination of such solutions, located on all the various lattice sites for which there are perturbative potential terms, and corresponding to the Wannier functions of all the bands concerned. The quantities $A_{n, \vec{R}_j}$ are the coefficients multiplying these various solutions, and the equations (2.66) are those determining these coefficients. We remember that the energy E appearing in (2.67) must be that determined from the secular equation found from (2.66). We now note also that the sums appearing in this secular equation are of the same form as those appearing in (2.67), so that as far as mathematical technique is concerned, the problems all reduce to the calculation of such sums, or to the solution of the problem of a perturbation at a single lattice point.

We thus have a formal solution for a very general type of perturbation, and the power of our method is shown in the fact that the resulting secular equation has an order equal only to the total number of atoms on which there is a perturbative potential, times the number of overlapping energy bands concerned. As we have mentioned earlier, symmetry can often lead to many relations between the $A_{n, \vec{R}_i}$'s, greatly reducing the order of the secular equation. The method just described is so new that no examples have yet been worked out using it, beyond the very simple ones which have already been described. But it should lead in time to a complete mastery of the difference equation method of dealing with perturbed potentials in terms of Wannier functions, so that the view expressed in earlier sections, to the effect that the difference equation method of dealing with such problems is better than the approximate differential equation method, seems entirely justified.

The method of the present section, in cases where we have overlapping bands, depends on having Wannier functions, and energies $E_{n, \vec{k}}$ for the unperturbed problem which really diagonalize the energy of that problem. Thus, if we are starting with the simplified tight binding approximation, as discussed in earlier sections, our first step in carrying through the present analysis must be to solve the periodic potential problem, and set up the Wannier functions and energy levels by that method. Then we can proceed to apply the present method. The details of doing this have not yet been worked out, but the general outlines of the method are clear.

7. Additional Types of Perturbative Potentials

In the preceding sections, we have been discussing the energy levels in the presence of a perturbation of the type which would be produced by an impurity atom substituted for one of the atoms of the crystal. Most of our analysis, however, is more general than this, and in the present section we shall mention a number of other types of perturbation which are important. Some extensions are obvious. For instance, there is the well-known problem of the F center in an alkali halide crystal. This arises when there is a missing halogen ion in the crystal. In such a case, the perturbative potential is essentially the negative of the potential around such a halogen ion, in a dielectric such as is provided by the crystal. Since the halogen ion is negatively charged, this is a positive potential, or negative potential energy for an electron, in all the region surrounding the missing ion, but a positive potential energy for the electron in the region from which the ion is missing. The Coulomb attraction surrounding the site of the missing ion will result in bound energy levels for an electron, and the lowest of these bound levels will correspond to the energy level concerned in the F center; an electron becomes trapped in such an energy level, and can be ejected to an excited level by absorption of energy.

It is now interesting to ask how the wave function of this lowest bound level is to be built up. We may, in accordance with the preceding sections, make it up by superposing atomic orbitals on the various atoms of the crystal, with amplitudes which will rapidly decrease as we go away from the missing ion. These amplitudes will be determined by our difference equation, which resembles a differential equation for a Schrödinger problem derived from the perturbative potential which is repulsive at the position of the missing ion, attractive at all other atomic sites, but rapidly decreasing in magnitude as we go away from the missing ion. A reasonable first approximation to this wave function is to suppose that there is no contribution at the site of the missing ion, on account of the large repulsive potential, but that the main contributions come on the alkali ions which are the nearest neighbors of this missing ion, with rapidly decreasing contributions on more distant ions. It has been shown by a number of workers[1] that we get quite an acceptable result by merely superposing atomic functions on the six alkali ions forming the nearest neighbors of the missing ion. We can, of course, make six possible linear combinations of these six atomic functions, whose coefficients are at once determined by requirements of symmetry. The sum of the six gives a wave function of the s-like type of symmetry in the cubic crystal, and we can in addition form three p-like functions transforming according to a three-dimensional irreducible

---

[1] T. Muto, Prog. Theoret. Phys. 4, 243 (1949)
T. Inui and Y. Uemura, Prog. Theoret. Phys. 5, 252, 395 (1950)
A. H. Kahn and C. Kittel, Phys. Rev. 89, 315 (L)(1953).

representation of the cubic space group, and two functions transforming according to a two-dimensional representation. The fact that we can get an acceptable wave function in this very simple way, agreeing with experiment in a number of respects, gives us courage to hope that in other cases an impurity wave function consisting of atomic functions on a very limited number of atoms surrounding an impurity atom may be fairly satisfactory. This is of course a case with a large perturbative energy, concentrated in a restricted neighborhood, for which in any case we should expect a wave function falling off rapidly with distance.

There is one feature met in this F-center calculation which is also encountered in any other real problem of impurity atoms. The atoms in the neighborhood of the impurity will be subjected to stresses, on account of the fact that the impurity atom will be of different size from the atoms of the lattice, and will therefore be displaced slightly from their normal positions. Thus, with a missing ion as in an F-center, the neighboring atoms will tend to be pushed into the hole left by the missing ion, provided an electron is occupying the impurity level to provide electrical neutrality; otherwise, there would be strong electrostatic forces which would tend to repel the neighboring alkali ions. The exact determination of the resulting forces and displacements is a very complicated problem. In principle, we should solve for the electronic energy of the crystal, as a function of the positions of the neighboring atoms, and vary the atomic positions to minimize this energy, in order to find the positions of equilibrium of all atoms. In practice, this is so hard to do that simplified approximate methods always have to be used to estimate the atomic displacements.

Similar questions come up, and have been considered, in problems of luminescence, where the sensitization of the luminescence of a crystal comes from the presence of an impurity atom. Thus Williams,[1] in considering the luminescence of a KCl crystal activitated by replacing a K^+ ion by a Tl^+ ion, has investigated the electronic energy levels surrounding the Tl^+ ion as a function of the position of the neighboring Cl^- ions, which are assumed capable of moving radially in or out. Williams's calculation of the electronic energy levels does not follow very closely the model which we have proposed, but we must notice that there are features of our method which are certainly not very appropriate in such a case. We have tacitly assumed that the potential perturbation, even at the position of the perturbing ion, was small enough so that it was reasonable to use the Wannier functions of the original atom to expand the wave function, even at the site of the perturbing ion. Thus, in this case, our method if used literally would involve building up an impurity wave function which would be made of K orbitals on the Tl^+ ion, as well as Cl orbitals on the neighboring chlorines. This is obviously absurd, in such a case as this, where the Tl^+ is very different from the K^+. It is clear that the spirit of our tight binding approximation would suggest in such a case that we build up a wave

---

[1] F. E. Williams, J. Chem. Phys. 19, 457 (1951).

function out of Tl atomic orbitals on the Tl^+ ion, together with Cl orbitals on the Cl^-'s, and K orbitals on the more distant K^+ ions. It seems likely, from the discussion which we have just given of F-centers, that quite a good approximation would be secured by combining a Tl orbital on the Tl^+ ions with Cl orbitals on the neighboring Cl^-'s, using a linear combination of these Cl orbitals having the same symmetry as the Tl orbital. This is not the method used by Williams, but suggests a useful approach to such problems.

Another type of problem which can be handled by the type of approach we have suggested is that of surface states. It was pointed out long ago by Tamm (references to papers on surface states are given in the bibliography at the end of this chapter) that if we have a one-dimensional periodic potential problem, consisting of a set of potential wells, but bounded at one end, and terminated by a region where the potential is constant, we often have one energy level split off from the continuum, associated with a wave function which falls off exponentially as we go away in either direction from the point of discontinuity between the periodic region and the constant potential. We can set up a similar problem for a one-dimensional chain, handling it by the tight-binding approximation. Such a treatment was given by Goodwin,[1] by a method equivalent to those taken up in the present chapter. His assumption was equivalent to the following modification of our problem of Section 2: we assume just the same behavior as in Eq. (2.13) for positive p, assume as in that equation a perturbative potential V(0) on the atom at the origin, but assume that there are no atoms for negative p, so that U(-1) is to be set equal to zero. In this case it is easy to solve the equations, and we find a bound state, similar to that discussed in Section 2, under some circumstances. This bound state is only present if V(0) is greater in magnitude than $\mathscr{E}(1)$, as we readily find, and if it is present, its energy depends on the magnitude of V(0). Somewhat similar, but by no means equivalent, results have been obtained using a one-dimensional periodic potential, not only by Tamm, as mentioned earlier, but by a variety of other writers, including Maue, Shockley, and others.

There have been some attempts to correlate the wave functions so found, falling off exponentially as we go away from the surface, and called surface states, with observed energy levels on the surfaces of certain crystals, such as germanium. Undoubtedly, in a general way, there is such correlation. Surely the surface atoms of a crystal are in a sufficiently different environment from those in the interior so that they act almost like different types of atoms, and a treatment based on handling them like impurity atoms will have elements of truth. However, the real situation is almost certainly so much more complicated than the cases which have been treated theoretically that it would be dangerous to draw any conclusions regarding real surface states from the existing theory. The reason for this is to a considerable extent experimental: the

---

[1] E. T. Goodwin, Proc. Cambridge Phil. Soc. 35, 221 (1939).

surfaces found on real crystals are almost certainly quite different from those assumed in the simple theories, where we merely assume that we have a periodic potential up to the surface, and then go into the vacuum. Real surfaces are almost certainly contaminated by impurity atoms very tightly held to the surface atoms, and the problem which should be solved is that of such a composite surface. Unfortunately, we do not generally know even what atoms are located on the surface, and their geometrical arrangement. Until this is better understood, from an experimental point of view, the theory can hardly make much progress.

The sort of thing to be expected can be seen by considering the case of germanium. If we cleaved a germanium crystal on a 111 face, for example, we could find a surface layer of atoms, each being held to the atoms below it by three of the four normal covalent bonds, but with the fourth bond sticking out from the surface in an unsaturated state; or alternatively each surface atom could be held to the atoms below by one bond, three being unsaturated. We can hardly assume that the surface would remain in this condition, in the presence of a gaseous atmosphere. Practically certainly the unsaturated bonds would attract atoms or radicals, such for instance as OH radicals, which would be held very tightly to the surface; we should expect a complicated surface chemistry involving the various materials which could be bound to a germanium surface. In practice, there seems often to be a rather extensive layer of oxide on a germanium surface, presumably a result of oxygen diffusing down into the crystal. No doubt the resulting surface material will be of complicated form. GeO_2 exists in two crystal structures, the rutile structure like TiO_2, and the quartz structure like SiO_2. Presumably the surface material will try to simulate one or the other of these structures, but on account of being tightly joined to the Ge crystal below, these oxide crystals will be distorted, and in practice the material is found to be practically amorphous, indicating that the crystals are too small to show structure by electron diffraction or other methods. With such a complicated surface, it is obvious that no simplified treatment assuming an ideal geometrical surface can possibly be valid.

When crystal surfaces can be produced with known impurity layers on them, it may be possible to set up a theory of surface states, having some analogy to our treatment of the present chapter, but also analogy to the theory of molecular orbitals in molecules. It is clear that our problem is half chemical, half physical. The bonds between the atoms will be similar to covalent bonds in molecules, and as such can be described fairly well by the molecular orbital method; but once we are inside the crystal of germanium or other material, we have a typical solid-state problem. Fortunately, the tight-binding approximation to the energy band problem, as considered in this chapter and the preceding one, is entirely analogous to the LCAO molecular orbital method, so that it will be possible to join these two treatments easily. But we

point out again that there is very little use in speculating about such problems, until there is a fairly good indication as to what atoms we really have in the surfaces we are considering.

Another type of crystal problem similar to that of surface states is that of the energy levels surrounding a dislocation. An edge dislocation in a crystal may be described as a split or crack in a crystal, ending in an extra plane of atoms which fills in the crack. In that region where the crack is of appreciable width, the atoms on the two surfaces of the crack are not in very different situations from the atoms on the surface of a crystal, and we should expect that something like surface states should arise. The problem is obviously much more complicated than that of a surface, and no attempts have been made so far to consider the detailed nature of the energy levels. There is another aspect which a dislocation presumably has in common with a surface: bonds are broken between neighboring atoms, and this results in a particularly favorable position for the adsorption of foreign atoms. It seems altogether likely that impurity atoms in crystals will tend for this reason to diffuse to dislocations and segregate there, and this possibility has occurred to many workers in the field. In this case, the wave functions concentrated in the neighborhood will have partly the characteristics of impurity states, and can be discussed by the same general methods, except of course that the geometry is much more complicated than that arising from a single substitutional atom.

The examples which we have mentioned will illustrate the breadth of the problems which can, in principle, be handled by the type of discussion taken up in this chapter. In every case, we shall expect the same general situation, though the detailed method of handling the problem can differ in different cases, and can be very complicated. The atoms in the neighborhood of the disturbance in the crystal would, if treated by themselves, have one-electron energy levels quite different from those of the crystal as a whole. If these energy levels happen to lie in an energy gap of the crystal, then the wave function of the composite system will tend to be large in the neighborhood of the disturbance, falling off exponentially as we go into the rest of the crystal, since the only type of solution of Schrödinger's equation in the crystal having this energy is an exponential one. Conversely, the wave functions corresponding to the energy bands in the crystal will fall off as we approach the lattice disturbance.

8. Periodic Perturbative Potentials

There is one type of perturbation which is quite different from the ones which we have taken up so far, but important enough to deserve special treatment. This is the case where the perturbative potential is a sinusoidal function of position in the lattice. We should only get a potential which was literally of this form if we had a sinusoidal electromagnetic wave in the crystal. If its wave length was comparable with the

interatomic distance, we should be dealing with x-rays, and would have to treat the problem by quantum electrodynamics. However, we have a physically equivalent problem if we have a sinusoidal wave of sound in the crystal. It produces perturbations which are equivalent to a sinusoidal perturbative potential, and its effect on the electrons of the lattice can be treated by the methods of this chapter. This problem is important because the thermal agitation in a crystal can be treated as a superposition of sound waves, mostly of very short wave lengths. We meet the effects of the scattering of electron waves by such sound waves when we study the effect of temperature agitation on electrical conductivity.

We shall find that the effect of a sinusoidal perturbative energy on the motion of electrons in a periodic lattice is very much like the effect of such a perturbation on free electrons; the resemblance is closer than would be expected merely from the possibility of replacing the difference equation by a differential equation. Accordingly first we shall remind the reader about the effect of a sinusoidal potential on the motion of free electrons. Then we shall consider a sinusoidal perturbation in our periodic potential problem. Finally, we shall take up the case of a sound wave in detail, and show why it is reasonable that its effect can be replaced by an equivalent periodic perturbation.

An electron moving in a sinusoidal field is of course a well-known problem, the simplest example of a periodic potential, and one leading to the Mathieu equation. If the potential has the form $V_0 \cos Kx$, we can separate variables in a three-dimensional problem, and the motion will be free as far as y and z are concerned. The Schrödinger equation for the motion along x is the Mathieu equation. Its energy levels fall into energy bands, the wave functions obeying Bloch's theorem. That is, the wave function can be written as a sum of exponential functions of form $\exp(i(k + nK)x)$, where k is a propagation constant, n an integer. For the lower energy levels, falling in narrow energy bands, the coefficients associated with the exponential functions of the type written above will fall off very slowly with n, in a complicated way, which has been discussed by the writer.[1] However, if V_0 is sufficiently small, there will be no really narrow energy bands, and the situation is quite different. This is the case in which there are no tightly bound states in the potential minima, or in which V_0 is a very small perturbation. In such cases, the coefficients will fall off very rapidly with n, only the cases with $n = 0, \pm 1$ being appreciable. It is only such cases that we shall consider; only for such a small amplitude is the effect of a sinusoidal perturbative potential simple. For larger perturbations bound states can arise, and all manner of complications.

For small V_0, the solution of the Mathieu problem has a simple form. The

---

[1] J. C. Slater, Phys. Rev. 87, 807 (1952)

coefficients of the terms with $n = \pm 1$ are power series in V_o, starting with linear terms; those with $n = \pm 2$ are power series starting with second power terms; and so on. Similarly the energy is a power series starting with a second power term. If V_o is small enough, we can consider only linear terms in the wave function, including thus only the terms with $n = \pm 1$, and can disregard the perturbation in the energy, since it is in V_o^2. But this is just the approximation made by first order perturbation theory. In other words, such a first order perturbation is adequate for the small amplitudes which we are considering. We regard the wave function as the sum of an unperturbed wave, a plane wave, of form exp(ikx), and two perturbed waves, of form exp(i(k $\pm$ K)x), with amplitudes proportional to V_o.

By perturbation theory, we can at once find the amplitudes of these scattered waves. The matrix component of the perturbative potential $V_o \cos Kx$ between the two wave functions exp(ikx) and exp(i(k $\pm$ K)x) is at once seen to be simply a constant (determined by normalization) times V_o. Thus by first order perturbation theory we know that the coefficient of the function exp(i(k $\pm$ K)x) is equal to the non-diagonal matrix component of energy between this function and the unperturbed function exp(ikx), which is proportional to V_o, divided by the energy difference between these waves. When we put the problem in three-dimensional vector language, this energy difference is that between waves with propagation vectors $\vec{k}$ and $\vec{k} \pm \vec{K}$. In the general case, this energy difference is large. However, there are special cases where the difference is small, or vanishes. In such a case the amplitude of the scattered wave becomes large, and in the limit where the difference vanishes, a power series expansion of the type we have been considering breaks down, and we are not allowed to use first-order perturbation calculations. This case is just that in which the two perturbed waves are related to the unperturbed wave by Bragg's law, as if we had x-ray scattering by planes of scattering material lying on the wave fronts of the function $V_o \cos Kx$. In such a case, we must use a solution consisting of equal contributions of the direct and scattered waves. If we solve a secular equation with two rows and columns between these two waves, in this neighborhood of the Bragg scattering, we arrive at the energy gap in the well-known way. In other words, Bragg scattering by the sinusoidal scattering potential introduces energy gaps, just as Bragg scattering by the periodic potential in the crystal does. But these energy gaps will be found across planes in k space determined by the wave vector K of the perturbing wave.

In this sketch, we have seen the general outlines of the theory of scattering of free electrons by a small sinusoidal potential. Now we see that the same sort of thing can be carried through for a sinusoidal perturbing potential superposed on a periodic potential. The reason why the problem is no more difficult is simply that we are contenting ourselves with a first order perturbation calculation, and this can be carried out just as well in the periodic potential case as with free electrons. We start with unperturbed functions of the form of $\sum(\vec{R}_i)\ U(\vec{R}_i)\ a(\vec{r} - \vec{R}_i)$, where the $U(R_i)$'s are of

the exponential form, $\exp(i\vec{k} \cdot \vec{R}_i)$. Let us assume that we have a periodic perturbation, such that there are no non-diagonal matrix components between Wannier functions in different bands, or on different atoms, and the diagonal matrix components are of the form $V_o \cos \vec{K} \cdot \vec{R}_i$. Then we find at once that there are non-vanishing matrix components of the perturbative potential only between two Bloch functions corresponding to $\exp(i\vec{k} \cdot \vec{R}_i)$ and $\exp(i(\vec{k} \pm \vec{K}) \cdot \vec{R}_i)$. As with the free electron case, these matrix components are simply normalization constants times V_o. We can then set up a perturbed wave function, in which these scattered waves again have amplitudes proportional to V_o divided by the difference between the energies of the waves with propagation vectors $\vec{k}$ and with $\vec{k} \pm \vec{K}$. The only difference between this case and the free electron case is that now we are to compute the energies of these waves using the proper expression for E_k, as determined by the energy band theory.

We note that this difference in the situation means that the scattering in the actual crystal will not in general follow the Bragg law. The derivation of this law is based on the assumption that the energy is a function only of the magnitude of the propagation vector, so that the quantities $|\vec{k}|^2$ and $|\vec{k} \pm \vec{K}|^2$ are equal. On the other hand, with a different dependence of the energy on $\vec{k}$, this will not be the case. The result of the Bragg law is that all waves of propagation vector $\vec{k}$ terminating in planes perpendicular to $\vec{K}$, and distant by the amount $|\vec{K}/2|$ from the origin, will be scattered strongly by the sinusoidal potential. With a more general dependence of energy on $\vec{k}$, the surfaces on which vectors must terminate to be scattered by this potential will no longer even be planes; they can have complicated forms. This fact does not seem to be recognized in the literature, for the assumption that the energy is a function of the magnitude of $\vec{k}$ is generally made at the very beginning of any discussion of scattering.

We can extend these results without trouble to a considerably more general case. Thus, let us assume that we have matrix components of the perturbative energy between Wannier functions on different atoms, and connected with different bands; but let us still assume that there is a sinusoidal dependence on position. That is, let us assume that the matrix component of perturbative energy between Wannier functions $a_n(\vec{r} - \vec{R}_i)$ and $a_m(\vec{r} - \vec{R}_j)$ depends in an arbitrary way on n and m, and on the vector displacement $\vec{R}_j - \vec{R}_i$ between the two atoms, but that it depends on the absolute position of the perturbed atoms in space only through an exponential factor $e^{i\vec{K} \cdot \vec{R}_i}$, or a corresponding cosine function. We shall still see, simply on account of the exponential dependence, that there will be matrix components of this perturbative energy only between plane wave solutions having propagation constants $\vec{k}$ and $\vec{k} \pm \vec{K}$. The matrix component will depend linearly on the various non-diagonal as well as diagonal matrix components of the perturbative energy, in a way which can be easily expressed mathematically. There are then only two differences between this case and the earlier one: (1) the formula for non-diagonal matrix component of energy between the incident and

scattered wave will be more complicated than before; and (2) there will be matrix components of energy between an incident wave set up from the Wannier functions a_n of one band, and a scattered wave set up from Wannier functions a_m of another band. It could well happen that a wave of propagation vector $\vec{k}$ in one band could have the same energy as one of propagation vector $\vec{k} \pm \vec{K}$ in another band. Thus there could be circumstances where the periodic potential perturbation could result in strong scattering of electrons from one band to another, provided the bands overlapped in energy.

We have now examined the effect of a periodic perturbation on the wave function; next let us ask why a sound wave will produce such a perturbation. We assume that this sound wave is, so to speak, frozen into the crystal; that is, we have displacements of the atoms, varying sinusoidally through the crystal, but independent of time. We shall later consider the time dependence which would be present in an actual sound wave. We must now consider the perturbing potential arising from the sinusoidal wave. This perturbing potential will be the difference between the potential of a self-consistent field for the undistorted crystal, and for the crystal with the sound wave in it. It will have quite a different form in those parts of the crystal where the sound wave is producing a compression or rarefaction, and in the other parts of the crystal where the atoms are displaced by a maximum amount, but without change of density. Let us first consider the regions of maximum strain, where there is change of density.

In a region of compression, for instance, each atom will find its neighbors closer to itself than in the undistorted crystal. The potential energy of an electron always tends to be depressed by the presence of neighbors; consequently, in such a region, there will be a net decrease of potential energy, felt at all atoms in the neighborhood. This will result in a modification of the diagonal and non-diagonal matrix components of energy between the various atomic orbitals, if we are using a tight-binding approximation. It is such a modification as would be produced by a compression of the whole crystal. The modification of the diagonal matrix components would result, if extended through the crystal, in a displacement of the band as a whole, and the modification of the non-diagonal matrix components of energy would result in changing the widths of the bands. Such changes of energy bands are known to arise from the compression of a lattice, and the effect of the deformation on the position of the energy bands has been described by Bardeen and Shockley[1] as a deformation potential. We see that if we look a little more carefully, we shall not have just one such potential, but rather shall have terms in all the diagonal and non-diagonal matrix components of perturbative energy between the various atomic orbitals or Wannier

---

[1] J. Bardeen and W. Shockley, Phys. Rev. 80, 72 (1950). For a discussion of this and related topics in the effect of a sinusoidal perturbative potential, see W. Shockley, Holes and Electrons in Electrons in Semiconductors (John Wiley and Sons, Inc.) 1951.

functions, arising from the deformation.

These matrix components of perturbative potential will, of course, be proportional to the magnitude of the compression, provided it is small, and hence will vary sinusoidally through the crystal, and will have the form we have already considered, taking on their maximum values at the positions of maximum compression and rarefaction. In addition, there will be other terms in the matrix components of perturbative potential, taking on their maximum values at the intermediate positions where the displacement of the atoms is a maximum, but where there is no compression. To understand these terms, let us ask how we should handle the problem which would arise if the whole crystal were to be displaced bodily by an infinitesimal amount. If we tried to expand the perturbed wave functions in terms of the unperturbed ones (located at the original atomic sites), we should of course merely end up with wave functions which were displaced by the same amount as the lattice, and with energy values which were unchanged. We could not accomplish this perfectly if we were merely using a finite set of atomic orbitals, rather than a complete set of Wannier functions. These matrix components of energy which have their maximum values at the positions of greatest displacement in the lattice will be trying to achieve just this result, of displacing the wave functions bodily without having an effect on the energy; they will achieve it as well as possible, consistent with the limited number of atomic orbitals or Wannier functions ordinarily used.

All these terms, as we see, will vary sinusoidally with position through the crystal, and hence will fall into the general scheme which we have described. Hence the effect of such a sound wave will be to produce scattered waves, of propagation vector $\vec{k} \pm \vec{K}$; and as we have seen, the amplitude of this scattered wave will be large only if it has almost exactly the same energy as the incident wave. As we have seen earlier, we can well have matrix components of the perturbative energy between wave functions connected with different bands, so that electrons can be scattered from one band to another, provided this is compatible with scattering without change of energy.

We have been talking about scattering by a frozen-in sound wave; but this is not what we encounter in scattering by thermal oscillations, which can be analyzed into a superposition of real sound waves, traveling through the crystal, and hence with a potential depending on time. This problem cannot be handled so simply, and we shall merely indicate the nature of its treatment (see Ref. 1, Page 62 for a good discussion of this problem). It has a close resemblance to problems in quantum electrodynamics; the interaction of the electrons with sound waves is not entirely different from the interaction with electromagnetic waves. In the first place, we can get a good approximation by treating the perturbation simply as an external perturbative potential varying with the time, using the method of variation of constants. We then find that, if we start at $t = 0$ with a single Bloch solution of the electronic problem, and a

perturbative potential varying sinusoidally with time with a frequency ν, a probability appears of transition to another Bloch wave, the amplitude being proportional to the amplitude of the perturbation. This scattered wave is the form in which the perturbed wave appears, when we are using the method of variation of constants, in which we are not setting up stationary states.

If $\vec{k}$ is the propagation constant of the initial Bloch wave, and if the sound wave has a propagation constant $\vec{K}$, then we find, as we should expect, that we have transitions only to Bloch waves with propagation constants $\vec{k} \pm \vec{K}$. The amplitude of the scattered wave does not have the difference between the energies of the Bloch waves with propagation constants $\vec{k}$ and $\vec{k} \pm \vec{K}$ in the denominator, however, as in the static case. Instead, it has this difference of energy, $\pm h\nu$. In other words, for a large scattering amplitude, the energy of the scattered Bloch wave must be greater or less than that of the incident wave, by the energy of a phonon of the appropriate sound wave. Ordinarily the electronic energy is so large in comparison with phonon energies that this modification of the results is unimportant. However, Shockley (see Ref. 1, Page 62) points out certain cases in which the phonon energy is important; it obviously can be when the electronic energy comes so close to the bottom of an energy band that we are dealing with very small electronic energies.

As in the case of quantum electrodynamics, this method of handling the time-dependent field by the method of variation of constants applied to a potential perturbation which varies with time is not completely adequate. We remember that in quantum electrodynamics, such a method can give us a treatment of absorption and forced emission, but not of spontaneous emission. To treat that, we must use the Dirac radiation theory, in which the electromagnetic field is treated as part of the system, as well as the atoms. We then set up an unperturbed problem in which the atoms, and the electromagnetic field, do not interact with each other. We set up their stationary states, which are described by giving the various quantum numbers of the atomic system, and the numbers of photons in the various electromagnetic vibrations. We then introduce the previously neglected interaction terms as perturbations, treat them by variation of constants, and find that there are probabilities of transition in which the atomic system gains or loses energy, with compensating loss or gain of energy by the electromagnetic field, so that the states which perturb each other, and which lead to the transitions, are states in which the whole energy, that of atoms and field, is unchanged in the transition.

The situation is somewhat similar with our acoustical case, though not identical. We treat the electronic system, and the nuclear vibrations, as independent to a first approximation, describing the nuclear vibrations by normal coordinates. We must here work in the framework of the Born-Oppenheimer approximation, in which the electronic problem is solved for fixed positions of the nuclei, with an energy

depending parametrically on the nuclear positions. This energy must then be used as a potential energy for the nuclear framework, describing its vibrations. We start with an unperturbed wave function which is a product of Hermite functions of the normal coordinates describing the nuclear displacements, and an electronic function containing the normal coordinates parametrically. Then we insert as perturbations the term in the Hamiltonian which is neglected in the Born-Oppenheimer approximation, terms in the cross derivatives of the wave function of the electronic motion with respect to both electronic and nuclear coordinates. When this perturbative term is inserted, we find as in the electromagnetic case that it introduces transitions between different approximate stationary states of the nuclear and electronic parts of the system. As before, these are transitions in which the total energy, electronic plus nuclear, does not change; so that as with the method of the time-dependent perturbative potential, we conclude that those transitions will occur in which the electron gains or loses the energy of a phonon in the scattering process.

These problems, whose treatment we have merely sketched, are really very intricate, and we shall not treat them further here. There is room for a really complete treatment of such questions, based on the Wannier function method of handling the scattering problem; such a treatment has not been given. We meet these questions in a practical way, of course, in the problem of electrical conductivity, where it is well known that the scattering of the electron waves by the lattice vibrations is one of the main sources of electrical resistance. We realize how intricate the problem can be, when we recall that the theories of superconductivity of Bardeen and Fröhlich are based on the treatment of the interaction between electronic motion and lattice vibrations, and that none of the existing methods of handling that problem are really satisfactory, in spite of the large amount of effort which has been put into the problem.

9. A Constant Accelerating Field

As a final example of perturbative potentials, we take the case of a constant accelerating field, which of course is required in any discussion of electrical conductivity. In such a problem our primary interest is in the time-dependent problem, in which the field continually accelerates an electron, rather than in the stationary states. We can handle this best by the method of wave packets. Let us consider an electron in a given energy band, with a given $\vec{k}$ vector, and ask how it is accelerated by the constant external field. We first build up a wave packet, by superposing Bloch waves whose $\vec{k}$'s are all near the one we are interested in, and all coming from the same energy band. This wave packet is supposed to be localized near a given point of space. We know, from the uncertainty principle, that complete localization is incompatible with using just waves of nearby $\vec{k}$'s. In fact, the closest we can come to a really localized function is a Wannier function, and this uses all the $\vec{k}$'s in the band.

If we use $\vec{k}$'s chosen from the immediate neighborhood of one point in $\vec{k}$ space, the wave packet will be much more spread out than a Wannier function. But this is satisfactory for the present purpose, for the accelerating fields met in the ordinary discussion of conductivity are small enough so that the potential changes only slightly in a distance of many lattice spacings.

When we have set up this wave packet, we ask how it moves under the action of the accelerating field. To answer this question, we use the well-known theorem stating that the center of gravity of a wave packet moves according to the classical Hamiltonian equations, provided the wave function satisfies a Schrödinger equation constructed from the same Hamiltonian. We use this theorem in connection with our solution of the perturbed periodic potential problem in terms of Wannier function, and in particular in connection with the Schrödinger-like equation given in Eq. (2.6). We express our wave packet by giving the expansion coefficients $U(\vec{r})$ in terms of Wannier functions, rather than in terms of Bloch functions, and then use the time-dependent form of Eq. (2.6) to determine the time variation of the U's. Before doing this, we must examine one thing: the justification of using the particular form of perturbative potential used in Eq. (2.6), in which we have matrix components $V_{nm}(\vec{R}_i, \vec{R}_j)$ of the perturbative potential between Wannier functions only if n = m, and $R_i = R_j$; for it is only in this case that an ordinary Schrödinger equation is appropriate. The justification in this case is that the perturbative potential, which is a linear function of position if we are dealing with a constant external field, is assumed to vary so slowly with position that over the region where one Wannier function $a_n(\vec{r} - \vec{R}_i)$ is appreciably different from zero, the potential may be treated as constant. The matrix component $V_{nm}(\vec{R}_i, \vec{R}_j)$, which is equal to $\int a_n^*(\vec{r} - \vec{R}_i)\, V\, a_m(\vec{r} - \vec{R}_j)\, dv$, where V is the perturbative potential, then becomes equal to an average value of V, times an overlap integral between two Wannier functions, which is zero unless n = m and $\vec{R}_i = \vec{R}_j$. We shall then have departures from our case of Eq. (2.6) only for very strong fields; we shall consider that case later.

From Eq. (2.6), we see that $U(\vec{r})$ satisfies a Schrödinger equation with Hamiltonian $E(\vec{p}) + V(\vec{r})$, where we have written $E(\vec{k})$ in terms of the equivalent $\vec{p} = \vec{k}\hbar$, where this $\vec{p}$ is the quantity to be converted into a differential operator. Thus the center of mass of a wave packet will move according to the Hamiltonian equations derived from this Hamiltonian. These Hamiltonian equations are

$$dp_x/dt = -\partial V/\partial x, \quad dx/dt = \partial E/\partial p_x \tag{2.68}$$

with similar equations for y and z. These equations have a very simple meaning. The first one means that the center of mass of the wave packet moves in $\vec{k}$ space according to simple classical mechanics, the time rate of change of the pseudomomentum $\vec{p}$

(which we remember is not really a momentum) equaling the external force. The second can be easily shown to be equivalent to the formula for group velocity of the wave packet. These simple results, which we have seen follow rigorously provided Eq. (2.6) is valid, are the basis of the ordinary quantum mechanical treatment of electrical conductivity.

It remains only, then, to consider the case of fields so large that this result is no longer correct. For a sufficiently large field, the components $\int a_n^*(\vec{r} - \vec{R}_i) V a_m(\vec{r} - \vec{R}_j) dv$ can become appreciable, connecting Wannier functions in two different bands. Under these circumstances, the Schrödinger-type equation (2.6) is no longer appropriate, and we must use the difference equations (2.5), or treat the interaction with other bands by some different method. The result will be, if we use the time-dependent treatment, that there will be a certain probability that an electron jumps from one band to another, under the action of the field. The wave packet, in other words, will not be preserved unchanged, but will gradually dissipate, with reduction of intensity in the original band, and a building up of intensity in the other band. Such an effect was discussed many years ago by Zener,[1] on the basis of elementary band theory, and is usually called the Zener effect. There is good indication that it can be experimentally observed, in conduction in germanium.[2]

We can consider the Zener effect in an alternative way, which is very instructive. Instead of asking about the motion of wave packets, let us consider the nature of stationary states in the presence of the constant accelerating field, or the potential energy varying linearly with a coordinate, say x. Let us then set up a figure

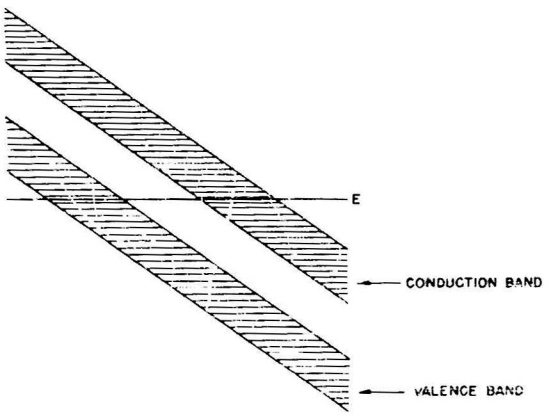


Fig. 7

Energy bands in presence of constant external field

---

[1] C. Zener, Proc. Roy. Soc. (London) A145, 523 (1934). For the application of Wannier functions to this problem, see Paula Feuer, Phys. Rev. 88, 92 (1952).

[2] K. B. McAfee, E. J. Ryder, W. Shockley and M. Sparks, Phys. Rev. 83, 650 (L) (1951).

like Fig. 7, in which the energy is plotted vertically, distance through the crystal horizontally. The energy bands are shown displaced upward by an amount equal to the perturbative potential V, which varies linearly with x. Now we consider a horizontal line of height E, as shown. This line falls inside each energy band only over a restricted region of space. If we set up a wave packet inside an energy band, as we have been doing, we see that we expect this packet to oscillate back and forth through the band, from one boundary to this restricted region to the other; it is only within this region that the classical velocity of the wave packet, as given by Eq. (2.68), will be real. This packet would be made up of wave functions whose energies were in the neighborhood of E, and they would be oscillatory (that is, $U(\vec{r})$ would be an oscillatory function of position) in the restricted region, but would fall off in a more or less exponential way in the energy gap.

Now we could set up solutions of this type, either in the form of wave packets or stationary states, in each of the two energy bands, in different regions of space. It is clear that if the exponential tails of these two solutions did not fall off rapidly enough, in the energy gap, the tails of the two solutions would overlap slightly, and this would result in transitions from one band to another, by a tunnel effect. There would be several different ways to treat this. We could set up stationary solutions $U(\vec{r})$ which held properly in both bands; these would be linear combinations of solutions in both regions, behaving in a symmetric or antisymmetric way in the region between. Then we could set up wave packets from these stationary solutions, and we should find a probability that the packet would leak from one band to another. Or we could handle it directly by wave packet methods. In any case, we should find that electrons could leak across the gap from one band to another, jumping in the process from one part of the crystal to another part in which the changed external potential had brought the conduction band down to the same energy as the valence band in the other region. Of course, such a jump, by a tunnel effect, is very unlikely if it corresponds to a distance of more than one or two atomic diameters. Thus we see intuitively that the Zener effect is not going to be important, unless the applied external field is so strong that it corresponds to an energy difference of the order of an electron volt (comparable with the gap widths) in an atomic spacing. This general approach can be made the basis of a quantitative treatment of the Zener effect in terms of Wannier functions, but we shall not go into it here.

BIBLIOGRAPHY

In this bibliography, we have included papers in the topics obviously related to this chapter, such as the theory of impurity levels and surface states. The bibliography is considerably broader than this, however. In spite of the small amount of discussion which we have given to problems of electrical conductivity, in the discussion of scattering of electron waves by thermal vibrations, and the acceleration of wave packets by an impressed external field, it has nevertheless seemed desirable to include the papers on conductivity in this bibliography. It is hoped that when this series of reports is finished, the combined bibliographies will cover fairly completely the subject of electronic structure of solids, and since we shall not touch on problems of conductivity again, we include these papers here. For completeness, we have also included the papers on electrical breakdown in insulators, those on superconductivity, and the closely related ones on free electron diamagnetism. Some other things have definitely not been included: ferro- and antiferro- and paramagnetism, cohesive energy of solids, the theory of excitons and optical absorption by solids (except such few papers as bear very closely on the topics of this chapter), thermal vibrations of solids, and various other topics. Most of these things will be included in the bibliography which will follow the final technical report in the present series, which will deal with various problems involving electron interactions in solids. Also, we have not included the rapidly growing literature on the transistor, and on the closely related problems of rectification in semiconductors.

1928

F. Bloch, Über die Quantenmechanik der Elektronen in Kristallgittern, Z. Physik 52, 555.

C. Eckart, Über die Elektronentheorie der Metalle auf Grund der Fermischen Statistik, insbesondere über den Voltaeffekt, Z. Physik 47, 38.

J. Frenkel, Elementare Theorie magnetischer und elektrischer Eigenschaften der Metalle beim absoluten Nullpunkt der Temperature, Z. Physik 49, 31.

J. Frenkel and N. Mirulobow, Zur wellenmechanischen Theorie der Metallischen Leitfähigkeit, Z. Physik 49, 885.

W. V. Houston, Die Elektronenemission kalter Metalle, Z. Physik 47, 33.

W. V. Houston, Elektrische Leitfähigkeit auf Grund der Wellenmehanik, Z. Physik 48, 449

L. W. Nordheim, On the kinetic method in the new statistics and its application in the electron theory of conductivity, Proc. Roy. Soc. (London) A119, 173.

A. Sommerfeld, W. V. Houston and C. Eckart, Zur Elektronentheorie der Metalle auf Grund der Fermischen Statistik, Z. Physik 47, 1.

A. Sommerfeld, Zur Elektronentheorie der Metalle auf Grund der Fermischen Statistik II. Thermo-eletrische, galvanomagnetische, und thermomagnetische Vorgänge, Z. Physik 47, 43.

1929

F. Bloch, Zur Susceptibilität und Widerstandsänderung der Metalle im Magnetfeld, Z. Physik 53, 216.

F. Bloch, Bemerkung zur Elektronentheorie des Ferromagnetismus und der elektrischen Leitfähigkeit, Z. Physik 57, 545.

W. V. Houston, Temperature dependence of electrical conductivity, Phys. Rev. 34, 279.

R. Peierls, Zur Theorie der galvanomagnetischer Effekte, Z. Physik 53, 255.

1930

F. Bloch, Zum elektrischen Widerstandsgesetz bei tiefen Temperaturen, Z. Physik 59, 208.

F. Bloch, Über die Wechselwirkung der Metallelektronen, Leipziger Vorgräge, 69.

N. H. Frank, Zur Theorie der galvanomagnetischen, thermomagnetischen, und thermoelektrischen Erscheinugten in Metallen, Z. Physik 63, 596.

S. Kikuchi and L. Nordheim, Über die kinetische Fundamentalgleichung in der Quantenstatistik, Z. Physik 60, 652.

L. Landau, Diamagnetismus der Metalle, Z. Physik 64, 629.

R. Peierls, Zur Theorie der elektrischen und thermischen Leitfähigkeit von Metallen, Ann. Physik 4, 121.

R. Peierls, Zwei Bemerkungen zur Theorie der Leitfähigkeit, Ann. Physik 5, 244.

1931

F. Bloch, Wellenmechanische Diskussion der Leitungs- und Photoeffekte, Physik Z. 32, 881

C. G. Darwin, The diamagnetism of the free electron, Proc. Cambridge Phil. Soc. 27, 86.

M. von Laue, Dynamical theory of x-rays, Ergeb. exakt. Naturwiss. 10, 133.

1931 (con'd)

L. Nordheim, Zur Elektronentheorie der Metalle, Ann. Physik 9, 607, 641.

R. Peierls, Zur Theorie der magnetischen Widerstandsänderung, Ann. Physik 10, 97.

A. Sommerfeld and N. H. Frank, Statistical theory of thermoelectric, galvanoelectric and thermomagnetic phenomena in metal, Revs. Modern Phys. 3, 1.

E. Teller, Der Diamagnetismus von freien Elektronen, Z. Physik 67, 311.

A. H. Wilson, The theory of electronic semiconductors, Proc. Roy. Soc. (London) A133, 458; A134, 277.

1932

L. Brillouin, Les electrons libres dans un reseau cristallin, equation ondulatoire et proprietes magnetiques, champs self-consistents et electrons metalliques, J. phys. et radium 7-3, 565.

E. Kretschmann, Beitrag zur Theorie des elektrischen Widerstandes und der Supraleitfähigkeit der Metalle, Ann. Physik 13, 564.

W. Kroll, Ein Beitrag zur Theorie der thermoelektrischen Effekte, Z. Physik 77, 322.

W. Kroll, Zur Theorie der thermoelektrischen Effekten bei tiefen Temperaturen, Z. Physik 80, 50.

R. Peierls, Zur Frage des elektrischen Widerstandgesetzes für tiefe Temperaturen, Ann. Physik 12, 154.

R. Peierls, Elektronentheorie der Metalle, Ergeb. exakt. Naturwiss. 11, 264.

I. Tamm, Über eine Mögliche Art der Elektronenbindung an Kirstalloberflächen, Physik Z. Sowjetunion 1, 733.

I. Tamm and D. Blochinzev, Über die Austrittsarbeit der Elektronen aus Metallen, Z. Physik 77, 774.

A. H. Wilson, A note on the theory of rectification, Proc. Roy. Soc. (London) A136, 487.

A. H. Wilson, The theory of metals I., Proc. Roy. Soc. (London) A138, 594.

1933

H. Bethe and H. Fröhlich, Magnetische Wechselwirking der Metallelektronen; Zur Kritik der Theorie der Supraleitung von Frenkel, Z. Physik 85, 389.

F. Bloch, Remarque sur un theoreme de conservation dans la theorie des metaux, J. phys. et radium 7-4, 486.

F. Bloch, Zur Bremsung rasch bewegter Teilchen beim Durchgang durch Materie, Ann. Physik 16, 285.

F. Bloch, Bremsvermögen von Atomen mit mehreren Elektronen, Z. Physik 81, 363.

D. Blochinzev and L. Nordheim, Zur Theorie der anomaler magnetischer und thermoelektrischer Effekte in Metallen, Z. Physik 84, 168.

L. Brillouin, Les electrons libres dans un reseau cristallin, equation ondulatoire et proprietes magnetiques, champs self-consistents et electrons metalliques, J. phys. et radium 7-4, 1.

L. Brillouin, Le champ self-consistent, pour les electrons lies, la supraconductibilite, J. phys. et radium 7-4, 333.

L. Brillouin, Sur la stabilite du courant dans un supraconducteur, J. phys. et radium 7-4, 677.

R. H. Fowler, Notes on some electronic properties of conductors and insulators, Proc. Roy. Soc. (London) A141, 56.

1933 (con'd)

J. Frenkel, On a possible explanation of superconductivity, Phys. Rev. 43, 907

W. Kroll, Zur Theorie der Druckabhängigkeit der elektrischen Leitfähigkeit der Metalle, Z. Physik 85, 398.

R. de L. Kronig, Zur Theorie der Supraleitfähigkeit, Z. Physik 78, 744; 80, 203.

L. Landau, Theorie der Supraleitfähigkeit, Physik Z. Sowjetunion 4, 43.

R. Peierls, Zur Theorie des Diamagnetismus von Leitungselektronen, Z. Physik 80, 763, 81, 186.

1934

L. Brillouin, Les electrons dans les metaux, du point de vue ondulatoire, Act Sci. et. Ind. 88.

L. Brillouin, Conductibilite electrique et thermique des metaux, Act. Sci. et Ind. 89.

L. Brillouin, Fluctuations du courant dans un conducteur, Helv. Phys Acta 7 (Sup. 2) 47.

L. Brillouin, Les bases de la theorie electronique des metaux et la methode des champs self-consistents, Helv. Phys. Acta 7 (Sup. 2), 33.

L. Brillouin, Le champ self-consistent de Fock pour les electrons des metaux, J. phys. et radium 7-5, 413.

H. Jones and C. Zener, A general proof of certain fundamental equations in the theory of metallic conduction, Proc. Roy. Soc. (London) A144, 101.

H. Jones and C. Zener, The theory of the change in resistance in a magnetic field, Proc. Roy. Soc. (London) A145, 268.

H. Jones, The theory of alloys in the γ-phase, Proc. Roy. Soc. (London) A144, 225.

H. Jones, Applications of the Bloch theory to the study of alloys and of the properties of bismuth, Proc. Roy. Soc. (London) A147, 396.

H. Jones, Application de la theorie electronique des metaux a l'etude des alliages, Helv. Phys. Acta 7 (Sup. 2), 84.

H. London, Production of heat in superconductors by alternating currents, Nature 133, 497.

A. Maue, Zur Frage der Oberflächenleitfähigkeit von Metallen, Naturwiss 38, 648.

A. Maue, Die Oberflächenwelle in der Elektronentheorie der Metalle, Helv. Phys. Acta 7 (Sup. 2), 68.

N. F. Mott, The conductivity of metals, Proc. Phys. Soc. (London) 46, 460.

N. F. Mott, The resistance of liquid metals, Proc. Roy. Soc. (London) A146, 465.

K. F. Niessen, Temperaturabhängigkeit eines Elektronengases, Physica 1, 979.

L. W. Nordheim, Kinetische Theorie des metallischen Zustandes, Müller-Pouillet's Lehrbuch der Physik 1, 293-876, Braunschweig.

L. W. Nordheim, Sur les limites de la theorie elementaire des electrons metalliques, Helv. Phys. Acta 7 (Sup. 2), 3.

R. Peierls, Über die statistischen Grundlagen der Elektronentheorie der Metalle, Helv. Phys. Acta 7 (Sup. 2), 24.

R. Peierls, Bemerkungen zur Theorie der Metalle, Z. Physik 88, 786.

S. Rijanow, Zur Frage nach der "Elektronenbewegung" im beschränkten Kristallgitter, Z. Physik 89, 806.

S. Schubin, On the theory of liquid metals, Physik Z. Sowjetunion 5, 81.

J. C. Slater, Electronic structure of metals, Revs. Modern Phys. 6, 209.

1934 (con'd)

A. H. Wilson, The electrical properties of semiconductors and insulators, Act. Sci. et Ind. 82.

C. Zener, A theory of the electrical breakdown of solid dielectrics, Proc. Roy. Soc. (London) A145, 529.

1935

N. H. Frank, Effect of pressure on the electrical conductivity of the alkalies, Phys. Rev. 47, 282.

F. Hund, Theorie der Bewegung der Elektronen in nichtmetallischen Kristallgittern, Z. tech. Physik 16, 331; Physik Z. 36, 725.

W. H. Keesom, Discussion on supraconductivity, Proc. Roy. Soc. (London) A152, 11.

M. H. Lensen and A. Michels, Contribution to the theory of the influence of pressure on the electrical resistivity of metals, Physica 2, 1091.

F. and H. London, The electromagnetic equations of the supraconductor, Proc. Roy. Soc. (London) A149, 71.

F. London, Discussion on supraconductivity, Proc. Roy. Soc. (London) A152, 24.

A. Maue Die Oberflächenwellen in der Elektronentheorie der Metalle, Z. Physik 94, 717.

N. F. Mott, Discussion on supraconductivity, Proc. Roy. Soc. (London) A152, 42.

J. H. Simpson, Charge distribution and energy levels of trapped electrons in ionic solids, Proc. Roy. Soc (London) A197, 269.

E Wigner and J. Bardeen, Theory of the work functions of monovalent metals, Phys. Rev. 48, 84.

C. Zener, Hall coefficients of alkali metals, Phys. Rev. 47, 636 (L).

1936

J. Bardeen, Theory of the work function. II. The surface double layer, Phys. Rev. 49, 653.

J. Bardeen, Electron exchange in the theory of metals, Phys. Rev. 50, 1098 (L).

F. Cernuschi, Elementary theory of the critical field of a dielectric, Proc. Cambridge Phil. Soc. 32, 276.

H. Fröhlich, Elektronentheorie der Metalle, Springer, Berlin.

H. Jones, The theory of the galvanomagnetic effects in bismuth, Proc. Roy. Soc. (London) A155, 653.

N. F. Mott, The electrical conductivity of transition metals, Proc. Roy. Soc. (London) A153, 699.

N. F. Mott, The electrical resistance of dilute solid solutions, Proc. Cambridge Phil. Soc. 32, 281.

N. F. Mott, The resistance and thermoelectric properties of the transition metals, Proc. Roy. Soc. (London) A156, 368.

N. F. Mott and H. Jones, The theory of the properties of metals and alloys, Oxford University Press.

E. P. Wigner, On the constant A in Richardson's equation, Phys. Rev. 49, 696.

A. H. Wilson, Theory of Metals, Cambridge University Press.

1937

J. Bardeen, Conductivity of monovalent metals, Phys. Rev. 52, 688.

H. Fröhlich, Theory of electrical breakdown in ionic crystals, Proc. Roy. Soc. (London) A160, 230.

R. W. Gurney and N. F. Mott, Trapped electrons in polar crystals, Proc. Phys. Soc. (London) 49, 32.

F. London, On a letter of Wick's about superconductivity, Phys. Rev. 52, 886 (L).

F. London, Une conception nouvelle de la supra-conductibilite, Act. Sci. et Indust. 458.

F. London, On the nature of the superconducting state, Phys. Rev. 51, 678 (L).

E. L. Peterson and L. W. Nordheim, Resistance of monovalent metals, Phys. Rev 51, 355.

J. C. Slater, Nature of the superconducting state, Phys. Rev. 51, 195; 52, 214.

J. C. Slater, Damped electron waves in crystals, Phys. Rev. 51, 840.

A. von Hippel, Electric breakdown of solid and liquid insulators, J. Appl. Phys. 8, 815.

G. H. Wannier, Structure of electronic excitation levels in insulating crystals, Phys. Rev. 52, 191.

G. C. Wick, Diamagnetism of superconducting bodies, Phys. Rev. 52, 526 (L).

A. H. Wilson, The second order electrical effects in metals, Proc. Cambridge Phil. Soc. 33, 371.

1938

R. W. Gurney and N. F. Mott, Conduction in polar crystals IV. On the color centres in alkali halide crystals, Trans. Faraday Soc. 34, 506.

M. von Laue, Zum Thermodynamik der Supraleitung, Ann. Physik 32, 71.

M. von Laue, Der magnetische Schwellenwert für Supraleitung, Ann. Physik 32, 253.

N. F. Mott, The conduction of electricity in polar crystals, Science Progress 32, No. 128.

R. J. Seeger and E. Teller, On the electrical breakdown of the alkali halides, Phys. Rev. 54, 515.

H. Welker, Über ein elektronentheoretisches Modell des Supraleiters, Physik Z. 39, 920.

A. H. Wilson, The electrical conductivity of the transition metals, Proc. Roy. Soc. (London) A167, 580.

1939

J. Bardeen and J. H. Van Vleck, Current in the Bloch approximation of "tight binding" for metallic electrons, Proc. Nat. Acad. Sci. 25, 82.

W. Franz, Zur Theorie des elektrischen Durchschlags fester Isolatoren, Z. Physik 113, 607.

W. Franz, Zur Theorie des elektrischen Durchschlags fester Isolatoren, Naturwiss. 27, 433.

H. Fröhlich, Dielectric breakdown in ionic crystals, Phys. Rev. 56, 349.

H. Fröhlich, Theory of electrical breakdown in ionic crystals, II. Proc. Roy. Soc. (London) A172, 94.

H. Fröhlich and N. F. Mott, The mean free path of electrons in polar crystals, Proc. Roy. Soc. (London) A171, 496.

1939 (con'd)

E. T. Goodwin, Electronic states at the surfaces of crystals I. The approximation of nearly free electrons; II. The approximation of tight binding, finite linear chain of atoms; III. The approximation of tight binding, further extensions, Proc. Cambridge Phil. Soc. 35, 205, 221, 232.

W. V. Houston, Conservation of momentum in electrical conductivity, Phys. Rev. 55, 1255.

R. J. Seeger and E. Teller, Remarks on the dielectric breakdown, Phys. Rev. 56, 352.

W. Shockley, On the surface states associated with a periodic potential, Phys. Rev. 56, 317.

S. C. Tibbs, Electron energy levels in NaCl, Trans. Faraday Soc. 35, 1471.

A. H. Wilson, Semiconductors and metals; an introduction to the electron theory of metals, Cambridge University Press.

1940

W. V. Houston, Acceleration of electrons in a crystal lattice, Phys. Rev. 57, 184

N. F. Mott and R. W. Gurney, Electronic processes in ionic crystals, Oxford University Press.

F. Seitz, Modern theory of solids, McGraw-Hill Book Company, Inc.

J. H. Van Vleck, Electronic conduction and the equilibrium of lattice oscillators, Rev. Univ. nac. Tucuman A1, 1, 81.

1941

H. Fröhlich, On the electric strength of mixed crystals, Proc. Roy. Soc. (London) A178, 493.

J. H. Van Vleck, Calculation of energy exchange between lattice oscillators, Phys. Rev. 59, 730.

1942

H. Fröhlich, Electric breakdown of ionic crystals, Phys. Rev. 61, 200.

1943

H. Fröhlich, Theory of dielectric breakdown, Nature 151, 339.

R. de L. Kronig and J. Korringa, Zur Theorie der Bremsung schneller geladener Teilchen in metallischen Leitern, Physica 10, 406; 10, 800.

V. Weisskopf, On the theory of the electrical resistance of metals, Am. J. Phys. 11, 1.

1945

M. von Laue, Supraleitung und Hertzsche Schwingungen, Z. Physik 124, 135.

R. Maurer, The electrical properties of semiconductors, J. Appl. Phys. 16, 653.

F. Seitz, The basic principles of semiconductors, J. Appl. Phys. 16, 553.

1946

S. Pekar, Local quantum states of an electron in an ideal ionic crystal, J. Phys. (U.S.S.R.) 10, 341.

S. Pekar, Autolocalization of an electron in a dielectric inertially polarizing medium, J. Phys. (U.S.S.R.) 10, 347.

F. Seitz, Color centers in alkali halide crystals, Revs. Modern Phys. 18, 384.

1947

J. Bardeen, Surface states and rectification at a metal semi-conductor contact, Phys Rev. 71, 717.

H. Fröhlich, On the theory of dielectric breakdown in solids, Proc. Roy. Soc. (London) A188, 521.

H. Fröhlich, Energy distribution and stability of electrons in electric fields, Proc. Roy. Soc. (London) A188, 532.

W. Heis berg, Zur Theorie der Supraleitung, Z. Naturforsch. 2a, 185.

W. Heisenberg, Die Übertragung elektromagnetischer Kräfte in Supraleiter, Nachr. Akad. Wiss. Göttingen, Math.-phys. Kl. 23.

V. A. Johnson and K. Lark-Horovitz, Transition from classical to quantum statistics in germanium semiconductors at low temperature, Phys. Rev. 71, 374 (L).

H. Koppe, Die spezifische Wärme der Supraleiter nach der Theorie von W. Heisenberg, Ann. Physik 1, 405.

H. A. Kramers, The stopping power of a metal for alpha particles, Physica 13, 401.

A. B. Pippard, Resistance of superconductors at microwave frequencies, Phys. Soc. Cambridge Conf. Report, 115.

A. B. Pippard, The surface impedance of superconductors and normal metals at high frequencies, Proc. Roy. Soc. (London) A191, 370, 385, 399.

E. H. Sondheimer and A. H. Wilson, The theory of the magneto-resistance effects in metals, Proc. Roy. Soc. (London) A190, 435.

1948

M. Born and K. C. Cheng, Theory of superconductivity, Nature 161, 968, 1017.

W. Heisenberg, Das elektrodynamische Verhalten der Supraleiter, Z. Naturforsch 3a, 65.

W. Heisenberg, On the electron theory of superconductivity. In: Two lectures, Cambridge University Press.

W. Heisenberg, Thermodynamische Betrachtungen zum Problem der Supraleitung, Ann. Physik 3, 289.

K. Huang, On Heisenberg's theory of superconductivity, Nat. Peking Univ. Semi-Cent. Vol., Math-Phys-Biol. Sci. 89.

H. Koppe, Zur Theorie der Supraleitung II: Berechnung der Sprungtemperature, Z. Naturforsch. 3a, 1.

R. Kubo, Interaction between electrons and ions in semiconductors, J. Phys. Soc. (Japan) 3, 254.

M. von Laue, Supraleitung und Kristallklasse, Ann. Physik 3, 40.

M. von Laue, London's Theorie für nicht-kubische Supraleiter, Ann. Physik 3, 31.

F. London, On the problem of the molecular theory of superconductivity, Phys. Rev. 74, 562.

J. J. Markham and F. Seitz, Binding energy for a self-trapped electron in NaCl, Phys. Rev. 74, 1014.

F. Möglich and R. Ronge, The magnetic threshold value in the theory of superconductivity, Ann. Physik 3, 322.

N. F. Mott and R. W. Gurney, Electric Processes in Ionic Crystals, 2nd Edition, Oxford University Press.

A. Radkowsky, Temperature dependence of electronic energy levels in solids, Phys. Rev. 73, 749.

1948 (con'd)

G. E. Reuter and E. H Sondheimer, The theory of the anomalous skin effect in metals, Proc. Roy. Soc. (London)A 195, 336.

F. Seitz, On the mobility of electrons in pure non-polar insulators, Phys. Rev. 73, 549.

E. P. Wohlfarth, Anomalous values of the thermionic emission constant, Phys. Rev. 74, 984 (L).

E. P. Wohlfarth, Thermionic emission constants and band overlap, Proc. Phys. Soc. (London) 60, 360.

1949

D. Bohm, Note on a theorem of Bloch concerning possible causes of superconductivity, Phys. Rev. 75, 502.

H. B Callen, Electric breakdown in ionic crystals, Phys. Rev. 76, 1394.

E. U. Condon, Superconductivity and the Bohr magnetron, Proc. Nat. Acad. Sci. 35, 588.

R. A. Hutner, E. S. Rittner, and F. K. Du Pre, Concerning the work of polarization in ionic crystals of the NaCl type, II. Polarization around two adjacent charges in the rigid lattice, J. Chem. Phys. 17, 204.

H. Ichimura, A statistical mechanical treatment of conduction electrons: an attempt to the theory of superconductivity, J. Phys. Soc. (Japan) 4, 265.

H. M. James, Energy bands and wave functions in periodic potentials, Phys. Rev. 76, 1602.

H. M. James, Electronic states in perturbed periodic systems, Phys. Rev. 76, 1611.

J. S. Koehler, Calculation of the changes in the conductivity of metals produced by cold work, Phys. Rev 75, 106.

M. Kohler, Allgemeine Theorie der Abweichungen von der Matthiessenschen Regel, Z. Physik 126, 495.

H. Koppe, Zur Theorie der unvollständigen Supraleitung, Ann. Physik 6, 375.

R. Kubo, Interaction between electrons and ions in semiconductors, II, III, J. Phys. Soc. (Japan) 4, 233; 4, 326.

M. von Laue, Eine nicht-lineare, phänomenologische Theorie der Supraleitung, Ann. Physik 5, 197.

M. von Laue, Supraleitung und elektrodynamisches Potential, Z. Physik 125, 517.

F. Möglich and R. Rompe, Zur Theorie der Supraleitung, Ann. Physik 6, 177.

F. Möglich and R. Rompe, Eindringtiefe magnetischer Felder in Supraleiter und Metallabsorption, Ann. Physik 4, 335.

N. F. Mott, Notes on the transistor and surface states in semiconductors, Elec. Res. Ass. (Ref. L/T216).

T. Muto, Theory of the F-centers of colored alkali halide crystals. I. Structure of the F-absorption bands, Prog. Theor. Phys. 4, 181.

T. Muto, Theory of the F-centers of colored alkali halides. II. Electronic structure of F-centers; general theory, Prog. Theor. Phys. 4, 243.

S. I. Pekar, Theory of polarons, J. Exp. Phys. (U.S.S.R.) 19, 796.

E. S. Rittner, R. A. Hutner, and F. K. du Pre, Concerning the work of polarization in ionic crystals of the NaCl Type. I. Polarization around a single charge in the rigid lattice, J. Chem. Phys. 17, 198.

G. V. Schubert, Der Energie-Impulstensor in der von Laue-Londonschen Elektrodynamik des Supraleiters, Ann. Physik 6, 163.

1949 (cont'd)

F. Seitz, On the theory of electron multiplication in crystals, Phys. Rev. 76, 1376.

J. H. Simpson, Charge distribution and energy levels of trapped electrons in ionic solids, Proc. Roy. Soc. (London) A197, 269.

J. C. Slater, Electrons in perturbed periodic lattices, Phys. Rev. 76, 1592.

G. H. Wannier, On the energy band structure of insulators, Phys. Rev. 76, 438 (L).

H. Welker, Über den Zusammenhang zwischen der Supraleitung und der gemischten Leitung, Ann. Physik 5, 1.

J. Yamashita, Theory of thermoelectric motive force in semiconductors, J. Phys. Soc. (Japan) 4, 310.

1950

J. Bardeen, Zero-point vibrations and superconductivity, Phys. Rev. 79, 167.

J. Bardeen, Wave functions for superconducting electrons, Phys. Rev. 80, 567.

J. Bardeen and W. Shockley, Scattering of electrons in crystals in the presence of large electric fields, Phys. Rev. 80, 69.

J. Bardeen and W. Shockley, Deformation potentials and mobilities in non-polar crystals, Phys. Rev. 80, 72.

G. Busch, Elektronenleitung in Nichtmetallen, Z. angew. Math. und Phys. 1, 3, 81.

H. B. Callen, Polaron states in ionic crystals, Phys. Rev. 79, 533.

E. Conwell and V. F. Weisskopf, Theory of impurity scattering in semiconductors, Phys. Rev. 77, 388.

C. A. Coulson and G. R. Baldock, On conditions for the existence of surface states, Disc. Faraday Soc. No. 8, 27.

F. K. du Pre, R. A. Hutner, and E. S. Pittner, Concerning the work of polarization in ionic crystals of the [illegible], III. Numerical results for a single charge in a rigid lattice, J. Chem. Phys. 18, 379.

H. Fröhlich and F. Seitz, Notes on the theory of dielectric breakdown in ionic crystals, Phys. Rev. 79, 526.

H. Fröhlich, H. Pelzer, and S. Zienau, Properties of slow electrons in polar materials, Phil. Mag. 41, 221.

H. Fröhlich, Isotope effect in superconductivity, Proc. Phys. Soc. (London) A63, 778.

H. Fröhlich, Conduction electrons in non-metallic solids, Research (London) 3, 202.

C. S. Hung, Theory of resistivity and Hall effect at very low temperatures, Phys. Rev. 79, 727.

T. Inui and Y. Uemura, Theory of color centers in ionic crystals, Prog. Theor. Phys. 5, 252, 395.

S. Katsura, T. Hatta, and A. Morita, On the conception of the energy band in the perturbed periodic field, Sci. Rep. Tohoku Univ. 34, 19.

H. Koppe, Theorie der Supraleitung, Ergeb. exakt. Naturwiss. 23, 283.

F. London, Superfluids, Vol. 1. Macroscopic theory of superconductivity, John Wiley and Sons, Inc., New York.

J. K. Mackenzie and E. H. Sondheimer, The theory of the change in the conductivity of metals produced by cold work, Phys. Rev. 77, 264.

T. Muto and S. Oyama, Theory of the temperature effect of electronic energy bands in crystals, Prog. Theor. Phys. 5, 833.

K. F. Niessen, On one of Heisenberg's hypotheses in the theory of specific heat of superconductors, Physica 16, 77.

1950 (con'd)

K. F. Niessen, The energy of the normal electrons in a superconductor as a function of temperature and thickness of the superconducting layer on the Fermi surface, Physica 16, 84.

A. B. Pippard, The surface impedance of superconductors and normal metals at high frequencies, Proc. Roy. Soc. (London) A203, 98, 195.

P. Rhodes, The Bloch integral equation and electrical conductivity, Proc. Roy. Soc. (London) A202, 466.

R. G. Sachs and D. L. Dexter, Quantum limits of the electrostatic image-force theory, J. Appl. Phys. 21, 1304.

F. Seitz, Note on the theory of resistance of a cubic semiconductor in a magnetic field, Phys. Rev. 79, 372.

W. Shockley, Electrons and holes in semiconductors, D. Van Nostrand Company.

W. Shockley, Effect of magnetic fields on conduction -- "tube integrals", Phys. Rev. Phys. Rev. 79, 191.

W. Shockley and J. Bardeen, Energy bands and mobilities in monatomic semiconductors, Phys. Rev. 77, 407 (L).

E. H. Sondheimer, The theory of the transport phenomena in metals, Proc. Roy. Soc. (London) A203, 75.

H. Statz, Zur Theorie der Oberflächenzuständen, Z. Naturforsch. 5a, 534.

1951

J. Bardeen, Choice of gauge in London's approach to the theory of superconductivity, Phys. Rev. 81, 469 (L).

J. Bardeen, Relation between lattice vibrations and London's theory of superconductivity, Phys. Rev. 81, 829.

J. Bardeen, Criterion for superconductivity, Phys. Rev. 82, 978.

J. Bardeen, Electron-vibration interactions and superconductivity, Revs. Modern Phys. 23, 261.

D. Bohm, Application of collective treatment of electron and ion vibrations to theories of conductivity and superconductivity, Phys. Rev. 84, 836.

S. D. Drell, On the interaction of conduction electrons and lattice vibrations, Phys. Rev. 83, 838 (L).

W. Franz, Der Mechanismus des elektrischen Durchschlags fester Isolatoren, Z. angew. Physik 3, 72.

H. Fröhlich, Theory of the superconducting state II, Magnetic properties at the absolute zero of temperature, Proc. Phys. Soc. (London) A64, 129.

E. P. Gross, Note on the interaction of an electron and a lattice oscillator, Phys. Rev. 84, 818.

W. R. Heller, Kinetic-statistical theory of dielectric breakdown in non-polar crystals, Phys. Rev. 84, 1130.

H. Jones, Hall coefficient of semiconductors, Phys. Rev. 81, 149.

W. Kohn and Vachaspati, Difficulty in Fröhlich's theory of superconductivity, Phys. Rev. 83, 462 (L).

R. Landauer, Conductivity of cold-worked metals, Phys. Rev. 82, 520.

J. M. Luttinger, Effect of a magnetic field on electrons in a periodic potential, Phys. Rev. 84, 814.

K. B. McAfee, E. J. Ryder, W. Shockley, and M. Sparks, Observations of Zener current in germanium p-n junctions, Phys. Rev. 83, 650 (L).

1951 (con'd)

T. Nishiyama, On the interaction of the lattice vibrations with the conduction electrons, Prog. Theor. Phys. 6, 897 (L).

C. Okasaki and T. Suita, Electronic conduction in ionic crystals. I, J. Phys. Soc. (Japan) 6, 200.

M. F. M. Osborne, Perfect diamagnetism of free electrons with application to superconductivity, Phys. Rev. 81, 147.

L. Pincherle, Energy levels of F-centers, Proc. Roy. Soc. (London) A64, 648.

F. Seitz, Color centers in additively colored alkali halide crystals containing alkaline earth ions, Phys. Rev. 83, 134.

G. Wentzel, Interaction of lattice vibrations with electrons in a metal, Phys. Rev. 83, 168.

F. E. Williams, Absolute theory of solid-state luminescence, J. Chem. Phys. 19, 457.

1952

E. N. Adams, II, Motion of an electron in a perturbed periodic potential, Phys. Rev. 85, 41.

Paula Feuer, Electronic states in crystals under large over-all perturbations, Phys. Rev. 88, 92.

J. Friedel, The distribution of electrons round impurities in monovalent metals, Phil. Mag. 43, 153.

R. Kubo, Thermal ionization of trapped electrons, Phys. Rev. 86, 929.

M. von Laue, Theory of superconductivity, trans. by L. Meyer and W. Band, Academic Press, Inc., New York.

F. London, Bemerkungen zu vorstehender Arbeit zur "Theorie der Supraleitung" von Max von Laue, Ann. Physik 10, 392, 314.

M. von Laue, Kritische Bemerkungen zur Theorie der Supraleitung, Ann. Physik 10, 296, 305.

T. Nagamiya, Theory of color centers, J. Phys. Soc. (Japan) 7, 354, 358.

M. F. M. Osborne, Number theory and the magnetic properties of an electron gas, Phys. Rev. 88, 438.

M. C. Steele, Application of the theory of numbers to the magnetic properties of an electron gas, Phys. Rev. 88, 451.

J. Yamashita and M. Watanabe, On the conductivity of non-polar crystals in strong electrostatic field, J. Phys. Soc. (Japan) 7, 334.

1953

H. Jones, The valency of the transition metals in some binary and ternary alloys, Phil. Mag. 44, 907.

A. H. Kahn and C. Kittel, F-center wave functions and electronic g-values in KCl crystals, Phys. Rev. 89, 315 (L).

H. Mori, The quantum statistical theory of transport phenomena, II. On the theory of metallic conductivity, Prog. Theor. Phys. 9, 473.

A. B. Pippard, An experimental and theoretical study of the relation between magnetic field and current in a superconductor, Proc. Roy. Soc. (London) A216, 547.