A Soluble Problem in Energy Bands*

J. C. SLATER†
Brookhaven National Laboratory, Upton, New York
(Received April 30, 1952)

The problem of an electron moving in a periodic simple cubic potential of the form $\cos x+\cos y+\cos z$ is investigated, with particular attention to the nature of the wave functions, Wannier functions, degeneracy of overlapping bands, etc. It is well known that this problem is separable, leading to Mathieu's equation for the functions of x, y, and z. This equation can be solved, making use of the Fourier expansion of the wave function (the momentum eigenfunction). Solutions are set up, for a number of interatomic distances. The momentum eigenfunctions are similar to Hermite functions for the tightly-bound levels but very different for free-electron-like levels. Those for the levels near the top of the potential barrier, corresponding to the valence and conduction bands of a real crystal, are very complicated; we illustrate their properties by means of a test of the completeness relation for orthogonal functions. It is proved that the Wannier function is the Fourier transform of the momentum eigenfunction, and special cases are discussed. The effect of various perturbations on the one-dimensional problem is considered. It is shown that the usual assumption made in band theory, connecting the width of the gap introduced by a perturbation with the matrix component of the perturbative potential, is not generally justified. The case of a perturbation of double periodicity is taken up, and it is shown that this results in a spreading out of the Wannier function, which sometimes has large contributions on many atoms. In the two- and three-dimensional cases, the interesting feature is the degeneracy between overlapping bands, for example, between the different components of a p or a d band. These are discussed in detail. For the unperturbed Mathieu case, the energy surfaces cut without interaction, but almost any reasonable perturbation will cause them to interact, resulting in perturbed bands of great complication, which will not cut each other, though they may adhere at certain points of crystal symmetry, as can also be proved by group theory. The Wannier functions in these cases become combinations of the Wannier functions of the unperturbed problem, located on many different atoms, and possessing the rotational symmetry of the crystal, which the unperturbed Wannier functions do not have. The type of perturbation is also taken up which would convert the simple cubic potential of the unperturbed problem into a face-centered or body-centered structure.

ONE of the important parts of the quantum theory of solids is the solution of Schrödinger's equation for an electron moving in a periodic potential. Everyone is familiar with the two limiting approximate methods which have proved useful for this problem: the bound-electron or atomic approximation appropriate for the tightly-bound atomic electrons whose wave functions hardly overlap, and the free-electron approximation appropriate for electrons moving so fast that their wave functions are not far from plane waves. Everyone is also familiar with the fact that the general properties of energy bands, arising merely from the periodicity of the potential, and the mathematics associated with Floquet's theorem are valid independent of any approximation and hold equally for both of the limiting methods mentioned above. Unfortunately, however, there is every reason for thinking that the electrons which we are most interested in—the valence and conduction electrons in crystals—have wave functions of a type intermediate between the range of validity of the two approximations, and probably not representable very well by either method. Both methods, of course, have been used to describe these wave functions, as has also the cellular method. Nevertheless, it would add greatly to our confidence if we had a problem, not too far from the realities of the crystal lattice, in which we could follow through rigorously between the regions in which the two limiting approximations are applicable and see what really happens in the intermediate case.

Fortunately such a problem exists: the potential energy of the general form $-\cos x-\cos y-\cos z$. This has minima at the points of a simple cubic lattice, $x=2\pi n_x$, $y=2\pi n_y$, $z=2\pi n_z$, where n_x, n_y, n_z are integers. Around each minimum, the potential energy can be expanded in power series, representing a three-dimensional linear oscillator for small amplitudes; around the origin it has the value $-3+r^2/2\cdots$, where $r^2=x^2+y^2+z^2$. Thus the wave functions of tightly-bound states will be like the solutions of the three-dimensional oscillator problem and can be written in terms of Hermite functions. On the other hand, the high energy levels of the problem are like plane waves. The great virtue of this potential is that, being the sum of a function of x, a function of y, and a function of z, the problem is separable, and the corresponding one-dimensional problem is Mathieu's equation, which has been the subject of much attention by the mathematicians.

It is, of course, nothing new to realize this situation. Brillouin[1] used this problem in some of his earliest work on the theory of energy bands. Morse[2] recognized its value very early and used it for study of electron diffraction; MacColl,[3] working recently on the same problem by similar methods, seems unaware of Morse's earlier work. As an illustration of a recent application of the method, we may mention the work of Statz[4] on surface states. Nevertheless, there is a good deal more information to be extracted from the problem than has been found so far, and the writer has examined particularly the nature of the intermediate cases between the free and bound states, with results to be described in the present paper. This work has been considerably

* Research carried on under the AEC.
† On leave from Massachusetts Institute of Technology, Cambridge, Massachusetts.

[1] L. Brillouin, J. phys. et radium **1**, 377 (1930).
[2] P. M. Morse, Phys. Rev. **35**, 1310 (1930).
[3] L. A. MacColl, Bell System Tech. J. **30**, 888 (1951).
[4] H. Statz, Z. Naturforsch. **5a**, 534 (1950).

simplified and facilitated by the recent appearance[5] of a very important tabulation of some of the numerical properties of the Mathieu functions; though we shall see that some of the things most needed for the present study are not tabulated and have had to be computed.

We make a study of the one-, two-, and three-dimensional cases; the one-dimensional case because it is obviously simplest, and the foundation of the others; the two-dimensional case because it presents many of the complicating features of the three-dimensional case, and yet can be visualized and shown graphically much more easily; and the three-dimensional case for the obvious reason that it most closely resembles a real crystal. We consider not only the separable Mathieu problem, but also modifications of it resulting from introducing other Fourier components of potential as perturbations, resulting in potentials closer to those met in real crystals. We make much use of the Wannier functions, and particularly of their Fourier transforms, which we show, by extension of existing theorems, to be identical with the momentum eigenfunctions. In connection with this, we throw light on such previously puzzling questions as the nature of the Wannier functions in a case of overlapping bands.

As for specific results of direct value in the theory of energy bands, it is perhaps best merely to say that the material presented here should be very valuable to one trying to set up methods of approximating the solution of the periodic potential problem in real crystals; and in particular, it shows that our misgivings are justified, that the real wave functions of the valence and conduction electrons must be rather widely different from the approximate functions set up by either of the two limiting processes, so that we must be much more on our guard against the use of these procedures than most writers have been in the past.

1. ENERGY BANDS FOR THE ONE-DIMENSIONAL MATHIEU PROBLEM

For convenience in intercomparison, we shall describe the potential energy of an electron in a crystal in the same notation used in an earlier reference by the writer.[6] In this case, we set up a set of vectors $\mathbf{Q}_j$ from the origin to points corresponding to the origin in the jth unit cell of the crystal; $\mathbf{Q}_j$ will be expressible as a linear combination of three fundamental vectors $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{Q}_3$, with integral coefficients. For a simple cubic lattice, these fundamental vectors are of magnitude Q_0, and point along the three coordinate axes. Then we set up a set of vectors $\mathbf{P}_i$, in a momentum space (proportional to a reciprocal space), by the equations

$$\mathbf{P}_i\cdot\mathbf{Q}_j=h\delta_{ij}, \tag{1}$$

where h is Planck's constant, and δ_{ij} the Kronecker symbol. Thus the vectors $\mathbf{P}_i$ can be written as linear combinations of three fundamental vectors $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$, with integral coefficients; in the simple cubic lattice, these vectors are all of magnitude h/Q_0 and are along the three coordinate axes. Then we can express any periodic potential energy U which repeats in each unit cell according to the Fourier expansion

$$U(r)=\sum(\mathbf{P}_i)W(\mathbf{P}_i)\exp[(-2\pi i/h)(\mathbf{P}_i\cdot\mathbf{r})], \tag{2}$$

where $\mathbf{r}$ is the radius vector to any point of the lattice. The periodicity, according to which $U(\mathbf{r}+\mathbf{Q}_j)=U(\mathbf{r})$, follows at once from (1) and (2). To insure having a real potential, we must have $W(-\mathbf{P}_i)=W(\mathbf{P}_i)^*$.

The general case of a simple cubic structure could then be expressed in the form

$$U(r)=\sum(lmn)W(lmn)\times\exp[(-2\pi i/Q_0)(lx+my+nz)], \tag{3}$$

where x, y, z are the components of r, and $W(lmn)$ represents the Fourier coefficient associated with the point in momentum space with the coordinate lP_0, mP_0, nP_0. The simplified case which we can handle by our Mathieu method is that for which the only non-vanishing components are for l, m, $n=0$ (a constant potential), and for the combinations (±100), (0 ± 10), (00 ± 1). We may then write

$$U(r)=W_0-2W_1(\cos2\pi x/Q_0+\cos2\pi y/Q_0+\cos2\pi z/Q_0), \tag{4}$$

where $W_0=W(0,0,0)$, and where we have chosen $W(\pm1,0,0)$, $W(0,\pm1,0)$, $W(0,0,\pm1)$ all to be equal, real, and equal to $-W_1$, where W_1 is positive; the negative sign has the effect of putting a minimum of potential at the origin. We shall find it convenient to choose W_0 so that the potential energy is zero at the potential minima, such as the origin; that is, we let

$$W_0=6W_1, \tag{5}$$

so that

$$U(r)=2W_1[(1-\cos2\pi x/Q_0)+(1-\cos2\pi y/Q_0)+(1-\cos2\pi z/Q_0)]. \tag{6}$$

Schrödinger's equation is then

$$-(h^2/8\pi^2m)\nabla^2u+U(r)u=Eu. \tag{7}$$

We can at once separate variables; if $u=u_x(x)u_y(y)u_z(z)$, corresponding to the energy $E=E_x+E_y+E_z$, we have

$$-(h^2/8\pi^2m)[d^2u_x(x)/dx^2]+2W_1(1-\cos2\pi x/Q_0)u_x(x)=E_xu_x(x). \tag{8}$$

This is Mathieu's equation. We wish to introduce dimensionless variables, to simplify notation. In doing

FIG. 1. Parabolic and sinusoidal potentials for two Q_0's, one twice the other. Note how the barrier height varies as Q_0^2.

[5] *Tables Relating to Mathieu Functions*, U. S. National Bureau of Standards (Columbia University Press, New York, 1951).

[6] J. C. Slater, Revs. Modern Phys. **6**, 209 (1934).

so, we shall try to make our equation resemble forms already existing in the literature. Unfortunately, there are many different notations in use. However, we gain a certain connection with other notations if we let

$$w=\pi x/Q_0, \quad s=32mQ_0^2W_1/h^2,$$
$$\epsilon=(2m/W_1)^{\frac{1}{2}}(Q_0E/h), \quad a=\epsilon s^{\frac{1}{2}}-s/2. \tag{9}$$

In terms of these quantities, we may express (8) in the alternative forms

$$-d^2u/dw^2+\tfrac{1}{2}s(1-\cos 2w)u=\epsilon s^{\frac{1}{2}}u, \tag{10}$$

$$d^2u/dw^2+(a+\tfrac{1}{2}s\cos 2w)u=0. \tag{11}$$

The form (11) is identical with Eq. (2.01) of reference 5, except that that equation has a minus sign instead of plus for the term $\frac{1}{2}s\cos 2w$; that is, the form in reference 5 corresponds to having the minima of potential at the points $w=\pm\pi/2, \pm 3\pi/2, \cdots$, instead of at $w=0, \pm\pi, \pm 2\pi, \cdots$, as we have in (10) and (11). On page xvii of reference 5 are given the rules for changing the sign of s, so that the many useful formulas of that reference are immediately applicable without change of notation.

Before going further, it is worth while making connection between our problem and a real case of energy bands in a crystal. The potential energy of Eq. (8) can be expanded about one of its minima (for instance, the minimum at $x=0$) as a quadratic, $W_1(2\pi x/Q_0)^2$. If we retained only this term, Eq. (8) would become the equation of a linear oscillator, with natural frequency ν_0, given by the relation

$$W_1(2\pi/Q_0)^2=2\pi^2m\nu_0^2. \tag{12}$$

By familiar methods, this linear oscillator would have wave functions

$$u_n=\exp(-z^2/2)H_n(z),$$
$$H_n(z)=(-1)^n\exp(z^2)d^n\exp(-z^2)/dz^n, \tag{13}$$
$$z=s^{\frac{1}{4}}w=s^{\frac{1}{4}}\pi x/Q_0,$$

corresponding to the energy

$$E_n=(n+\tfrac{1}{2})h\nu_0, \quad \epsilon_n=2n+1. \tag{14}$$

Thus we see that ϵ measures the energy, in units $h\nu_0/2$. We see furthermore that, since the charge density u_n^2 of the ground state of the oscillator falls to $1/e$ of its maximum value when z [of Eq. (13)] equals 1, the quantity s may be interpreted according to the equation

$$s^{\frac{1}{4}}=w_0^{-1}=Q_0/\pi x_0, \tag{15}$$

where w_0, x_0 represent the values of these quantities for which the charge density of the ground state of the oscillator will have fallen to $1/e$ of its maximum value. We may now visualize a situation in which "atoms" with a quadratic potential $2\pi^2m\nu_0^2x^2$ are located at the lattice points $x=0, \pm Q_0, \pm 2Q_0$, etc. The sinusoidal potential energy $m\nu_0^2Q_0^2(1-\cos 2\pi x/Q_0)$ reduces properly to these "atomic" potentials near each atomic position, but joins smoothly from one to the next, with a potential maximum between atoms. We picture ν_0, and x_0, which equals $(2\pi)^{-1}(h/m\nu_0)^{\frac{1}{2}}$, or the properties of an individual "atom," as staying constant while Q_0, or the lattice spacing, is changed. We see from Fig. 1 how it is that the height of the maximum (given by $4W_1=2m\nu_0^2Q_0^2$) increases as the lattice spacing increases at constant ν_0. And we see from (15) that, keeping x_0 fixed, the quantity Q_0 is equal to $\pi x_0 s^{\frac{1}{4}}$. Hence, if we wish to indicate the splitting of the energy bands as a function of Q_0, the lattice spacing, we can use $s^{\frac{1}{4}}$ as a dimensionless measure of the lattice spacing.

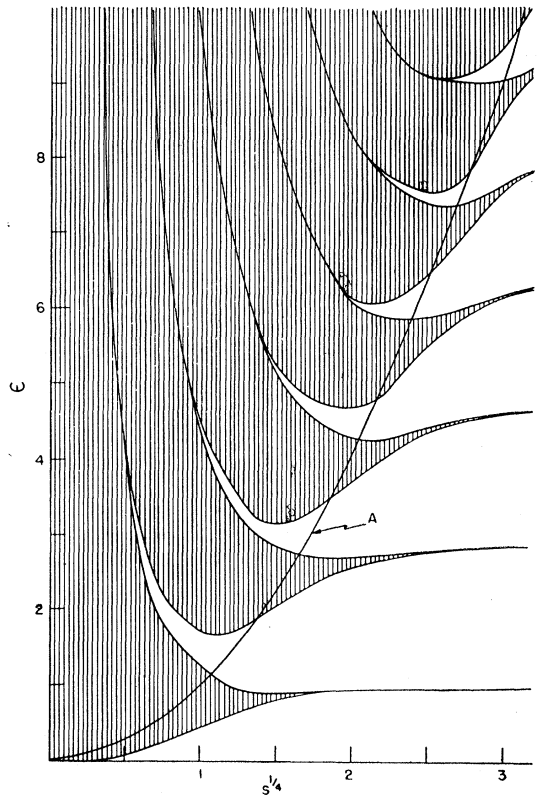

FIG. 2. Energy bands ϵ as function of lattice energy $s^{\frac{1}{4}}$, one-dimensional Mathieu problem. Curve A represents the height of the potential barrier between atoms.

We can now use the results of reference 5 to give Fig. 2, showing the widths of the energy bands and energy gaps of the one-dimensional problem as a function of lattice spacing. We plot ϵ as a function of $s^{\frac{1}{4}}$, in accordance with the discussion of the preceding paragraph. The curves are taken out to $s=100$, the limiting value given in reference 5. There are many interesting features of these energy bands. In the first place, for large enough values of Q_0, where the bands become indefinitely narrow and the wave functions become atomic and do not overlap appreciably, the energies approach the values $\epsilon_n=2n+1$ of the linear oscillator. As the interatomic distance decreases, there is a depression of the energy levels, which sets in before they start to broaden. It is not hard to find the origin of this

depression. We remember that the potential energy [expressed in the dimensionless form of Eq. (10)] is

$$\tfrac{1}{2}s(1-\cos 2w)=\tfrac{1}{2}s[(2w)^2/2-(2w)^4/24+\cdots]$$
$$=s(w^2-w^4/3+\cdots).$$

The terms $-sw^4/3$ and following terms represent a perturbation applied to the linear oscillator problem. It is just the mean value of these perturbative energy terms over the unperturbed linear oscillator wave functions which are responsible for the depression. Thus, for example, let us take the ground state, whose wave function is proportional to $\exp(-s^{\frac{1}{2}}w^2/2)$. The mean value of $-sw^4/3$ over this wave function is

$$-\tfrac{1}{3}s\int_{-\infty}^{\infty}w^4\exp(-s^{\frac{1}{2}}w^2)dw\Big/\int_{-\infty}^{\infty}\exp(-s^{\frac{1}{2}}w^2)dw=-\tfrac{1}{4}.$$

Thus we expect that the energy of the ground state, for large s, will be given by the equation, $\epsilon s^{\frac{1}{2}}=s^{\frac{1}{2}}-\frac{1}{4}$, or $\epsilon=1-1/(4s^{\frac{1}{2}})$. This gives correctly the first two terms of the asymptotic expression for the energy as a function of s for large s, given in Eq. (2.35), reference 5, which for this case gives $\epsilon=1-1/(4s^{\frac{1}{2}})-1/(16s)-\cdots$. We shall see later, when we take up the two- and three-dimensional cases, that terms of this sort in the potential represent the crystalline field, the departure of the true potential from the linear-oscillator approximation, and having a cubic symmetry, arising from the approach of neighboring atoms. And we shall see that the cubic field, in the three-dimensional case, can remove the degeneracy arising from the spherical symmetry, just as the crystalline field removes the degeneracy of certain atomic states in problems of magnetic ions in crystals.

As the atoms approach more and more closely, in Fig. 2, the energy levels begin to broaden into bands, in the familiar way, the energy gaps disappearing as the interatomic distance goes to zero. We should expect that the broadening of a level would begin to be large when the barrier is lowered enough so that this level is at the top of the barrier; those levels below the barrier will be relatively sharp; those above will be broadened into bands with narrow gaps between. From Eq. (10), the top of the barrier comes for $\epsilon s^{\frac{1}{2}}=s$, $\epsilon=s^{\frac{1}{2}}$. We draw in Fig. 2 the parabola $\epsilon=s^{\frac{1}{2}}=(s^{\frac{1}{4}})^2$, marked A, and we see that it does roughly divide the figure into two regions, that below the parabola corresponding to the levels below the barrier, which are only slightly broadened, and that above corresponding to the levels above the barrier, which are very broad, with very narrow gaps. It is interesting to compare the relation of this curve A to the broadening of the bands, and the corresponding curve A in Fig. 9 of reference 6, where similar curves are drawn for the energy bands in a sodium crystal, as computed by the cellular method. The general resemblance is very strong.

2. WAVE FUNCTIONS, MOMENTUM EIGENFUNCTIONS, WANNIER FUNCTIONS, AND ENERGIES FOR THE ONE-DIMENSIONAL MATHIEU PROBLEM

The material concerning Mathieu functions given in reference 5 consists of numerical tables, and approximate formulas, for the wave functions and energies connected with the edges of the bands, for values of s up to 100, and for the lowest 15 bands. No information is given about the intermediate values between the edges of the bands, so that we cannot plot a curve of energy as a function of momentum directly from the results of that reference. We shall next therefore describe the method used for computing these quantities and show how these have been used in the present work to find the behavior of the bands.

The method used to discuss the Mathieu equation is to compute the momentum eigenfunction, or Fourier representation of the wave function; the reason why Mathieu's equation is so important is the unique simplicity of the difference equation governing the momentum eigenfunction. In reference 6, Sec. 20 ff, the momentum eigenfunction $v(p_x, p_y, p_z)$ is introduced. For convenience we describe the principal properties of this function; we state our results for the general three-dimensional problem and then specialize for our one-dimensional case. We start by considering a lattice of points in the momentum space, consisting of all the vectors $\mathbf{p}$ which can be made from a given $\mathbf{p}_0$ by adding any one of the vectors $\mathbf{P}_i$ described in Eq. (1). We know that a wave function $u(\mathbf{r})$ can be written as a sum of plane waves $\exp[(2\pi i/h)(\mathbf{p}_0+\mathbf{P}_i)\cdot\mathbf{r}]$, associated with some particular $\mathbf{p}_0$; for such a sum is the product of the plane wave $\exp[(2\pi i/h)\mathbf{p}_0\cdot\mathbf{r}]$ and a periodic function with the periodicity of the lattice. We can associate the Fourier coefficient connected with the wave of momentum $\mathbf{p}_0+\mathbf{P}_i$ with the point $\mathbf{p}_0+\mathbf{P}_i$ in momentum space. Let the Fourier coefficient be called $v(\mathbf{p}_0+\mathbf{P}_i)$. That is, we can write

$$u_{n,\,p0}(\mathbf{r})=\sum(\mathbf{P}_i)v_n(\mathbf{p}_0+\mathbf{P}_i)$$
$$\times\exp[(2\pi i/h)(\mathbf{p}_0+\mathbf{P}_i)\cdot\mathbf{r}]. \quad (16)$$

The function $u_{n,\,p0}(\mathbf{r})$ of course is to be associated with the particular $\mathbf{p}_0$, or rather with the lattice of points $\mathbf{p}_0+\mathbf{P}_i$; a different wave function is associated with each point of a unit cell in the momentum space. And of course we have different wave functions associated with the various energy bands, indicated by the subscript n on u and v. It is immaterial for most purposes whether we choose $\mathbf{p}_0$ to be in the central unit cell of momentum space or not, though we shall generally do so.

We now notice that a single wave function out of an energy band, associated with a given $\mathbf{p}_0$, defines the function v only at a discrete lattice of points in $\mathbf{p}$ space. However, the set of u's associated with an energy band, and all $\mathbf{p}_0$'s, defines this function at all points of $\mathbf{p}$ space. In other words, for each energy band, there is a single function $v(\mathbf{p})$ which gives complete information about

all the wave functions of this band. Before writing down the equation determining $v(\mathbf{p})$, the equivalent of Schrödinger's equation in momentum space, we may state some of the properties of $v(\mathbf{p})$. First, it is normalized, in the following sense: if $u(\mathbf{r})$ is normalized so that the average value of its product with its conjugate is unity, $v(\mathbf{p})$ will be normalized so that the sum of the quantity v^*v for all points of a single lattice of points equals unity.[7] Furthermore, there is an orthogonality relation between v's associated with different bands; we may write the normalization and orthogonality relations in the form

$$\sum(\mathbf{P}_i)v_m^*(\mathbf{p}_0+\mathbf{P}_i)v_n(\mathbf{p}_0+\mathbf{P}_i)=\delta_{mn}, \tag{17}$$

where the subscript denotes the band. There is another orthogonality and normalization condition (see reference 6, Sec. 22),

$$\sum(n)v_n^*(\mathbf{p}_0+\mathbf{P}_i)v_n(\mathbf{p}_0+\mathbf{P}_j)=\delta_{ij}, \tag{18}$$

whose special case for $i=j$ gives the very important completeness relation, which we shall use later: the sum of the squares of the absolute values of all the momentum eigenfunctions associated with all bands, at a given point of momentum space, equals unity. There is in addition a relation inverse to (16), expressing v in terms of u:

$$v_n(\mathbf{p})=(1/\Omega)\int u_{n,p}(\mathbf{r})\exp[(-2\pi i/h)(\mathbf{p}\cdot\mathbf{r})]dxdydz, \tag{19}$$

where the integration is to be taken over a unit cell, Ω is the volume of the cell, and $\mathbf{p}$ equals one of the vectors $\mathbf{p}_0+\mathbf{P}_i$ associated with the lattice of points for which $u(\mathbf{r})$ is a wave function.

These relations are all discussed in reference 6. There is a further relation, of great importance, not given in that reference: $v(\mathbf{p})$ is the Fourier integral transform of the Wannier function,[8] a function which had not been proposed at the time reference 6 was written. Let us discuss this important theorem.

Wannier showed that there existed a function $a(\mathbf{r}-\mathbf{Q}_j)$, localized in the neighborhood of the $\mathbf{Q}_j$'th cell, such that we can write the wave function $u_{n,p}(\mathbf{r})$ associated with a given momentum vector $\mathbf{p}$ by the equation

$$u_{n,p}(\mathbf{r})=(\Omega)^{\frac{1}{2}}\sum(\mathbf{Q}_j)\exp[(2\pi i/h)(\mathbf{p}\cdot\mathbf{Q}_j)]a_n(\mathbf{r}-\mathbf{Q}_j). \tag{20}$$

That is, we can make a combination of a's in each cell, multiplied by an appropriate modulating factor, just as Bloch[9] proposed to do approximately using atomic orbitals. Here Ω is the volume of the unit cell, as before, and the factor $(\Omega)^{\frac{1}{2}}$ insures that if the a's are normalized to unity, the u's will be normalized on the average, as we have assumed earlier. Wannier, furthermore, showed howto set up the functions a. In our present normalization, his formula is

$$a_n(\mathbf{r}-\mathbf{Q}_j)=\frac{(\Omega)^{\frac{1}{2}}}{h^3}\int u_{n,p}(r)\times\exp[(-2\pi i/h)(\mathbf{p}\cdot\mathbf{Q}_j)]dp_xdp_ydp_z, \tag{21}$$

where the integration is to be taken over a single unit cell in momentum space. He proved that the a's so set up are orthogonal, in the sense that

$$\int a_n^*(\mathbf{r}-\mathbf{Q}_j)a_n(\mathbf{r}-\mathbf{Q}_k)dv=\delta_{jk}.$$

Now let us express $u_{n,p}(\mathbf{r})$ in terms of $v_n(\mathbf{p})$, by Eq. (16). Then (21) becomes

$$a_n(\mathbf{r}-\mathbf{Q}_j)=\frac{(\Omega)^{\frac{1}{2}}}{h^3}\sum(\mathbf{P}_i)\int v_n(\mathbf{p}+\mathbf{P}_i)\times\exp[(2\pi i/h)(\mathbf{p}+\mathbf{P}_i)\cdot(\mathbf{r}-\mathbf{Q}_j)]dp_xdp_ydp_z, \tag{22}$$

where we have used Eq. (1), and where still the integration is over a unit cell in momentum space. The combination, however, of integrating the $\mathbf{p}$'s over unit cell, and summing over $\mathbf{P}_i$, amounts to integrating over all momentum space. Thus

$$a_n(\mathbf{r}-\mathbf{Q}_j)=\frac{(\Omega)^{\frac{1}{2}}}{h^3}\int v_n(\mathbf{p})\times\exp[(2\pi i/h)\mathbf{p}\cdot(\mathbf{r}-\mathbf{Q}_j)]dp_xdp_ydp_z, \tag{23}$$

where now the integration is over all $\mathbf{p}$'s. This expresses a as the Fourier integral transform of v. Similarly we can express the inverse transformation. We use Eq. (19) to express $v_n(\mathbf{p})$ in terms of $u_{n,p}(\mathbf{r})$, and (20) to express $u_{n,p}(\mathbf{r})$ in terms of the a's. Here in a similar way we have an integration over a unit cell in coordinate space, coupled with a summation over all cells, amounting to an integration over all space. In this way we arrive at the result

$$v_n(\mathbf{p})=(\Omega)^{-\frac{1}{2}}\int a_n(\mathbf{r})\exp[(-2\pi i/h)(\mathbf{p}\cdot\mathbf{r})]dxdydz. \tag{24}$$

We note that in (23) where the integral is extended over all $\mathbf{p}$'s the integrand falls off for large $\mathbf{p}$'s, on account of the behavior of $v(\mathbf{p})$, just as the integrand in (24) falls off for large x's on account of the nature of a. From the results of this paragraph, we see that a very practical way to find a Wannier function is to find the momentum eigenfunction $v_n(\mathbf{p})$ and then its Fourier transform.

We have now examined the properties of the momentum eigenfunction $v(\mathbf{p})$, and we are ready to write down

[7] This normalization determines $v(\mathbf{p})$ at a lattice of points uniquely, except for an arbitrary phase factor. This factor must be properly chosen, for each lattice of points, to secure continuity of the function $v(\mathbf{p})$. Ordinarily the phase can be chosen by a requirement that $v(\mathbf{p})$ be real, or pure imaginary, leaving only a sign to be determined.

[8] G. Wannier, Phys. Rev. **52**, 191 (1937).

[9] F. Bloch, Z. Physik **52**, 555 (1928).

the equation it satisfies. From reference 5, Secs. 20 and 21, we see that the equivalent of Schrödinger's equation for w is the set of difference equations

$$[(p_x^2+p_y^2+p_z^2)/2m]v(\mathbf{p}) +\sum(\mathbf{P}_i)W(\mathbf{P}_i)v(\mathbf{p}+\mathbf{P}_i)=Ev(\mathbf{p}). \quad (25)$$

These equations, of course, may be considered as derived by the usual perturbation methods of quantum mechanics, in the process of expanding the correct wave function $u_p(\mathbf{r})$ of the problem, in a series of plane waves, which form a complete orthogonal set. The quantity $(p_x^2+p_y^2+p_z^2)/2m+W(0)$ is the diagonal matrix component of energy, associated with the wave of momentum $\mathbf{p}$, and $W(\mathbf{P}_i)$ is the nondiagonal matrix component of energy between the waves of momentum $\mathbf{p}$ and of $\mathbf{p}+\mathbf{P}_i$. Equation (25) will lead to a secular equation of the ordinary sort for the determination of the energy E.

We may now specialize for the one-dimensional Mathieu problem in which the W's have been given already. We have

$$(p^2/2m+2W_1-E)v(\mathbf{p}) -W_1[v(\mathbf{p}+\mathbf{P}_0)+v(\mathbf{p}-\mathbf{P}_0)]=0. \quad (26)$$

Here again let us introduce dimensionless quantities. It is convenient to introduce

$$g=2p/P_0=2pQ_0/h, \quad (27)$$

where as before $P_0=Q_0/h$. Then, using the abbreviation (9), Eq. (26) becomes

$$(g^2+\tfrac{1}{2}s-\epsilon s^{\frac{1}{2}})v(g)-\tfrac{1}{4}s[v(g+2)+v(g-2)]=0. \quad (28)$$

This forms a very simple difference equation for the v's. For preliminary orientation, we may observe that if $v(g)$ varies only slowly with g, this difference equation can be replaced by a differential equation. For then we should have approximately $[v(g+2)-v(g)]/2=dv/dg$ computed for $g+1$; $[v(g)-v(g-2)]/2=dv/dg$ for $g-1$; and $\{[v(g+2)-v(g)]-[v(g)-v(g-2)]\}/4\sim d^2v/dg^2$. Hence Eq. (28) is approximately equivalent to

$$-sd^2v(g)/dg^2+(g^2-\epsilon s^{\frac{1}{2}})v(g)=0. \quad (29)$$

This would be a Schrödinger-type equation for a particle in a parabolic potential, proportional to g^2. Thus its solutions would be proportional to

$$v_n=\exp(-z^2/2)H_n(z), \quad z=s^{-\frac{1}{4}}g, \quad (30)$$

corresponding to the same energy $\epsilon_n=2n+1$ as in (14). We shall find later that for the low energy levels, which are hardly broadened at all, it actually is true that $v(g)$ varies only slowly with g, so that (30) gives the correct solution for the momentum eigenfunction. We could have predicted in advance that this would be the case. For in this case the problem is essentially the linear oscillator problem, its coordinate wave functions are Hermite functions for the various atoms, the atoms overlap so little that the Wannier functions are equivalent to the atomic functions, and we know that the Fourier transforms of the Hermite functions of x are Hermite functions of p, or of g, just as we have found in Eq. (30).[10]

The situation is completely different, however, for the high energy levels, the free-electron-like solutions of (28). The extreme case would be one in which only one of the quantities $v(g')$ was different from zero. This would be found only when the energy was so great that s could be neglected compared to ϵ. Then we should have $(g^2-\epsilon s^{\frac{1}{2}})v(g)=0$, so that we should have only one equation, for the nonvanishing component $v(g')$, which would then determine the energy by the equation $\epsilon s^{\frac{1}{2}}=g'^2$, the free-electron value. The important intermediate cases of electrons which are partly free correspond to cases where the v's vary very rapidly from one value of g to the next value, but where, nevertheless, many components are different from zero. For these cases we must use numerical methods to solve Eq. (28).

One simple numerical method will immediately occur to the reader. From (28) we can immediately solve for $v(g+2)$ in terms of $v(g)$ and $v(g-2)$, or alternatively for $v(g-2)$ in terms of $v(g)$ and $v(g+2)$. Thus if we know E, and two successive entries in the table of v's, we can compute the next entry, and so on indefinitely. This is just like the standard method of numerical integration of a differential equation, which we replace approximately by a difference equation; only here the process is exact. We know that if we start with two arbitrary successive entries in the table, and an arbitrary E, and integrate both ways to $\pm\infty$, the function $v(g)$ will in general become exponentially infinite. We can always choose the ratio of the two initial successive entries so as to make it go to zero exponentially in one direction, but then it will still go infinite in the other direction. If the energy is properly chosen, it will go to zero both at $+\infty$ and $-\infty$. This determines the energy values. Instead of integrating from a finite point out to infinity, we can reverse the procedure, if we can estimate the limiting behavior of the function for large positive and negative g's, corresponding to the exponentially decreasing functions. We can then integrate in from infinity in both directions, toward $g=0$, for a given energy. In general we shall find that the integrations from the two limits do not match when they join; that is, we can determine the ratio of two successive entries (say $v(g_0)/v(g_0+2)$) from the integrations from both $+\infty$ and $-\infty$ and can then try various energy parameters until these ratios agree. This determines both the momentum eigenfunctions (which of course must then be normalized) and the energy, and it is an entirely practical method of calculation.

A slight modification of this method, however, the method of continued fractions, is essentially equivalent

[10] Oddly enough, these simple relations between Mathieu functions and the Hermite functions, and the theta-functions and their derivatives which arise when we make Bloch-type linear combinations of them, do not seem to be widely recognized, though my colleague Professor P. M. Morse informs me that he is familiar with them and will discuss them in a forthcoming book.

but more convenient in practice, and it is the method ordinarily used. This method operates with the ratios of successive v's, rather than with their differences. Thus let

$$G_g = v(g)/v(g-2). \tag{31}$$

Then [using the abbreviation a given in Eq. (9)], Eq. (28) becomes

$$(g^2-a)-\tfrac{1}{4}s(G_{g+2}+1/G_g)=0. \tag{32}$$

If we let

$$V_g=(4/s)(g^2-a), \tag{33}$$

this is

$$V_g-G_{g+2}-1/G_g=0, \tag{34}$$

which we can solve in the form

$$G_g=(V_g-G_{g+2})^{-1}. \tag{35}$$

This is a very convenient equation to use for integrating in from $+\infty$. For a large positive g, the quantity V_g becomes very large. On the other hand, since we are looking for a solution falling off exponentially with increasing g, the ratio G_g or G_{g+2} will be very small. Thus G_{g+2} will be negligible compared to V_g for large enough g, and we can commence an inward integration by letting $G_g=V_g^{-1}$, for some particular large g. Then we have $G_{g-2}=(V_{g-2}-G_g)^{-1}$, etc., and we can continue integrating inward very easily. Here as before we assume an energy, integrate inward until we get to an entry near $g=0$. We set up an equivalent expression for integrating in from $-\infty$, continue to the same point, and see if the two solutions join smoothly. If they do not, we modify the energy until they do.

Once we know the energy, the actual numerical work involved in the calculation, using a desk computer, is very small. It is in this way that the calculations in reference 5 were performed; but these were done only for the edges of energy bands, corresponding to integral values of the g's. We have made similar calculations, to be described in the next sections, for a number of non-integral g's, enough to determine the general form of the continuous function $v(g)$ for the various energy bands, for several values of s. In the actual calculations, it is a great help to know in advance the approximate energy; if this is known, we can merely calculate for several energy values in that neighborhood, and interpolate until we get the correct energy, for which the ratio G calculated by integration for both directions agrees. In estimating the energy, the results of reference 5 are of great value. For the tightly bound bands, for which the band width is small, and the energies at the edges of the bands are given in reference 5, an interpolation between the values given there gives a very accurate preliminary value for the energy. For the higher energies, for which there is not much deviation from the free-electron energy, there is a series representation of the energy as a function of g which is quite accurate, except very close to the edges of the band (where the values are already given in reference 5). This series is given in Eq. (2.34) of reference 5, though for a slightly different purpose; it comes originally from Mathieu. It is

$$a=\epsilon s^{\frac{1}{2}}-s/2=g^2+\frac{s^2}{32(g^2-1)}+\frac{(5g^2+7)s^4}{8192(g^2-1)^3(g^2-4)} +\frac{(9g^4+58g^2+29)s^6}{262144(g^2-1)^5(g^2-4)(g^2-9)}+\cdots. \tag{36}$$

Clearly, as g approaches an integer (the edge of a band), the later terms of the series become infinite; nevertheless, the earlier terms give a very good approximation to the energy, even rather close to the edge of the band.

It is perhaps worth while indicating how Eq. (36) is derived, since it throws considerable light on the nature of the Mathieu problem. Let us start with a free-electron approximation, in which the solution is approximately $\exp(ig_0w)$, for some particular g_0, for which we should have $a=g_0^2$. We then set up the series

$$u_{g_0}(w)=\sum(n)v(g_0+2n)\exp[i(g_0+2n)w], \tag{37}$$

equivalent to (16) but expressed in dimensionless form. We now postulate that we can set up a solution for small s's in which each successive v, going away from $v(g_0)$ in either direction, is smaller by a power of s. That is, if we set $v(g_0)=1$ (for an unnormalized function), we assume that $v(g_0\pm2)$ are of order s, $v(g_0\pm4)$ of order s^2, etc. Further we assume that a can be expanded in a series in s, of the nature of (36). We then inset these assumptions into Eq. (28), using that equation to work outward in both directions from g_0. We find that at each step we have enough equations to determine coefficients up to an additional power of s, so that the process actually can be carried through step by step. When we do this, we not only arrive at Eq. (36), for the energy, but also at the following frequently useful

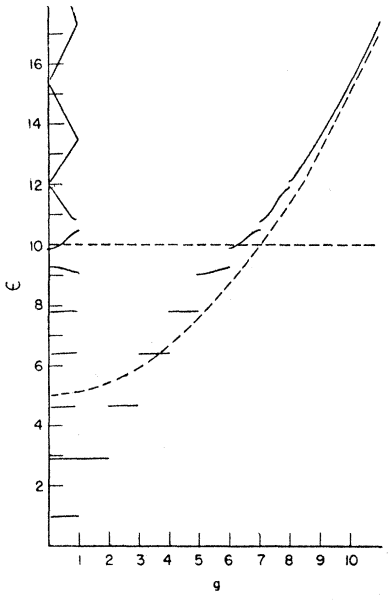

FIG. 3. Energy ϵ as function of momentum g, in dimensionless units, for the case $s=100$. Parabola shows energy of free electron in mean potential.

equations for the (unnormalized) momentum eigenfunction:

$$v(g_0)=1,$$

$$v(g_0\pm 2)=\frac{\pm s}{16(g_0\pm 1)}\pm\frac{(g_0^2\pm 4g_0+7)s^3}{8192(g_0\pm 1)^3(g_0\pm 2)(g_0\mp 1)}\pm\frac{(g_0^6\pm 5g_0^5+8g_0^4\pm 8g_0^3+47g_0^2\mp 13g_0+232)s^5}{3{,}145{,}728(g_0\pm 1)^5(g_0\pm 2)(g_0\pm 3)(g_0\mp 1)^3(g_0\mp 2)}\pm\cdots,$$

$$v(g_0\pm 4)=\frac{s^2}{512(g_0\pm 1)(g_0\pm 2)}+\frac{(g_0^3\pm 7g_0^2+20g_0\pm 20)s^4}{196{,}608(g_0\pm 1)^3(g_0\pm 2)^2(g_0\pm 3)(g_0\mp 1)}+\cdots,$$

$$v(g_0\pm 6)=\pm\frac{s^3}{24{,}576(g_0\pm 1)(g_0\pm 2)(g_0\pm 3)}\pm\cdots,$$

$$v(g_0\pm 8)=\frac{s^4}{2^{16}(4!)(g_0\pm 1)(g_0\pm 2)(g_0\pm 3)(g_0\pm 4)}+\cdots,$$

. . . . . .

$$v(g_0\pm 2n)=(\pm 1)^n\frac{s^n}{2^{4n}(n!)(g_0\pm 1)(g_0\pm 2)\cdots(g_0\pm n)}\pm\cdots. \tag{38}$$

We now see from (38) why our series representation fails near the edges of the energy bands. As g_0 approaches an integer, say a positive integer, the v's for negative g's will become large. As an example, let g_0 approach 3; then $v(g_0-2)$ will have a factor g_0-3 in the denominator of its third term, $v(g_0-4)$ in the second term, and $v(g_0-6)$ and all successive v's in its first term. The meaning of this is of course clear from our knowledge of Bragg reflection. We know that as g_0 approaches 3, for instance, there will come in another component, $v(-3)$, equal just at the edge of the band to $v(3)$. In other words, near the edge of the band, we are not justified in assuming that the v's fall off as successively higher powers of s; $v(-g_0)$ will be equal to $v(g_0)$, and a series which attempts to expand $v(-g_0)$ as a series starting with a term in s^{g_0} cannot be expected to converge.

Just at the edge of the band, however, a different procedure is possible. There we know that the component $v(-g)$ must be equal or opposite to $v(g)$, depending on whether the function is even or odd. We can then set up the assumption that the v's for positive g's fall off as we assumed above, but combine this with the condition that $v(g)$ be an even or odd function of g, so that we do not continue the integration to negative g's. In this case we find quite different series, similar to (36) and (37) but of course differing for each g_0 and not involving g_0 explicitly. These formulas, for the energy, are given in Eqs. (2.24)$\cdots$(2.33) of reference 5; similar formulas can be set up for the momentum eigenfunctions as well.

By use of a combination of the approximate formulas discussed above, and the numerical solution using (35), it is not hard to arrive at a good understanding of the energy and wave functions arising from Mathieu's equation for any value of s and as a function of g. We shall give in the next section some typical examples of the results, paying particular attention to the change in properties of the solutions as we go from the low, bound states to the higher states resembling free electrons.

3. NUMERICAL EXAMPLES FOR THE ONE-DIMENSIONAL CASE

Calculations have been made for $s=1$, 16, and 100, or $s^{\frac{1}{2}}$ (which we remember is proportional to the interatomic distance) equal to 1, 2, and 3.16. This gives a good range of properties. We remember that the top of the barrier comes from $\epsilon=s^{\frac{1}{2}}$ and that the energy levels of the corresponding linear oscillator problem are given by $\epsilon=2n+1$. Thus for $s=1$, the lowest level of the oscillator will lie just at the top of the barrier, so that even that lowest level is greatly broadened; for $s=16$, the energy levels for $n=0$ and $n=1$, corresponding to $\epsilon=1$ and 3, will lie below the barrier, which has a height of 4 units; for $s=100$, the lowest five levels, for $\epsilon=1$, 3, 5, 7, 9, will all lie below the barrier at $\epsilon=10$ (actually, since the energies are depressed as compared to those of the linear oscillator case, six energy levels come below the barrier in this case). In some ways the case $s=100$ is the most informing of the three, since in this the change from the properties of the bound states to the free states is more gradual, and we have very typical examples of the limiting sorts of states. Consequently we shall describe this case of $s=100$ first.

In Fig. 3, we show the energy as a function of g, for the various energy bands for $s=100$. We plot this in two ways. First, we plot it according to a free electron correspondence; that is, each energy band is plotted in the unit cell where it would be located if the barrier height were to be reduced to zero. Secondly, we plot the whole thing in the central cell of momentum space. We show the top of the barrier, at $\epsilon=10$; and we plot a parabola, whose equation is $\epsilon s^{\frac{1}{2}}=g^2+s/2$, giving the

energy of a free electron in the mean potential. There are many interesting features to be derived from this figure. Perhaps the first and most striking one is the very rapid way in which the bands become broadened as we go from the levels below the barrier, where they are quite narrow, to those above the barrier, where very soon the energy gaps almost disappear.

Next, we may consider how closely the energy approximates the atomic values for the lower bands, the free-electron values for the higher ones. We recall that the atomic levels would lie at $\epsilon=1, 3, 5, 7, 9$; we actually find them at approximately 0.97, 2.87, 4.65, 6.30, 7.79, 9.09. These depressions compared to the free atom case, as was mentioned in Sec. 1, arise because the sinusoidal potential energy is really lower than the parabolic potential describing the free atom, and the depression in energy is approximately the average of this reduction in potential energy, averaged over the unperturbed wave function. The comparison with the free-electron energy $\epsilon s^{\frac{1}{3}}=g^2+s/2$ is informing. We observe that the tightly bound levels lie below the parabola; the wave functions are concentrated near the potential minima, so that there is a decrease of potential energy as compared to the average through the crystal. For the higher bound states, on the contrary, and for the free electron states, the energy is higher than the parabola; and when we examine the wave functions, we find that at least for the free electron states the wave functions tend to be more concentrated around the potential maxima than the minima, giving a potential energy greater than the average through the crystal. At first sight this seems peculiar; but when we consider it more closely we see that it is not. For an electron whose energy is considerably above the top of the barrier, we can use the WKB approximation with good accuracy. We then of course find that the amplitude of the wave function is greater in regions of low velocity, and this is precisely at the potential maxima. A simple WKB approximation of this sort gives a good value for the excess of $\epsilon s^{\frac{1}{3}}$ over the parabola $g^2+s/2$ [given analytically in Eq. (36)], for the states of high energy. This phenomenon, by which the wave functions in the lower free electron states tend to concentrate near the potential maxima, of course has its analog in real crystals, where for instance in the conduction bands of ordinary metals the wave functions tend to be large in the region between atomic centers, being less concentrated near the nuclei. From Fig. 3, and its interpretation in terms of the WKB method, we may anticipate that this effect will be less marked for really fast electrons, in energy bands far above the barrier.

Now that we have considered the energy, we may examine the behavior of the momentum eigenfunction $v(g)$, for the case $s=100$. The values for integral values of g are given in reference 5, though the normalization must be changed to agree with our present conventions; they have been computed for half-integral values for a number of bands. These values are given in Table I.[11] The functions are plotted in Figs. 4 and 5. While the details are not completely obvious from the integral and half-integral values of Table I, they were made clear by comparison with the cases of $s=1$ and $s=16$, for which more points were computed, and by calculations for quarter-integral values for some of the more complicated cases. It is clear from examination of Figs. 4 and 5 that the curves have an extraordinary complexity, which needs a careful interpretation.

In the first place, the momentum eigenfunctions of the five or six lowest bands look very much like the first five or six Hermite functions, as we should expect from Eq. (30). These are the atomic levels; the Wannier functions are rather close to the atomic functions of the linear oscillator, and consequently the momentum eigenfunctions are similar to Hermite functions. But as soon as we go above these levels, the simplicity vanishes, and no real simplicity of any other sort appears, even as high as the fifteenth band, not shown in Fig. 5, but similar to the eleventh band. Even that high, there is no resemblance to a real free electron case, in which one component only of the wave function would be large. There is one interesting feature which comes in with these higher bands, however: the momentum eigenfunction is large for an interval of unity along g (say from g equal to an even integer out to the next odd integer) but is practically zero in the next interval of unity; the same thing repeats itself for negative g's, all momentum eigenfunctions being either even or odd functions of g, as we readily find out. This alternation has the effect that the solution is almost precisely a traveling rather than a standing wave. Thus in Fig. 6 we show such a momentum eigenfunction, and the lattice of points corresponding to two values of g, one (labeled 1) corresponding to a number between 0 and 1 plus an even integer, one (labeled 2) to a number between 0 and -1 plus an even integer; that is, remembering that the central Brillouin zone for this problem consists of the range of g from -1 to 1, the first lattice corresponds to a point in the right half of this central zone, the second lattice to a point in the left half. We now observe that for the first lattice, the only Fourier contributions of any significance come from negative g's, so that this represents a progressive wave in one direction; for the other lattice the only Fourier components come for positive g's. As the lattice of points approaches the edge of a band, however, (for instance as the lattice approaches the numbers $\cdots -5, -3, -1, 1, 3, 5, \cdots$) each point will lie in an almost vertical transition section of the curve, and there will, on account of the symmetry of the $v(g)$ curve, be equal contributions from positive and negative g's. This is the standing wave arising from Bragg reflection, and

[11] We shall later find it more convenient to regard the momentum eigenfunctions for the even-numbered states to be pure imaginary, for this leads to real Wannier functions. These imaginary functions are i times the functions given in Table I and in Figs. 4 and 5.

TABLE I. $v(g)$ for $s=100$. Note that 1st, 3d, and odd-numbered bands have functions $v(g)$ which are even functions of g, while 2d, 4th, and even-numbered bands have odd $v(g)s$.

g	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
0.0	0.60775	0.00000	−0.46792	0.00000	0.47293	0.00000	−0.42290
0.5	0.59953	0.14453	−0.43431	−0.20659	0.40506	0.31047	−0.33629
1.0	0.57557	0.27666	−0.33910	−0.36670	0.21296	0.46428	−0.09001
1.5	0.53785	0.38586	−0.19797	−0.44809	−0.03886	0.47987	0.20969
2.0	0.48932	0.46488	−0.03296	−0.43738	−0.26270	0.28905	0.41424
2.5	0.43356	0.51053	0.13245	−0.34293	−0.40968	−0.02519	0.47618
3.0	0.37428	0.52360	0.27778	−0.19121	−0.44350	−0.26093	0.28443
3.5	0.31496	0.50823	0.38859	−0.01436	−0.36452	−0.43031	−0.07452
4.0	0.25848	0.47074	0.45800	0.15722	−0.22225	−0.43706	−0.32233
4.5	0.20700	0.41838	0.48608	0.29885	−0.04978	−0.27952	−0.52050
5.0	0.16186	0.35826	0.47823	0.39702	0.12521	−0.14021	−0.43338
5.5	0.12365	0.29649	0.44312	0.44851	0.26749	0.01142	−0.09342
6.0	0.09233	0.23774	0.39061	0.45633	0.36765	0.16201	0.01092
6.5	0.06744	0.18507	0.32974	0.42989	0.42507	0.25191	0.16892
7.0	0.04821	0.14011	0.26780	0.38235	0.42816	0.34533	0.23568
7.5	0.03375	0.10331	0.21000	0.32427	0.42507	0.42545	0.15291
8.0	0.02315	0.07430	0.15950	0.26285	0.34276	0.39947	0.31666
8.5	0.01558	0.05217	0.11763	0.20455	0.29326	0.29060	0.45990
9.0	0.01027	0.03581	0.08437	0.15411	0.23337	0.26383	0.37679
9.5	0.00665	0.02405	0.05895	0.11281	0.17356	0.25164	0.12409
10.0	0.00423	0.01581	0.04020	0.08003	0.12860	0.19270	0.17938
10.5	0.00264	0.01019	0.02679	0.05510	0.09483	0.11376	0.21771
11.0	0.00162	0.00644	0.01745	0.03706	0.06600	0.08682	0.15614
11.5	0.00098	0.00399	0.01114	0.02441	0.04331	0.07236	0.04076
12.0	0.00058	0.00243	0.00696	0.01569	0.02860	0.04912	0.04945
12.5	9.00034	0.00146	0.00427	0.00983	0.01889	0.02546	0.05404
13.0	0.00020	0.00086	0.00257	0.00605	0.01198	0.01729	0.03540
13.5	0.00011	0.00050	0.00152	0.00356	0.00715	0.01308	0.00800
14.0			0.00088	0.00217	0.00432	0.00812	0.00859
14.5			0.00050	0.00126	0.00264	0.00382	0.00868
15.0					0.00152	0.00235	0.00529

the range of g values over which the curve is changing abruptly from one value to another is precisely the range where Bragg reflection is important, a narrower and narrower range as the energy gaps get narrower. There is no physical significance in the fact that the lattice of points in the right half of the central Brillouin zone, in Fig. 6, correspond to negative g's, and vice versa; this is because we are dealing with an even-numbered band, with odd momentum eigenfunction. In an odd-numbered band the situation would be reversed, points in the right half of the central Brillouin zone corresponding to positive g's.

We see, then, that this very sharp alternation of $v(g)$ between large and practically vanishing values in successive integral ranges of g is a necessary feature of the free electron wave functions; in the low-lying bands, where it is not found, the wave functions correspond almost exactly to standing waves, which is another way of saying that they carry practically no current. The momentum eigenfunction corresponding to a completely free electron would have just one integral range of g, for positive g's, for which it would be different from zero; there it would be equal to unity, and in the corresponding range of negative g's it would equal ± 1 depending on whether we were dealing with an even or odd state. We may now ask why it is that even the 15th band of our problem $s=100$ is so far from a free electron state; why are there a number of successive intervals of g where $v(g)$ is large? The answer is just what we have already mentioned in a preceding paragraph: the WKB method indicates that there will be a very considerable variation of amplitude, and also of effective wavelength,

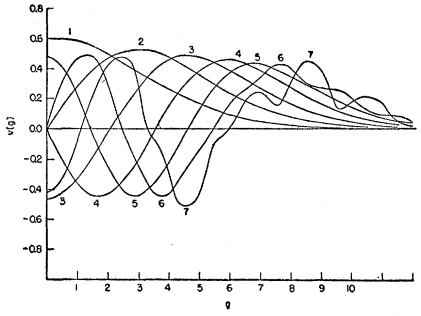

FIG. 4. Momentum eigenfunctions $v(g)$, for $s=100$, first seven bands.

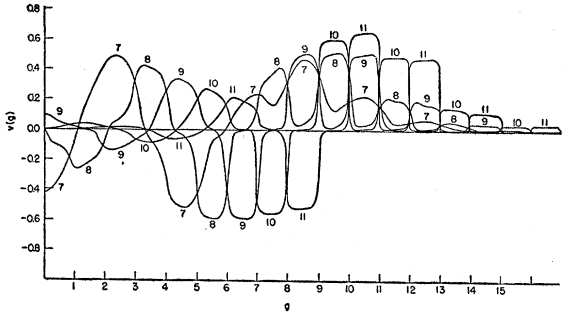

FIG. 5. Momentum eigenfunctions $v(g)$, for $s=100$, seventh to eleventh bands.

TABLE I—*Continued.*

g	Band 8	Band 9	Band 10	Band 11	Band 12	Band 13	Band 14
0.0	0.00000	0.09824	0.00000	−0.00710	0.00000		
0.5	−0.11098	0.04889	0.00749				
1.0	−0.26900	0.01637	0.03276	−0.00139	−0.00215		
1.5	−0.21239	−0.01961	0.02561				
2.0	−0.09420	−0.13661	0.01273	0.01468	−0.00094		
2.5	0.06384	−0.13019	−0.00242				
3.0	0.33931	−0.07138	−0.07742	0.00821	0.00839		
3.5	0.40063	0.00961	−0.08642				
4.0	0.24513	0.25980	−0.05053	−0.05114	0.00533		
4.5	−0.03259	0.31879	0.00100				
5.0	−0.38974	0.20069	0.20283	−0.03627	−0.03632		
5.5	−0.59244	−0.00523	0.25646				
6.0	−0.42605	−0.41961	0.16356	0.16377	−0.02681	−0.02681	
6.5	−0.00881	−0.59155	−0.00053				
7.0	0.16789	−0.41046	−0.40999	0.13491	0.13492	−0.02036	
7.5	0.36297	0.00014	−0.56603				
8.0	0.31824	0.30288	−0.38929	−0.38928	0.11286	0.11286	
8.5	0.03971	0.49995	0.00010				
9.0	0.33242	0.39026	0.38882	−0.36653	−0.36653	0.09566	
9.5	0.50210	0.00512	0.59488				
10.0	0.36167	0.35268	0.44725	0.44709	−0.34431	−0.34431	
10.5	0.02434	0.49954	0.00045				
11.0	0.14413	0.34758	0.34658	0.49028	0.49027	−0.32352	−0.32351
11.5	0.19492	0.00260	0.48034				
12.0	0.12938	0.12705	0.33159	0.33151	0.52354	0.52354	−0.30439
12.5	0.00680	0.16741	0.00020				
13.0	0.03346	0.10942	0.10919	0.31470	0.31471	0.54974	0.54974
13.5	0.04193	0.00064	0.14289				
14.0	0.02618	0.02581	0.09355	0.09354	0.29808	0.29808	0.57075
14.5	0.00117	0.03215	0.00004				
15.0	0.00500	0.01950	0.01955	0.08068	0.08068	0.28231	0.28231

of the function in going from a potential maximum to a potential minimum. Even in the 15th band, the energy ϵ is only about 28 units, while the top of the barrier is at 10 units, so that the kinetic energy varies from 28 units at the potential minimum to 18 units at the maximum. When we set up a WKB solution under these conditions, take account of the periodic change of amplitude and phase associated with the sinusoidal potential and expand the resulting wave function in Fourier series, we get in fact a good approximation to the exact momentum eigenfunction. There is then nothing mysterious about its departure from the free electron case; but we are put on our guard against thinking that we shall find a really free electron like case until the energy goes far above the top of the barrier.

There is a very informing way to plot the momentum eigenfunction, and that is to make a graphical test of the completeness relation given in Eq. (18). This relation tells us that if we square the v's associated with all the various energy bands, and sum them, the sum will equal unity for each value of p or of g. We demonstrate this in Fig. 7, where we plot $\sum(n)v_n^*(g)v_n(g)$ against g, for the various bands. We give first a curve for the first band; then for the first and second; then the first, second, and third; and so on. It is clear that the relation is verified (this gives a good check of the numerical calculations, checking all bands with a single operation). But there is much more information to be derived from this graph, which really contains within itself complete information about all the wave functions of all the energy bands of the problem (except that the sign of the wave function of course is not given, since Fig. 7 involves only the square of the wave function). Thus we see the great contrast in properties between the low levels, with their smooth momentum eigenfunctions, and the high levels, with their curious alternation of properties between successive integral ranges of g, but still with a certain orderliness which becomes more apparent in the figure than in the functions themselves. It is particularly interesting to see how, in the eighth and higher bands, the Fourier component begins to emerge which becomes the leading one in the free

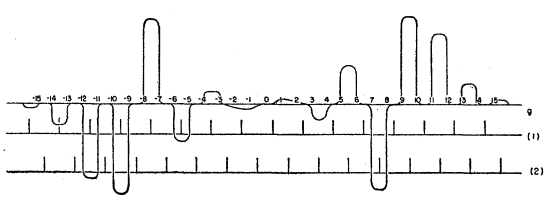

FIG. 6. Momentum eigenfunction for 10th band, $s=100$, with lattices of points (1) corresponding to $g_0=\frac{1}{2}$ and (2) corresponding to $g_0=-\frac{1}{2}$.

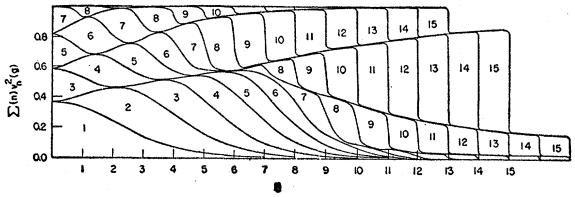

FIG. 7. Test of completeness relation, for $s=100$. Numbers indicate the bands fitting into each segment of the chart.

electron limit; that is, the component of the nth band in the range from $g=n-1$ to $g=n$, or the nth zone in momentum space. We note, as we have already pointed out, that even for the 15th band this component is still far from being the only one, though it is the largest.

It is also interesting in Fig. 7 to see the very complicated behavior in the intermediate range between the low and the high energies; and we cannot fail to realize that these intermediate states, the seventh and eighth bands in this case, represent precisely those bands which will be important in a real crystal, since they will represent the valence and conduction bands. Their properties are determined in a way by their being "squeezed in" between the very different patterns of the bound electrons and free electrons in Fig. 7. It is perhaps this observation—the remarkable complication of the wave functions for these bands near the top of the barrier—which is the most significant result of the present study. It is extraordinary that the energy levels, given in Fig. 3, can be so relatively simple, in spite of the complications of the wave functions. Certainly the fact that the energy of the conduction electrons in a metal varies in a simple fashion with the momentum, reminding us of a free electron case, is no reason for supposing that there is any corresponding simplicity in the wave functions.

A further use of Fig. 7 can perhaps suggest a rather powerful tool in considering the energy bands of a real crystal. Suppose we know the momentum eigenfunctions of some of the lower states—say the lowest six states in our present case. These represent bound states, and we could get a rather good approximation to their momentum eigenfunctions by actually using atomic functions, so that this represents a type of information which we could get in a corresponding case of a real crystal. Now we consider the next energy band, the seventh in our case. We have one specific piece of information: the component connected with any g value cannot be greater than the amount required to bring the sum of the $v^2(g)$'s for the g value up to unity. Thus for instance, in Fig. 7, we see that the component for $g=\frac{1}{2}$ could not be greater than approximately $(0.15)^{\frac{1}{2}}$. Such information could be very useful in actual cases. Thus it is well known that the wave functions in the conduction band for sodium correspond rather closely to a plane wave whose g value lies in the lowest Brillouin zone; the amplitude of this corresponds to more than $(0.90)^{\frac{1}{2}}$, and it is a result of this fact that there has been

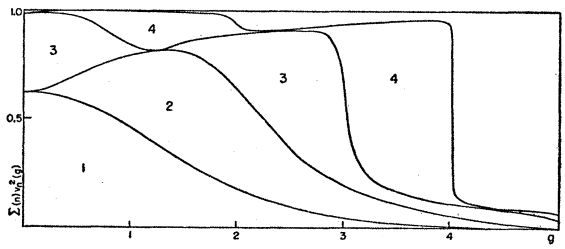

FIG. 8. Test of completeness relation, for $s=16$.

such a general assumption that the conduction electrons in a metal have wave functions like plane waves. This means, of course, that the momentum eigenfunctions of the bound electrons for sodium have a sum of $v^2(g)$ in the central Brillouin zone which is less than (0.10). We might now want to know whether a similar situation could hold in quite a different metal, say one in the iron group. We could throw light on this by finding the momentum eigenfunctions of the bound electrons, using an atomic approximation; squaring and summing, and seeing to what extent these filled the central zones of momentum space. The writer suspects that the case of the alkali metals will be found to be an exception.

It is worth while noticing, while we are considering the relation of Fig. 7 to a real crystal problem, that there is a very important difference between our Mathieu case and the case of a real crystal, in the matter of "filling up" the momentum space with the $v^2(g)$'s of the various bands. In a linear oscillator problem, the mean kinetic energy is half the total energy. Thus the lowest state has the lowest kinetic energy. Now the kinetic energy is of course merely the average of $p^2/2m$, weighted by the square of the momentum eigenfunction, so that this means that the states of low energy will have their momentum eigenfunctions concentrated in the zones of low g, successive bands coming at higher and higher g. This behavior is clear from Fig. 7. On the other hand, a tightly bound electron in an atom is in a potential similar to an inverse square field, in which the kinetic energy is the negative of the total energy, being large for the lowest energy level (which has a large negative value of energy) and decreasing as we go to higher energy levels. In other words, the momentum eigenfunction of the lowest state will extend out to high g values, and those of higher and higher bands will fill in at smaller and smaller momenta. The lowest kinetic energy of any energy bands will come approximately at the top of the barrier, or for the valence and conduction bands; it is for this reason that the $v(g)$ for the conduction band of sodium can lie largely in the central zone of momentum space. Then as we go to the higher free electron bands, of course the kinetic energy will increase again (these electrons are in a very different field from an inverse square field, so that there is nothing inconsistent in the kinetic and total energies increasing together), and the v's will lie further and further out. It is clear that, since the bound levels have already partly filled up the momentum space, the free electron wave functions cannot correspond to contributions of unity at any single momentum value, and hence cannot be really free electron like. This situation was pointed out in detail in reference 6, Secs. 21 and 22, but it is not generally recognized, and it seems worth while referring to it again here.

We have now considered the behavior of the momentum eigenfunction $v(p)$ or $v(g)$; let us ask what we can deduce about the nature of the Wannier function, its Fourier transform. We could get at this from our $v(g)$

only by making the Fourier analysis numerically, and this has not been done. We can make some deductions, however, fairly simply. In the first place, for the very tightly bound states, the Wannier function will be identical with the atomic function, a Hermite function. As the bands begin to broaden slightly, we can proceed as Wannier did in reference 8, to estimate the form of the function. The more interesting situation arises in the opposite limit, that of approximately but not completely free electrons, where the Wannier function has not been previously discussed. We have seen in this limit, for instance in Fig. 6, that $v(g)$ is approximately zero for one interval of unity along the g axis, approximately a constant in the next unit interval, zero along the next, another constant along the next, etc. If we assume that this approximate behavior is exact, then we can easily perform the Fourier transformation to determine the Wannier function analytically. Thus for instance let us consider an even momentum eigenfunction, which is equal to one constant, say $v(1/2)$, from $g=0$ to 1; zero from 1 to 2; another constant, say $v(5/2)$, from 2 to 3; zero from 3 to 4; and so on. The values for negative g's are the same. Then we can at once compute the Wannier function, by Eq. (23), or by the corresponding one-dimensional formula expressed in terms of our dimensionless variables, which is

$$a(w)=[(\pi)^{-\frac{1}{2}}/2]\int_{-\infty}^{\infty} v(g)\exp(igw)dg, \qquad (39)$$

where this a is normalized so that the integral of its square with respect to dw is equal to unity. If we substitute our assumption for v and carry out the integration, we find

$$a(w)=(\pi)^{-\frac{1}{2}}[(\sin\tfrac{1}{2}w)/(\tfrac{1}{2}w)][v(1/2)\cos w/2 + v(5/2)\cos(5w/2)+\cdots]. \qquad (40)$$

The term $v(1/2)\cos w/2+\cdots$ represents the real part of the wave function for $g=1/2$, and we see that it is to be multiplied by the function $(\sin\frac{1}{2}w)/(\frac{1}{2}w)$, equal to unity at $w=0$ but decreasing as w increases. The first zero of this function comes for $w=2\pi$, which we remember is two atomic distances away from $w=0$ (the lattice spacing in units of w equals π). In other words, $a(w)$ has the periodic oscillations of the actual wave function (which means that there are many oscillations in one lattice spacing, if we are dealing with a high energy), its amplitude shows the variation of the real wave function through the unit cell (having its maximum at the potential barriers, as we mentioned earlier), but superposed on this is the $\sin x/x$ type of function which makes the Wannier function gradually decrease as we go away from $w=0$. But we see that the function has considerable values not only in the cell at $w=0$, but also in the adjacent cells at $w=\pm\pi$. This slow falling off of the Wannier function for a free electron band is typical of situations we shall find in a number of other cases. We must realize, then, that though the Wannier functions have all the analytic properties mentioned earlier (orthogonality, usefulness in expressing the wave function by a Bloch-type sum, etc.), they are not always as concentrated in a single unit cell as has been generally assumed. A formula similar to (40) can be set up for momentum eigenfunctions which are odd functions of g. In this case it is interesting to note, as mentioned earlier, that in order to have the Wannier function real, the momentum eigenfunction $v(g)$ must be pure imaginary. We shall hence make this assumption in the future; it does not affect the results so far given, except that in Figs. 4 and 5 we must assume that we are plotting the odd momentum eigenfunctions divided by i.

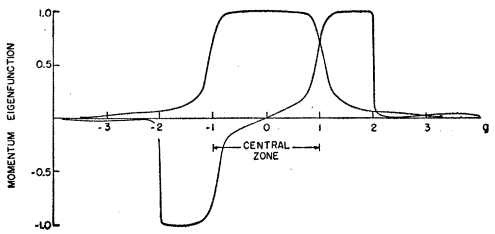

FIG. 9. Momentum eigenfunctions for two lowest bands, $s=1$.

We have now examined the case of $s=100$, considering the energy, momentum eigenfunctions, completeness relation, and Wannier functions. The cases of smaller s do not give us a great deal of new information, but they are interesting for comparison. In Fig. 8 we give a figure, like Fig. 7, for $s=16$, calculations from which these curves were plotted having been made at intervals of g 0.25 unit apart. We see that the lowest and second lowest bands have distributions much like the Hermite functions, as we expect, since they are bound states; the third is a transition type, and by the time we have reached the fourth band, it is very nearly free electron like, more so than even the 15th band for $s=100$. This situation is even more strikingly true for $s=1$. In Fig. 9 we give the momentum eigenfunctions for the two lowest states for $s=1$. (Calculations from which these curves were plotted were made at intervals of g 0.1 unit apart.) We see that even for these bands, the momentum eigenfunctions are almost free electron like; if they were exactly so, $v(g)$ for the lowest band would be equal to unity from $g=-1$ to $+1$, zero elsewhere, and for the second band it would be unity from $g=1$ to 2, -1 from -2 to -1, and zero elsewhere. And for the higher bands, the functions are very close to those for free electrons. We have given the momentum eigenfunctions themselves in Fig. 9, instead of their squares, as concerned in the test of the completeness relation, since the departures from the free electron case are more obvious in the functions themselves than in the squares.

4. THE EFFECT OF VARIOUS PERTURBATIONS ON THE ONE-DIMENSIONAL PROBLEM

Before we go on to consider the two- and three-dimensional cases, it is very much worth while to con-

TABLE II. Matrix components of sinusoidal perturbative energy, for free electron wave functions.

$0<g_0<1$ and $-1<g_0<0$: l and m both even, or l and m both odd:
$v_{lm}(g_0, n)=1$ if $2n=|m-l|$, unless $|m-l|=0$, when $v_{lm}(g_0, 0)=2$
$=0$ for other n's
$0<g_0<1$, l even, m odd: $v_{lm}(g_0, n)=i$ if $2n=m+l-1$; 0 for other n
l odd, m even: $v_{lm}(g_0, n)=-i$ if $2n=m+l-1$; 0 for other n
$-1<g_0<0$, one index even, other odd: conjugates of case $0<g_0<1$
$g_0=0$, l odd, m even, or vice versa: $v_{lm}(0, n)=0$
l and m both even: $v_{lm}(0, n)=1$ if $2n=|m-l|$, unless $|m-l|=0$, when $v_{lm}(0, 0)=2$
$v_{lm}(0, n)=-1$ if $2n=m+l$; $=0$ otherwise
l and m both odd: $v_{lm}(0, n)=1$ if $2n=|m-l|$, unless $|m-l|=0$, when $v_{lm}(0, 0)=2$
$v_{lm}(0, n)=1$ if $2n=m+l-2$; $=0$ otherwise
$g_0=1$, l odd, m even, or vice versa: $v_{lm}(1, n)=0$
l and m both even: $v_{lm}(1, n)=1$ if $2n=|m-l|$, unless $|m-l|=0$, when $v_{lm}(1, n)=2$
$v_{lm}(1, n)=-1$ if $2n=m+l-2$; $=0$ otherwise
l and m both odd: $v_{lm}(1, n)=1$ if $2n=|m-l|$, unless $|m-l|=0$, when $v_{lm}(1, n)=2$
$v_{lm}(1, n)=1$ if $2n=m+l$; $=0$ otherwise

sider the effect of certain small perturbations on the one-dimensional Mathieu problem. We have seen in Sec. 1 that a general one-dimensional potential would have Fourier components of wavelength Q_0, $2Q_0$, $\cdots$, but that the Mathieu case is that in which the amplitudes of all but one of these components were zero. We shall first consider the effect of adding a small Fourier component of wavelength nQ_0, where n is an integer [of course, if we are to retain a potential symmetrical in x, this implies setting up $W_n=W_{-n}\neq 0$, in the notation of Eq. (3)]. We must first find the nondiagonal matrix component of energy resulting from this perturbation, between any two unperturbed states. This is most conveniently handled in the momentum space; and for generality, let us consider the problem of the nondiagonal matrix component of an arbitrary Fourier component $W(\mathbf{P}_i)\exp[(-2\pi i/h)(\mathbf{P}_i\cdot\mathbf{r})]$ between two unperturbed wave functions expressed in the general form (16). We can show easily that there is no matrix component except between two wave functions u_l and u_m corresponding to the same lattice of points in p space (that is, corresponding to the same p_0, if p_0 is chosen to be in the central cell of momentum space). The matrix component between two such functions is then easily found to be

$$W(\mathbf{P}_i)\sum(\mathbf{P}_k)v_l^*(\mathbf{p}_0+\mathbf{P}_k)v_m(\mathbf{p}_0+\mathbf{P}_k+\mathbf{P}_i). \quad (41)$$

In our case of the one-dimensional problem with perturbative potential terms associated with W_n and W_{-n}, this nondiagonal component, expressed now in terms of our dimensionless variables, is

$$W_n\sum(k)v_l^*(g_0+2k)[v_m(g_0+2k+2n)+v_m(g_0+2k-2n)]. \quad (42)$$

We are interested, then, in the behavior of the sum

$$v_{lm}(g_0, n)=\sum(k)v_l^*(g_0+2k)[v_m(g_0+2k+2n)+v_m(g_0+2k-2n)], \quad (43)$$

where for simplicity we may take $-1\leq g_0\leq 1$.

As a first guide in our calculations, let us ask what this quantity, giving a matrix component of energy for a perturbative potential $2W\cos(2\pi nx/Q_0)$ for $W=1$, would equal for the case of really free electrons. We have already stated the nature of the free electron wave function. The function v_l is unity over the unit range of g from $l-1$ to l, and from $-(l-1)$ to $-l$, if l is odd, so that the wave function is even (bands are numbered upward from the lowest band, $l=1$). Just at $g=\pm(l-1)$, $\pm l$, the function must equal $2^{-\frac{1}{2}}$ in this case, to keep normalization; at these points, we are superposing traveling waves to form standing waves, or are at the edges of the band. The points $g=\pm(l-1)$ correspond to the lowest energy, $\pm l$ to the highest energy, in the band. Remembering that we are discussing odd l's, we see that the lowest energy of the band then corresponds to $g_0=0$, the highest energy to $g_0=1$. For even l, so that the wave function is odd, the v_l equals i (we remember that in this case v_l is imaginary, so as to yield a real Wannier function) from $l-1$ to l, $-i$ from $-(l-1)$ to $-l$, $2^{-\frac{1}{2}}i$ at $l-1$ and l, $-2^{-\frac{1}{2}}i$ at $-(l-1)$ and $-l$. Using these values and examining (43) with a little care, we find that $v_{lm}(g_0, n)$ takes on the values shown in Table II.

These values are somewhat complicated only on account of the way in which the momentum of the wave is expressed in terms of g_0 and an integer. We can explain them in words very simply. Except at the edge of a band, the wave function is a single wave, of the form $\exp(igw)$. If we wish to find the matrix component of a potential $\exp(\pm 2inw)$ between two waves with g's equal to g_l, g_m, we have to find the integral over w of $\exp i(-g_l+g_m\pm 2n)$, which is zero unless $2n=|g_l-g_m|$. At the edge of a band, however, we have standing waves of the form $2^{-\frac{1}{2}}[\exp(igw)\pm\exp(-igw)]$, so that we have an integral of a quantity like

$$\tfrac{1}{2}[\exp(-ig_lw)\pm\exp(ig_lw)]\times[\exp(-ig_mw)\pm\exp(ig_mw)]\exp(\pm 2inw),$$

which gives nonvanishing terms not only if $2n=|g_l-g_m|$, but also if $2n=|g_l+g_m|$. All the components in Table II fall into these simple categories.

It is now interesting to see to what extent the matrix components $v_{lm}(g_0, n)$ for the real wave functions agree with these free electron values. Accordingly we give,

in Table III, the components $v_{lm}(g_0, n)$ for the wave functions of the 9th and 10th bands, for the Mathieu case with $s=100$. Thus we have $v_{9,9}$, $v_{9,10}$, and $v_{10,10}$. We have computed these for $g_0=0$, $\frac{1}{2}$, and 1; but the behavior for all values of g_0 between 0 and 1 does not differ much from the values at $g_0=\frac{1}{2}$, the change of properties as we approach the edge of the band at $g_0=0$, 1 being quite sudden. Our rules of Table II would indicate that in this case if we had free electrons $v_{9,9}(g_0, n)$ should be 2 if $n=0$, for each g_0. There should be no further nonvanishing components except at the edge of the band, where at $g_0=0$ there should be a component equal to unity for $n=8$, and at $g_0=1$, where there should be a component of unity for $n=9$. Similarly $v_{10,10}(g_0, n)$ should be 2 if $n=0$ for each g_0, and in addition should have a value of -1 for $g_0=0$, $n=9$, and for $g_0=1$, $n=10$. The nondiagonal matrix component $v_{9,10}(g_0, n)$ should be $-i$ for $n=9$, zero for other n's, except at the edges of the band; it should be zero at both band edges. When we examine Table III, we see that the actual values are very far from these free electron cases. A few of the predictions of the free electron case must hold here too; thus for $n=0$ the result must come out the same, for then they depend only on the normalization of the wave function, and the vanishing of the nondiagonal component $v_{9,10}$ at the edges of the band follows from symmetry alone. (It is to be noted that this vanishing of the nondiagonal component holds only because one of our states has an even momentum eigenfunction, one an odd function; it would not occur if both had even eigenfunctions, or if both had odd functions.) But in all the cases where only one component would be different from zero in the free electron case, we actually find a wide range of components different from zero, and the one which would appear in the free electron case is not even the biggest component. This situation is, of course, a result of the fact that the individual wave functions have large contributions connected with many sinusoidal waves, as we have already pointed out.

The effect of perturbations is particularly interesting as far as it concerns the widening of gaps between bands; for here there has been widespread use,[12] particularly in the theory of alloys, of the assumption that a single gap was widened by the action of a single Fourier component of potential, as it would be for free electrons. Let us remind ourselves of how this would work out for the case we are considering, the 9th and 10th bands, if the wave functions were really free electron like. There would be a nondiagonal matrix component of energy coming from the perturbation $\cos 2nw$, with $n=9$. This would have a maximum effect as g approached the value 9, the boundary between the 9th and 10th zones in momentum space, at which the diagonal energies of the two states would become equal to each other. Just at this value $g=9$, the gap would appear, the perturbed wave functions being the sum and difference of the progressive waves, the gap width being proportional to the nondiagonal matrix component; the depression of the lower energy level and the elevation of the upper one, on account of the perturbation, would rapidly decrease as g departed from 9. If, as a matter of fact, the unperturbed wave functions were set up, as we have been doing it, to correspond to a standing wave just at the edge of the zones, the perturbation will not have to produce a linear combination of direct and reflected waves just at this point, since the linear combinations will already have been set up in the unperturbed wave functions. Hence the gap will be produced, just at $g=9$, by the diagonal, rather than the nondiagonal, matrix component of perturbative energy. As soon as g deviated from 9, however, the perturbation would produce linear combinations, so that the final result would be just as in the usual treatment of perturbations of sinusoidal wave functions by sinusoidal potentials. No matrix component of perturbative energy, except that for $n=9$, would be effective in setting up the gap at $g=9$.

We can now see how completely the situation is altered when we have the real wave functions, from the figures of Table III. The gap comes at $g_0=1$. Here $v_{9,9}(g_0, n)$ and $v_{10,10}(g_0, n)$ give a measure of the perturbation produced at the two edges of the gap by the diagonal matrix components of the various Fourier components of potential. We see that for all the larger n values, the two terms are almost equal and opposite. That is, each Fourier component pushes the energy at the two edges of the gap in opposite directions. As g_0 deviates slightly from unity, we have nondiagonal matrix components of energy; and we see from $v_{9,10}(g_0, n)$ for $g_0=\frac{1}{2}$ that these components are very nearly equal in absolute magnitude to the diagonal component for $g_0=1$, which is necessary in order to have a continuity in behavior between the edge of the gap, where our

TABLE III. Matrix components of sinusoidal perturbative potential, for Mathieu wave functions of 9th and 10th bands, $s=100$.

	$g_0=0$		$g_0=\frac{1}{2}$			$g_0=1$	
n	(9/9)	(10/10)	(9/9)	(9/10)	(10/10)	(9/9)	(10/10)
0	2.000	2.000	2.000	0.000	2.000	2.000	2.000
1	−0.516	−0.264	−0.386	−0.010i	−0.292	−0.392	−0.350
2	0.282	0.060	0.120	0.017i	0.066	0.064	0.112
3	−0.288	−0.016	−0.046	−0.051i	−0.020	0.018	−0.060
4	0.380	−0.010	0.024	0.116i	0.006	−0.100	0.120
5	−0.496	0.080	−0.008	−0.237i	0.002	0.228	−0.238
6	0.508	−0.166	0.002	0.421i		−0.416	0.424
7	−0.196	0.312		−0.521i		0.562	−0.554
8	−0.029	−0.532		0.383i		−0.396	0.380
9	0.240	0.496		0.188i		−0.180	0.186
10	0.360	0.056		−0.376i		0.380	−0.372
11	0.212	−0.460		−0.380i		0.380	−0.376
12	0.068	−0.364		−0.179i		0.180	−0.180
13	0.012	−0.148		−0.054i		0.056	−0.056
14		−0 042		−0.011i		0.012	−0.012
15		−0.008		−0.003i			

[12] H. Jones, Proc. Roy. Soc. (London) **144**, 225 (1934); L. Pauling and F. J. Ewing, Revs. Modern Phys. **20**, 112 (1948); and many other references.

wave functions represent standing waves, and different values of g_0, where they are approximately traveling waves. But the important fact is that almost any Fourier component of potential is effective in broadening the gap at $g=9$, not just the component $n=9$; and the interesting additional feature that some of the components are of one sign, some of the other, so that some components of potential can actually narrow the gaps, rather than broaden them.

There is, in other words, simply nothing like a one-to-one correspondence between Fourier components of potential and their effect on broadening the gaps between energy levels; and one must look with a great deal of reservation on such applications as those mentioned in reference 12. It is notable that the gap between the 9th and 10th bands, which we have used for illustration, is in a region of the energy *vs* g curve, shown in Fig. 3, where this curve resembles a free electron parabola very strongly, and the gap is very narrow; thus we might very easily be under the impression that a real free electron treatment was appropriate, whereas it obviously is not. It is interesting to point out another aspect of this situation, by which one Fourier component of potential can be effective in broadening other gaps, besides that which the free electron approximation would suggest. This aspect arises from our Mathieu problem itself. This is the problem in which there is only one Fourier component of potential, which on the free electron approximation would produce only an energy gap at $g=1$, between the first and second bands. On the other hand, as we see from Fig. 3, it has actually produced wide gaps between all the lower energy bands. If we try to explain the existence of these gaps from perturbation theory, we find that we must do it by using higher-order perturbations, the gap at $g=2$ coming from second-order perturbations, that at $g=3$ from third-order, and so on. The great width of the actual gaps shows the inappropriateness of perturbation method for handling the problem. This does not affect, however, the correctness of the application of perturbation theory which we are making in the present section, where we investigate the effect of an additional small perturbative potential of nature cos2*nw*, superposed on the large potential of form cos2*w* already taken into account by our exact solution of the Mathieu problem

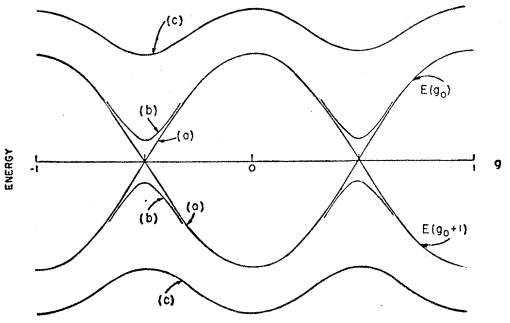

FIG. 10. Energy *versus* momentum for problem of band with perturbative potential of double periodicity, showing splitting of band. Unit cell in momentum space, unperturbed problem, extends from $g=-1$ to 1; for perturbed problem, from $g=-\frac{1}{2}$ to $\frac{1}{2}$. Curve (a), vanishing perturbation; (b), larger perturbation; (c), still larger perturbation.

The perturbation cos2*nw* which we have discussed does not interfere with the periodicity of the structure, or the size of the unit cell. Another interesting type of perturbation, however, is that produced by a potential of for instance cos$w$, which has a longer period (in this case 2π) than the unperturbed potential (with its period of π). Such a perturbative potential has the effect of producing an alternation in the depth of adjacent potential wells. It is this type of alternating potential which has been considered by the writer[13] in discussions of antiferromagnetism and order-disorder effects in certain alloys.

We easily find that such a perturbative potential has nondiagonal matrix components between two states of the same energy band, one with a given g_0, the other g_0 a unit greater. These nondiagonal components are more important than those between different energy bands. For a potential 2 cos$w$, the nondiagonal matrix component of energy between these two states proves to be

$$\sum(k)v_l^*(g_0+2k)[v_l(g_0+2k+1)+v_l(g_0+2k-1)], \quad (44)$$

showing a close resemblance to Eq. (43). We can understand the situation better from Fig. 10. Here we plot the energy of the lth band as a function of g_0, in the central cell of momentum space, between $g_0=-1$ and 1. We also plot the energy, as a function of g_0+1. Then our rule states that there will be a nondiagonal matrix component of energy between states on the two curves lying vertically above each other. These nondiagonal components will cause the correct wave functions to be linear combinations of the two unperturbed functions, one of given g_0, the other of g_0+1. Just at $g_0=\pm\frac{1}{2}$, where the two curves cross, the unperturbed problem is degenerate, the perturbation will cause the levels to separate by amounts proportional to the nondiagonal matrix component, and the resulting wave functions will be the sum and difference of the unperturbed wave functions. As g_0 departs from $\pm\frac{1}{2}$, and the unperturbed energy levels deviate from each other, the situation will depend on the relative magitude of the nondiagonal matrix component of energy, and the separation of the unperturbed states; if the nondiagonal component is small compared to the separation of the unperturbed states, the displacements of the perturbed states, as seen from second-order perturbation theory, will be proportional to the square of the nondiagonal matrix components, divided by the difference of unperturbed energies. In Fig. 10(a), we show the unperturbed energy levels; in (b) the perturbed levels with a small perturbations; in (c) the perturbed levels with a

[13] J. C. Slater, Phys. Rev. **82**, 538 (1951); **84**, 179 (1951).

large perturbation. It is clear how the curves of Fig. 10 are to be interpreted. Once we have our perturbation $\cos w$, the perturbed problem really has a periodicity of 2π along w, rather than of π. Thus we should use a Brillouin zone half as great, the central zone extending from $g_0=-\frac{1}{2}$ to $\frac{1}{2}$ rather than from -1 to 1. The original energy band of the unperturbed problem has split into two, each holding one electron per unit cell of each spin; but since the unit cell is twice as great, the two bands taken together hold the same number of electrons per atom as the original unperturbed band. The net effect of the perturbation is to introduce a gap in the middle of the original band.

It is interesting to ask what happens to the perturbed wave functions as a result of this perturbation. If the perturbation is small—that is, if the separation of the unperturbed states is large compared to the nondiagonal matrix component—of course the effect is small; a little of one wave function is mixed in with the other. As the perturbation becomes large, however, the unperturbed wave functions are combined with almost the coefficients ± 1. In this case the combination becomes

$$\sum(k)\{v_l(g_0+2k)\exp[i(g_0+2k)w] \pm v_l(g_0+2k+1)\exp[i(g_0+2k+1)w]\}.$$

This can equally well be written

$$\sum(k)\{v_l(g_0+2k)\exp[i(g_0+2k)w] \pm(\tfrac{1}{2})[v_l(g_0+2k+1)\exp[i(g_0+2k+1)w] + v_l(g_0+2k-1)\exp[i(g_0+2k-1)w]]\},$$

since the latter two terms, when summed over k, are identical. It is now particularly interesting to consider how this behaves in the case of tightly-bound states. For this, we remember that $v_l(g)$ is a slowly varying function of g, not changing much in a unit interval along g. Thus we can approximately replace $v_l(g_0+2k\pm 1)$ by $v_l(g_0+2k)$. If we do this, the function becomes

$$(1\pm\cos w)\sum(k)v_l(g_0+2k)\exp[i(g_0+2k)w]. \quad (45)$$

This is the wave function in the absence of perturbations, multiplied by the factor $1\pm\cos w$. If we take the $+$ sign, the function equals 2 when $w=0, 2\pi, \cdots$, and equals 0 when $w=\pi, 3\pi, \cdots$. That is, in this case, the wave function will correspond to having the wave functions on the atoms at $0, 2\pi, \cdots$, much as in the absence of perturbations, but to having the wave functions on the alternate atoms practically equal to zero. With the $-$ sign, the wave function is large on the intermediate atoms. The interpretation of the situation is simple. With a large nondiagonal matrix component of energy, the alternate potential minima are of quite different depths, as shown in Fig. 11(a). If we are dealing with tightly bound states, below the potential barrier, so that these states lie not far from the corresponding states of the isolated atoms, it is clear that the band will split into two quite separate bands, one corresponding to a state where the wave function is large in one type of atom, the other where it is large in the others. It is this case which we have when the wave functions for all g_0's are the sums or differences of the unperturbed wave functions, as in (45). The other case, however, where this occurs only in a narrow range of g_0's about $g_0=\pm\frac{1}{2}$, is that in which the nondiagonal matrix component of energy is small compared to the band width; this will be the case in Fig. 11(b), where the difference in heights of the two types of minima is small compared to the width of the band. In this case most of the wave functions extend with approximately equal amplitudes into both types of atoms; it is only those in the immediate neighborhood of the gaps which are concentrated largely in one type of atom or the other (a concentration which is not easy to show in a diagram like Fig. 11(b)).

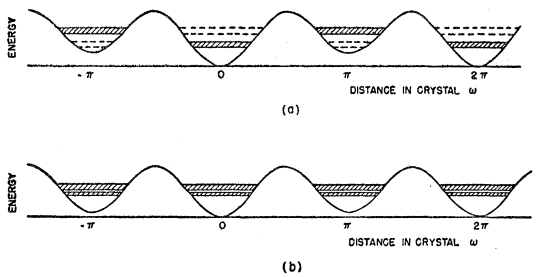

FIG. 11. Energy band split by an alternating component of potential. Curve (a), fairly large perturbation; dotted lines indicate region where the wave function is small. Curve (b), small perturbation, which merely introduces energy gaps.

Now that we have considered the general nature of the perturbation produced by the potential $\cos w$, it is interesting to consider the nondiagonal matrix component (44) involved in the perturbation. If we were dealing with perfectly free electrons, we could show easily that this expression would be zero, except for the first band, for which it would be unity; no higher bands would be split by this perturbation. For contrast, we give in Table IV the values computed for $s=100$, for a series of bands. We see that for the bound states, the matrix component is between 1 and 2 (it would be just 2 if our approximation that $v_l(g_0+2k\pm 1)$ equaled $v_l(g_0+2k)$ were justified, as we see from (44), using the fact that the wave functions are normalized). It begins to decrease rapidly as we get to the top of the barrier, and by the time we have gone several states above the barrier, it has become small. For the states near the top of the barrier, then, the component is not far from unity; and this shows us that, in a real crystal, we may well expect that this type of perturbation, applied to the valence or conduction band, may produce a splitting of the band, much as it would do in the lowest band with really free electrons.

It is now interesting to go further and ask about the behavior of the Wannier functions when we impose a perturbation of the type we have been considering on an

TABLE IV. Matrix components of sinusoidal perturbative potential $2\cos w$, for Mathieu wave functions of various bands, $s=100$, for $g_0=\frac{1}{2}$.

Band	Matrix component
1	1.944
2	1.837
3	1.714
4	1.574
5	1.404
6	1.118
7	0.629
8	0.184
9	0.026

originally sinusoidal potential. Our approach is naturally to consider the momentum eigenfunction, and then to find the Wannier function by Fourier transformation. Let us first start with the unperturbed problem and ask what happens to the momentum eigenfunction when we describe the problem in terms of a unit cell twice as great (that is, of period 2π in w), and correspondingly a Brillouin zone half as great (extending from $g=-\frac{1}{2}$ to $\frac{1}{2}$). For definiteness, let us take the lowest band of the unperturbed problem, for $s=100$. Here the momentum eigenfunction for the unperturbed problem, as given in Fig. 4, is reproduced in Fig. 12(a). Next, in Fig. 12(b), we show the momentum eigenfunction of the lower half-band (in the sense of Fig. 10) arising from the single band of the unperturbed problem, when we change the cell size. We must now write our wave functions in terms of a sum over all g values differing by an integer, rather than all g values differing by an even integer. Yet the final wave function, of course, must be the same as in our original description. Thus for instance if g_0 is $\frac{1}{4}$, the components of the momentum eigenfunction at $1\frac{1}{4}, 3\frac{1}{4}, 5\frac{1}{4}, \cdots, -\frac{3}{4}, -2\frac{3}{4}, \cdots$, must all be zero; but those at $2\frac{1}{4}, 4\frac{1}{4}, \cdots, -1\frac{3}{4}, -3\frac{3}{4}, \cdots$, must have the same values as in the original description. A similar situation holds if g_0 is between 0 and $-\frac{1}{2}$, say $-\frac{1}{4}$. In other words, $v(g)$ must be just as in Fig. 12(a) for g in the range $-\frac{1}{2}$ to $\frac{1}{2}$, $1\frac{1}{2}$ to $2\frac{1}{2}$, $3\frac{1}{2}$, $\cdots$, and $-1\frac{1}{2}$ to $-2\frac{1}{2}$, $-3\frac{1}{2}$ to $-4\frac{1}{2}$, etc., but it must be zero in between; that is, it must be as in Fig. 12(b). For the upper half-band, the components in alternate unit intervals of g must be different from zero, as in Fig. 12(c).

We may now consider the Wannier function, obtained by Fourier transformation of the functions of Fig. 12. These functions may be obtained from the original momentum eigenfunction of Fig. 12(a), which we may call $v_0(g)$, by multiplying by a function which, in (b), equals 1 from $-\frac{1}{2}$ to $\frac{1}{2}$, $1\frac{1}{2}$ to $2\frac{1}{2}$, etc., and is zero between; or which, in (c), equals 1 from $\frac{1}{2}$ to $1\frac{1}{2}$, $2\frac{1}{2}$ to $3\frac{1}{2}$, etc., and is zero between. These step functions are easily expressed as Fourier series. Thus the function which is unity in the region $-\frac{1}{2}$ to $\frac{1}{2}$, $1\frac{1}{2}$ to $2\frac{1}{2}$, etc., zero between, is

$$f_1(g)=\sum(n=-\infty\cdots\infty)c_n\exp(\pi ing),$$

where

$$c_0=\tfrac{1}{2},\quad c_1=c_{-1}=1/\pi,\quad c_2=c_{-2}=0,$$
$$c_3=c_{-3}=-1/3\pi,\quad c_4=c_{-4}=0,\quad c_5=c_{-5}=1/5\pi,\cdots. \quad (46)$$

The corresponding function $f_2(g)$, equal to unity when g is between $\frac{1}{2}$ and $1\frac{1}{2}$, $2\frac{1}{2}$, and $3\frac{1}{2}$, etc., and zero between, is $1-f_1(g)$, or

$$f_2(g)=\sum(n)d_n\exp(\pi ing),\quad d_0=\tfrac{1}{2},$$
$$d_1=d_{-1}=-1/\pi,\quad d_3=d_{-3}=1/3\pi,\text{ etc.} \quad (47)$$

Thus the momentum eigenfunction given in Fig. 12(b) can be written as $v_0(g)f_1(g)$, and that in Fig. 12(c) as $v_0(g)f_2(g)$. It is now simple to substitute the expressions into the formula (23) for finding the Wannier function, or the corresponding formula (39) in terms of our dimensionless variables.

From Eq. (39), we have

$$a(w)=\tfrac{1}{2}\pi^{-\frac{1}{2}}\int_{-\infty}^{\infty}v(g)\exp(igw)dg.$$

When we substitute $v(g)=v_0(g)f_1(g)$ or $v_0(g)f_2(g)$, we have

$$a(w)=\sum(n)c_n\int_{-\infty}^{\infty}v_0(g)\exp[i(w+n\pi)g]dg$$
$$=\sum(n)c_na_0(w+n\pi),\quad\text{or}\quad\sum(n)d_na_0(w+n\pi), \quad (48)$$

where

$$a_0(w)=\tfrac{1}{2}\pi^{-\frac{1}{2}}\int_{-\infty}^{\infty}v_0(g)\exp(igw)dg. \quad (49)$$

That is, from (48), the Wannier function of the lower half-band, in the new description in terms of the doubled unit cell, consists of the original Wannier function a_0 located at the atom at the origin, with an amplitude $\frac{1}{2}$; but also similar Wannier functions with amplitudes $1/\pi$ located on the adjacent atoms at $\pm\pi$, functions with amplitudes $-1/3\pi$ at the atoms at $\pm3\pi$,

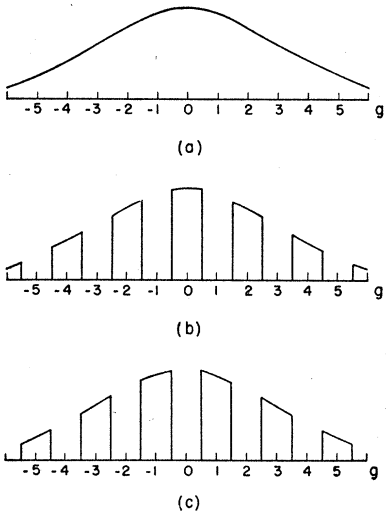

FIG. 12. Momentum eigenfunction of lowest band, $s=100$, for (a) original problem; (b) and (c), same problem expressed in terms of unit cell twice as great; (b), lower half-band, and (c), upper half-band.

and so on. A similar situation holds for the upper half-band. This composite Wannier function, in other words, extends out much further than the original Wannier function, and we note that this change has been brought about, not by a change in the physical situation, for the perturbation is still vanishingly small, but merely by a change in the method of description. This is an example of a situation that has not been sufficiently appreciated: the method of setting up the Wannier functions is not unique.

Now let us ask how these Wannier functions will change when the amplitude of the perturbative potential builds up from zero to a large value. First we consider the limiting case where the perturbation is so great that for all g values the perturbed wave functions are the sum and difference of the original wave functions, so that the perturbed momentum eigenfunctions are the sum and difference of the original momentum eigenfunctions. For the sum, the problem is simple. The sum of the functions shown in Figs. 12(b) and 12(c) is obviously the original function shown in Fig. 12(a); thus the corresponding Wannier function must be just the original Wannier functions located at the origin. This is also clear from Eq. (48), in which the sum of c_n and d_n equals unity for $n=0$, zero for any other n. But there is a difference between this problem and the unperturbed one, nevertheless. For now our unit cells are twice as great as before. Wannier functions are to be located only at the atoms at $w=0, \pm 2\pi, \pm 4\pi$, etc., in setting up the Bloch sum (20) and are to be omitted from the atoms at $w=\pm\pi, \pm 3\pi$, etc. This is the case shown in Fig. 11(a), where the potential minima at $w=0, \pm 2\pi, \pm 4\pi$, etc., are much lower than those at $\pm\pi, \pm 3\pi$, etc., and we are speaking of the wave functions which have their intensities located almost entirely in these lower potential wells. It is natural that the Wannier functions are concentrated about these potential wells in which the final wave function is large.

The situation is really not as simple as this, however, as we see when we consider the upper half-band. There the wave functions, and momentum eigenfunctions, in the limit of large perturbations, are the differences rather than the sums of the original functions. The momentum eigenfunction is then $v_0(g)$ multiplied by a function which is unity from $-\frac{1}{2}$ to $\frac{1}{2}$, -1 from $\frac{1}{2}$ to $1\frac{1}{2}$, 1 from $1\frac{1}{2}$ to $2\frac{1}{2}$, -1 from $2\frac{1}{2}$ to $3\frac{1}{2}$, etc. And the Wannier function is determined by the difference c_n-d_n in Eq. (48), and hence corresponds to nothing on the atom at $w=0$, functions of amplitude $2/\pi$ at $w=\pm\pi$, nothing at $w=\pm 2\pi$, functions of amplitude $-2/3\pi$ at $w=\pm 3\pi$, and so on. This is an unexpected situation, for physically we expect the wave functions in this case to be essentially like those in the lower half-band, only concentrated now on the atoms at odd multiples of π, rather than at even multiples, and we might suppose that the Wannier functions would be entirely concentrated at one atom at $w=\pi$. The reason why it is not is that we introduce at the outset the postulate that the

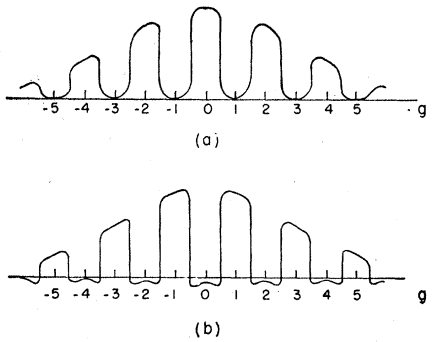

FIG. 13. Momentum eigenfunction of cases shown in Fig. 12(b) and Fig. 12(c), but for nonvanishing but small perturbative potential. (a), same case as Fig. 12(b); (b), same case as Fig. 12(c).

Wannier functions were to be located at $w=0, \pm 2\pi$, etc., by our original equations (39) and (48). The Wannier functions which we have found, with contributions at $w=\pm\pi, \pm 3\pi$, etc., lead in fact to the same final wave functions, when we combine them into Bloch sums, as we should get if we had a Wannier function concentrated only on the atom at $w=\pi$, with a similar one on $w=3\pi, 5\pi$, etc.; this is shown in Appendix I. But the description is different, and this is another illustration of the fact that the Wannier function is not unique.

We have considered the Wannier function, for our problem of the perturbative potential $2W\cos w$, for two limiting cases: first, for infinitesimal W; next for very large W. Let us now ask qualitatively how they behave for intermediate W. It is clear in a general way what will happen: the Wannier function for the lower band, which has components at $w=0, \pm\pi, \pm 3\pi$, etc., when $W=0$, and which has only a component at $w=0$ when W is large, will naturally gradually lose the components at $\pm\pi, \pm 3\pi$, etc., as W increases. We can say something more than this, however, from our general knowledge of Fourier series, coupled with our knowledge of how the perturbation problem behaves. The rather slow falling off of the c_n's of Eq. (48) with increasing n is a result of the discontinuities of the step function given in Eq. (46); if the function were rounded off at the half-integral values of g, instead of being discontinuous, the c_n's would decrease more rapidly with n. But this is just what we are to expect as a result of our perturbations. With a relatively small value of W, we may expect Fig. 12(b) and (c) to be modified as indicated in Fig. 13 (which is not a result of quantitative calculation). The modification comes in, for small W, mostly near the half-integral values of g, resulting in each case in a decreasing of the amplitude of the function as given in Fig. 12, in the region where that function is different from zero, with a compensating build-up in the region where the function of Fig. 12 is equal to zero. Thus we may suppose that the Wannier function corresponding to the lower band begins to lose its contributions corresponding to large w values quite rapidly as W increases, the contributions from $w=\pm\pi$ persist-

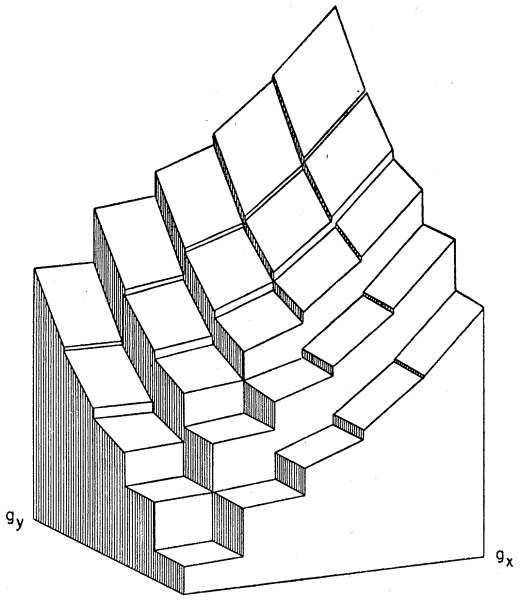

FIG. 14. Energy ϵ as function of components of momentum, g_x and g_y, in dimensionless units, for the case $s=16$.

ing until quite a large value of W is reached. For the upper band, on the other hand, where these contributions for large w's are present even for large W, the Wannier function will naturally be extended over many unit cells for any value of W.

5. THE UNPERTURBED MATHIEU PROBLEM IN TWO DIMENSIONS

The unperturbed Mathieu problem in two dimensions follows directly from the one-dimensional problem, and at first glance there seems to be very little about it which we have not already met in our one-dimensional problem; we find later, however, that certain degeneracies between energy levels, which would be removed by almost any sort of perturbation, really make an essentially new feature of the problem. First we consider the unperturbed case. The energy as a function of momentum is of course simply the sum of the energies of two one-dimensional problems. Thus if we let g_x, g_y be quantities like our earlier g, measuring momentum in dimensionless variables in the x and y direction, we can show the energy as a function of g_x and g_y by a surface. Such a surface, for the case $s=16$, is shown in Fig. 14. A section of this surface, but by a plane $g_x=$constant or $g_y=$constant, would be like Fig. 3, except that we show the case $s=16$, instead of $s=100$ as in Fig. 3. In Fig. 14, we have shown the energy in an extended momentum space, rather than showing all sheets of the energy surface in the central zone of momentum space, since this makes it easier to visualize the situation. We can also display the energy bands as a function of lattice spacing, as shown in Fig. 15. This can be computed directly from Fig. 1, the corresponding one-dimensional case, by simply adding the energies of two one-dimensional cases. We now observe the interesting fact that for small interatomic distances we no longer have any energy gaps; the bands overlap, as in a real crystalline conductor. As for the momentum eigenfunction, it simply is a function of g_x times a function of g_y, each being of the form discussed in Sec. 3.

The problem becomes more interesting, however, as we begin to consider the various energy bands, and their degeneracies, more in detail. Let us consider the limit of large interatomic distances. In this limit, the potential energy is proportional to r^2, where r is the distance from a lattice point, so that the solutions have the properties characteristic of the two-dimensional analog of the central field, and hence can be described by an azimuthal quantum number and a principal quantum number. We can also, of course, describe the solutions by the quantum numbers n_x and n_y associated with the one-dimensional linear oscillator problems in terms of which the two-dimensional linear oscillator can be expressed. At infinite separation, we have seen that the energy of the one-dimensional problem is $2n+1$, in units of $h\nu_0/2$, ν_0 being the oscillation frequency of the oscillator, and n taking on the values 0, 1, 2, $\cdots$. Thus the energy of the two-dimensional problem, at infinite separation, is $2(n_x+n_y)+2$. The atomic wave functions, which represent the real wave functions at infinite separation, are the products of Hermite functions, as given in Eq. (13), for x and y. Thus let us consider some of the lowest wave functions. Using Eq. (13), we have functions as given in Table V. We see that the ground state of the problem, with $n_x=0$, $n_y=0$, wave function $\exp(-r^2/2)$, has the properties of an s state, the wave function depending on r only. The next two functions, which can be written $\cos\theta[r\exp(-r^2/2)]$ and $\sin\theta[r\exp(-r^2/2)]$, where θ is the angle in a polar coordinate system where z_x, z_y are rectangular coordinates, are p-like functions, corresponding to the p_x and p_y type of atomic orbitals. Of the next three, the one with $n_x=1$, $n_y=1$, whose wave function can be written $\sin2\theta[r^2\exp(-r^2/2)]$, is of a d type. The other two do not correspond to any definite azimuthal quantum number as they stand; but their sum and difference do. The sum is $-2(1-r^2)\exp(-r^2/2)$, an s-like function, which we might call $2s$, and the difference is $\cos2\theta[2r^2\exp(-r^2/2)]$, the other function of the d type. We can make similar identification of the cases with higher n_x, n_y in terms of the quantum numbers of the

TABLE V. Unnormalized atomic wave functions for two-dimensional linear oscillator, for different quantum numbers n_x, n_y. Here $r^2=z_x^2+z_y^2$, $z=s^{\frac{1}{4}}w$.

n_x	n_y	ϵ	u
0	0	2	$\exp(-r^2/2)$
1	0	4	$z_x\exp(-r^2/2)$
0	1	4	$z_y\exp(-r^2/2)$
2	0	6	$(-1+2z_x^2)\exp(-r^2/2)$
0	2	6	$(-1+2z_y^2)\exp(-r^2/2)$
1	1	6	$2z_xz_y\exp(-r^2/2)$

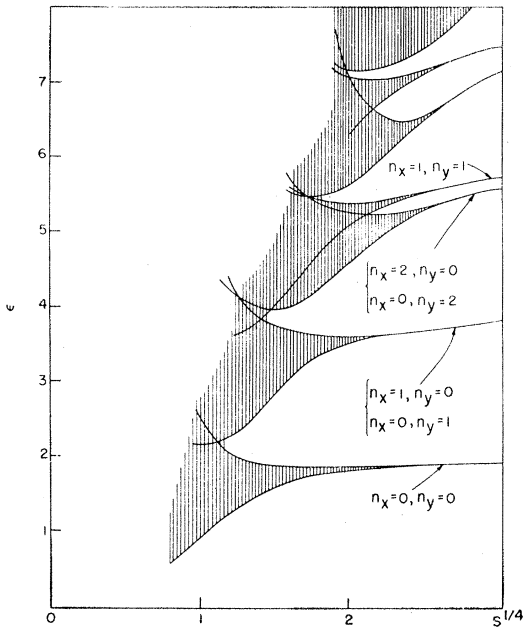

FIG. 15. Energy bands as functions of lattice spacing, two-dimensional Mathieu problem. Quantum numbers n_x, n_y are given as indicated in Table V.

central field problem. It is then convenient to denote the various bands, as we do in the crystal formed from real atoms, by the quantum numbers of the corresponding states in the separated atoms. Thus the band arising from $n_x=0$, $n_y=0$ can be considered to be the $1s$ band; that arising from $n_x=1$, $n_y=0$, and $n_y=1$, $n_x=0$ is the $2p$ band, showing the twofold degeneracy characteristic of a p state in a two-dimensional problem; that arising from the sum of $n_x=2$, $n_y=0$ and $n_x=0$, $n_y=2$ is the $2s$ band; that from the difference of these states and that from $n_x=1$, $n_y=1$ are the two components of the first d band, which we may call $3d$; like the p states, it shows only a twofold degeneracy in the two-dimensional problem. We observe that there is degeneracy between the $2s$ and $3d$, in this linear oscillator problem.

Now that we have looked at the behavior of our energy levels and wave functions at infinite separation, we can ask what happens to the states as the interatomic distance decreases. From Fig. 2, or Fig. 15, we know that the first effect is to displace each of the energy levels of the one-dimensional problem, before the band begins to broaden significantly. The displacement is greater, in the one-dimensional problem, the greater the value of n, but it is not a simple function of n. We then see that the displacement will be the same for the two p-like solutions $n_x=1$, $n_y=0$, and $n_x=0$, $n_y=1$, so that the degeneracy of these p states will not be removed by the effect of the "crystalline field" mentioned in Sec. 1. However, the displacement of the two states $n_x=2$, $n_y=0$, and $n_x=0$, $n_y=2$, will be different from that of the state $n_x=1$, $n_y=1$, so that the degeneracy of the two d-like states will be removed by the crystalline field, though the degeneracy between the one with wave function like $\cos 2\theta r^2 \exp(-r^2/2)$ and the $2s$-like function, will not be removed by this field. This removal of the degeneracy of the d states by the square, or cubic, crystalline field is entirely analogous to the corresponding situation in real three-dimensional crystals. The degeneracy between one of these d states and the s state is, however, peculiar to our special problem.

Let us now reduce the interatomic distance still more, to a point where the bands begin to broaden. Nothing remarkable happens to the $1s$-like band. The $2p$-like band, however, immediately involves us in a very interesting problem of degeneracy, the simplest case of the general problem of overlapping bands which one meets in real crystals. We can make this problem plain by drawing the surfaces giving energy ϵ as a function of momentum g_x and g_y, not as in Fig. 14, but in the central cell of momentum space. This is done for the two p-like states in Fig. 16. First, in Figs. 16(a) and (b), we draw separately the surfaces for the case $n_x=1$, $n_y=0$, and $n_x=0$, $n_y=1$. Then in Fig. 16(c) we superpose these two surfaces. It is clear from the figure that the two surfaces intersect along the 45° diagonals in the momentum space, given by $g_y=\pm g_x$. That is, at any point of these two diagonals, the two energy bands are degenerate with each other. This degeneracy, as we shall show later, is characteristic only of our specialized Mathieu problem. We shall see in the next section that if we introduce as perturbations the type of additional Fourier components of potential which would automatically be present in a real crystal, there will be nondiagonal matrix components of energy between the two states shown by the energy surfaces of Fig. 16, acting to repel the upper sheet from the lower; and this will cause the two degenerate states to separate along the 45° lines so that there will really be two energy surfaces, one lying above the other at every g value and formed by displacement of the upper surfaces of Fig. 16, and the other lying everywhere below. The only exceptions come at certain special symmetry points, namely, the center and corners of the cell in momentum space; at these points the nondiagonal matrix component of energy is automatically zero, by symmetry, and the two perturbed energy surfaces adhere to each other. These interesting facts, and their effect on the Wannier functions, will be taken up in the next section.

Similarly in Fig. 17 we show the three energy sur-

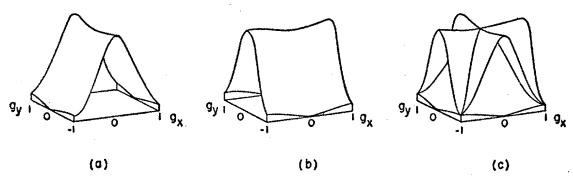

FIG. 16. Energy as a function of momentum, p-like states, two-dimensional Mathieu problem. (a) energy surface for case $n_x=1$, $n_y=0$; (b) for case $n_x=0$, $n_y=1$; (c) combination of cases (a) and (b), showing degeneracy, or intersection of surfaces, along diagonals $g_y=\pm g_x$.

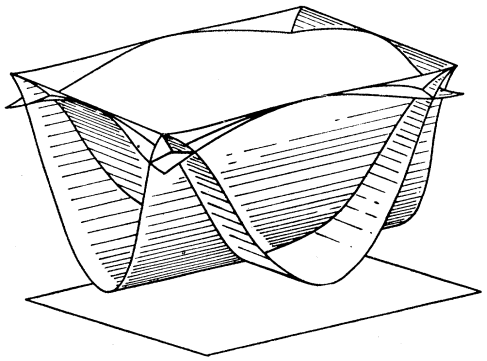

FIG. 17. Energy surfaces as function of momentum, for the states $n_x=2$, $n_y=0$; $n_x=0$, $n_y=2$; and $n_x=1$, $n_y=1$, for two-dimensional Mathieu problem.

faces for the states for $n_x=2$, $n_y=0$; $n_x=0$, $n_y=2$; and $n_x=1$, $n_y=1$, which are degenerate with respect to each other at infinity, at a smaller interatomic distance at which they are overlapping each other, but where they do not overlap any other band. The two states $n_x=2$, $n_y=0$, and $n_x=0$, $n_y=2$, have surfaces intersecting along the 45° diagonal, for the same symmetry reason found in the case of the p functions of Fig. 16. The remaining state, however, has an entirely different energy surface, which cuts the other two, but along lines in the momentum space which have no simple geometrical significance, and which vary with internuclear distance. Here as before when we investigate the nondiagonal matrix components of perturbative energy these components do not vanish for the type of perturbation found in a real crystal; in this case, unlike the preceding one, there are not even any special points at which symmetry demands that they vanish. Thus really the perturbed energy surfaces of all these three states will not cut each other, and we shall have three perturbed surfaces, one resembling the upper sheet of the surfaces on Fig. 17, for each g_x, g_y value, one resembling the middle sheet, one the lower sheet. Since the intersections are now partly along curves in momentum space with no simple geometrical description we see how the real energy surfaces in the perturbed problem can have great complication, even in this simple problem. We therefore pass on, in the next section, to a consideration of the effect of various perturbations on the two-dimensional Mathieu problem.

6. THE PERTURBED MATHIEU PROBLEM IN TWO DIMENSIONS

The type of perturbative potential which is consistent with a general cubic two-dimensional structure is a sum of terms like $W_{m_1m_2}\exp(\pm 2im_1w_x\pm 2im_2w_y)$. In the general case where m_1 and m_2 are different from each other, there are four such terms, and four others with m_1 and m_2 interchanged, and the combination possessing cubic symmetry is a sum of these eight terms. Then by methods like those used in deriving Eqs. (41) and (42) we find for the nondiagonal matrix component of this perturbative potential the quantity

$$\begin{aligned}W_{m_1m_2}\{&\textstyle\sum(k_x)v_{n_x}{}^*(g_{0x}+2k_x)[v_{n_x}{}'(g_{0x}+2k_x+2m_1)\\&+v_{n_x}{}'(g_{0x}+2k_x-2m_1)]\\&\times\textstyle\sum(k_y)v_{n_y}{}^*(g_{0y}+2k_y)[v_{n_y}{}'(g_{0y}+2k_y+2m_2)\\&+v_{n_y}{}'(g_{0y}+2k_y-2m_2)]\\&+\textstyle\sum(k_x)v_{n_x}{}^*(g_{0x}+2k_x)[v_{n_x}{}'(g_{0x}+2k_x+2m_2)\\&+v_{n_x}{}'(g_{0x}+2k_x-2m_2)]\\&\times\textstyle\sum(k_y)v_{n_y}{}^*(g_{0y}+2k_y)[v_{n_y}{}'(g_{0y}+2k_y+2m_1)\\&+v_{n_y}{}'(g_{0y}+2k_y-2m_1)]\}.\end{aligned}\quad(50)$$

This is the component between an initial wave function characterized by the momentum eigenfunction $v_{n_x}(g_x)$ along the x direction, $v_{n_y}(g_y)$ along y, and a final wave function with momentum eigenfunctions $v_{n_x}{}'(g_x)$, $v_{n_y}{}'(g_y)$. In other words, this nondiagonal matrix component can be built up out of the nondiagonal matrix components of the one-dimensional problem, which we have already discussed in Sec. 4. Let us then use the results of that section to discuss the effects to be expected here.

The general result of Sec. 4, and in particular of Table III, was that the matrix components of the type given in Eq. (43), out of which the expression (50) is built up, are different from zero for the actual wave functions for all values of the m's, and for all g_0's except $g_0=0$ and ± 1, corresponding to the edges of the bands in the one-dimensional case. At these special points, the nondiagonal matrix component of energy between a state with an even momentum eigenfunction and one with an odd eigenfunction must be zero; no such situation holds between two states with even eigenfunctions, or two states with odd eigenfunctions. We can then conclude that the type of perturbation potential of the sort considered in Eq. (50), provided both m_1 and m_2 are different from zero, will produce nonvanishing, nondiagonal matrix components between any two unperturbed wave functions, except in some cases on the lines $g_{0x}=0$, ± 1, and $g_{0y}=0$, ± 1 in momentum space. In a real crystal of two-dimensional cubic symmetry we may expect all components of the type $W_{m_1m_2}$ to be different from zero; thus certainly there will be perturbations which will cause intersecting energy surfaces derived from the Mathieu problem, of the type discussed in Sec. 5, to interact with each other as discussed in that section. If they intersect along the 45° diagonals, as discussed there, they will interact everywhere except at the center of the central zone in momentum space, $g_{0x}=g_{0y}=0$, and at the corners of the central cell, $g_{0x}=\pm 1$, $g_{0y}=\pm 1$. At those points, in case the momentum eigenfunctions have suitable symmetry, the nondiagonal matrix component of energy is zero, and the two energy surfaces will not repel each other at these points. There will, of course, be diagonal matrix components of perturbative energy as well as nondiagonal components, just as in Sec. 5, and they can also be found from Eq. (50), by setting $n_x'=n_x$, $n_y'=n_y$. We note,

however, that if we are considering the interaction of two states which intersect along the 45° diagonal on account of symmetry, such as the states $n_x=1$, $n_y=0$, and $n_x=0$, $n_y=1$, the diagonal components of energy of these two states, as computed by Eq. (50), will be equal by symmetry at the points on the 45° diagonal. Thus neither the diagonal nor the nondiagonal matrix component of energy will cause these states to separate from each other at the points $g_{0x}=g_{0y}=0$ or $g_{0x}=\pm 1$, $g_{0y}=\pm 1$, so that the energy surfaces of the two-dimensional cubic problem will show a degeneracy at these special points in the momentum space. This is the sort of result concerning the adhering of energy surfaces at special points in the momentum space which can also be proved by group theory.[14]

The type of perturbation which we have just discussed, which is present in a real cubic field but is absent in a separable problem like the Mathieu problem, produces profound changes in the energy surfaces, wave functions, and Wannier functions. Let us take the p-like functions as an example, those with $n_x=1$, $n_y=0$ and $n_x=0$, $n_y=1$. For our illustration, we shall choose a fairly large interatomic distance so that the band is only slightly broadened and the wave functions are almost atomic. We have already shown the energy surfaces, in Fig. 16. Let us now show the energy by a different graphical method, plotting energy contours in a two-dimensional momentum space. Such a plot, for a relatively small value of nondiagonal perturbative energy, is shown in Fig. 18. Here (a) shows the lower band, (b) the upper band. The method of calculating the energy contours is very simple: we merely use cosine-type functions for expressing the diagonal and nondiagonal matrix components of energy for the unperturbed functions and solve the quadratic secular equation between the two unperturbed states. Specifically, we take the diagonal energies to be $a\cos\pi g_x - b\cos\pi g_y$ and $a\cos\pi g_y - b\cos\pi g_x$ for the $n_x=1$, $n_y=0$ and $n_x=0$, $n_y=1$ states, respectively, and take the nondiagonal component to be $c\sin\pi g_x \sin\pi g_y$, where a, b, c are constants; these are the simplest functional forms which depend in the right way on the momentum. The general result is that the energy of the lower band is approximately equal to that of one of the unperturbed wave functions in half the unit cell in momentum space, enclosed between the 45° diagonals, but is approximately equal to that of the other in the other half of the unit cell, changing rapidly from the one to the other near the 45° diagonal. As a result of this, the analytic expression for energy as a function of momentum takes on a complicated form, expressible in terms of square roots, near the 45° diagonal, and hence at the center and corners of the unit cell. There are singularities in the energy at these points, of such a nature that the energy cannot be expanded in power series in the momentum.

[14] Bouckaert, Smoluchowski, and Wigner, Phys. Rev. **50**, 58 (1936).

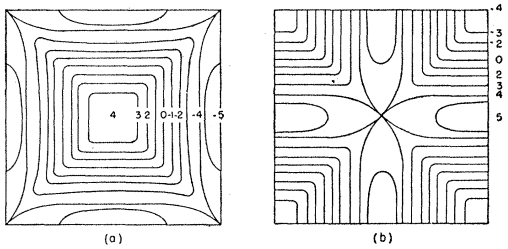

FIG. 18. Energy contours illustrating energy surfaces arising from p-like states, $n_x=1$, $n_y=0$, and $n_y=0$, $n_y=1$, for two-dimensional Mathieu problem. Energy contours are indicated in the unit cells in momentum space. Numbers on contours indicate energies, in arbitrary units, so that energy of 5 units is the highest in the composite energy band, -5 the lowest. Case (a), lower band; (b), upper band. These surfaces arise by letting a horizontal plane cut the surfaces of Fig. 16(a), modified by perturbations.

The maxima and minima of energy in the composite band, however, do not come at these points, but rather at the points $g_{0x}=0$, $g_{0y}=\pm 1$ and $g_{0x}=\pm 1$, $g_{0y}=0$; at these points the energy is a regular function of the momentum, though it is clear from Fig. 18 that it is a highly anisotropic function, varying much more strongly with g_y than with g_x, or vice versa. It is interesting that this rather complicated type of energy surface results from the very simplest case of overlapping energy bands which we can set up.

Now let us consider the momentum eigenfunctions and Wannier functions arising from these perturbed p-like states. We take first the case where the perturbation is nonvanishing, but of negligible magnitude. In Fig. 19 we show the unit cell in momentum space, divided into four regions, numbered 1, 2, 3, 4, by the diagonals. Then the wave function of the lowest band will be almost exactly that of the state $n_x=1$, $n_y=0$ through the regions 1 and 3 of the unit cell, and that of $n_x=0$, $n_y=1$ in regions 2 and 4. Closer examination shows that for the perturbed function to be continuous as we go across the diagonals, the wave function of the lowest band should actually be that of the state $n_x=1$, $n_y=0$ in region 1, but its negative in region 3; and similarly it should be that of $n_x=0$, $n_y=1$ in region 2, but its negative in region 4. We also find that to have the resulting Wannier function real, these values must be multiplied by i. In other words, the momentum eigenfunction is equal to that of the state $n_x=1$, $n_y=0$, multiplied by a function $f(g_x, g_y)$ which is i in region 1, 0 in region 2, $-i$ in region 3, 0 in region 4; plus that of the state $n_x=0$, $n_y=1$, multiplied by the same function $f(g_y, g_x)$. This function f can be expanded in Fourier series. It is easily shown that it is

$$f(g_x, g_y)=\sum(n, m)c_{nm}\exp[\pi i(ng_x+mg_y)], \quad (51)$$

where the quantities c_{nm} are given in Table VI.

In terms of this function, we can now set up the Wannier functions for the two bands. Let $v_0(g)$ be the momentum eigenfunction of the state $n=0$ of the one-dimensional problem, and $v_1(g)$ for the state $n=1$; $v_0(g)$ is a real even function of g, $v_1(g)$ an imaginary odd

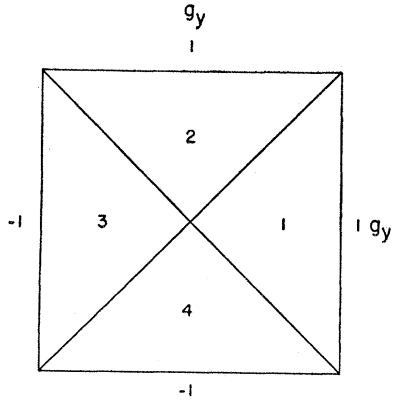

FIG. 19. Unit cell in momentum space, subdivided into regions 1, 2, 3, 4 in which momentum eigenfunction of p states has different behavior.

function. The one-dimensional Wannier function will be given in terms of the v's by Eq. (39), where we note that $a_0(w)$ will be a real even function of w, $a_1(w)$ a real odd function. Then if we set up the two-dimensional analog of Eq. (39) and determine the Wannier function of our lower p-band, we find that by methods like those used in deriving Eq. (48) it is

$$a(w_x, w_y)=\sum(n, m)[c_{nm}a_1(w_x+n\pi)a_0(w_y+m\pi) + c_{mn}a_0(w_x+n\pi)a_1(w_y+m\pi)], \quad (52)$$

while in a similar way for the upper band we have

$$a(w_x, w_y)=\sum(n, m)[c_{nm}a_0(w_x+n\pi)a_1(w_y+m\pi) - c_{mn}a_1(w_x+n\pi)a_0(w_y+m\pi)]. \quad (53)$$

We can now describe these results, both for the momentum eigenfunctions and the Wannier functions, in words. In the lower band, if we take the propagation vector in the central cell of momentum space, the wave function is made up largely of that type of p atomic function which points as closely as possible along the propagation vector, with its nodal plane as closely as possible perpendicular to that vector. In the upper band, we use the p function in each case which points as nearly as possible perpendicular to the propagation vector. The Wannier function, for the lower band, has no contribution at all on the atom for which it is being computed. It has contributions, however, on all atoms displaced from the central atom along the plus or minus x or y axes, or along the 45° diagonals. The contributions in each case fall off in magnitude inversely proportional to the distance from the central atom. For the atoms along the x direction, this contribution is like a Wannier function $a_1(w_x)a_0(w_y)$, a function with p_x-like symmetry; for those along the $-x$ direction, it is like $-a_1(w_x)a_0(w_y)$, the same thing but pointing the other way; along the y axis, it is like $a_1(w_y)a_0(w_x)$, a p_y-like function; and so on. Along the 45° diagonal, $w_x=w_y$, we have the combination $a_1(w_x)a_0(w_y)+a_0(w_x)a_1(w_y)$, which resembles a p-like function pointing along the same 45° direction; and so on. In other words, starting with original Wannier functions which have the symmetry properties of a vector, pointing along the x or y directions, we have constructed a function having a square symmetry, consisting of functions pointing out (or in) in a radial direction. We can indicate this schematically as in Fig. 20(a), where we represent a p-like Wannier function in a neighboring atom by a vector, pointing along the axis of the p-function. We indicate the amplitudes of the functions by the lengths of the vectors. We show in Fig. 20(a) the lower band, which we have described above, and in Fig. 20(b) the upper band.

As the strength of the perturbation is increased, and we approach the type of potential which we should find in a real cubic two-dimensional crystal, the situation just described changes quantitatively, but not qualitatively. The energy surfaces will still resemble those of Fig. 16 and Fig. 18, but more smoothed in the neighborhood of the 45° diagonals. The momentum eigenfunctions will still resemble the previous case qualitatively, but will simply change rapidly, rather than discontinuously, across the 45° diagonal. This will leave Wannier functions of the general type of Fig. 20, but falling off somewhat more rapidly as we go away from the central atom.

To set up these resulting Wannier functions, we may proceed along the lines already described. Our treatment is a little neater, however, if we prove an important general theorem about the Wannier functions of perturbed energy bands. Let us suppose that we have a set of unperturbed wave functions $u_{n,p}{}^0$, which for instance might be the functions determined from the unperturbed Mathieu problem. From these unperturbed wave functions we construct unperturbed Wannier functions $a_n{}^0(\mathbf{r}-\mathbf{Q}_i)$ by Eq. (21). Now we introduce a perturbation having the periodicity of the lattice; this might for instance be the perturbative potential having a nondiagonal matrix component like that given in Eq. (50). We assume that we solve the perturbation problem resulting from this potential and that the resulting wave functions $u_{n,p}$ are related to the unperturbed functions by the equations

$$u_{n,p}=\sum(m)s_{mn}(\mathbf{p})u_{m,p}{}^0, \quad (54)$$

where on account of the periodicity of the perturbing potential our sum involves only unperturbed wave

TABLE VI. $2\pi c_{mn}$, from Eq. (51).

m \ n	−4	−3	−2	−1	0	1	2	3	4
4	−1/4	0	0	0	0	0	0	0	1/4
3	0	−1/3	0	0	0	0	0	1/3	0
2	0	0	−1/2	0	0	0	1/2	0	0
1	0	0	0	−1	0	1	0	0	0
0	2/4	−2/3	2/2	−2	0	2	−2/2	2/3	−2/4
−1	0	0	0	−1	0	1	0	0	0
−2	0	0	−1/2	0	0	0	1/2	0	0
−3	0	−1/3	0	0	0	0	0	1/3	0
−4	−1/4	0	0	0	0	0	0	0	1/4

Note: The c_{nm}'s continue for greater n and m, in an obvious manner.

functions of the same p value. Now we wish to compute the perturbed Wannier functions. As a first step, we note that the s_{mn}'s are periodic functions of $\mathbf{p}$ in the reciprocal space, and hence can be expanded in the form

$$s_{mn}(\mathbf{p})=\sum(\mathbf{Q}_k)S_{mn}(\mathbf{Q}_k)\exp[(-2\pi i/h)(\mathbf{p}\cdot\mathbf{Q}_k)]. \quad (55)$$

When we insert (55) in (54), and use Eq. (21) to find the correct Wannier functions, we find at once

$$a_n(\mathbf{r}-\mathbf{Q}_j)=\sum(m,\ \mathbf{Q}_k)S_{mn}(\mathbf{Q}_k)a_m{}^0(\mathbf{r}-\mathbf{Q}_j-\mathbf{Q}_k). \quad (56)$$

This shows that the situation which we have encountered a number of times in the present work, by which the Wannier function of a perturbed problem is a sum of unperturbed Wannier functions located on different atoms, is a perfectly general situation; it also shows us directly how to find the coefficients of the expansion in series of unperturbed Wannier functions. The expansions (52) and (54) were special cases of this general theorem, though the notation was a little different. Unfortunately, in most actual cases the Fourier expansion (55) is too difficult to carry out analytically, as we were able to do in our special case in Table VI; but this does not diminish the validity of the general theorem (56), and the insight which it gives into the nature of the Wannier functions of overlapping bands.

The general nature of the problem of the d-like levels, and of the s-like band degenerate with them, as shown in Fig. 17, is not different in principle from the p bands which we have just considered. The interaction of the two states with $n_x=2$, $n_y=0$ and $n_x=0$, $n_y=2$ is very much like the problem of the p bands, except that symmetry no longer requires that the nondiagonal matrix component of a general cubic perturbation should vanish at the center and corners of the unit cell; it will be nonvanishing everywhere, so that the perturbed energy surfaces will not adhere to each other. Here in addition, however, as we pointed out in the preceding section, the band with $n_x=1$, $n_y=1$ overlaps these. We see from Eq. (50) that there will be nondiagonal matrix components of energy between these two states everywhere where they intersect each other, so that the perturbed energy surfaces will not intersect. The problem of the momentum eigenfunction and Wannier function are more difficult to handle here than in the earlier case, however, on account of the fact that the regions of momentum space where the different unperturbed wave functions are approximately correct are no longer bounded by lines as simple as the 45° diagonals. It would thus be a complicated procedure to find the Wannier functions; but the result would still be qualitatively much like that of Fig. 20, the Wannier function consisting of the Wannier functions of the unperturbed problems, suitably combined on atomic sites distributed at considerable distances from the place where the Wannier function is centered, with amplitudes falling off as we depart from that point. In all these cases, there is one general and interesting result to be noted: after we have applied the perturbations, the energy surfaces, momentum eigenfunctions, and Wannier functions all show a fourfold symmetry; rotation through 90° brings all these functions back to their original values. This is not true in general of the unperturbed solutions of the Mathieu problem.

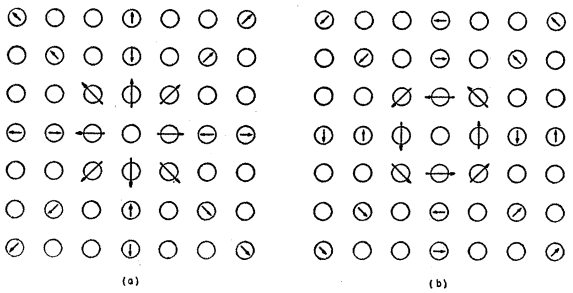

FIG. 20. Wannier function for lower (a) and upper (b) bands arising from degeneracy of two p states in two-dimensional Mathieu problem with small perturbations. The circles indicate atoms; the Wannier function is being computed for the central atom in each figure. They consist of p-like unperturbed Wannier 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.

There is another type of perturbative potential which is of interest in the two-dimensional case, in addition to those already considered. This is a term which would convert the problem from a two-dimensional analog to the simple cubic crystal, to the two-dimensional analog to the body-centered or face-centered cubic. In Fig. 21, we show the positions of the atoms in a two-dimensional square lattice. If we lowered the potential minima at the atoms surrounded by circles in Fig. 21, raised those not so surrounded, then the atoms of either type would form a two-dimensional face-centered structure, with twice the original lattice spacing in each direction. We can accomplish this by superposing a perturbative potential of the form made up by adding the four terms $W\exp(\pm iw_x\pm iw_y)$. This is a problem similar to the one-dimensional one considered in Sec. 4, where we took up the potential $W\cos w$. We shall not really achieve our face-centered structure unless the perturbative potential is great enough in magnitude so that it entirely splits the energy bands, as in Fig. 11(a), so that some bands consist of wave functions entirely concentrated around the atoms surrounded by circles in Fig. 21, other bands of wave functions concentrated around the remaining atoms, so that bands of either type can be considered as belonging to two interpenetrating face-centered lattices of different sorts of atoms. Let us consider what will happen to the energy surfaces, momentum eigenfunctions, and Wannier functions when such a perturbation is applied.

When we investigate the nondiagonal matrix components of energy arising from this perturbation, we find that we have such components between an original state characterized by values g_{0x}, g_{0y}, and states with $g_{0x}'=g_{0x}\pm1$, $g_{0y}'=g_{0y}\pm1$; the matrix component of this

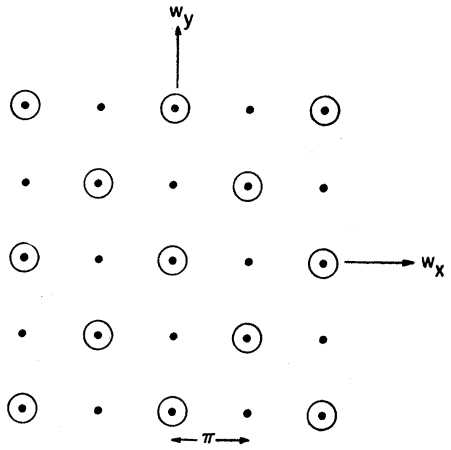

FIG. 21. Position of atoms in a two-dimensional square lattice, with modified potential lowering potential minima at those atoms indicated by circles, so that atoms of latter type lie on a two-dimensional analog to a face-centered cubic three-dimensional structure.

potential has the form

$$\sum(k_x)v_{n_x}^*(g_{0x}+2k_x)\times[v_{n_x}'(g_{0x}+2k_x+1)+v_{n_x}'(g_{0x}+2k_x-1)],$$
$$\sum(k_y)v_{n_y}^*(g_{0y}+2k_y)\times[v_{n_y}'(g_{0y}+2k_y+1)+v_{n_y}'(g_{0y}+2k_y-1)]. \quad (57)$$

We start by plotting the energy of the original simple cubic problem as a function of g_{0x}, g_{0y}. This energy is periodic, with a square unit cell which we may choose as extending from $g_{0x}=-1$ to 1, $g_{0y}=-1$ to 1. If we already have introduced the type of perturbation leading to a general simple cubic potential, as discussed earlier in this section, there will be a fourfold symmetry of energy in the square unit cell; it may be, for instance, that there will be an energy minimum at the center of the cell, maximum in the corners. Now let us repeat the same energy surface shifted along each axis by unit distance, so that the positions that were at the centers of the original unit cells will lie at the corners, and vice versa. Then we shall find that the two states represented by points vertically above each other on the two resulting sheets will have nondiagonal matrix components of energy between them, similar to the case of Fig. 10. As in that figure, the perturbations will have a maximum effect where the two energy surfaces intersect each other; and by symmetry, the intersection will occur along the diagonal lines drawn in Fig. 22(a). The nondiagonal matrix component of energy will be different from zero at each point of these lines, as we see from Eq. (57). Thus the two surfaces will split apart, resulting in an upper and lower sheet which nowhere touch each other. Furthermore, it is clear that the pattern found inside the shaded tilted square of Fig. 22(a) will repeat itself outside the square. This square, in other words, is a suitable unit cell in momentum space, for this perturbed problem; it is, in fact, precisely the central Brillouin zone, as defined in the usual way, for this face-centered lattice. If we ask what is the unit cell in coordinate space for which this is the unit cell in momentum space, we find the tilted square shown in Fig. 22(b), which is precisely the Wigner-Seitz cell for this problem. We note that this unit cell contains two of the original atoms per unit cell, one at the center, another at the corner (each of the four corners counts $\frac{1}{4}$ in this unit cell). The tilted unit cells of Fig. 22 are the natural ones to use for this problem, rather than those shown by dotted lines; the dotted unit cell in coordinate space, in Fig. 22(b), contains twice as many atoms as the Wigner-Seitz cell, that in Fig. 22(a) being correspondingly half as large as the Brillouin zone, and if we use them we run into complications regarding the momentum eigenfunction and Wannier function which do not correspond to any physical reality, but arise merely from using a unit cell twice as large as necessary.

We may now ask, following closely the model of Sec. 4, what is the effect of this perturbation on the momentum eigenfunction and the Wannier function. For a small perturbation, such as we should have if the two types of sites in Fig. 21 corresponded to only very slightly different potential wells, we must introduce the new unit cells in momentum space, corresponding to Fig. 22(a), but without making appreciable changes in the real wave functions. A given wave function will have a momentum eigenfunction which has a component in each of the square unit cells shown in Fig. 22(a); for instance, there may be a lattice of points consisting of g_0, as shown in Fig. 22(a), and corresponding points in each square unit cell. As in Sec. 4 and in Fig. 12, we may call the function made up by letting g_0 range through the square unit cell $v_0(g)$. But with our perturbed problem, we wish to use the shaded unit cell in Fig. 22(a) half as large as the original square one. We shall then have twice as many points in our momentum lattice as before; thus the point g_0' in Fig. 22(a) will correspond to g_0. But the real unperturbed wave function, which has a nonvanishing component $v_0(g_0)$, will have to be zero at the point g_0'. In other words, the $v(g)$ for this function, using our new unit cells, will equal $v_0(g)$ within the shaded part of the square unit cell in Fig. 22(a), but will be zero in the unshaded region. Thus its momentum eigenfunction will be $v_0(g)f_1(g)$, where $f_1(g)$ is unity in the shaded region of Fig. 22(a), zero in the unshaded region. Similarly for the upper band, the momentum eigenfunction will be $v_0(g)f_2(g)$, where $f_2(g)$ is zero where $f_1(g)$ is unity, and vice versa. We can now expand these functions $f_1(g)$ and $f_2(g)$ in Fourier series, as in Eqs. (46) and (47). We find that

$$f_1(g_x, g_y)=\sum(n, m)c_{nm}\exp[\pi i(ng_x+mg_y)],$$

where

$$c_{nm}=-\pi^{-2}[(-1)^n-(-1)^m]/(n^2-m^2) \quad \text{if} \quad n\neq m,$$
$$c_{nn}=0 \quad \text{if} \quad n\neq 0; \quad c_{00}=\tfrac{1}{2}. \quad (58)$$

For $f_2(g_x, g_y)$, the components are the negatives of the c_{nm}'s above, except c_{00}, which equals $\frac{1}{2}$. Thus when we set up the Wannier functions, corresponding to small perturbations, following the model of Eqs. (52) and (53), we find components in many cells of coordinate space, falling off slowly with the distance.

The case which we have considered is that of a vanishing perturbation. We have already noted, however, that we are interested in the case of a large perturbation; this is the case where the sites of the two sorts are entirely different, so that a wave function can be concentrated around sites of one type or the other. In such a case, as in Sec. 4, the correct wave function will consist approximately of the sums or differences of the two types of functions, and hence the Wannier functions will be the sum or difference of those indicated above. The sums correspond to the wave functions concentrated about the atoms at the centers of the tilted unit cells shown in Fig. 22(b), the differences to atoms at the corners of the cell. And now we see from Eq. (58) that the sums, but not the differences, of the Wannier functions are equivalent to Wannier functions of the unperturbed problem located at the origin. The differences, on the contrary, are distributed through many unit cells.

7. THE THREE-DIMENSIONAL MATHIEU PROBLEM

The three-dimensional case resembles very strongly the two-dimensional one, which we have already treated in Secs. 5 and 6; for that reason our treatment of it will be very brief. The energy bands as functions of lattice spacing are very similar to those of Fig. 15. We can describe each energy band at large distances in terms of the quantum numbers of the spherically symmetrical field around a potential well, as in Table V. Thus the lowest level is an s-like level, with energy $\epsilon=3$. Next come three degenerate p-like levels, similar to the ordinary functions p_x, p_y, p_z, with $\epsilon=5$. Next, with $\epsilon=7$, there are six degenerate levels, coming from the quantum numbers n_x, n_y, n_z equal to the combinations (200), (020), (002), (011), (101), (110). From the first three, we can make linear combinations to form an s-like and two d-like functions, having symmetry like the functions $(x^2-y^2)f(r)$, and $(2z^2-x^2-y^2)f(r)$; the last three are d-like functions as they stand, with wave functions like $yzf(r)$, $zxf(r)$, $xyf(r)$. As the interatomic distance decreases, the crystalline field separates the latter three d states from the former two; but the degeneracy or hybridization between these two and the s-like state persists as long as we have the unperturbed Mathieu problem. As in the two-dimensional case, the really interesting problems arise from the degeneracies between the p and d bands, and the effect of perturbations on these degeneracies. The situation is so similar to that in the two-dimensional case that we can state the results in many cases in a qualitative form.

First, as in Sec. 6, let us consider those perturbations

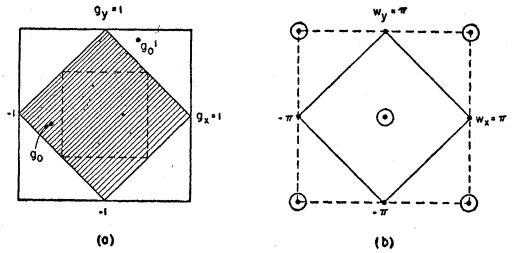

FIG. 22. Unit cells in momentum space (a) and coordinate space (b), for problem of square lattice perturbed to convert it to face-centered lattice. In (a), large square is original unit cell of unperturbed problem; shaded tilted square is central Brillouin zone for face-centered structure, corresponding to tilted square in (b), which is Wigner-Seitz cell; small square in (a) corresponds to large square in (b), square unit cell two atomic distances on a side.

arising from a general simple cubic potential, but absent in the separable Mathieu problem. We consider first the p-like energy bands. We can no longer plot energy as a function of position in the unit cell of momentum space. We can, however, investigate the planes over which the energies of the unperturbed bands, formed from the functions of the type p_x, p_y, and p_z, will become equal to each other. Let us write things down analytically. We may let $\epsilon_0(g)$ be the energy of the one-dimensional problem for the first band, for $n=0$, as a function of g; $\epsilon_1(g)$ for the second band; and so on. Then the energy of the p_x function is $\epsilon_1(g_x)+\epsilon_0(g_y)+\epsilon_0(g_z)$; of the p_y, $\epsilon_0(g_x)+\epsilon_1(g_y)+\epsilon_0(g_z)$; and similarly for p_z. The function $\epsilon_0(g)$ has a minimum at $g=0$, maximum at $g=\pm 1$; $\epsilon_1(g)$ on the contrary has a maximum at $g=0$, minimum at $g=\pm 1$, and it varies much more with g than does $\epsilon_0(g)$. Clearly the energies of the p_x and p_y functions will be equal when $g_x=\pm g_y$ (the $\pm$ arising because the ϵ's are even functions of g), and so on. We then may subdivide the unit cube in momentum space (extending from $g_x=-1$ to 1, g_y from -1 to 1, etc.) by the planes $g_x=\pm g_y$, $g_y=\pm g_z$, $g_z=\pm g_x$. These planes are shown in Fig. 23. We may now examine the various segments of the cube, and ask which order the various energy surfaces lie in, in each segment. On a plane where two of the bands have the same energy, we may expect a general cubic perturbation to result in the bands separating from each other.

There are 24 equivalent prisms making up the cube, four extending out to each of the six faces of the cube. Let us consider the prism indicated by heavy lines in Fig. 23, bounded by the plane $g_x=1$, the planes $g_y=\pm g_z$, and the plane $g_x=g_y$, with $|g_x|>|g_y|>|g_z|$. Within this prism, the energy $\epsilon_1(g_x)+\epsilon_0(g_y)+\epsilon_0(g_z)$ will everywhere lie lowest; that $\epsilon_0(g_x)+\epsilon_1(g_y)+\epsilon_0(g_z)$ will be next; and $\epsilon_0(g_x)+\epsilon_0(g_y)+\epsilon_1(g_z)$ will lie highest. In Fig. 24(a), we show the way the lowest energy surfaces fit together in the various prisms, on the assumption that we have introduced slight perturbations of the cubic variety, to make the problem nondegenerate. For the lowest level, in other words, we have roughly cubical energy surfaces, rounded on the edges, and slightly concave. The

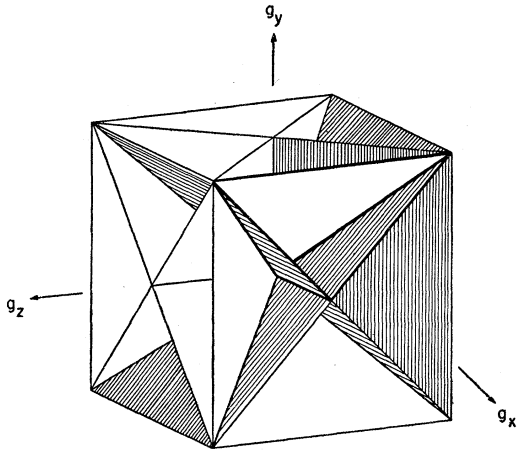

FIG. 23. Unit cube in momentum space, with planes $g_x=\pm g_y$, $g_y=\pm g_z$, $g_z=\pm g_x$, on which the various p-like wave functions are degenerate with each other in the three-dimensional Mathieu problem.

highest energy corresponds to the energy surface at the center of the cell, the lowest to the surface about the boundary. In Fig. 24(b) we show similarly the form of the energy surface for the middle energy band, and in Fig. 24(c) for the highest band; in each case the lowest energies correspond to the surfaces nearest the boundaries or corners of the cube, the highest to those nearest the center. The resemblance of these figures to the two-dimensional case of Fig. 18 is obvious. Here again, as in the two-dimensional case, we can show that the three sheets of the energy surface adhere at the center and the corners of the cube forming the unit cell in momentum space.

For the d-like states, with the s state degenerate with them, the problem as in the two-dimensional case is much more complicated. The three unperturbed states corresponding to n_x, n_y, n_z equal to (200), (020), (002), respectively, will have three surfaces which will intersect, with each other, very much like the three p states, which we have discussed in Figs. 23 and 24, resulting in similar perturbed energy surfaces. Similarly the three states (011), (101), (110) will have a similar perturbation problem, with similar energy surfaces. But now we must consider the perturbations between these two sets of energy surfaces. The one set will be separated from the other by the effect of the crystalline field, and for large interatomic distance, the two sets of energy bands will not overlap. As the distance decreases, however, they will cut, and obviously the intersections of the surfaces will not be given by any simple considerations. As in the two-dimensional case, a general cubic perturbative potential will separate the surfaces at these intersections, so that the final perturbed surfaces will not cut each other, but the resulting forms will be very complicated.

We can next, as in two dimensions, introduce further perturbations designed to convert the problem from simple cubic to face-centered or body-centered cubic. The method of handling it is essentially just as in two dimensions. We set up a perturbative potential of the proper sort to depress those cubic lattice points forming the face-centered or body-centered cubic structure with respect to the remaining lattice points. We then find that we can carry out a construction, as in the one- and two-dimensional cases, by which we superpose energy surfaces with suitable displacements of ± 1 along g_x, g_y, and g_z, such that there are perturbations between the different sheets of the energy surface at the same points of momentum space. These perturbations will cause the energy surfaces to be pushed away from each other in the neighborhood of the intersections of these energy surfaces. But now we find that the planes on which the surfaces intersect, by symmetry, are precisely those planes bounding the central Brillouin zone for the appropriate type of structure. Thus we arrive at these Brillouin zones from the perturbation of the Mathieu problem, rather than by perturbation of a plane wave solution as is ordinarily done. Within the central Brillouin zone, the energy surfaces of the three p states, or of the six $d-s$ states, will not be unlike their values in the corresponding part of the unit cell of the simple cubic momentum space; but the energy will always have a vanishing normal derivative at the surface of the Brillouin zone.

As in one and two dimensions, we can also set up momentum eigenfunctions and Wannier functions for these more complicated three-dimensional cases; and as before, we shall find Wannier functions which generally are rather broadly extending out from one atom to another. Also as in two dimensions, the final Wannier functions will generally show the rotational symmetry of the lattice, as well the energy bands. It does not seem worth while, at present, to go more in detail into the nature of these functions. By the time we have introduced all the various perturbations which we have considered, we have departed so widely from the original Mathieu case that we really have a new problem, which could appropriately be considered separately, in a study of the actual behavior of the energy bands, momentum eigenfunctions, and Wannier functions of p and d bands in real crystals. It is hoped that this discussion of the Mathieu case, however, will be useful in pointing the way toward this more complicated future investigation.

APPENDIX 1

Let $a(w)$ be the original Wannier function of the unperturbed band. We have found that for the per-

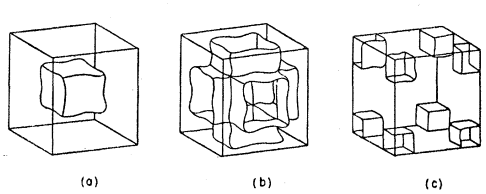

FIG. 24. Surfaces of constant energy, in momentum space, for (a) lowest band, (b) middle band, (c) highest band, resulting from removal of degeneracy between the overlapping p bands, in the three-dimensional Mathieu problem.

turbed upper half-band the Wannier function located at the atom at position $2n\pi$ may be written

$$a'(w-2n\pi)=(2/\pi)[a(w-2n\pi-\pi)+a(w-2n\pi+\pi)]$$
$$-(2/3\pi)[a(w-2n\pi-3\pi)+a(w-2n\pi+3\pi)]\cdots.$$

We now wish to consider the sum $\sum(n)\exp(2\pi ing)\times a'(w-2n\pi)$, found by locating these Wannier functions on the atoms at positions $2n\pi$. When we insert $a'(w-2n\pi)$, as defined above, in this sum, we may rearrange the resulting double sum, so that it may be written

$$\sum(n)\exp[2\pi i(n+\tfrac{1}{2})g]a[w-2(n+\tfrac{1}{2})\pi]$$
$$\times[(2/\pi)(\exp\pi ig+\exp-\pi ig)$$
$$-(2/3\pi)(\exp 3\pi ig+\exp-3\pi ig)\cdots].$$

From Eqs. (46) and (47), however, we know that the function $(2/\pi)(\exp\pi ig+\exp-\pi ig)\cdots$ equals unity when g is between $-\frac{1}{2}$ and $\frac{1}{2}$, which includes the whole first unit cell of our lattice of double periodicity. Thus the wave function is

$$\sum(n)\exp[2\pi i(n+\tfrac{1}{2})g]a[w-2(n+\tfrac{1}{2})\pi],$$

just as if we had the original Wannier functions located at the points $\pm\pi$, $\pm 3\pi$, etc. Thus we verify the statement made in the text and show a simple example of a case where the identical wave functions can be expressed in terms of two types of Wannier functions, one, $a[w-2(n+\frac{1}{2})\pi]$, concentrated on a given atom, the other, $a'(w-2n\pi)$, extended over many atoms.

PHYSICAL REVIEW VOLUME 87, NUMBER 5 SEPTEMBER 1, 1952

Statistics of the Recombinations of Holes and Electrons

W. SHOCKLEY AND W. T. READ, JR.
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received April 29, 1952)

The statistics of the recombination of holes and electrons in semiconductors is analyzed on the basis of a model in which the recombination occurs through the mechanism of trapping. A trap is assumed to have an energy level in the energy gap so that its charge may have either of two values differing by one electronic charge. The dependence of lifetime of injected carriers upon initial conductivity and upon injected carrier density is discussed.

SECTION 1. INTRODUCTION

IN connection with studies of transistor physics, the recombination of holes and electrons plays an important role. The lifetime of injected carriers in germanium has been found to be a structure sensitive property of the material. This suggests that the recombination process takes place through the medium of imperfections of some sort in the germanium crystal.[1] It is the purpose of this paper to investigate the mathematical consequences of a particular type of imperfection. We shall accordingly suppose that the crystal contains a density N_t of traps which contribute to the recombination process.

In Fig. 1 we illustrate the way in which holes and electrons may be recombined through the traps. The figure illustrates a trap which may exist in either of two states differing by one electronic unit of charge, being either negative or neutral; similar treatments may be applied to other possibilities, such as neutral or positive, or cases in which the charge changes between -1 and -2 units. If the trap is neutral it may capture an electron from the conduction band. The energy loss of the electron is then converted into heat or light or both depending upon the nature of the trapping process. It may also capture an electron from the valence band represented by part (d) of the figure, in which case it acquires a negative charge and leaves a hole in the valence band. Parts (b) and (c) represent the emission of an electron and the capture of a hole.

The effects which we shall consider in this study arise from the statistics of the processes shown in Fig. 1, and the limitation in rate is assumed to be due to the availability of electrons and holes to enter the traps. We shall neglect another possible limiting factor in the recombination process, the time of readjustment of the electron in the trap once it is trapped: Thus, the electron in part (a) of the figure might be trapped in an excited state in the trap and require some time before falling to the ground state. While the electron was in the excited state, the ability of the trap to emit an electron or to capture a hole would be different from the conditions represented in part (b) and (c) of the figure and there would, therefore, be a time lag before the trap reached its normal state. We shall assume that the readjustment time for a trapped electron is negligible compared to time required on the average for the trap to emit the electron or to capture a hole.

[1] See W. Shockley, *Electrons and Holes in Semiconductors* (D. van Nostrand Company, Inc., New York, 1950), p. 347. The methods of measuring lifetime and its role in transistor electronics are also discussed in this reference. This process of recombination has also been discussed by R. N. Hall, Phys. Rev. **83**, 228 (1951) and **87**, 387 (1952).