Formalisms of Band Theory

E. I. BLOUNT

Bell Telephone Laboratories, Murray Hill, New Jersey

It is well known that in crystalline solids the energy levels are arranged in "bands," and that, strictly speaking, one must take account of the multiplicity of these bands in solving problems in solid state physics. It is equally well known that, nevertheless, one can obtain considerable information, frequently sufficient to interpret experiments, by considering only one band and treating the electrons in this band by methods similar to those applicable to free electrons. Considerable progress has been made in the last decade in understanding this fact, in developing techniques for exploiting it, and in the actual application to problems. This subject has not been reviewed recently, however, and the published accounts stop considerably short of what can actually be done. The purpose of this paper is

to fill this gap at least partially. The various procedures used will be developed, not in the historical order, but from a point of view closely akin to that of Adams. We shall develop the Crystal Momentum Representation (CMR) *ab initio* and relate the other methods to it. These include the Kohn-Luttinger or Modified Crystal Momentum Representation (MCMR) and the Wannier Representation (WR) as well as a new procedure developed by the author.

The procedures discussed in this article have their roots in a paper by Wannier in 1937.[1] Wannier was interested in the problem of the large exciton, which is similar to but more difficult than the usual type of problem to be attacked by the methods we will discuss. He introduced Wannier functions in order to have localized states to describe the system, and later derived an equation for the relative motion of the hole and electron which is very similar to the equations we shall discuss here.

Slater[2] directed attention to this formalism in 1949 and rederived the Hamiltonian and equations of motion for an electron in a perturbed periodic lattice.

Adams[3] further generalized the Wannier representation approach to include some corrections, and later[4] showed that the whole procedure can be carried out exactly and more easily in the CMR. He also initiated the development of a compact operator formalism which greatly facilitates calculations. The present paper follows Adams in regarding the CMR as the logical basis for discussing these problems.

I. Crystal Momentum Representation

1. INTRODUCTION

The basic concept of the crystal-momentum formalism is to use a representation whose basis consists of the energy eigenstates of the electrons in the crystal. When such a representation is used, it matters little what the precise nature of the original Hamiltonian may have been. Thus our results will be valid for any of the following three Hamiltonians:

(a) Schrödinger

$$H = \frac{p^2}{2m} + U$$

[1] G. H. Wannier, *Phys. Rev.* **52,** 191 (1937).
[2] J. C. Slater, *Phys. Rev.* **76,** 1592 (1949).
[3] E. N. Adams, *Phys. Rev.* **85,** 41 (1952).
[4] E. N. Adams, *J. Chem. Phys.* **21,** 2013 (1953).

where m is the mass of the electron and U is the periodic crystalline potential energy.

(b) Pauli (we shall use this term to indicate the Schrödinger Hamiltonian with spin-orbit coupling included)

$$H = \frac{p^2}{2m} + \frac{e^2}{4m^2c^2} \mathbf{p} \cdot \mathbf{d} \times \nabla U + U$$

where $\mathbf{d}$ refers to the Pauli matrices.

(c) Dirac

$$H = c\,\boldsymbol{\alpha} \cdot \mathbf{p} + U$$

where $\boldsymbol{\alpha}$ is a four-component spinor. For each of these Hamiltonians we can define a velocity operator $\mathfrak{B}$ such that

$$\mathfrak{B} = -\frac{i}{\hbar}[\mathbf{x}, H] = \nabla_{\mathbf{p}} H. \tag{1.1}$$

For our three Hamiltonians, $\mathfrak{B}$ takes the forms

(a)
$$\mathfrak{B} = \frac{\mathbf{p}}{m}$$

(b)
$$\mathfrak{B} = \frac{\mathbf{p}}{m} + \frac{e^2}{4m^2c^2} \mathbf{d} \times \nabla U \tag{1.2}$$

(c)
$$\mathfrak{B} = c\,\boldsymbol{\alpha}.$$

For the first two, we have the commutator

$$-\frac{i}{\hbar}[\mathbf{x}, \mathfrak{B}] = \frac{1}{m} \tag{1.3}$$

while for the Dirac equation

$$-\frac{i}{\hbar}[\mathbf{x}, \mathfrak{B}] = 0. \tag{1.3a}$$

It is well known that when U is a periodic potential, the eigenfunctions of these Hamiltonians can be classified by the quantum numbers $(\mathbf{k}, n)$, where $\mathbf{k}$ is a vector lying in the so-called first Brillouin zone (BZ) (or in any unit cell in reciprocal space), and n is a discrete index labeling the so-called bands [see below, Eq. (1.6)]. The eigenfunctions can be written in the form

$$\psi_{n\mathbf{k}}(\mathbf{x}) = \exp(i\mathbf{k} \cdot \mathbf{x}) u_{n\mathbf{k}}(\mathbf{x})$$

where $u_{n\mathbf{k}}$ is a periodic function with the same periodicity as U. That the

eigenfunctions have this form is seen most easily by constructing the matrix element of H for two such functions belonging to different values of k. This is easily shown to vanish.

It now follows that instead of the equation for ψ

$$H\psi_{n\mathbf{k}}(\mathbf{x}) = E_n(\mathbf{k})\psi_{n\mathbf{k}}(\mathbf{x}) \tag{1.4}$$

we can write an equation for u in the form

$$H(\mathbf{k})u_{n\mathbf{k}}(\mathbf{x}) = E_n(\mathbf{k})u_{n\mathbf{k}}(\mathbf{x}) \tag{1.5}$$

where

$$H(\mathbf{k}) \equiv \exp(-i\mathbf{k}\mathbf{x})H \exp(i\mathbf{k}\cdot\mathbf{x}). \tag{1.6}$$

The periodic solutions of Eq. (1.5) form a complete set of periodic functions. Thus the functions $\psi_{n\mathbf{k}}$ and the functions $\psi_{n,\mathbf{k}+\mathbf{K}}$, where $\mathbf{K}$ is a reciprocal lattice vector, span the same space, and only one of the sets should be included in our catalog of eigenfunctions. This is the reason for restricting $\mathbf{k}$ to a unit cell in $\mathbf{k}$ space. For the first two Hamiltonians (a) and (b), $H(\mathbf{k})$ has the simple form

$$H(\mathbf{k}) = H + \hbar\mathbf{k}\cdot\mathfrak{B} + \frac{\hbar^2 k^2}{2m} \tag{1.7}$$

while for the Dirac equation, the last term is omitted.

From Eq. (1.6) we also obtain

$$H(\mathbf{k} + \mathbf{q}) = H(\mathbf{k}) + \hbar\mathbf{q}\cdot\mathfrak{B}(\mathbf{k}) + \frac{\hbar^2 q^2}{2m} \tag{1.8}$$

where

$$\mathfrak{B}(\mathbf{k}) = \exp(-i\mathbf{k}\cdot\mathbf{x})\mathfrak{B} \exp(i\mathbf{k}\cdot\mathbf{x}) = \frac{\partial H(\mathbf{k})}{\partial \mathbf{k}}$$

$$= \mathfrak{B} + \frac{\hbar\mathbf{k}}{m}$$

where again the last term is omitted for the Dirac Hamiltonian. From Eq. (1.8) it is apparent that if the $u_{n\mathbf{k}}$ are known for some $\mathbf{k}$, they can be found for nearby $(\mathbf{k} + \mathbf{q})$ by perturbation theory. The question of the convergence of this expansion is discussed in Appendix C.

2. Representation of Wave Functions and Operators

We have now stated some of the important properties of our Hamiltonians and can proceed to consider the representation of wave functions and some important operators in the CMR.

Any wave function can be written as a superposition of Bloch functions

$$f(\mathbf{x}) = \sum_n \int d^3k f_n(\mathbf{k}) \psi_{n\mathbf{k}}(\mathbf{x}).$$

$f_n(\mathbf{k})$ is then the wave function in the CMR. One important point should be emphasized; we frequently must solve differential equations for $f_n(\mathbf{k})$. Strictly speaking, $\mathbf{k}$ is limited to one Brillouin zone—any points outside it provide no new information. On the other hand, we have the option of continuing $f(\mathbf{k})$ in $\mathbf{k}$ space as we see fit. As one might suppose, the simplest procedure is to impose periodic boundary conditions in $\mathbf{k}$ space, requiring $f_n(\mathbf{k} + \mathbf{K}) = f_n(\mathbf{k})$; also for operators, $O_{nn'}(\mathbf{k} + \mathbf{K}_1, \mathbf{k}' + \mathbf{K}_2) = O_{nn'}(\mathbf{k}, \mathbf{k}')$ where $\mathbf{K}$ is a reciprocal lattice vector. We shall always assume that such conditions have been imposed.

The simplest operator is the crystal momentum itself $\mathbf{p}_c = \hbar\mathbf{k}$, the operator whose matrix elements are

$$\mathbf{p}_{c,nn'}(\mathbf{k}, \mathbf{k}') = \hbar\mathbf{k}\, \delta(\mathbf{k} - \mathbf{k}')\, \delta_{nn'}. \tag{2.1}$$

The true momentum has the form

$$\mathbf{p}_{nn'}(\mathbf{k}, \mathbf{k}') = \delta(\mathbf{k} - \mathbf{k}')\left(\hbar\mathbf{k}\, \delta_{nn'} - i\hbar \int u_n{}^* \frac{\partial u_{n'}}{\partial \mathbf{x}}\, d\tau\right)$$

where $d\tau$ is integrated over a unit cell.

The momentum, however, is less important, in general, than the velocity, which is also diagonal in $\mathbf{k}$, with matrix elements $\mathfrak{V}_{nn'}(\mathbf{k})$. We shall sometimes find it convenient to split $\mathfrak{V}$ into two operators: $\mathbf{v}$ with only intraband matrix elements $\mathbf{v}_n = \mathfrak{V}_{nn}$, and $\mathbf{V}$, with only interband elements $\mathbf{V}_{nn'} = \mathfrak{V}_{nn'}(n \neq n')$.

The representation of $\mathbf{x}$ is somewhat more subtle. Thus if we simply try to evaluate the following integral

$$I_{n'\mathbf{k}'n\mathbf{k}} = \int \psi^*_{n'\mathbf{k}'}\mathbf{x}\psi_{n\mathbf{k}}\, d^3\mathbf{x}$$

we see that it has no well-defined value. For instance, we might try to break the integral up into a sum of integrals over unit cells

$$I_{n'\mathbf{k}',n\mathbf{k}} = \sum_R \exp\left[i(\mathbf{k} - \mathbf{k}')\cdot\mathbf{R}\right] \int \exp\left[i(\mathbf{k} - \mathbf{k}')\cdot(\mathbf{x} - \mathbf{R})\right]u^*_{n'\mathbf{k}'}\mathbf{x}u_{n\mathbf{k}}\, d\tau$$

$$= \delta_{nn'} \sum_R \exp\left[i(\mathbf{k} - \mathbf{k}')\cdot\mathbf{R}\right]\mathbf{R} + \sum \exp\left[i(\mathbf{k} - \mathbf{k}')\cdot\mathbf{R}\right]\xi_{n'n}(\mathbf{k})$$

$$\xi_{n'n}(k) = \int u^*_{n'\mathbf{k}}\mathbf{x}u_{n\mathbf{k}}\, d\tau \tag{2.2}$$

where the integral is over the unit cell with $\mathbf{R} = 0$. We see that not only is the first sum not very well defined mathematically speaking (a problem familiar from free electrons), but the second term depends on the particular choice of unit cell.[5]

A different approach to this problem is to write $I_{n\mathbf{k},n'\mathbf{k}'}$ in the following form[4]:

$$\int \psi^*_{n'\mathbf{k}'} x^\mu \psi_{n\mathbf{k}} \, d^3x = -i \frac{\partial}{\partial k^\mu} \int \psi^*_{n'\mathbf{k}'} \psi_{n\mathbf{k}} \, d^3x$$

$$+ \int u^*_{n'\mathbf{k}'} \exp\left[i(\mathbf{k} - \mathbf{k}') \cdot \mathbf{x}\right] \frac{i \, \partial u_{n\mathbf{k}}}{\partial k^\mu} \, d^3x$$

$$= -i \frac{\partial}{\partial k^\mu} \Delta_{n'n}(\mathbf{k}', \mathbf{k}) + \delta(\mathbf{k} - \mathbf{k}') \mathcal{X}^\mu_{n'n}(\mathbf{k}) \qquad (2.3)$$

where

$$\Delta_{n'n}(\mathbf{k}', \mathbf{k}) = \int \psi^*_{x'\mathbf{k}'} \psi_{n\mathbf{k}} \, d^3x$$

$$(2.3a)$$

$$\mathcal{X}_{n'n}(\mathbf{k}) = \int u^*_{n'\mathbf{k}} i \frac{\partial u_{n\mathbf{k}}}{\partial \mathbf{k}} \, d\tau.$$

We have not written $\Delta_{n'n}(\mathbf{k}', \mathbf{k}) = \delta(\mathbf{k} - \mathbf{k}') \delta_{nn'}$ because it is phase-sensitive in the sense to be discussed shortly. This does not matter unless it is to be differentiated, but that is precisely what we are doing, and some care is, therefore, advisable. The procedure in Eq. (2.3) is somewhat more acceptable because $\mathcal{X}_{n'n}$ is not sensitive to the choice of unit cell, but it is open to the nontrivial objection that $\psi_{n\mathbf{k}}$ is not differentiable with respect to $\mathbf{k}$. For no matter how small $\delta\mathbf{k}$ may be, $\psi_{n,\mathbf{k}+\delta\mathbf{k}} - \psi_{n\mathbf{k}}$ is not small and $\psi_{n,\mathbf{k}+\delta\mathbf{k}}$ does not approach $\psi_{n\mathbf{k}}$. Therefore, we investigate this subject more carefully in Appendix A. The conclusion is that Eq. (2.3) gives the right answer in practically all cases; the limitations are discussed in the appendix. From Eq. (A12) in the appendix we find that the action of $\mathbf{x}$ on a wave function $f_n(\mathbf{k})$ is given by

$$\mathbf{x}f = i \frac{\partial f_n}{\partial \mathbf{k}} + \sum \mathcal{X}_{nn'} f_{n'}. \qquad (2.4)$$

[5] K. Fuchs, *Proc. Roy. Soc.* **A176**, 214 (1940) has used this representation of x in a treatment of energy levels in a perfect crystal.

There is, however, an indeterminacy in the $\mathfrak{X}$'s also. This arises from the fact that the u's are not completely specified but can be multiplied by an arbitrary phase factor exp $[-i\varphi(\mathbf{k})]$. If this change is made, $i\,\partial u/\partial \mathbf{k}$ becomes exp $(-i\varphi)(i\,\partial u/\partial \mathbf{k} + u\,\partial\varphi/\partial \mathbf{k})$ and the $\mathfrak{X}$'s become

$$\mathfrak{X}'_{nn'} = \exp(i\varphi_n)\mathfrak{X}_{nn'}\exp(-i\varphi_{n'}) \qquad n \neq n'$$

(2.5)

$$\mathfrak{X}'_{nn} = \mathfrak{X}_{nn} + \frac{\partial\varphi_n}{\partial \mathbf{k}}.$$

Thus, while $\mathfrak{X}_{nn'}$ transforms like an operator, under such a transformation $\mathfrak{X}_{nn}$ does not. A compensatory change occurs in $i\,\partial f/\partial \mathbf{k}$ so that the sum $\mathbf{x}_c = (i\,\partial/\partial \mathbf{k} + \mathfrak{X}_{nn})$ does transform like an operator. Thus our final form for $\mathbf{x}$ in the CMR is

$$\mathbf{x} = \mathbf{x}_c + \mathbf{X}$$

(2.6)

where $\mathbf{X}$, like $\mathbf{V}$, is an operator with only interband matrix elements, in this case equal to the $\mathfrak{X}_{nn'}$. If we consider the matrix elements (2.3), this phase-invariance follows from the fact that $\Delta_{nn'}(\mathbf{k}, \mathbf{k}')$ is altered by the phase transformation so that the sum (2.3) is merely multiplied by exp $[i(\varphi_{n'} - \varphi_n)]$, the intraband terms, therefore, being unaffected.

While $\mathfrak{X}_{nn}$ is indeterminate, its curl is a well-defined quantity invariant under the phase transformations and is thus a characteristic of the band structure, which we shall designate as Ω_n.

The reader may have noticed that there is a close similarity here to the representation of momentum in the Schrödinger representation, where there is also an arbitrary phase factor related to gage invariance. The operator $-i\,\partial/\partial x$ is not gage-invariant, but $(-i\,\partial/\partial x + (e/c)\mathbf{A})$ is. $\mathfrak{X}_{nn}$ thus plays a role analogous to that of A. In the present case also, only $\Omega_n = \operatorname{curl}\mathfrak{X}_{nn}$ has any invariant meaning as only $\mathbf{B} = \operatorname{curl}\mathbf{A}$ has physical meaning.

When spin degeneracy is taken into account, with or without spin-orbit coupling, the energy bands are doubly degenerate throughout $\mathbf{k}$ space if a center of inversion exists in a crystal. This means that the eigenfunctions have not only an arbitrary phase, but also that arbitrary linear combinations of the pairs can be taken. The most general arbitrary transformation can be conveniently written in the form exp $(i\varphi)$ exp $(i\sigma_i\alpha_i)$, where the σ_i are the Pauli matrices ($\sigma_1 = \sigma_x$, $\sigma_2 = \sigma_y$, $\sigma_3 = \sigma_z$), which in the present case do not refer to the spin of an electron but are purely algebraic symbols. If only α_3 is nonzero, this transformation simply consists of phase transformations for the two degenerate states. When α_1 and α_2 are also nonzero, the transformation can be thought of as analogous to a rotation of the spin axis.

When this situation exists, it is convenient to consider a band as consisting of the two degenerate states. Wave functions in the CMR will then still be written in the form $f_n(\mathbf{k})$, but for each band f_n will be a two-component function,

$$\begin{pmatrix} f_{n1}(\mathbf{k}) \\ f_{n2}(\mathbf{k}) \end{pmatrix},$$

just as in the Pauli representation for spin-orbit problems. Intraband matrix elements of operators will be taken to be 2×2 matrices with elements $\mathfrak{X}_{nl,nm}$ ($l, m = 1, 2$); interband elements will be 2×2 blocks of infinite-rowed matrices. With this economical notation, it will be possible to discuss this type of problem in exactly the same way as problems for which the bands are not degenerate.

In particular, $\mathfrak{X}_{nn}$ will now have the form

$$\mathfrak{X}_{nn} = \begin{vmatrix} \mathfrak{X}_{n1,n1} & \mathfrak{X}_{n1,n2} \\ \mathfrak{X}_{n2,n1} & \mathfrak{X}_{n2,n2} \end{vmatrix}. \tag{2.7}$$

Frequently it will be convenient to write this as follows

$$\mathfrak{X}_{nn} = \sum_i \mathfrak{X}_{nn}{}^{(i)}\sigma_i + \mathfrak{X}_{nn}{}^{(0)}I \tag{2.7a}$$

where I is the identity matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\mathfrak{X}_{nn}{}^{(0)} = \tfrac{1}{2}(\mathfrak{X}_{n1,n1} + \mathfrak{X}_{n2,n2}) \qquad \mathfrak{X}_{nn}{}^{(1)} = \tfrac{1}{2}(\mathfrak{X}_{n1,n2} + \mathfrak{X}_{n2,n1})$$

$$\mathfrak{X}_{nn}{}^{(2)} = \tfrac{1}{2}i(\mathfrak{X}_{n1,n2} - \mathfrak{X}_{n2,n1}) \qquad \mathfrak{X}_{nn}{}^{(3)} = \tfrac{1}{2}(\mathfrak{X}_{n1,n1} - \mathfrak{X}_{n2,n2}).$$

Let us now consider the effect of a general arbitrary transformation of the form $S = \exp(i\varphi) \exp(i\sigma_i\alpha_i)$ on $\mathfrak{X}_{nn}$.

$$\mathfrak{X}_{nl,nm} = \int u_{nl}^* i \frac{\partial u_{nm}}{\partial \mathbf{k}} d\tau$$

$$\mathfrak{X}'_{nl,nm} = \int S_{ll'} u_{nl'}^* i \frac{\partial}{\partial \mathbf{k}} u_{nm'} S_{m'm}^\dagger d\tau$$

$$= S_{ll'}\mathfrak{X}_{nl',nm'} S_{m'm}^\dagger + S_{lm'} i \frac{\partial S_{m'm}^\dagger}{\partial \mathbf{k}}. \tag{2.8}$$

This can be written as

$$\mathfrak{X}' = S\mathfrak{X}S^\dagger + S\,\delta\mathfrak{X}S^\dagger$$

(2.8a)

$$\delta\mathfrak{X}_{nn} = \left(i\,\frac{\partial S^\dagger}{\partial \mathbf{k}}\right)S = -iS^\dagger\,\frac{\partial S}{\partial \mathbf{k}}.$$

We now consider the derivatives of $\mathfrak{X}_{mm}$ which we shall now designate as $\mathfrak{x}$ for convenience, using (2.8a)

$$\frac{\partial \mathfrak{x}'^{\mu}}{\partial \mathbf{k}^{\nu}} = S\,\frac{\partial}{\partial \mathbf{k}^{\nu}}\,(\mathfrak{x}^{\mu} + \delta\mathfrak{x}^{\mu})\,S^\dagger + iS[\delta\mathfrak{x}^{\nu},\ \mathfrak{x}^{\mu} + \delta\mathfrak{x}^{\mu}]S^\dagger$$

Thus

$$\nabla_k\cdot\mathfrak{x} = S\nabla\cdot\mathfrak{x}S^\dagger + iS[\delta\mathfrak{x},\ \cdot\ \mathfrak{x}]S\dagger + S\nabla\cdot\delta\mathfrak{x}S^\dagger \qquad (2.9)$$

$$\nabla_k \times \mathfrak{x}' = S(\nabla_k \times \mathfrak{x} + \nabla_k \times \delta\mathfrak{x} + i\,\delta\mathfrak{x} \times \mathfrak{x} + i\mathfrak{x} \times \delta\mathfrak{x} + 2i\,\delta\mathfrak{x} \times \delta\mathfrak{x})\,S^\dagger.$$

(2.10)

Hence even $\nabla_k \times \mathfrak{x}$ is not invariant under our more general transformation, applicable to bands which are everywhere degenerate. However, let us now consider $\mathfrak{x}' \times \mathfrak{x}'$

$$\mathfrak{x}' \times \mathfrak{x}' = S(\mathfrak{x} \times \mathfrak{x} + \delta\mathfrak{x} \times \mathfrak{x} + \mathfrak{x} \times \delta\mathfrak{x} + \delta\mathfrak{x} \times \delta\mathfrak{x})\,S\dagger. \quad (2.11)$$

Now, observing that $\nabla \times \delta\mathfrak{x}_{nn} = -i\delta\mathfrak{x} \times \delta\mathfrak{x}$ we see that

$$\nabla \times \mathfrak{x}_{nn} - i\mathfrak{x}_{nn} \times \mathfrak{x} = S(\nabla \times \mathfrak{x}_{nn} - i\mathfrak{x} \times \mathfrak{x})\,S\dagger. \qquad (2.12)$$

Thus $\nabla \times \mathfrak{x} - i\mathfrak{x} \times \mathfrak{x}$ does behave like an operator under S. When the bands are not degenerate so that the $\mathfrak{x}$ are numbers, not matrices, $\mathfrak{x} \times \mathfrak{x} = 0$; hence the operator behavior of $\nabla \times \mathfrak{x} - i\mathfrak{x} \times \mathfrak{x}$ includes the simpler case and we now designate it as Ω_n.

We can now consider the commutator $[x_c{}^{\mu}, x_c{}^{\nu}]$ where $\mathbf{x}_c$ is in general a 2×2 matrix operator

$$\mathbf{x}_c = i\,\frac{\partial}{\partial \mathbf{k}}\,I + \mathfrak{x} - i[x_c{}^{\mu}, x_c{}^{\nu}]$$

$$= (\nabla \times \mathfrak{x})^{\mu\nu} - i(\mathfrak{x} \times \mathfrak{x})^{\mu\nu} \equiv \Omega_n{}^{\lambda}(\mathbf{k})\,\epsilon_{\lambda\mu\nu}. \qquad (2.13)$$

The quantity $\Omega_n(\mathbf{k})$ plays a role analogous to that of the magnetic field in real space. The commutator [Eq. (2.13)] is very similar to $[v^{\mu}, v^{\nu}] = [p^{\mu} + (e/c)A^{\mu},\ p^{\nu} + (e/c)A^{\nu}] = H^{\lambda}\epsilon^{\lambda\mu\nu}.$

We have obtained the forms of the most important operators in the CMR. We now consider their commutators.[3,4] First we have

$$\mathfrak{B} = -\frac{i}{\hbar} [\mathbf{x}, H]$$

which yields

(a)
$$\mathbf{v}_n = \mathfrak{B}_{nn} = \frac{1}{\hbar} \frac{\partial E_n}{\partial \mathbf{k}}$$

$$(2.14)$$

(b)
$$\mathbf{V}_{nn'} = \mathfrak{B}_{nn'} = +\frac{i}{\hbar} X_{nn'} (E_n - E_{n'}).$$

Next we consider

$$-\frac{i}{\hbar} [\mathbf{x}, \mathfrak{B}] = -\frac{i}{\hbar} [\mathbf{x}_c, \mathbf{v}] - \frac{i}{\hbar} [\mathbf{X}, \mathbf{V}] - \frac{i}{\hbar} [\mathbf{X}, \mathbf{v}] - \frac{i}{\hbar} [\mathbf{x}_c, \mathbf{V}].$$

The left side is $1/m$ for the Schrödinger and Pauli Hamiltonians, 0 for the Dirac.

We focus on the diagonal terms and obtain

$$-\frac{i}{\hbar} [x_c^\mu, v^\nu]_n = \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k^\mu \, \partial k^\nu} = -\frac{i}{\hbar} [x^\mu, \mathfrak{B}^\nu] + \frac{i}{\hbar} [X^\mu, V^\nu]. \quad (2.15)$$

For the Schrödinger and Pauli Hamiltonians we obtain

$$\alpha^{\mu\nu} \equiv \frac{1}{\hbar^2} \frac{\partial^2 E_n}{\partial k^\mu \, \partial k^\nu} = \frac{1}{m} + \sum_{n'} \frac{(V_{nn'}^\mu V_{n'n}^\nu + V_{nn'}^\nu V_{n'n}^\mu)}{E_n - E_{n'}} \quad (2.15')$$

while the Dirac Hamiltonian yields

$$\frac{1}{\hbar^2} \frac{\partial^2 E_n}{\partial k^\mu \, \partial k^\nu} = \sum_{n'} \frac{(V_{nn'}^\mu V_{n'n}^\nu + V_{nn'}^\nu V_{n'n}^\mu)}{E_n - E_{n'}} \quad (2.15'')$$

where, however, the sum must now extend over negative-energy states as well as positive-energy states. For low-lying states n, the sum over negative-energy states is very nearly equal to $\delta_{\mu\nu}/m$ since for all states n' with significant $V_{nn'}$, $E_n - E_{n'} \approx 2mc^2$ if $E_n - mc^2 \ll mc^2$. Furthermore, the velocity operator matrix between positive and negative states is just $c\alpha^\mu$. The sum is then approximately $(1/m)\{\alpha^\mu, \alpha^\nu\} = \delta_{\mu\nu}/m$.

From (2.14) and (2.15), we find a rough relation between $|\mathbf{X}|^2$ and $\alpha^{\mu\nu}$

$$\tfrac{1}{2}\{X^\mu, X^\nu\}_{nn} = \frac{\hbar^2}{2} \sum_{n'} \frac{(V^\mu_{nn'}V^\nu_{n'n} + V^\nu_{nn'}V^\mu_{n'n})}{(E_n - E_{n'})^2}$$

$$= \frac{(\alpha^{\mu\nu} - \delta^{\mu\nu}/m)}{2} \left(\overline{\frac{1}{E_n - E_{n'}}}\right) \qquad (2.16)$$

where the final factor is an appropriate averaged reciprocal energy gap. In the particular case where only one band contributes significantly to the sum, we have

$$\tfrac{1}{2}\{X^\mu, X^\nu\}_{nn} \approx \frac{\hbar^2}{2} \frac{(\alpha^{\mu\nu} - \delta^{\mu\nu}/m)}{E_n - E_{n'}} \qquad \text{(one band case)}. \qquad (2.16')$$

Finally we consider $[x^\mu, x^\nu] = 0$:

$$0 = [x_c{}^\mu, x_c{}^\nu] + [X^\mu, X^\nu] + [x_c{}^\mu, X^\nu] + [X_c{}^\mu, x_c{}^\nu]. \qquad (2.17)$$

We see in particular that the intraband terms give

$$[x_c{}^\mu, x_c{}^\nu]_n = -[X^\mu, X^\nu]_n. \qquad (2.17')$$

All of these relations can also be obtained by detailed calculation, but the foregoing derivations have the advantage of conciseness and closer contact with the physics of the situation. From a practical point of view, these formulas imply that if we can obtain theoretical or experimental information about the V's, it can be turned into knowledge of the quantities $\mathbf{X}$, $\alpha^{\mu\nu}$, and $\Omega^{\mu\nu}$, and vice versa.

Some properties of $\boldsymbol{\Omega}$, $\boldsymbol{\alpha}$, $\mathbf{v}$, $\mathbf{V}$, $\mathbf{X}$, which result from time reversal and inversion symmetry, are discussed in Appendix B.

Further insight into the operators $\mathbf{x}_c$ and $\mathbf{X}$ may be obtained by considering a normalized wave packet $f_1(\mathbf{k})$ composed entirely of wave functions belonging to band 1. The center of this packet is at the point $\bar{\mathbf{x}}$

$$\bar{\mathbf{x}} = \int f^* \mathbf{x}_c f \, d^3k$$

$$= \int f^* i \frac{\partial f}{\partial \mathbf{k}} \, d^3k + \int |f|^2 \, \boldsymbol{\xi} \, d^3k.$$

Thus $\mathbf{x}_c$ rather than $i(\partial/\partial \mathbf{k})$ represents the center of the packet. The spread of the packet is given by

$$\overline{\mathbf{x}^2} - \bar{\mathbf{x}}^2 = \int f^* \mathbf{x}_c{}^2 f \, d^3k + \int f^* \mathbf{X}^2 f \, d^3k - \bar{\mathbf{x}}^2. \qquad (2.18)$$

Let us now choose a definite form for f

$$f = \sqrt{\frac{\lambda^3}{(2\pi)^{3/2}}} \exp\left[-(\lambda^2|\mathbf{k} - \mathbf{k}_0|^2)/2\right] \exp\left[-i\varphi(\mathbf{k})\right].$$

Then

$$\bar{\mathbf{x}} = \int |f^2|(\mathfrak{x} + \nabla\varphi)\,d^3k = \mathfrak{x} + \overline{\nabla\varphi}$$

and

$$\overline{\mathbf{x}^2} - \bar{\mathbf{x}}^2 = \tfrac{3}{2}\lambda^2 + \overline{\mathbf{X}^2} + \overline{(\mathfrak{x} + \nabla\varphi)^2} - \overline{(\mathfrak{x} + \nabla\varphi)}^2$$

$$+ \int |f|^2\,\nabla\cdot(\mathfrak{x} + \nabla\varphi)\,d^3k. \qquad (2.19)$$

φ can be chosen so that $\nabla\cdot(\mathfrak{x} + \nabla\varphi) = 0$ and so that $\mathfrak{x} + \nabla\varphi = 0$ at $\mathbf{k} = \mathbf{k}_0$. If we suppose λ large enough so that we can be content with the linear part of $\mathfrak{x} + \nabla\varphi$, we can further choose φ so that $\mathfrak{x} + \nabla\varphi = \tfrac{1}{2}\Omega_1 \times (\mathbf{k} - \mathbf{k}_0)$ in the neighborhood of $\mathbf{k}_0$. Then

$$\overline{\mathbf{x}^2} - \bar{\mathbf{x}}^2 = \tfrac{3}{2}\lambda^2 + (|\mathbf{X}(k_0)|^2)_{11} + \frac{\Omega_1^2}{4\lambda^2}. \qquad (2.19')$$

Thus $\sqrt{\overline{X^2}}$ is a limit to the spread of a wave packet. The $\mathbf{X}_{nn'}$ may be said to play a role analogous to the Compton wavelength for free electrons. This point will be pursued in the following in a detailed discussion of the Dirac theory regarded from the standpoint of band theory. $\mathbf{\Omega}_n$ also represents a restriction on the localization, but no more than we already have, since $\tfrac{1}{4}|\Omega|^2 < |X^2|^2$. This fact can be shown easily as follows. From Eq. (2.19)

$$\mathbf{\Omega}_n = \nabla \times \mathfrak{x}_n = i\sum_{n'} \mathbf{X}_{nn'} \times \mathbf{X}_{n'n}.$$

Thus

$$|\Omega_n^{(1)}|^2 < 4\sum_{n'n''} |X_{nn'}^{(2)}|^2\,|X_{nn''}^{(3)}|^2 = 4(|X^{(2)}|^2)_{nn}(|X^{(3)}|^2)_{nn}$$

$$< (|X^2|^2 + |X^3|^2)^2$$

$$\tfrac{1}{4}|\Omega|^2 < \tfrac{1}{4}(|\Omega^{(1)}| + |\Omega^{(2)}| + |\Omega^{(3)}|)^2 < |X^2|^2.$$

It is tempting to say that the $\mathfrak{X}_{nn'}$ represent the position of the electron in the unit cell and to infer that the $\mathfrak{X}$'s must be approximately the size of a lattice vector. The $\mathfrak{X}_{nn}^{\mu}$ may in fact be kept small by phase transformations, but the $\mathbf{X}_{nn'}$ may be very large in special cases. Thus in Bi one

can infer, from the small effective mass and small energy gap,[6] that the $\mathbf{X}_{nn'}$ at the electron valleys between the valence and conduction bands is of the order of 100 A. In general, one can expect $\mathbf{X}_{nn'}$ to be large between bands whose energy separation is small, unless $\mathbf{V}_{nn'}$ is also small for some reason.

We previously introduced the quantities $\xi_{nn'}$ which do indeed represent the position of the electron in the unit cell. One might suppose that some relation exists between them and the $\mathfrak{X}_{nn'}$. This relation cannot be simple, however. The $\mathfrak{X}_{nn'}$ are independent of the choice of unit cell but depend on the choice of phases, whereas the $\xi_{nn}{}'$ do just the reverse. We can show that there is no choice of unit cells and phases such that $\xi_{nn'} = \mathfrak{X}_{nn'}$. Thus from (2.2a)

$$+ i \frac{\partial}{\partial k^{\mu}} \xi_{nn'}^{\nu} = \mathfrak{X}_{nn''}^{\mu} \times \xi_{n''n'}^{\nu} - \xi_{nn''}^{\nu} \times \mathfrak{X}_{n''n'}^{\mu}$$

$$i\nabla_{\mathbf{k}} \times \xi = \mathfrak{X} \times \xi + \xi \times \mathfrak{X}$$

while from Eq. (2.17)

$$i\nabla \times \mathfrak{X} = \mathfrak{X} \times \mathfrak{X}$$

also

$$+ i\nabla \cdot \xi = \mathfrak{X} \cdot \xi - \xi \cdot \mathfrak{X}.$$

Thus $\xi = \mathfrak{X}$ is incompatible with these equations unless both vanish.

Using the operators discussed so far it would be possible in principle to construct the representative of any operator. This would, however, be quite cumbersome except for operators which are low-order polynomials in $\mathbf{k}$, $\mathbf{x}$, $\mathfrak{B}$, $\mathbf{p}$, etc. For most functions, it will be simpler to construct the matrix elements directly; no particularly simple form is possible. The crystal momentum representation is most useful, however, when one is dealing with slowly varying perturbations. Thus if we have $V(\mathbf{x})$ such that

$$V(\mathbf{x}) = \int d^3q V(\mathbf{q}) \exp(i\mathbf{q}\cdot\mathbf{x})$$

where $V(q) \neq 0$ only for small q, we can expand the matrix elements in powers of q, obtaining

$$V_{n'n}(\mathbf{k'},\ \mathbf{k}) = \int d^3q V(\mathbf{q}) \int \psi_{n',\mathbf{k}+\mathbf{q}}^{*}(\mathbf{x}) \exp(i\mathbf{q}\cdot\mathbf{x})\psi_{n\mathbf{k}}(\mathbf{x})$$

$$= V(\mathbf{k'} - \mathbf{k}) \int d\tau u_{n',\mathbf{k'}}^{*} u_{n\mathbf{k}} \equiv V(\mathbf{k'} - \mathbf{k}) S_{n'n}(\mathbf{k'},\ \mathbf{k})$$

$$= V(\mathbf{k'} - \mathbf{k}) \delta_{nn'} + iV(\mathbf{k'} - \mathbf{k})(\mathbf{k'} - \mathbf{k})\cdot\mathfrak{X}_{n'n} + \cdots.$$

[6] M. H. Cohen, *Phys. Rev.* **121**, 387 (1961).

To this order we can write the result in the form

$$V(\mathbf{k'} - \mathbf{k}) - \mathbf{F}(\mathbf{k'} - \mathbf{k}) \cdot \mathfrak{X}_{nn'}$$

where   (2.20)

$$\mathbf{F} = -\frac{\partial V}{\partial \mathbf{x}}.$$

Further, if we are willing to keep diagonal-in-band terms to first order, and throw away interband terms of this order, we have

$$V_{nn'}(\mathbf{k'}, \mathbf{k}) \cong \delta_{nn'}(V(\mathbf{k'} - \mathbf{k}) - F(\mathbf{k'} - \mathbf{k}) \cdot \mathfrak{X}_{nn}) \qquad (2.20')$$

or, as an operator, to first order in $\mathbf{F}$,

$$V \cong V(\mathbf{x}_c).$$

Thus if V is a slowly varying perturbing potential, the complete approximate Hamiltonian is

$$H_0(\mathbf{k}) + V(\mathbf{x}_c). \qquad (2.21)$$

This discussion could be carried considerably further, but we shall not do so because we will develop much more convenient techniques for handling a wider class of perturbation in Part III. The main point is that we have been able to eliminate interband matrix elements for the restricted accuracy in which we are interested. We also have a clear procedure for including higher order terms in $\mathbf{q}$. With these, however, the convenience is much reduced. The use of x_c instead of merely $i\partial/\partial k$ introduces effects analogous to spin-orbit coupling (see also Part IV).

As an example of the role played by $\mathbf{\Omega}$, we will consider the velocity of a Bloch electron in an external field V. For this purpose we will not simply discard the interband matrix elements of (2.20), but transform them away to first order in V by the transformation e^T,

$$T_{n'\mathbf{k'},n\mathbf{k}} = \mathbf{F}(\mathbf{k'} - \mathbf{k}) \cdot \mathfrak{X}_{n'n}/(E_n - E_{n'}) \qquad n \neq n'.$$

The entire Hamiltonian to first order in V or $\mathbf{F}$ then becomes (2.21), while the intraband part of x is

$$\mathbf{x} = \mathbf{x}_c + [T, \mathbf{X}]_{nn}.$$

The in-band velocity operator is then, again to first order in $\mathbf{F}$,

$$\mathbf{v} = -\frac{i}{\hbar}[\mathbf{x}_c, H_0 + V(\mathbf{x}_c)]$$

$$= \frac{i}{\hbar}\frac{\partial H}{\partial \mathbf{k}} - \mathbf{F} \times \mathbf{\Omega}. \qquad (2.22)$$

This extra term in the velocity has been called an anomalous velocity by Adams and Blount[7] and plays an important role in the Hall effect in ferromagnetic materials.[8]

3. THE MODIFIED CMR

So far we have discussed the CMR, in which the basic functions are Bloch functions. A closely related procedure, used particularly by Kohn and Luttinger,[9] is the modified CMR (MCMR). In this representation the crystal momentum is still diagonal. Instead of using Bloch functions, however, one uses what we shall call Kohn-Luttinger (K-L) functions

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \exp\,(i\mathbf{k}\cdot\mathbf{r})u_{n\mathbf{k}_0}(\mathbf{r}).$$

Here $u_{n\mathbf{k}_0}(\mathbf{r})$ is the periodic part of the Bloch function at a particular conveniently chosen point $\mathbf{k}_0$ in k space. In this case H is diagonal only at $\mathbf{k}_0$, but the representative of $\mathbf{x}$ is simply $i\,\partial/\partial\mathbf{k}$. This is simpler for some purposes than having to be concerned with $\mathfrak{X}$'s. In general, however, one must expect that the MCMR will be useful only when one is interested in a very small area of $\mathbf{k}$ space, which is indeed the case for which it was created. In many problems, one can achieve the same results with the same accuracy in the CMR by setting $\mathbf{x}_c = i\,\partial/\partial\mathbf{k}$, that is, neglecting ξ_{nn}. In some other cases, computation can probably be performed more simply in the MCMR. This is most likely to be true when one cannot ignore the interband matrix elements, as in the neighborhood of degeneracies, which was also true for the problem Kohn and Luttinger were interested in.

II. Wannier Representation

4. THE WANNIER HAMILTONIAN

In the preceding section we have discussed the CMR, which is evidently analogous to the momentum representation of ordinary continuum mechanics. In this section, we shall discuss the Wannier representation[1] (WR), which is analogous to the coordinate representation. This representation is obtained by making the unitary transformation $\exp\,(i\mathbf{k}\cdot\mathbf{R})$. The basis functions are the Wannier functions $a_n(\mathbf{r} - \mathbf{R})$:

$$a_n(\mathbf{r} - \mathbf{R}) = \int d^3k\psi_{n\mathbf{k}}(\mathbf{r}) \exp\,(-i\mathbf{k}\cdot\mathbf{R}). \tag{4.1}$$

[7] E. N. Adams and E. I. Blount, *Phys. and Chem. Solids* **10**, 286 (1959).
[8] J. M. Luttinger, *Phys. Rev.* **112**, 739 (1958).
[9] W. Kohn and J. M. Luttinger, *Phys. Rev.* **97**, 869 (1955).

(All integrations d^3k are to be taken over one BZ.)

$$\psi_{nk}(\mathbf{r}) = \sum_{\mathbf{R}} \exp{(i\mathbf{k}\cdot\mathbf{R})} a_n(\mathbf{r} - \mathbf{R}).$$

Conversely, the wave function $f(\mathbf{k})$ is transformed to

$$f(\mathbf{R}) = \int \exp{(i\mathbf{k}\cdot\mathbf{R})} f(\mathbf{k}) \, d^3\mathbf{k}. \qquad (4.2)$$

The advantage of this representation is that the Wannier functions are localized in a sense which will be clarified in the following. This can be seen roughly, however, as follows: consider $a_n(\mathbf{R}_i)$ the value of a_n at the ith lattice site. Then

$$a_n(\mathbf{R}_i - \mathbf{R}) = \int d^3k u_{nk}(0) \exp{[i\mathbf{k}\cdot(\mathbf{R}_i - \mathbf{R})]}.$$

Except for $\mathbf{R}_i$ near $\mathbf{R}$, the exponential is rapidly varying, so that the integral over $\mathbf{k}$ tends to be small. Thus $a_n(\mathbf{R}_i - \mathbf{R})$ becomes small at large distance.

An operator $O_{nn'}(\mathbf{k}, \mathbf{k}')$ becomes

$$O_{nn'}(\mathbf{R}, \mathbf{R}') = \int d^3\mathbf{k} \, d^3\mathbf{k}' \exp{[i(\mathbf{k}\cdot\mathbf{R} - \mathbf{k}'\cdot\mathbf{R}')} O(\mathbf{k}, \mathbf{k}'). \qquad (4.3)$$

In particular, we have the following changes:

(1) $\quad \mathbf{p}_c = \int \hbar\mathbf{k} \exp{[i\mathbf{k}\cdot(\mathbf{R} - \mathbf{R}')]} \, d^3\mathbf{k} = -i\hbar \frac{\partial}{\partial\mathbf{R}} \, \delta(\mathbf{R} - \mathbf{R}') \quad (4.4)$

$$\mathbf{p}_c f(\mathbf{R}) = -i \frac{\partial f}{\partial\mathbf{R}}$$

(2) $$\mathbf{p}_c^n f(\mathbf{R}) = \left(-i \frac{\partial}{\partial R}\right)^n f(\mathbf{R}). \qquad (4.5)$$

(3) In general, any function of $\mathbf{k}$, $O(\mathbf{k})$ becomes

$$O(\mathbf{k}) \to O(\mathbf{R} - \mathbf{R}') = \int d^3\mathbf{k} \exp{[i\mathbf{k}\cdot(\mathbf{R} - \mathbf{R}')]} O(\mathbf{k}) \qquad (4.6)$$

$O(\mathbf{R} - \mathbf{R}')$ is the Fourier transform of $O(\mathbf{k})$.
(3a) $\exp{(i\mathbf{L}\cdot\mathbf{k})}$ ($\mathbf{L}$ a lattice vector) becomes

$$\delta(\mathbf{R} - \mathbf{R}' - \mathbf{L}) = \exp{(\mathbf{L}\cdot\partial/\partial\mathbf{R})}. \qquad (4.7)$$

Thus (2) and (3) show that for any function $O(\mathbf{k})$ which can be expanded in either a power or Fourier series, we can write $O(-i\,\partial/\partial\mathbf{R})$.

(4)
$$i\,\frac{\partial}{\partial\mathbf{k}} \quad \text{becomes} \quad \mathbf{R} \tag{4.8}$$

(5)
$$\mathbf{x}_c \rightarrow \mathbf{R} + \mathbf{\xi}(\mathbf{R} - \mathbf{R}'). \tag{4.9}$$

Thus, if we consider the one-band Hamiltonian of Eq. (2.21), we find in the WR the alternative forms

$$H = H_0\left(-i\,\frac{\partial}{\partial\mathbf{R}}\right) + V(\mathbf{R}) \tag{4.10a}$$

$$H(\mathbf{R}, \mathbf{R}') = H_0(\mathbf{R} - \mathbf{R}') + V(\mathbf{R})\,\delta(\mathbf{R}, \mathbf{R}') \tag{4.10b}$$

where we have neglected $\mathbf{\xi}$. Equation (4.10a) is the form obtained by Slater by a straightforward WR derivation. Adams[4] observed that this equation could be obtained more readily from the CMR. This Hamiltonian can be used to obtain either a difference equation for $f(\mathbf{R})$

$$\sum_{\mathbf{R}'} H_0(\mathbf{R} - \mathbf{R}')f(\mathbf{R}') + V(\mathbf{R})f(\mathbf{R}) = Ef(\mathbf{R}) \tag{4.11a}$$

or a differential equation

$$H_0\left(-i\,\frac{\partial}{\partial\mathbf{R}}\right)f(\mathbf{R}) + V(\mathbf{R})f(\mathbf{R}) = Ef(\mathbf{R}). \tag{4.11b}$$

The first of these equations speaks for itself. Rigorously, the second has no meaning other than the first. It has been used by assuming that a slowly varying $f(R)$ can be treated as a continuous function. The second equation can then be treated as a differential equation. This still produces no useful simplification without the further assumption that $H(-i\,\partial/\partial R)$ can be replaced by an approximate form, usually quadratic in the argument. The first step is usually made rather intuitively with no clear statement defining its validity, while the second is made without justification, as if obvious. This may lead to the supposition that the first step leads to the major approximation. We shall see that, in fact, the first step can be made rigorously, and that all the approximation occurs in the second. Moreover, we shall find a measure of the approximation involved.

We start by *defining* the continuum Hamiltonian corresponding to a CMR $H(\mathbf{k}, \mathbf{k}')$—including both H_0 and V—as

$$H(\mathbf{x}, \mathbf{x}') = \int_{\text{BZ}} d^3k\, d^3k' \exp\,(i\mathbf{k}\cdot\mathbf{x})H(\mathbf{k}, \mathbf{k}')\exp\,(-i\mathbf{k}'\mathbf{x}') \tag{4.12}$$

where the integrations can be taken over any BZ. This quantity evaluated at lattice points is equal to $H(\mathbf{R}, \mathbf{R}')$. Now suppose that we have an eigenfunction $\psi(\mathbf{x})$ of $H(\mathbf{x}, \mathbf{x}')$

$$\int d^3x' H(\mathbf{x}, \mathbf{x}')\psi(x') = E\psi(x).$$

We multiply on the left by $\exp(-i\mathbf{k}\cdot\mathbf{x})$ and integrate over all $\mathbf{x}$. If k lies in the chosen BZ, we obtain

$$\int d^3k' H(\mathbf{k}, \mathbf{k}')\varphi(\mathbf{k}') = E\varphi(\mathbf{k}); \qquad \mathbf{k} \text{ in BZ} \qquad (4.13)$$

where

$$\varphi(\mathbf{k}) = \int \exp(-i\mathbf{k}\cdot\mathbf{x})\psi(\mathbf{x}) \, d^3x. \qquad (4.14)$$

If k is outside the BZ, on the other hand, the left side yields zero, so

$$\varphi(\mathbf{k}) = 0 \qquad \mathbf{k} \text{ outside BZ.} \qquad (4.15)$$

Thus, $\varphi(\mathbf{k})$ satisfies the CMR equation and

$$\psi(\mathbf{x}) = \int_{BZ} d^3k \varphi(\mathbf{k}) \exp(i\mathbf{k}\cdot\mathbf{x}).$$

The values of $\psi(\mathbf{x})$ at the lattice points give the WR solution $\psi(\mathbf{R})$ which obeys (4.11). Thus we have established that the replacement of the difference equations by differential equations is rigorous. $\psi(\mathbf{x})$ must not, of course, be interpreted as the actual wave function. It is simply a continuous function whose value at a lattice point is the WR wave function. It should be noted that the preceding discussion shows that for $V(\mathbf{x})$, we should use

$$V(\mathbf{x}) = \sum_R \int V(\mathbf{R}) \exp[i\mathbf{k}\cdot(\mathbf{x} - \mathbf{R})] \, d^3k. \qquad (4.16)$$

We now turn to the problem of approximating $H_0(\mathbf{k})$ by some simple polynomial, usually a quadratic. This is most easily considered in the CMR. Suppose that in a range of energy near E, $H_0(\mathbf{k})$ is well approximated by $H'(\mathbf{k})$, and that we have a solution f for $H'(k)$:

$$H'(\mathbf{k})f(\mathbf{k}) + V(\mathbf{x}_c)f(\mathbf{k}) = Ef(\mathbf{k}). \qquad (4.17)$$

If $H'(k) = H_0(k)$ over a region in k space and $f(\mathbf{k})$ is zero outside this region, $f(k)$ is an eigenfunction of $(H_0 + V)$ with energy E. These con-

ditions are, of course, never satisfied in practice, but they may be nearly satisfied, in which case we expect that f and E will be a good approximation. We can indeed formulate a simple test.

Suppose we have solutions φ_i for H'

$$(H' + V)\varphi_i = E_i'\varphi_i \qquad (4.18)$$

and that the solutions for H_0 are ψ_j

$$\psi_j^*(H_0 + V) = E_j\psi_j^*. \qquad (4.19)$$

Then, multiplying by ψ_j^* and φ_i, respectively, we have

$$(E_i' - E_j)(\psi_j, \varphi_i) = (\psi_j, (H' - H_0)\varphi_i). \qquad (4.20)$$

Squaring and summing over j,

$$\sum_j (E_i' - E_j)^2 \mid (\psi_j, \varphi_i) \mid^2 = (\varphi_i, (H' - H_0)^2\varphi_i). \qquad (4.21)$$

The left side of (4.21) is greater than the minimum value of $(E_i - E_j)^2$, which is, therefore, less than the right side. Thus we can say that there is a level of H_0 which lies closer to E_i' than $(\varphi_i, (H' - H_0)^2\varphi_i)^{1/2}$. If the latter is much less than the separation of levels, we can also be confident that the wave function is quite close to φ_i. In some cases in which the accuracy is not as good as might be desired, it could be improved by treating $(H_0 - H')$ as a perturbation.

Finally, it will usually be easier to solve (4.18) by transforming from **k** space to **x** space. We can then write

$$H'\left(-i\frac{\partial}{\partial \mathbf{x}}\right)\varphi(\mathbf{x}) + V(\mathbf{x})\varphi(\mathbf{x}) = E\varphi(\mathbf{x}). \qquad (4.22)$$

As with other $\psi(\mathbf{x})$, only the values $\varphi(\mathbf{R})$ at lattice points are meaningful. Here also we must be careful that φ can be written with Fourier components in only one BZ. This will always be less restrictive than the requirement that $(\varphi_i(H' - H_0)^2\varphi_i)$ be small, however, provided we choose the BZ in such a way as to keep the surface $H_0(\mathbf{k}) = E$ as far from the zone boundaries as possible.

Our emphasis on the derivation of (4.22) should not obscure the fact that the difference equations (4.11a) may in some cases be more convenient to work with. This will tend to be the case, for instance, when dealing with deep energy levels. The question, in practice, will generally be whether the difference equation, approximated by using only nearest neighbors, is more accurate than the best quadratic H'. In a rough sort of way one expects that the quadratic approximation will be quite good, so long as the

energies involved are small compared to vertical band separations. When this condition is violated, however, one must consider the effect of other bands on the energy levels.

5. PROPERTIES OF WANNIER FUNCTIONS

Aside from their role in justifying the effective mass theory, the Wannier functions are of some interest in themselves. In particular one is concerned with their localization. Before discussing this last subject at some length, we shall consider a variational approach used by Parzen[10] and by Koster.[11]

These authors observed that the Wannier function has the following property: for functions subject to the restrictions

$$\int \psi^*(\mathbf{x})\psi(\mathbf{x} + \mathbf{R}_i)\, d^3x = 0 \qquad \mathbf{R}_i \neq 0$$

$$(5.1)$$

$$\int |\psi|^2\, d^3x = 1$$

the integral

$$I = \int \psi^* H \psi\, dx \tag{5.2}$$

has an extremum for the Wannier function. Let us consider this condition in detail, in the CMR, setting

$$\psi(\mathbf{x}) = \sum_n \int d^3k f_{nk}\psi_{nk}(x)$$

$$I = \sum_n \int H_n(\mathbf{k}) |f_n(\mathbf{k})|^2\, d^3k \tag{5.3}$$

$$\int \psi^*(\mathbf{x})\psi(x + R_i)\, d^3x = \sum_n \int |f_n|^2 \exp(i\mathbf{k}\cdot\mathbf{R}_i)\, d^3k = 0$$

$$(5.3')$$

$$\int |\psi|^2\, d^3x = \sum_n \int |f_n|^2\, d^3k = 1.$$

These two restrictions then mean that $\sum_n |f_n|^2 = 1$, independent of k. Thus evidently the absolute minimum of I is obtained when $|f_n(\mathbf{k})|^2 = 1$ for the lowest band at each $\mathbf{k}$. But this corresponds exactly to the Wannier

[10] G. Parzen, *Phys. Rev.* **89**, 257 (1953).
[11] G. F. Koster, *Phys. Rev.* **89**, 67 (1953).

function for this band. We can examine the general extremal properties by means of Lagrangian multipliers. The restraints are taken in the form

$$\sum_n |f_n(\mathbf{k})|^2 = 1 \qquad \text{for each } k. \tag{5.4}$$

The equations obtained are

$$[H_n(\mathbf{k}) - \lambda(\mathbf{k})]f_n(\mathbf{k}) = 0. \tag{5.5}$$

This can be satisfied only by

$$f_n(\mathbf{k}) = 1 \qquad \text{if } H_n(\mathbf{k}) = \lambda(\mathbf{k})$$

$$f_n(\mathbf{k}) = 0 \qquad \text{for all other } n. \tag{5.6}$$

Thus $\lambda(\mathbf{k})$ can, in general, be any function which is, at each k, equal to some $H_n(\mathbf{k})$; $|f_n(\mathbf{k})|$ is equal to 1 for the same n, and 0 for all others. For any $\mathbf{f}(\mathbf{k})$, $\mathbf{f}(\mathbf{k}) \exp(i\mathbf{k}\cdot R_i)$ describes, for each i, an orthogonal function, whose center is removed from that of f by $\mathbf{R}_i$. For another function g to be orthogonal to this whole family of functions generated by f, we must have $f_n^*(\mathbf{k})g_n(\mathbf{k}) = 0$ for all $\mathbf{k}$. Thus if we generate a complete set by the Schmidt process, requiring that each function minimize I subject to being orthogonal to all previously obtained functions, we obtain precisely the Wannier function for each band (see next paragraph) in which the bands are defined as follows: the nth band has at each $\mathbf{k}$ the nth energy level, counting from the bottom. There are some special cases in which other choices of bands may be preferable. In such cases, the Wannier functions correspond to extrema of I, for they have the form mentioned above, but not to minima.

Actually the variational procedure determines only the magnitude of $f(\mathbf{k})$, not the phase $\varphi(\mathbf{k})$. We shall see in the following that an additional variational procedure can be used to determine $\varphi(\mathbf{k})$, namely minimization of the spread of the Wannier function.

We now turn to the location and spread of the Wannier functions. First we observe that in the CMR the Wannier function $a_n(\mathbf{r} - \mathbf{R}_i)$ is represented by

$$f_{n\mathbf{R}_i}(\mathbf{k}) = \exp(-i\mathbf{k}\cdot\mathbf{R}_i) \sqrt{\frac{\Omega}{(2\pi)^3}} \tag{5.7}$$

where Ω is the volume of the unit cell.

On the basis of the discussion in Appendix A, the operation of $\mathbf{x}$ on a WF can be represented as follows

$$(x^\mu)^l a_n(\mathbf{x}) = \sqrt{\frac{\Omega}{(2\pi)^3}} \int d^3k \exp(i\mathbf{k}\cdot\mathbf{x}) \left(\frac{i\partial}{\partial k^\mu}\right)^l f u_n \tag{5.8}$$

except for the case where there are planes of degeneracy. The matrix elements for $R_i = 0$ are given by

$$\int a_n{}^*(\mathbf{x} - \mathbf{R})(x^\mu)^l a_m(\mathbf{x})\ dV = \frac{\Omega}{(2\pi)^3} \int u_{n\mathbf{k}}^* e^{-i\mathbf{k}\cdot\mathbf{R}} \left(\frac{i\partial}{\partial k^\mu}\right)^l u_{m\mathbf{k}}\ d^3k\ d^3x$$

$$= \frac{\Omega}{(2\pi)^3} \int u_{n\mathbf{k}}^* e^{-i\mathbf{k}\cdot\mathbf{R}} \frac{i^l \partial^l}{\partial k^l} u_{m\mathbf{k}}\ d^3k\ d^3x$$

$$= \frac{\Omega}{(2\pi)^3} \int e^{-i\mathbf{k}\cdot\mathbf{R}}\ \delta(\mathbf{k} - \mathbf{k}') \frac{i^l \partial^l}{\partial k'^l} S_{nm}(\mathbf{k}, \mathbf{k}')$$

$$\cdot d^3k'\ d^3k \qquad (5.9)$$

where

$$S_{nm}(\mathbf{k}, \mathbf{k}') = \int u_{n\mathbf{k}}^* u_{m\mathbf{k}'}\ d\tau.$$

S is the transformation matrix between the u's at $\mathbf{k}$ and $\mathbf{k}'$.

In particular, the matrix elements of x and x^2 are as follows:

$$\int a_n{}^*(\mathbf{x} - \mathbf{R}) x a_m(\mathbf{x})\ dV = \frac{\Omega}{(2\pi)^3} \int d^3k\, \mathfrak{X} e^{-i\mathbf{k}\cdot\mathbf{R}}$$

$$\int a_n{}^*(\mathbf{x} - \mathbf{R}) x^2 a_m(\mathbf{x})\ dV = \frac{\Omega}{(2\pi)^3} \int d^3k\, ((\mathfrak{X}^2)_{nm} + i\nabla\cdot\mathfrak{X}_{nm}) e^{-i\mathbf{k}\cdot\mathbf{R}}$$

$$= \frac{\Omega}{(2\pi)^3} \int d^3k\, ((\mathfrak{X}^2)_{nm} - \mathbf{R}\cdot\mathfrak{X}_{nm}) e^{-i\mathbf{k}\cdot\mathbf{R}}. \qquad (5.10)$$

If we are concerned with bands everywhere degenerate, there are two Wannier functions for each degenerate band. Thus, we can consider each of the foregoing matrix elements as being a 2 × 2 matrix.

Let us now consider the position and spread of a given WF, $a_n(\mathbf{x})$.

$$\mathfrak{X} = \int |a_n|^2\, \mathbf{x}\, dV = \frac{\Omega}{(2\pi)^3} \int d^3k\, \mathfrak{X}_n = \overline{\mathfrak{X}_n} \qquad (5.11)$$

$$x^2 = \int |a_n|^2\, x^2\, dV = \frac{\Omega}{(2\pi)^3} \int d^3k (\mathfrak{X}^2)_{nn} = \overline{(\mathfrak{X}^2)_{nn}} = \overline{(X^2)_{nn}} + \overline{\mathfrak{X}_n{}^2}. \qquad (5.12)$$

$\overline{\mathfrak{X}_n}$ and $\overline{(\mathfrak{X}_n{}^2)}$ are sensitive to generalized phase transformations whereas

the first term in $\overline{x^2}$ is not. Thus if we make a phase transformation $e^{i\varphi(\mathbf{k})}$, $\bar{x}$ is modified as follows

$$\mathbf{\bar{x}'} = \mathbf{\bar{x}} + \frac{\Omega}{(2\pi)^3} \int \frac{\partial \varphi}{\partial \mathbf{k}} \, d^3k$$

$$= \mathbf{\bar{x}} + \frac{\Omega}{(2\pi)^3} \int d\mathbf{S} \, \Delta\varphi. \qquad (5.13)$$

Here the surface integral is over the planes bounding the BZ in the direction indicated, and $\Delta\varphi$ is the change in φ across the zone. The periodicity of $e^{i\varphi}$ requires that $\Delta\varphi$ be a multiple of 2π and the integral is, therefore, a lattice vector. Thus

$$\mathbf{\bar{x}'} = \mathbf{\bar{x}} + \mathbf{L}. \qquad (5.14)$$

$\bar{x}$ can in this way be made to be in the unit cell including the origin.

$$\overline{x^{2'}} = \overline{x^2} + \frac{\Omega}{(2\pi)^3} \int d^3k (2\boldsymbol{\xi}\cdot\nabla\varphi + (\nabla\varphi)^2)$$

$$= \overline{x^2} + \frac{\Omega}{(2\pi)^3} \int d^3k (-2\varphi\nabla\cdot\boldsymbol{\xi} + (\nabla\varphi)^2) + \frac{\Omega}{(2\pi)^3} \int d\mathbf{S}\cdot\boldsymbol{\xi}\varphi. \qquad (5.15)$$

The surface integral is to be taken over the BZ boundary and surfaces surrounding any degeneracies, as discussed in the Appendix. The former integrals give zero if φ is periodic, that is, if it does not move $\mathbf{\bar{x}}$. The integrals around degeneracies also vanish unless φ has such a strong singularity at the degeneracy that the integral of $(\nabla\varphi)^2$ diverges and dominates (5.15). If the degeneracy occurs on a curve, it is possible to choose φ to be multiple-valued in which case (5.15) is not really valid unless we also introduce a branch cut. In that case, the surface integral may diverge, but then $\overline{x^2}$ diverges anyway as shown in the following paragraphs.

Excluding this last situation, $\overline{x^2}$ can be minimized, the minimum occurring when $\nabla\cdot\boldsymbol{\xi}_{nn}$ is zero everywhere. This is an absolute minimum. The condition $\nabla\cdot\boldsymbol{\xi} = 0$ together with the value of $\nabla \times \boldsymbol{\xi}$, which is given everywhere suffices to determine $\boldsymbol{\xi}$ completely save for a constant equal to a lattice vector, provided the phase is so chosen near degeneracies as not to add any surface terms to $\overline{x^2}$. This is then a natural choice of phases and corresponds to the choice $\nabla\cdot A = 0$ for the vector potential. Similarly, we could have added a phase factor to (5.7) with the same result.

It should be noted that even if $\nabla \times \boldsymbol{\xi} = 0$ everywhere as when the crystal has a center of inversion, $\boldsymbol{\xi}$ may not necessarily be taken to be zero everywhere, unless the band in question has no lines or curves of degeneracy with other bands. If such lines exist, and the energies of the intersecting bands separate linearly away from the line, $\oint \boldsymbol{\xi}\cdot d\mathbf{k} =$

$(2n + 1)\pi$ where n depends on the particular choice of phases. If the bands diverge quadratically, $\oint \mathfrak{x} \cdot d\mathbf{k} = 2n\pi$. If all lines are of the latter type, all n's can be made zero and $\mathfrak{x}$ can then be made zero. This cannot be done in the former case.

If there is inversion symmetry and spin-orbit coupling, we have the added possibility of generalized transformations. Using the form

$$\mathfrak{x} = \sum_i \mathfrak{x}^{(i)}\sigma_i + \mathfrak{x}^{(0)}I \qquad (5.16)$$

we have

$$\mathfrak{x}^2 = (\sum_i \mathfrak{x}^{(i)2} + \mathfrak{x}^{(0)2})I + 2 \sum_i \mathfrak{x}^{(i)} \cdot \mathfrak{x}^{(0)}\sigma_i. \qquad (5.17)$$

It is convenient to treat these two parts separately. First we consider the I term. $\mathfrak{x}^{(0)}$ is invariant under the rotations $e^{iA \cdot \sigma}$ and $\mathfrak{x}^{(i)}$ is invariant under the phase transformations $e^{i\varphi}$. Thus the effects of the two transformations can be considered separately. The phase transformations are treated exactly as above. The rotations have the effect of changing $\overline{\mathfrak{x}^2}$ to

$$\overline{\mathfrak{x}'^2} = \overline{\mathfrak{x}^2} + \{\mathfrak{x} \cdot \delta\mathfrak{x}\} + (\delta\mathfrak{x})^2. \qquad (5.18)$$

This problem is complicated by the fact that $\delta\mathfrak{x}$ is no longer a gradient. A convenient form for $\delta\mathfrak{x}$ is found by writing $S = \exp(i\theta\sigma_i A_i/2)$, where $\mathbf{A}$ is a unit vector in the "spin space." Then $\delta\mathfrak{x}$ becomes

$$\delta\mathfrak{x} = \nabla\sigma_i A_i \sin\theta + (1 - \cos\theta)(\sigma_i A_i \nabla\theta + \epsilon_{ijl}\sigma_i A_i \nabla A_l). \qquad (5.19)$$

The I term is then

$$\tfrac{1}{2}Tr(\overline{\mathfrak{x}'^2} - \overline{\mathfrak{x}^2}) = -\frac{2\Omega}{(2\pi)^3} \int \nabla \cdot \mathfrak{x}^{(i)} A_i \sin\theta$$

$$+ \frac{2\Omega}{(2\pi)^3} \int \mathfrak{x}^{(i)} \cdot (1 - \cos\theta)(A_i \nabla\theta + \epsilon_{ijl}A_j \nabla A_l)$$

$$+ \frac{\Omega}{(2\pi)^3} \int \{(\nabla\theta)^2 + 2(1 - \cos\theta)[(\nabla\varphi)^2 + \sin^2\varphi(\nabla)^2]\} d^3k. \qquad (5.20)$$

The vanishing of $\nabla \cdot \mathfrak{x}$ is necessary and sufficient for the first term to vanish. Since the second and third terms are of at least second order in θ, this is also the requirement for a stationary value of $\overline{\mathfrak{x}^2}$. It is not obviously sufficient although it is necessary for an absolute minimum. This constitutes an existence proof that $\nabla \cdot \mathfrak{x}^{(i)}$ can be made to vanish everywhere, since $\overline{\mathfrak{x}^2}$ is positive definite. The actual solution of this problem appears to be very complicated and we have not solved it. It is worthwhile to point out, however, that if we restrict ourselves to transformation with constant A_n, then $\nabla \cdot \mathfrak{x} = 0$ is a sufficient condition. Also if we are interested in

minimizing the spread of WF, composed from BF's whose diagonal components of magnetic moment are pointing in a given direction, the $\nabla \cdot \mathbf{\underline{r}} = 0$ condition is sufficient.

Turning our attention to the σ_i terms in (5.17), we note that in Appendix B it is shown that $\mathbf{\underline{r}}^{(0)}$ must be of the form $\nabla_\mathbf{k}\varphi$. On integrating the second term in (5.17), we obtain, therefore,

$$\int \mathbf{\underline{r}}^{(i)} \cdot \mathbf{\underline{r}}^{(0)} \, d^3k = \int \mathbf{\underline{r}}^{(i)} \cdot \nabla_\mathbf{k}\varphi \, d^3k = \oint d\mathbf{S} \cdot \mathbf{\underline{r}}^{(i)} \varphi - \int \varphi \nabla_\mathbf{k} \cdot \mathbf{\underline{r}}^{(i)} \, d^3k. \quad (5.20a)$$

The same remarks apply to this surface integral as to that in (5.15) Again excluding the case of lines of additional degeneracy—beyond the twofold degeneracy everywhere—we see that if $\nabla_\mathbf{k} \cdot \mathbf{\underline{r}}^{(i)} = 0$, $\overline{\mathbf{\underline{r}}^{(i)} \cdot \mathbf{\underline{r}}^{(0)}}$ vanishes, so that $\overline{x^2}$ has no σ^i terms.

So far we have not included the treatment of bands with planes of degeneracy. It is shown in the Appendix that the $\overline{x^2}$ is not finite in this case. This is a rather unsatisfactory situation and it is desirable to remove the divergence if possible. The difficulty has its root in the fact that if we follow a path along one of the bands from one plane to the next one in $\mathbf{k}$ space, we come back to a different band. This circumstance, however, enables us to define new WF's which do not have the divergence in x^2. This is done by using both intersecting bands to define the WF. Thus, if we follow one band from $k = -K/2$ to $K/2$ and continue into the next zone without changing the slope of the energy at the degeneracy, we have no discontinuities. If we suppose that we have thus defined one band throughout a double zone (DBZ), we can form WF's defined by

$$a_n{}'(\mathbf{x}; \mathfrak{R}) = \sqrt{\frac{\Omega}{2(2\pi)^3}} \int_{\text{DBZ}} d^3k e^{-i\mathbf{k} \cdot \mathfrak{R}} \psi_{n\mathbf{k}}(\mathbf{x}). \quad (5.21)$$

Here we can permit $\mathfrak{R}$ to be either a lattice vector $\mathbf{R}$ or $\mathbf{R} + \mathbf{a}$ where $\mathbf{a}$ is a vector extending from a lattice point to that point to which it is carried by the twofold screw axis which is present in all crystals with planes of degeneracy. In terms of the original bands

$$a_n{}'(\mathbf{x}; \mathfrak{R}) = \sqrt{\frac{\Omega}{(2\pi)^3}} \int d^3k e^{-i\mathbf{k} \cdot \mathfrak{R}} \frac{(\psi_{n1,\mathbf{k}} + e^{-i\mathbf{K} \cdot \mathfrak{R}} \psi_{n2,\mathbf{k}})}{\sqrt{2}}. \quad (5.22)$$

If

$$\mathfrak{R} = \mathbf{R}, \ a_n{}' = \frac{a_{n1} + a_{n2}}{2}.$$

If

$$\mathfrak{R} = \mathbf{R} + \mathbf{a}, \ a_n{}' = \frac{a_{n1} - a_{n2}}{\sqrt{2}}.$$

The second of these relations is slightly inconsistent since the a's are not defined for $\mathbf{R} + \mathbf{a}$. In this case we define the functions a_{n1}, a_{n2} by substituting $(\mathbf{R} + \mathbf{a})$ for $\mathbf{R}$ in (4.1) anyway. The representation of a' in the CMR is

$$f_1 = \frac{1}{\sqrt{2}} e^{-i\mathbf{k}\cdot\mathfrak{R}}$$

(5.23)

$$f_2 = \frac{1}{\sqrt{2}} e^{-i(\mathbf{k}+\mathbf{K})\cdot\mathfrak{R}}.$$

Thus if $\mathfrak{R} = \mathbf{R}$, f_1 and f_2 are periodic and continuous at the plane, while, if $\mathfrak{R} = \mathbf{R} + \mathbf{a}$, each can be taken periodic, but is then discontinuous. $f_1\psi_1 + f_2\psi_2$ is continuous and periodic in $\mathbf{k}$ space in both cases; the surface contribution discussed in the Appendix accordingly vanishes. It is easily verified that $a'(\mathbf{x}; \mathfrak{R})$ and $a'(\mathbf{x}; \mathfrak{R}')$ are orthogonal.

We can now investigate $\mathbf{x}$ and $\overline{x^2}$ for a'.

$$\mathbf{x} = \mathbf{R} + \frac{\Omega}{(2\pi)^3} \int f_i \mathfrak{X}_{ij} f_j \, d^3k$$

$$= \mathbf{R} + \tfrac{1}{2}(\mathfrak{X}_{11} + \mathfrak{X}_{22} + \mathfrak{X}_{12} + \mathfrak{X}_{21})$$

(5.24)

$$\overline{x^2} = \bar{x}^2 + \sum_{i,j=1,2} \int f_i (\mathfrak{X}^2)_{ij} f_j \, d^3k$$

$$= \bar{x}^2 + \sum_{\substack{n\neq 1,2 \\ i,j=1,2}} f_i \overline{\mathfrak{X}_{in}\cdot\mathfrak{X}_{nj}} f_j + \sum_{i,j,l=1,2} f_i \overline{\xi_{il}\cdot\xi_{lj}} f_j.$$

(5.25)

The phase-sensitive part of this is

$$\overline{\mathfrak{X}_{11}^2} + \overline{\mathfrak{X}_{22}^2} + \overline{(\mathfrak{X}_{22} + \mathfrak{X}_{11})\cdot(\mathfrak{X}_{22} + \mathfrak{X}_{21})}.$$

If we make the phase transformations $e^{i\varphi_1}$, $e^{i\varphi_2}$, we obtain

$$\overline{x'^2} = \overline{x^2} - \frac{\Omega}{(2\pi)^3}$$

$$\cdot \int \left[\varphi_1 \nabla\cdot\left(\mathfrak{X}_{11} + \frac{\mathfrak{X}_{12} + \mathfrak{X}_{21}}{2}\right) + \varphi_2 \nabla\cdot\left(\mathfrak{X}_{22} + \frac{\mathfrak{X}_{12} + \mathfrak{X}_{21}}{2}\right) \right] d^3k.$$

(5.26)

Thus a minimum is achieved if

$$\nabla\cdot\mathfrak{X}_{11} = \nabla\cdot\mathfrak{X}_{22} = -\frac{\nabla\cdot(\mathfrak{X}_{12} + \mathfrak{X}_{21})}{2}.$$

(5.27)

(There is no question of a generalized phase transformation here, because bands which are doubly degenerate throughout k space never have additional degeneracies over entire planes.) Thus the new WF are similar in all their important properties to those in a medium without plane degeneracies. They do not satisfy the Parzen-Koster variational principle, however, since more than one f is finite everywhere.

We have devoted time to these functions because an important class of material has this plane degeneracy. In general, this happens to all bands in crystals having twofold screw axes perpendicular to primitive reciprocal lattice vectors, except if inversion symmetry is present *and* spin-orbit splitting is not neglected. This includes, in particular, close-packed hexagonal crystals. The new functions are clearly better suited to these materials than the normal WF, for the latter are assigned only to atoms in planes separated by the lattice spacing, whereas in these materials there are always other identical atoms in identical environments on planes spaced twice as closely. There is an a' for each of these atoms also. Thus the a' conform much more to our notion of localized functions about each atom than do the standard WF.

We now turn to the question of the asymptotic behavior of WF's. This discussion will be divided in two parts. The first is concerned with WF's of bands which do have degeneracies at a given $\mathbf{k}$ with other bands, the second with bands with no such degeneracies. To discuss the former we must see how the X's behave in the neighborhood of degeneracies.

The general behavior is fairly trivial. Near a degeneracy, the Bloch function is, in lowest approximation, a linear combination of the BF's at the degeneracy, depending only on the direction of the shortest vector from the point to the degeneracy. That is $S(k_0, \mathbf{k}) = S(\theta, \varphi)$. Since $X = S^{-1} \partial S/\partial \mathbf{k}$, $\mathbf{X} \times \mathbf{k}/|k|$ varies as $|k|^{-1}$ while $\mathbf{X} \cdot \mathbf{k}/|k|$ is zero in this approximation and, in general, varies as $|\mathbf{k}|$ or $|k|^2$. Thus $|X_{nn'}|^2$ behaves as $|k|^{-2}$. Its integral diverges logarithmically for a curve of degeneracy and is finite for a point of degeneracy. The integral of $|X^2|^2$ diverges as k^{-1} for a point. Formally, X^3 would diverge logarithmically; however, the Hamiltonian must have twofold symmetry about a degeneracy to lowest orders. Thus the divergent contribution to the X^3 integrals vanishes, in fact. Hence if a band has a curve of degeneracy with another band, $\overline{x^2}$ diverges for its WF. In contrast only, $\overline{x^4}$ diverges if it has a point of degeneracy. In fact, the WF falls off like x^{-2} if there is a line of degeneracy, and like x^{-3} if there is point degeneracy. This follows from the fact that a logarithmic divergence of the integral over $\mathbf{k}$ corresponds to a logarithmic divergence of the spatial integral if we consider periodic boundary conditions.

The integrals of all powers of x would appear to converge if there are no degeneracies, and one would anticipate an exponential falloff of the WF. To investigate this point more closely, we recall that

$$\overline{x^l} = \frac{\Omega}{(2\pi)^3} \int d^3k \lim_{\mathbf{k}' \to \mathbf{k}} \left(\frac{i\partial}{\partial \mathbf{k}'}\right)^l S(\mathbf{k}, \mathbf{k}'). \tag{5.28}$$

Hence

$$e^{\mathbf{\kappa} \cdot \mathbf{x}} = \lim_{\mathbf{k}' \to \mathbf{k}} \frac{\Omega}{(2\pi)^3} \int d^3k \sum_l \frac{1}{l!} \left(i\mathbf{\kappa} \cdot \frac{\partial}{\partial \mathbf{k}'}\right)^l S(\mathbf{k}, \mathbf{k}'). \tag{5.29}$$

Thus the integral has a finite value if $\mathbf{\kappa}$ is less than the minimum radius of convergence of the power series expansion of $S(\mathbf{k}, \mathbf{k}')$ for $\mathbf{k} - \mathbf{k}'$ parallel to $\mathbf{\kappa}$. This implies that $S(\mathbf{k}, \mathbf{k}')$ must be defined and analytic for complex values of $\mathbf{k}$ if convergence is to occur for any $\mathbf{\kappa}$, no matter how small. Another approach to this point is more direct. Thus

$$\int |a|^2 e^{-i\mathbf{q} \cdot \mathbf{x}} \, dV = \frac{\Omega}{(2\pi)^3} \int d^3k \, S\left(\mathbf{k} - \frac{\mathbf{q}}{2}, \mathbf{k} + \frac{\mathbf{q}}{2}\right). \tag{5.30}$$

Therefore if $S(\mathbf{k}, \mathbf{k}')$ can be extended into the complex plane as far as κ, the same result may be obtained simply by letting $\mathbf{q} = i\mathbf{\kappa}$. We demonstrate in the Appendix that this is possible for nondegenerate bands, the limits arising from a surface of branch points. This has also been shown by Kohn[12] for one-dimensional bands by a different method not easily generalized to three dimensions. Thus we can say that $\int |a|^2 e^{2\mathbf{\kappa} \cdot \mathbf{x}}$ is finite if κ is less than κ_0, the minimum distance of the branch surface from the real k space in the direction of κ. It is easily seen that this means that $|a|$ falls off as $e^{-|\kappa_0||x|}$ in this direction.

6. Problems Caused by Degeneracies

Just as the asymptotic behavior of the a's depends on the presence of degeneracies, so does that of $H(R)$ itself. Let us consider first the effect of degeneracies near which the bands separate linearly.

(1) The sum of the squares of the Fourier transform of the first derivative converges, if the degeneracy occurs over a plane, while this is not true for the second derivative with respect to the component of $\mathbf{\kappa}$ perpendicular to the plane. This, in fact, diverges linearly, so that $H(\mathbf{R})$ goes as R^{-2} for R in this same direction, and becomes small exponentially for R in directions parallel to the plane.

[12] W. Kohn, *Phys. Rev.* **115**, 809 (1959).

(2) For line degeneracies, the divergence is weaker. The first derivative of $H(\mathbf{k})$ approaches a constant which depends on the direction of approach to the line, so the second derivative becomes large as $|k|^{-1}$. Thus the integral of the square of $\nabla^2 H$ diverges logarithmically. Correspondingly, the sum $\Sigma_R |R^2 H|^2$ diverges logarithmically, which implies that $|H(\mathbf{R})|$ goes as $|R|^{-3}$ for $\mathbf{R}$ in the plane perpendicular to the degeneracy.

(3) For point degeneracy, the foregoing arguments show that $\Sigma |R^2 H(\mathbf{R})|^2$ converges. In this case we must consider the third derivatives, some of which clearly go as $|k|^{-2}$ near the degeneracy. Thus the integral of $|\nabla^3 H|^2$ diverges as $|k|^{-1}$, corresponding to a linear divergence in the sum $\Sigma |R^3 H|^2$. This implies that $|H|$ goes asymptotically as $|R|^{-4}$.

In the case of line degeneracies at which the bands separate quadratically, the second derivatives of the energy approach a constant near the line, while the third derivatives go as $|k|^{-1}$. Thus in this case $\Sigma |R^3 H|^2$ diverges logarithmically and $|H(R)|$ decreases like $|R|^{-4}$ at large $|R|$. This behavior again applies only in directions normal to the line of degeneracy, the decrease being exponential in the direction of this line.

Finally, in order to see how the representation of slowly varying perturbations are affected by band degeneracies, we shall consider the matrix elements of $e^{i\mathbf{q}\cdot\mathbf{x}}$. Its matrix elements are

$$S_{nn}'(\mathbf{R}', \mathbf{R}; \mathbf{q}) = \int d^3k \, d^3k' e^{i\mathbf{k}'\cdot\mathbf{R}'} e^{-i\mathbf{k}\cdot\mathbf{R}} S_{nn}''(\mathbf{k}', \mathbf{k}) \, \delta(\mathbf{k}' - \mathbf{k} - \mathbf{q})$$

$$= e^{i\mathbf{q}\cdot\mathbf{R}} \int d^3k \, S_{nn}'(\mathbf{k}, \mathbf{k} - \mathbf{q}) e^{-i\mathbf{k}\cdot(\mathbf{R}-\mathbf{R}')}. \tag{6.1}$$

If there are no degeneracies, $S_{nn}'(\mathbf{k}, \mathbf{k} - \mathbf{q})$ approaches S_{nn}' as $\mathbf{q}$ approaches zero, for all $\mathbf{k}$, so that for small $\mathbf{q}$

$$S_{nn}'(\mathbf{R}', \mathbf{R}; q \to 0) \sim \delta_{nn}' e^{i\mathbf{q}\cdot\mathbf{R}} \, \delta(\mathbf{R} - \mathbf{R}'). \tag{6.2}$$

However, S_{nn}' does not approach δ_{nn}' in the neighborhood of degeneracies. As a result, there are contributions to S which not only have interband elements but also connect WF's centered at different R's. These fall off with $(R - R')$ only as small inverse powers of $|R - R'|$. Such contributions may be small numerically but can seriously affect the asymptotic behavior of the wave functions.

Thus, the WF's, H_0, and V all acquire undesirable characteristics when the bands have degeneracies, even though these may occur for energies far from those of interest. In particular, the use of the difference equations involving only a few nonzero terms $H_0(\mathbf{R} - \mathbf{R}')$ and $V(\mathbf{R})$ is clearly

unjustified in localized problems and could lead to improper asymptotic behavior of the solutions. On the other hand, the results obtained seem physically correct and one may assume that the concomitant bad behavior of the WF's somehow compensates for this. Nevertheless, one would like a more justifiable procedure.

One such procedure is to use not Wannier functions, but combinations of them defined as follows

$$\alpha_n(\mathbf{x} - \mathbf{R}) = \sqrt{\frac{\Omega}{(2\pi)^3}} \int d^3k e^{-i\mathbf{k}\cdot\mathbf{R}} \psi_{n'k}(\mathbf{x}) \, T_{n'n}(\mathbf{k})$$

in which $T(\mathbf{k})$ is some unitary matrix. The quantities $\Sigma_{n'} \psi_{n'} T_{n'n} = \varphi_n$ define new functions which are not energy eigenfunctions but could be chosen so as to have no singularities when regarded as functions of $\mathbf{k}$. For instance, one could add a periodic perturbation represented by $\Delta_{nn'}(\mathbf{k})$ which would remove all degeneracies between bands. The WF's for the newly defined bands would then be well behaved, as would H_0 and V. To be useful at an energy E, one would want to be sure that $T_{n'n} \cong \delta_{n'n}$ for $\mathbf{k}$'s such that $H(\mathbf{k}) \sim E$; this would require only that $\Delta_{nn'}'(\mathbf{k})$ be very small at such $\mathbf{k}$. This change in the energy spectrum can be justified by the arguments used to justify the approximate Hamiltonian $H'(\mathbf{k})$. It is worthwhile to note that in actual problems attacked with the use of the WR, one does not know the WF's anyway, and attempts only to get the general shape of the envelope. The α's are quite as useful as the a's for this purpose. This argument also justifies the common use of no more than nearest neighbor values of $H(R - R')$, for, near the bottom of the band where the effective mass Hamiltonian is valid, these terms are, or can be chosen to be, sufficient to describe the band shape. This will be a good approximation if the resulting levels are reasonably close to the band edge, which can be tested as in Eq. (4.21).

Another procedure for getting well-behaved pseudo-WF's will be applicable even when the degeneracies are close to the energy of interest. This situation is possible when a group of bands is isolated, or can be isolated, as in the last paragraph, from the other bands. One can then choose the $T_{nn'}$ not to be $\delta_{nn'}$ at the energy of interest, but to make WF's which are very well localized, and have desirable symmetry properties. Our treatment of structures with planes of degeneracy is an example of this procedure. Another example is provided by the valence band of Ge and Si, where four bands apparently intersect no other bands. In this case, we assert without demonstration that one can define bond functions, one for each of the four valence bands in a unit cell. They correspond to the bonding orbitals used by chemists. They have been used by the author in

a very crude discussion of energy levels at vacancies and interstitials in Ge.[13]

Either of these procedures will result not only in well-behaved WF's and $H(R - R')$, but also in better behaved values of $V(R)$ because of the improved localization. In particular, the first procedure constitutes a justification of the results of WR calculations which have been made for crystals with bands which have degeneracies.

III. A "Mixed Representation"

7. BASIC PROPERTIES

In this section, we shall develop a new procedure for working with slowly varying perturbations, which is less restricted by the strength of the perturbations than the CMR and WR procedures. It will be seen that the formalism is in a sense intermediate between these two, although it is not, properly speaking, a representation in the quantum-mechanical sense at all.

We begin by considering an operator $O_{mn}(\mathbf{k}, \mathbf{k}')$ in the CMR, and rewrite it as follows.

(1) We rewrite

$$O_{mn}(\mathbf{k}', \mathbf{k}'') = O_{mn}\left(\frac{\mathbf{k}' + \mathbf{k}''}{2}; \mathbf{k}' - \mathbf{k}''\right)$$

$$= O_{mn}(\mathbf{k}; \mathbf{q}) \tag{7.1}$$

where the semicolon is used to distinguish the function $O(\mathbf{k}; \mathbf{q})$ from $O(\mathbf{k}, \mathbf{k}')$. $O_{mn}(\mathbf{k}', \mathbf{k}'')$ is periodic in $\mathbf{k}'$ and $\mathbf{k}''$. $O_{mn}(\mathbf{k}; \mathbf{q})$ on the other band is not periodic in $\mathbf{q}$, but has the following properties

$$O(\mathbf{k} + \mathbf{K}; \mathbf{q}) = O(\mathbf{k}; \mathbf{q})$$

$$\tag{7.2}$$

$$O\left(\mathbf{k} + \frac{\mathbf{K}}{2}; \mathbf{q} + \mathbf{K}\right) = O(\mathbf{k}; \mathbf{q})$$

where $\mathbf{K}$ is any reciprocal lattice vector.

(2) We define

$$O_{mn}(\mathbf{k}, \mathbf{R}) = \int d^3q \, O_{mn}(\mathbf{k}; \mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{R}} \tag{7.3}$$

[13] E. I. Blount, *Phys. Rev.* **113**, 995 (1959).

which is then analogous to a classical quantity. (The reverse transform

$$O_{mn}(\mathbf{k}; \mathbf{q}) = \sum_R O_{mn}(\mathbf{k}, \mathbf{R})e^{-i\mathbf{q}\cdot\mathbf{R}} \tag{7.4}$$

must be used with caution, for it is periodic in $\mathbf{q}$, and can thus be correct only in the first BZ of $\mathbf{q}$ space.) The quantum nature of $O(\mathbf{k}, \mathbf{R})$ lies in the multiplication rules which are derived in Appendix D, where final result for $P(\mathbf{k}, \mathbf{R})$ if $P = NO$ is given, with certain limitations, by

$$P(k, R) = \int d^3K_1\, d^3K_2 \sum_{\substack{i,j \\ R_1 R_L}} \frac{\partial^j N[(k + (K_1/2), R)]}{\partial R^j} \frac{\partial^l O(k + (K_2/2), R)]}{\partial R^l}$$

$$\times \frac{i^{j+l}(-1)^j}{j!l!} \frac{\partial^j}{\partial K_2^j} \frac{\partial^l}{\partial K_1^l} \exp\left[i(K_1\cdot R_2 - K_2\cdot R_1)\right]$$

$$= \sum_{j+l} \frac{(-1)^j(-i)^{j+l}}{2^{j+l}j!l!} \frac{\partial^{j+l}N}{\partial k^l\, \partial R^j} \frac{\partial^{j+l}O}{\partial k^j\, \partial R^l}$$

$$P(k, R) = N(k, R)O(k, R) - \frac{i}{2}\left(\frac{\partial N}{\partial k}\frac{\partial O}{\partial R} - \frac{\partial N}{\partial R}\frac{\partial O}{\partial k}\right)$$

$$- \frac{1}{8}\left(\frac{\partial^2 N}{\partial k^2}\frac{\partial O}{\partial R^2} - \frac{2\partial^2 N}{\partial k\, \partial R} + \frac{\partial^2 N}{\partial R^2}\frac{\partial^2 O}{\partial k^2}\right) + \cdots. \tag{7.5}$$

(To avoid complication of the equation, this is written as if for one dimension, but the generalization to three dimensions is trivial.)

The first term is just the classical, local product. The second is half the Poisson bracket (P.B.), which we shall designate by angular brackets,

$$\frac{\partial N}{\partial \mathbf{R}}\frac{\partial O}{\partial \mathbf{k}} - \frac{\partial N}{\partial \mathbf{k}}\frac{\partial O}{\partial \mathbf{R}} \equiv \langle N, O \rangle. \tag{7.6}$$

The remaining terms have no classical analogies, but may be thought of as higher order P.B.'s.

When we are concerned with functions whose variation with $\mathbf{R}$ is slow, characterized by a maximum wave number $\mathbf{q}$ which is small, the terms in Eq. (7.5) are of order $|\mathbf{q}|^{l+j}$. In this sense we may speak of them as successive orders in an expansion whose parameter is the speed of variation of the functions with $\mathbf{R}$. This is the point of view we shall adopt; when we speak of functions of some "order," it will be meant in this sense. We might also say that the expansion parameter is $(i\, \partial/\partial R)$.

8. Unitary Operators and Perturbation Theory

It is easy to show that a Hermitian operator $O_{mn'}(\mathbf{k}, \mathbf{k'})$ yields a Hermitian matrix $O_{mn'}(\mathbf{k}, \mathbf{R})$. The identity $\delta_{mn}\, \delta(\mathbf{k} - \mathbf{k'})$ is simply 1. On the other hand, the criterion for unitarity does not have a simple form. We require, in the general case [see (D5)],

$$\sum \int_I U\left(k + \frac{K_1}{2}, R + R_1\right) U^\dagger\left(k + \frac{K_2}{2}, R + R_2\right) e^{i(K_1 \cdot R_2 - K_2 \cdot R_1)} = 1.$$

A case where the criterion is simple is that of a matrix independent of $\mathbf{R}$. Then, $U(k)$ must be a unitary matrix. Similarly, if U is independent of $\mathbf{k}$, $U(\mathbf{R})$ must be a unitary matrix. For U's which vary slowly with R, the expansion [Eq. (7.5)] is useful

$$U(k, R)\, U^\dagger(k, R) + \frac{i}{2}\left(\frac{\partial U}{\partial k}\frac{\partial U^\dagger}{\partial R} - \frac{\partial U}{\partial R}\frac{\partial U^\dagger}{\partial k}\right) + \cdots = 1.$$

It will be convenient to consider unitary operators to be generated by unitary matrices by an iterative procedure. We write such an operator in the following form

$$U = \prod_{i=1}^{\infty} (1 + G^{(i)})\, U^{(0)} \tag{8.1}$$

$$^{(n)}U = \prod_{i=1}^{n} (1 + G^{(i)})\, U^{(0)} \tag{8.2}$$

$$U^{(0)}(\mathbf{k}, \mathbf{R})\, U^{(0)\dagger}(\mathbf{k}, \mathbf{R}) = 1.$$

The superscript on the right will signify a quantity of nth order, while the superscript on the left means that all orders up to n may be present. It will also be necessary to distinguish between matrix multiplication and operator multiplication. For the rest of the section, operator products will be indicated by an asterisk $P = N * O$.

$G^{(1)}$ can be partially determined by the equation

$$G^{(1)} + G^{(1)\dagger} + \frac{i}{2}\langle U^{(0)}, U^{(0)\dagger}\rangle = 0. \tag{8.3}$$

For the nth order we will obtain

$$^{(n)}U * {}^{(n)}U^\dagger = 1 + \Delta^{(n+1)} + \Delta_R \tag{8.4}$$

where $\Delta^{(n+1)}$ is of order $(n + 1)$ and Δ_R is of higher order. Then $G^{(n+1)}$ is chosen so that

$$G^{(n+1)} + G^{(n+1)\dagger} + \Delta^{(n+1)} = 0. \tag{8.5}$$

So far we have determined only the Hermitian part of the G's. This leaves us with the freedom needed to diagonalize any particular operator O. We seek U such that

$$U * O * U^\dagger = O_D$$

where O_D is diagonal. Thus we require

$$U^{(0)} O U^{(0)\dagger} = O_D^{(0)}$$

while

$$U^{(0)} * O * U^{(0)\dagger} = O_D^{(0)} + O^{(1)} + O_R.$$

Hence the determination of $G^{(1)}$ is completed by

$$G^{(1)} O_D^{(0)} + O_D^{(0)} G^{(1)\dagger} + O^{(1)} = 0.$$

In general order

$$^{(n)}U * O * {}^{(n)}U^\dagger = {}^{(n)}O_D + O^{(n+1)} + O_R \qquad (8.6)$$

and

$$G^{(n+1)} O_D^{(0)} + O_D^{(0)} G^{(n+1)\dagger} + O^{(n+1)} = O_D^{(n+1)}.$$

The off-diagonal elements yield

$$G^{(n+1)} O_D^{(0)} + O_D^{(0)} G^{(n+1)\dagger} = -O^{(n+1)}.$$

Combining with (8.5)

$$\tfrac{1}{2}[(G^{(n+1)} - G^{(n+1)\dagger}), O_D^{(0)}] = -O^{(n+1)} - \tfrac{1}{2}\{\Delta^{(n+1)}, O_D^{(0)}\} \qquad (8.7)$$

where the curly brackets mean the anticommutator. The diagonal elements of (8.6) yield

$$O^{(n+1)} = O_d^{(n+1)} - \Delta^{(n+1)} O_D^{(0)} \qquad (8.8)$$

where $O_d^{(n+1)}$ is the diagonal part of $O^{(n+1)}$. We now have a complete iteration procedure for the diagonalization of O. Of course we cannot expect that it will converge, a failing it shares with normal perturbation theory.

We also note that the imaginary parts of the diagonal elements of $G^{(i)}(k, R)$ are not determined. This simply means that we can at each stage make an intraband canonical transformation, as in classical physics. The only important difference is that when a time-independent canonical transformation is made in classical physics, the Hamiltonian is generally considered to be the same function of p, q which, however, are expressed in terms of the transformed variables P, Q. The effect of a one-band unitary operator in our present formalism is to leave $\mathbf{k}$, $\mathbf{R}$ invariant while changing the functional form of the Hamiltonian and also the relation between actual observables to $\mathbf{k}$, $\mathbf{R}$. The procedures are equivalent, but the labeling is different. In the application of our procedure, this freedom can be used to make the most convenient choice of $G^{(i)}$.

The procedure given in the foregoing is the most direct and naive perturbation theory, and could undoubtedly be made more efficient as is the case with the Rayleigh-Schrödinger perturbation theory.

We will now go through this procedure for a Hamiltonian to first order in addition to setting the Hamiltonian up in the mixed representation in the first place. We want the matrix elements $H_{nn'}(\mathbf{k}, \mathbf{R})$ of the Hamiltonian $H(\mathbf{p}, \mathbf{x})$

$$H_{nn'}(\mathbf{k}, \mathbf{R})$$

$$= \int d^3q e^{i\mathbf{q}\cdot\mathbf{R}}(\psi_{n,\mathbf{k}+(q/2)}, H(\mathbf{p}, \mathbf{x})\psi_{n,\mathbf{k}-(q/2)})$$

$$= \int d^3q(u_{n,\mathbf{k}+(q/2)}, [e^{-(i/2)q\cdot(\mathbf{x}-\mathbf{R})}e^{-i\mathbf{k}\cdot\mathbf{x}}H(\mathbf{p}, \mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}e^{-i(q/2)\cdot(\mathbf{x}-\mathbf{R})}]u_{n,\mathbf{k}-(q/2)}).$$

$$(8.9)$$

We define the operator

$$\mathfrak{H}(\mathbf{k}, \mathbf{R}) \equiv \mathfrak{H}(\mathbf{k}, \mathbf{R}; \mathbf{p}, \mathbf{x})$$

$$= \int d^3q e^{-(i/2)(q/2)\cdot(\mathbf{x}-\mathbf{R})}e^{-i\mathbf{k}\cdot\mathbf{x}}H(\mathbf{p}, \mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}e^{-i(q/2)\cdot(\mathbf{x}-\mathbf{R})}. \qquad (8.10)$$

$\mathfrak{H}$ is an operator function of $\mathbf{p}$ and $\mathbf{x}$ with indices $\mathbf{k}$ and $\mathbf{R}$. For $\mathbf{x}$ far from $\mathbf{R}$, $\mathfrak{H}$ is very small. If $H(\mathbf{p}, \mathbf{x})$ is nearly periodic, it will be very nearly the same as the $\mathfrak{H}$ defined for a periodic H which, near $\mathbf{R}$, closely approximates $H(\mathbf{p}, \mathbf{x})$. The matrix elements of $\overline{\mathfrak{H}(\mathbf{k}, \mathbf{R})}$ between the $u_{n\mathbf{k}}$ will be called $\mathfrak{H}_{nn'}(\mathbf{k}, \mathbf{R})$, so on expanding $u_{n,\mathbf{k}\pm(q/2)}$ we find

$$H_{nn'}(\mathbf{k}, \mathbf{R}) = \mathfrak{H}_{nn'}(\mathbf{k}, \mathbf{R}) + \tfrac{1}{2}\left\{\frac{\partial\mathfrak{H}}{\partial\mathbf{R}}\cdot\mathfrak{X}(\mathbf{k})\right\}_{nn'} \qquad (8.11)$$

when the second term is of first order in our sense.

We now apply the transformation $U^{(0)}(\mathbf{k}, \mathbf{R})$ which we immediately choose to make unitary to first order by multiplying on the left by

$$[1 + \tfrac{1}{2}(\partial U^{(0)\dagger}/\partial\mathbf{R})\cdot(\partial U^{(0)\dagger}/\partial\mathbf{k})].$$

Keeping only first-order terms we obtain

$$^{(1)}H_0 = E(\mathbf{k}, \mathbf{R}) - \tfrac{1}{2}\{F(\mathbf{k}, \mathbf{R})\cdot\mathfrak{X}(\mathbf{k}, \mathbf{R})\} + \tfrac{1}{2}\{\mathbf{v}(\mathbf{k}, \mathbf{R})\cdot\mathbf{a}(\mathbf{k}, \mathbf{R})\} \qquad (8.12)$$

where $E(\mathbf{k}, \mathbf{R})$ is the diagonal form of the matrix $\mathfrak{H}_{nn'}(\mathbf{k}, \mathbf{R})$,

$$\mathbf{v}(\mathbf{k}, \mathbf{R}) = \frac{1}{\hbar} \frac{\partial E}{\partial \mathbf{k}}$$

$$F(\mathbf{k}, \mathbf{R}) = -U^{(0)} \frac{\partial \mathfrak{H}}{\partial \mathbf{R}} U^{(0)}(\mathbf{k}, \mathbf{R})$$

$$\mathfrak{X}(\mathbf{k}, \mathbf{R}) = U^{(0)}(\mathbf{k}, \mathbf{R}) \mathfrak{X}(\mathbf{k}) U^{(0)\dagger}(\mathbf{k}, \mathbf{R}) + \frac{i}{2} \left(U^{(0)} \frac{\partial U^{(0)\dagger}}{\partial \mathbf{k}} - \frac{\partial U^{(0)\dagger}}{\partial \mathbf{k}} U^{(0)\dagger} \right)$$

$$\mathbf{a}(\mathbf{k}, \mathbf{R}) = \frac{-i}{2} \left(U^{(0)} \frac{\partial U^{(0)}}{\partial \mathbf{R}} - \frac{\partial U^{(0)}}{\partial \mathbf{R}} U^{(0)\dagger} \right). \tag{8.13}$$

Thus $\mathbf{v}$ is, of course, the expectation value of the velocity, F is essentially a force, $\mathfrak{X}(\mathbf{k}, \mathbf{R})$ is the required modification of $\mathfrak{X}(\mathbf{k})$, and $\mathbf{a}$ acts like a vector potential. If $H(\mathbf{p}, \mathbf{x})$ represents a constant electric field applied to the crystal, $\mathbf{a}$ vanishes, F is just the electric field, and all remaining quantities are independent of $\mathbf{R}$. The first-order interband terms could now immediately be removed, as in perturbation theory. The intraband term in $\mathbf{a}$ can be modified by the addition of any gradient by means of the transformation $e^{i\varphi(\mathbf{k}, \mathbf{R})}$. In particular it can be made solenoidal as can $\mathfrak{x}(\mathbf{k}, \mathbf{R})$, though it is not clear that they can, in general, be made solenoidal simultaneously.

    The meaning of our results so far is clear. If a perturbation varies slowly in space an electron adjusts adiabatically to the changes and sees only the changes in its energy. This amounts to a generalization of the "deformation potential" idea to an arbitrary nearly periodic perturbation; similarly it is related to the common procedure of drawing "bent" bands in regions of high field such as junctions. Previous treatments have been valid only for weak perturbations or for nearly constant fields. The present derivation will apply as well to strong nearly periodic perturbations, and it provides a straightforward means of proceeding to higher accuracy. It is particularly useful for the treatment of electrons in a magnetic field, as is shown in a forthcoming paper by the author.

    We have presented this formalism as a means of manipulating operators, which is particularly well-suited to nearly periodic perturbations, and have treated the objects $O_{nn'}(\mathbf{k}, \mathbf{R})$ as numbers with certain multiplication rules. It should be pointed out, however, that at the same time this is the natural way to obtain objects which are functions of the operators $\mathbf{k}$ and $\mathbf{R}$, just as quantum mechanics customarily deals with functions of the operators $\mathbf{p}$ and $\mathbf{x}$. The difference is that in most elementary quantum-

mechanical problems, the operators are given originally as functions of **p** and **x**, which are then given explicit quantum-mechanical representations, possibly as matrix elements. Our situation is the reverse. From the Schrödinger representation, we find matrix elements which we then convert to operator form as functions of **k** and **R**. We take the matrix multiplication as basic and derive the rule (7.5) which is functional but only approximate. While the analog of (7.5) is valid in regular quantum mechanics, it is of little use. Its primary use to us is to remove interband matrix elements in the perturbation theory. This situation does not normally occur in normal quantum-mechanical problems (though it does occur in the Dirac equation, and our formalism might well be useful in the removal of matrix elements between positive- and negative-energy states, as an alternative to that of Foldy and Wouthuysen.[14]) There, one is usually concerned with finding states which diagonalize the Hamiltonian. This implicitly involves finding a new momentum variable P which is in general of a different character from **p**, so that no useful one-to-one correspondence can be found between **p** and **P**. This is the same situation which confronts us in the treatment of the one-band Hamiltonian, once we have it.

So far in this section we have developed a procedure for manipulating operators and removing the interband matrix elements of the Hamiltonian to any desired order. Let us now suppose that we have in this way obtained a one-band Hamiltonian as outlined. We can then proceed in two different fashions. The first is simply to treat **k**, **R** as quantum-mechanical operators and determine energy levels or scattering effects by standard quantum-mechanical methods in either the **k** or **R** representations. This seems to require no further discussion.

Frequently, however, we will be satisfied to obtain pseudoclassical information from the equations of motion of **k** and **R**. We shall proceed to set up equations like Hamilton's equations for **k**, **R**.

We suppose that by a transformation U we have obtained a diagonal Hamiltonian $H(k, R)$. As a result of U, **x** is no longer represented by **R** + $\mathfrak{X}$, but by

$$\mathbf{x} = \mathbf{R} + U * \mathfrak{X} * U\dagger + \frac{i}{2}\left(U * \frac{\partial U\dagger}{\partial \mathbf{k}} - \frac{\partial U}{\partial \mathbf{k}} * U\dagger\right).$$

To lowest order this is **R**. The complete intraband part of **x** is written as **R** + $\mathfrak{x}$(**k**, **R**), with

$$\mathfrak{x}(\mathbf{k}, \mathbf{R}) = U * \mathfrak{X} * U\dagger + \frac{i}{2}\left(U * \frac{\partial U\dagger}{\partial \mathbf{k}} - \frac{\partial U}{\partial \mathbf{k}} * U\dagger\right). \qquad (8.14)$$

[14] L. L. Foldy and S. A. Wouthuysen, *Phys. Rev.* **78**, 29 (1950).

The intraband velocity becomes

$$\mathbf{v}(\mathbf{k}, \mathbf{R}) = -\frac{i}{\hbar} [\mathbf{R} + \mathfrak{x}(\mathbf{k}, \mathbf{R}) *, H(\mathbf{k}, \mathbf{R})]$$

or, to first order,

$$= \frac{1}{\hbar} \frac{\partial H}{\partial \mathbf{k}} - \frac{i}{\hbar} \langle \mathfrak{x}, H \rangle. \tag{8.15}$$

The second term is a generalization of the anomalous velocity discussed in Part I, Eq. (2.22). The entire motion can be determined from the equations

$$\dot{\mathbf{R}} = \frac{1}{\hbar} \frac{\partial H}{\partial k}$$

$$\tag{8.16}$$

$$\dot{\mathbf{k}} = -\frac{1}{\hbar} \frac{\partial H}{\partial R}$$

which are identical in form to Hamilton's equations.

In many quantum-mechanical problems one is concerned with the density matrix $\rho(k, k')$. This is a Hermitian matrix which immediately is seen to take the form $\rho(k, R)$ in our formalism. It is the precise analog of the distribution function in classical mechanics. The equation of motion of ρ is

$$\frac{\partial \rho}{\partial t} = \frac{i}{\hbar} [\rho, *H].$$

The operator product can be evaluated by (7.5) and we obtain

$$\frac{\partial \rho}{\partial t} = i[\rho, H] = \frac{i}{2} \langle \rho, H \rangle - \frac{i}{2} \langle H, \rho \rangle + \cdots \text{ (matrix products)}. \tag{8.17}$$

If we suppose that we have removed interband components of H, ρ will be intraband and we have

$$\frac{\partial \rho}{\partial t} = i \langle \rho, H \rangle \tag{8.17a}$$

as in classical problems.

The formalism in this section is closely related to Wigner's[15] phase space density and the developments based on it due to Moyal.[16]

[15] E. P. Wigner, *Phys. Rev.* **40**, 749 (1932).
[16] J. E. Moyal, *Proc. Cambridge Phil. Soc.* **45**, 99 (1949).

A different procedure for obtaining some degree of localization in both
k and **R**, has been proposed by McIrvine and Overhauser.[16a] Their pro-
cedure uses a true quantum-mechanical representation which they call the
"superlattice representation." The basis functions are wave packets which
are partially localized in both **k** space and **R** space, the relative localization
in each being quite arbitrary. Wannier functions and Bloch functions are
both special cases. It is our opinion that the complicated nature of the basis
functions leads to similar complication of matrix elements, with the result
that the procedure we have outlined above is considerably more con-
venient to use in most cases.

IV. The Dirac Equation

In order to convey a more intuitive feeling for some of the subjects
discussed in this paper, we shall discuss the Dirac equation in the ter-
minology of band theory. While there is no lattice involved in the Dirac
theory, the positive and negative energy states do form a pair of bands. In
the CMR the only effect of absence of a lattice is to remove the periodic
character of **k** space.

In order to bring the Dirac equation into a form convenient for this
discussion, it is necessary to transform from the usual representation to a
representation whose basic functions—analogs of BF's—are eigenfunctions
of the Hamiltonian. This is analogous to transforming from the K-L
functions to Bloch functions. This transformation is elementary and has
been given by Foldy and Wouthuysen.[14] It is

$$U = \frac{(E_p + \beta H)}{\sqrt{2E_p(E_p + mc^2)}}$$

$$E_p = \sqrt{m^2c^4 + p^2c^2}.$$

$$\text{(IV.1)}$$

The Hamiltonian takes the form

$$E = \beta E_p.$$

The momentum is unchanged, while x becomes

$$x^\mu = x_c^\mu + X^\mu$$

$$x_c^\mu = i\hbar \frac{\partial}{\partial p_\mu} - \frac{\hbar(\mathbf{v}_p \times \mathbf{d})_\mu}{2(E_p + mc^2)} \sim i\hbar \frac{\partial}{\partial p_\mu} - \frac{\hbar(\mathbf{v}_p \times \mathbf{d})_\mu}{4mc^2}$$

$$\text{(IV.2)}$$

$$X^\mu = \frac{i\hbar c}{2E_p} \beta\alpha_\mu - \frac{i\hbar c^3\beta(\mathbf{\alpha}\cdot\mathbf{p})p_\mu}{2E_p^2(E_p + mc^2)} \sim \frac{i\hbar}{2mc} \beta\alpha_u$$

[16a] E. C. McIrvine and A. W. Overhauser, *Phys. Rev.* **115**, 1531 (1959).

and the velocity is

$$\mathfrak{B}^\mu = v^\mu + V^\mu$$

$$v^\mu = \frac{\beta}{\partial p^\mu}\frac{\partial E_p}{} = \beta v_p{}^\mu \tag{IV.3}$$

$$V^\mu = -c\alpha_\mu + \frac{c^2(\boldsymbol{\alpha}\cdot\mathbf{p})p_\mu}{E_p(E_p + mc^2)} \sim -c\alpha_\mu.$$

We readily verify Eq. (2.14), while the commutator of $\mathbf{x}$ and $\mathfrak{B}$ yields an inverse effective mass of

$$\frac{\partial^2 E}{\partial p^\mu\,\partial p^\nu} = \frac{i}{\hbar}\,[X^\mu, V^\nu] = \frac{\delta^{\mu\nu}\beta c^2}{E_p} - \frac{\beta v^\mu v^\nu}{E_p} \sim \frac{\beta}{m}. \tag{IV.4}$$

This yields the familiar transverse and longitudinal masses applicable perpendicular and parallel to the direction of motion, respectively. We see that they arise in a way entirely similar to the effective mass in band theory.

The term $\hbar(\mathbf{v} \times \boldsymbol{\sigma})/2(E_p + mc^2)$ corresponds to $\boldsymbol{\mathfrak{x}}$ in band theory and represents an electric moment arising from the motion of the magnetic moment in the Dirac theory. It is phase dependent and the form given seems to be the most convenient. The operator Ω has the form

$$\Omega = \hbar\boldsymbol{\sigma}\left(\frac{c^2}{E_p(E_p + mc^2)} + \frac{mc^2 v^2}{2(E_p + mc^2)^2 E_p}\right)$$

$$+ \frac{\boldsymbol{\sigma}\cdot\mathbf{vv}}{2E_p(E_p + mc^2)} \sim \frac{\sigma}{2}\left(\frac{\hbar}{mc}\right)^2 \tag{IV.5}$$

while X^2 is

$$\mathbf{X}^2 = \frac{3}{4}\left(\frac{\hbar c}{E_p}\right)^2 - \frac{\hbar^2 v^2}{2E_p(E_p + mc^2)} + \frac{\hbar^2 v^4}{C^2(E_p + mc^2)^2}$$

$$= \frac{1}{4}\left(\frac{\hbar c}{E_p}\right)^2\left[2 + \left(1 - \frac{v^2}{c^2}\frac{E_p}{E_p + mc^2}\right)\right]^2 \sim \frac{3}{4}\left(\frac{\hbar}{mc}\right)^2. \tag{IV.6}$$

This result confirms our remarks on the similarity between X and the Compton wavelength.

Another operator of interest in magnetic problems is the diagonal-in-band contribution to the magnetic moment from the nonsingular part of $\mathbf{x}$. In band theory this has the form $(e/2c)[(\mathbf{X} \times \mathbf{V}) + 2\boldsymbol{\mathfrak{x}} \times \mathbf{v}]$.

The presence of the factor 2 in the second term is surprising at first sight. A rigorous derivation is outside the proper scope of this paper, and it will appear in a forthcoming paper by the author on the behavior of Bloch electrons in a magnetic field. We shall give a simple-minded derivation here. A charge cloud whose center of charge and mass is at $\mathbf{r} + \bar{\mathbf{x}}$ moves in a magnetic field H with an angular velocity $\omega = (eH/mc)$. It also has internal motion characterized by a current density $\mathbf{j}(x)$ whose integral is zero. Its magnetic moment is then

$$\mathbf{\mu} = \frac{e}{2c} \left\{ [\omega \times (r + \bar{x})] \times [\mathbf{r} + \bar{\mathbf{x}}] + \int j(x) \times \mathbf{x}\, d^3x \right\}$$

$$\mu \cdot H = \frac{e\mathbf{H}}{2c} \cdot \left\{ (\omega \times \mathbf{r}) \times \mathbf{r} + 2(\omega \times \mathbf{r}) \times \mathbf{x} \right.$$

$$\left. + \int \mathbf{j}(\mathbf{x}) \times (\mathbf{x} - \bar{\mathbf{x}})\, d^3x + (\omega \times \bar{\mathbf{x}})^2 \right\}.$$

Its total velocity is $\dot{\mathbf{r}} + \dot{\bar{\mathbf{x}}} = \omega \times (\mathbf{r} + \bar{\mathbf{x}})$ and $\mathbf{\mu} \cdot H$ can be written

$$\mathbf{\mu} \cdot \mathbf{H} = \frac{e\mathbf{H}}{2c} \left\{ \dot{\mathbf{r}} \times \mathbf{r} + 2\dot{\mathbf{r}} \times \bar{\mathbf{x}} + \int \mathbf{j}(\mathbf{x}) \times (\mathbf{x} - \bar{\mathbf{x}})\, d^3x + (\omega \times \bar{\mathbf{x}})^2 \right\}.$$

Thus the term $\dot{\mathbf{r}} \cdot \bar{\mathbf{x}}$ has the extra factor 2, while the third term does not. Now if we identify $\mathbf{r}$ with $i(\partial/\partial\mathbf{k})$, which like $\mathbf{r}$ does not represent the center of the wave packet for an electron, $\dot{\mathbf{r}} = (i/\hbar)(\partial E/\partial\mathbf{k}) = \mathbf{v}$ and $\bar{\mathbf{x}}$ corresponds to $\mathbf{\xi}$. The third term in the last equation then corresponds to the contribution of nondiagonal velocity and position operator, for the average of $\mathbf{j}(\mathbf{x})$ and of $(\mathbf{x} - \mathbf{x})$ are both zero. Thus the second and third terms above correspond to our expression for the magnetic moment and the factor 2 is present in the same way.

In the Dirac theory the diagonal elements of the magnetic moment take the form

$$\frac{e}{2c} (2\mathbf{\xi} \times \mathbf{v} + \mathbf{X} \times \mathbf{V}) = \frac{e\hbar c}{2E_p} \beta \mathbf{\sigma}$$

$$\sim \frac{e\hbar}{2mc} \beta \mathbf{\sigma}.$$

Thus we see that in this theory the spin magnetic moment is the analog of the contribution of $\mathbf{\xi}$ to the magnetic moment in band theory. We also see that the g factor is given at all energies by $e/2m^*c$ where m^* is the

transverse mass. In addition to this term there are interband terms of the magnetic moment, which vanish at $p = 0$.

If we consider a perturbation $V(x)$, the Hamiltonian is

$$H = E_p\beta + V(\mathbf{x}_c + \mathbf{X})$$

$$\cong E_p\beta + V(\mathbf{x}_c) + \frac{\partial V}{\partial \mathbf{x}} \cdot \mathbf{X}$$

$$\cong E_p\beta + V\left(\frac{i\partial}{\partial \mathbf{k}}\right) + \frac{\partial V}{\partial \mathbf{x}} \cdot \frac{\hbar \mathbf{\sigma} \times \mathbf{v}}{2(E_p + mc^2)} + \frac{\partial V}{\partial \mathbf{x}} \cdot \mathbf{X} \sim E_p\beta + V\left(\frac{i\partial}{\partial \mathbf{k}}\right)$$

$$+ \frac{\partial V}{\partial \mathbf{x}} \cdot \frac{\hbar \mathbf{\sigma} x \mathbf{v}}{4mc^2}. \qquad (IV.7)$$

The third term, analogous to the term $\mathbf{F} \cdot \xi_{nn}$ in (2.20′) is the spin-orbit energy. The X term can be removed to higher order by perturbation theory as in the discussion preceding (2.22).

In the presence of the perturbing potential, the velocity has the form (after removing the interband terms)

$$\mathbf{v} = -\frac{i}{\hbar}[\mathbf{x}, H] \cong \beta \frac{\partial E_p}{\partial \mathbf{p}} + \frac{1}{2}\left(\frac{\hbar}{mc}\right)^2 \frac{\partial V}{\partial \mathbf{x}} \times \sigma. \qquad (IV.8)$$

This is not equal to $\partial H/\partial \mathbf{p}$, whose second term would be only half that above. The difference is the term $(i/\hbar)[V, \mathbf{x}_c]$ which can be interpreted as a change in the electric dipole moment as the electron is accelerated. The corresponding term in band theory arises from the variation in $\xi(\mathbf{k})$ with $\mathbf{k}$. This term cannot contribute to the current in a steady-state situation as pointed out by Adams and Blount.[7]

When using the Pauli form for the Hamiltonian including spin-orbit coupling, ξ is customarily disregarded and the velocity is taken to be simply $\partial H/\partial \mathbf{p}$. We see that this leads to no difficulties for steady-state situations, but might in other cases if velocities of this order are observable at all. It is clear that the anomalous velocities (2.22) are entirely analogous to the spin-orbit contribution to the velocity of the Dirac electron, which can be written

$$\mathbf{v}_{s.o.} = +\frac{\partial V}{\partial \mathbf{x}} \times \mathbf{\Omega}.$$

As in the band case we can write the eigenfunctions of the unperturbed Hamiltonian in the Bloch form

$$\psi_{n\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{x}}u_{n\mathbf{k}} \qquad (IV.9)$$

where u_k is now independent of x and is, of course, a four-component symbol describing the Foldy-Wouthuysen one-band functions in terms of the original Dirac basis. We can also establish the relation

$$\mathbf{X}_{nn'} = u_n{}^* i \frac{\partial u_{n'}}{\partial \mathbf{k}} \qquad (\text{IV.10})$$

analogous to Eq. (2.3).

To form localized functions from one band we proceed as in Part II

$$A_n(\mathbf{x} - \mathbf{R}) = \int e^{i\mathbf{k} \cdot (\mathbf{x} - \mathbf{R})} u_{n\mathbf{k}} \, d^3k$$

where R is now a continuous variable. Because of the $\mathbf{k}$ dependence of $u_{n\mathbf{k}}$, these functions are not delta functions, but are spread out over a Compton wavelength as pointed out by Foldy and Wouthuysen. This is also related to the fact that no one-band wavepacket can be more tightly localized than $\sqrt{X^2}$ which in the Dirac theory is equivalent to $\hbar/m^*c$. Foldy and Wouthuysen split x into two parts, a mean position equal to our $i\partial/\partial k$ and a remainder equal to our $\mathfrak{X}_{nn'}$. The diagonal part of the velocity is the motion of $i\partial/\partial k$ or x_c, while the $X_{nn'}$ lead to the off-diagonal velocity elements which at low $\mathbf{p}$ are equal to c. This is the "Zitterbewegung." The fact that it is not completely random gives rise to the spin (IV.7). The diagonal part of X contributes no velocity in the absence of acceleration, but does contribute to the magnetic moment, or vice versa.

We see that there is a perfect analogy between band theory and the Dirac theory, which we hope may clarify the meaning of some of the unfamiliar aspects of both. In a more practical vein, one occasionally has situations in crystals where only two bands seem to matter.[17] In such cases it will sometimes be possible to take over the results of the Dirac theory directly or with only some scale changes required.

V. Self-Consistent Field Theory

In any actual crystal, forces between the electrons cannot be ignored, of course. The band theory is commonly said to be based on a treatment of these forces by the Hartree-Fock method. On the other hand, nearly all formal treatments, including the preceding, use the properties of a strictly one-electron Hamiltonian.

We suppose, therefore, that the electronic wave functions are solutions of the equation

$$H\psi = (H_0 + H_{sc})\psi = \epsilon\psi \qquad (\text{V.1})$$

[17] M. H. Cohen and E. I. Blount, *Phil. Mag.* **5**, 115 (1960).

Here H_{sc} is the self-consistent field, which can be written as

$$H_{sc} = U_D(x) + U_E(x, x')$$

$$U_D(x) = e^2 \int \frac{\rho_D(\mathbf{x})\, d^3x}{|\,x - x'\,|} \qquad\qquad \text{(V.2)}$$

$$U_E(x, x') = e^2 \frac{\rho(x, x')}{|\,x - x'\,|}$$

where $\rho(x, x')$ is the total density matrix of the electrons in the crystal, and $\rho_D(x)$ is its diagonal part. U_D is the "direct" part of the self-consistent potential, while U_E is the exchange term represented as a nonlocal potential. The relation (1.1) does not hold since the velocity operator is still given by (1.2).

On the other hand, the operator $w = -(i/h)[x, H]$ is a very useful one as will be seen in the following. With it one can set up a formalism nearly identical to that in the first part of Part I.

This has been done in part by Kane.[18] The important changes from Part I are twofold. First, we no longer have the relation (1.3), so that the expression for the effective mass in terms of W and X (2.18) contains an additional term. Related to the removal of relation (1.3) is the fact that $H(k + q)$ is no longer a quadratic in q but, in general, an infinite series. These matters have no serious consequences for the formalism. The second change is that the perturbing potential must include a self-consistent term. This will be discussed below. When such a formalism is used, the relation between X and W is the same as that between X and V in Part I, and the effective masses are likewise defined in terms of X and W, the inverse effective mass tensor being now

$$\alpha = \frac{1}{h^2} \frac{\partial^2 \epsilon}{\partial \mathbf{k}\, \partial \mathbf{k}}. \qquad\qquad \text{(V.3)}$$

The representation of operators is the same as in Part I.

When a perturbation is applied to this system, not only is the wave function in which one is interested changed but also all the others. Hence the self-consistent potentials are perturbed and H becomes

$$H + V = H + V_0 + V_{sc} \qquad\qquad \text{(V.4)}$$

where V_0 is the applied perturbation and V_{sc} is

$$V_{sc}(x, x') = \delta(x, x')e^2 \int \frac{\rho_D^{(1)}(x'')}{|\,x - x'\,|}\, d^3x'' + e^2 \frac{\rho^{(1)}(x, x')}{|\,x - x'\,|}. \qquad \text{(V.5)}$$

[18] E. O. Kane, *Phys. and Chem. Solids* **6**, 236 (1958); **8**, 38 (1959).

Here the superscripts indicate the perturbed values of the density matrix. We can now write this in the formalism of Part III as

$$V_{sc}(\mathbf{k}, \mathbf{R}) = \int \sum_{R'} \rho(1, \mathbf{R}') V_{sc}(1, \mathbf{R}'; \mathbf{k}, \mathbf{R}) \, d^3l \qquad (\text{V.6})$$

$$V_{sc}(1, \mathbf{R}; \mathbf{k}, \mathbf{R}) = V_D(1, \mathbf{R}'; \mathbf{k}, \mathbf{R}) - V_E(1, \mathbf{R}'; \mathbf{k}, \mathbf{R})$$

$$V_D(1, \mathbf{R}'; \mathbf{k}, \mathbf{R}) = \sum_{\mathbf{K}} \int d^3t \, S\left(\mathbf{k} + \frac{\mathbf{t}}{2} + \mathbf{K}, \mathbf{k} - \frac{\mathbf{t}}{2}\right)$$

$$\cdot S\left(1 - \frac{\mathbf{t}}{2} - \mathbf{K}, 1 + \frac{\mathbf{t}}{2}\right) \frac{e^{i\mathbf{t}\cdot(\mathbf{R}-\mathbf{R}')}}{|\mathbf{t} + \mathbf{K}|^2} \qquad (\text{V.7})$$

$$V_E(1, \mathbf{R}', \mathbf{k}, \mathbf{R}) = \sum_{\mathbf{K}} \int d^3t \, S\left(\mathbf{k} + \frac{\mathbf{t}}{2} + \mathbf{K}, 1 + \frac{\mathbf{t}}{2}\right)$$

$$\cdot S\left(1 - \frac{\mathbf{t}}{2} - \mathbf{k}, \mathbf{k} - \frac{\mathbf{t}}{2}\right) \frac{e^{i\mathbf{t}\cdot(\mathbf{R}-\mathbf{R}')}}{|\mathbf{k} - 1 + \mathbf{K}|^2}.$$

The S's have been defined in Eq. (5.9). (Actually the V's should have four band indices; one should be associated with each $\mathbf{k}$ and 1 in the expression on the right. These have been suppressed to avoid extra complication in the expression.) There is a natural division of V_D into two parts

$$V_{LR} = \int \frac{d^3t}{t^2} \, S\left(\mathbf{k} + \frac{\mathbf{t}}{2}, \mathbf{k} - \frac{\mathbf{t}}{2}\right) S\left(1 - \frac{\mathbf{t}}{2}, 1 + \frac{\mathbf{t}}{2}\right) e^{i\mathbf{t}\cdot(\mathbf{R}-\mathbf{R}')} \qquad (\text{V.8})$$

$$V_D' = \sum_{\mathbf{K}\neq 0} \int \frac{d^3t}{|\mathbf{t} + \mathbf{K}|^2} \, S\left(\mathbf{k} + \frac{\mathbf{t}}{2} + \mathbf{K}, \mathbf{k} - \frac{\mathbf{t}}{2}\right) S\left(1 - \frac{\mathbf{t}}{2} - \mathbf{K}, 1 + \frac{\mathbf{t}}{2}\right)$$

$$\cdot e^{i\mathbf{t}\cdot(\mathbf{R}-\mathbf{R}')}. \qquad (\text{V.9})$$

V_D' and V_E lead to short-range effects. This is apparent for nonintersecting bands since for them S can be extended a finite distance into the complex k space as in Appendix C. Thus V_D' and V_E fall off exponentially at large $(R - R')$. Even if there are discontinuities in the band, however, this will be compensated by corresponding discontinuities in the ρ's, so that no long-range effects actually result. V_{LR}, however, has additional singularities due to the t^2 in the integrand. In fact, the leading term in V_{LR} goes as $|R - R'|^{-1}$ at large values of $|R - R'|$. This term is just the long-range Coulomb interaction arising from charge density fluctuation. Additional terms of higher order in t can be interpreted in terms of multipole density fluctuations.

For a sinusoidal $\rho^{(1)}$ of wave number $\mathbf{q}$, we require only $V_{sc}(1, \mathbf{k}; \mathbf{q})$:

$$V_{sc}(1, \mathbf{k}; \mathbf{q})$$

$$= \sum_{R,R'} e^{i\mathbf{q}\cdot(\mathbf{R}-\mathbf{R'})} V_{sc}(1, \mathbf{R'}; \mathbf{k}, \mathbf{R})$$

$$V_D'(1, \mathbf{k}; \mathbf{q})$$

$$= \sum_{\mathbf{K} \neq 0} \frac{S[\mathbf{k} + (\mathbf{q}/2) + \mathbf{K}, \mathbf{k} - (\mathbf{q}/2)] \, S\,[1 - (\mathbf{q}/2) - \mathbf{K}, 1 + (\mathbf{q}/2)]}{|\mathbf{q} + \mathbf{K}|^2}$$

$$V_E(1, \mathbf{k}; \mathbf{q})$$

$$= \sum_{\mathbf{K}} \frac{S[\mathbf{k} + (\mathbf{q}/2) + \mathbf{K}, 1 + (\mathbf{q}/2)] \, \dot{S}\,[1 - (\mathbf{q}/2) - \mathbf{K}, \mathbf{k} - (\mathbf{q}/2)]}{|\mathbf{k} - 1 + \mathbf{K}|^2}$$

$$V_{LR}(1, \mathbf{k}; \mathbf{q})$$

$$= \frac{S[\mathbf{k} + (\mathbf{q}/2), \mathbf{k} - (\mathbf{q}/2)] \, S\,[1 - (\mathbf{q}/2), 1 + (\mathbf{q}/2)]}{|q|^2}. \qquad (V.10)$$

If we consider a crystal with the static lattice-periodic perturbation represented by $V_0(\mathbf{k})$, its total perturbed H is

$$H(\mathbf{k}) = H_0 + V_0(\mathbf{k}) + \sum_{R'} \rho^{(1)}(\mathbf{k}) V_{sc}(1, \mathbf{R'}; \mathbf{k}, \mathbf{R}) \, d^3k. \qquad (V.11)$$

Comparing (V.11) with (V.4) and (V.6), we see that, for a given $\mathbf{R}$, $H(\mathbf{k}, \mathbf{R})$ is the same as $H(\mathbf{k})$ in a crystal with the periodic perturbation $V_0(\mathbf{k}) = V_0(\mathbf{k}, \mathbf{R})$ and $\rho^{(1)}(\mathbf{k}) = p^{(1)}(\mathbf{k}, \mathbf{R})$. However, $\rho^{(1)}$ and V_0 cannot be given independently but are related. $\rho^{(1)}(\mathbf{k})$ is determined by the requirement that

$$\rho + \rho^{(1)} = f(H - \zeta) \qquad (V.12)$$

where ζ is determined by the total density of electrons. This requirement implies that

$$[(\rho + \rho^{(1)}), H] = 0. \qquad (V.13)$$

For a static $V_0(k, R)$, $\rho^{(1)}(k, R)$ is determined by the same equation, which in lowest order (in the sense of Part III) has the consequence that

$$\rho(\mathbf{k}, \mathbf{R}) + \rho^{(1)}(\mathbf{k}, \mathbf{R}) = f(H(\mathbf{k}, \mathbf{R}) - \zeta(R)). \qquad (V.14)$$

This equation is similar but not identical to (V.12). It differs because $\zeta(R)$ may not be constant and because $H(k, R)$ may contain terms from V_{LR} which do not occur in the periodic case. Two important special cases exist, however.

(1) In a crystal all of whose bands are full except one, which is nearly empty, the density fluctuations and V_{LR} may be negligible. Then ζ will be so far removed from the filled bands that its variations do not affect V. In

this case, applicable to nearly pure semiconductors, $\rho^{(1)}(\mathbf{k}, \mathbf{R}) = \rho^{(1)}(\mathbf{k})$ if $V(\mathbf{k}, \mathbf{R})$ is substituted for $V(\mathbf{k})$. Applying this result to acoustic waves leads to the deformation potential.[18a]

(2) In a metal or semimetal, neutrality is maintained so closely that the difference between the long-range potential and $\zeta(\mathbf{R})$ is constant and is determined by the constant density. This condition again restores the equality of $\rho^{(1)}(\mathbf{k}, \mathbf{R})$ and $\rho^{(1)}(\mathbf{k})$. The result is that V_0 plus the short-range self-consistent field are just as in the uniform perturbation. The lack of neutrality, however, does lead to a significant long-range potential because the factor $1/q^2$ for long wavelengths greatly magnifies the importance of even small density fluctuations. This effect has been discussed by Bardeen in connection with the electron-phonon interaction and by the author in relation to acoustic attenuation.[19]

In other cases, the $\rho^{(1)}(k, R)$ can be calculated by straightforward perturbation theory. This will frequently lead to essentially similar results, but the procedure now centers about a perturbation treatment involving the strength of the interaction, rather than only its effective wave number.

In the case of perturbations which are not constant in time but vary slowly compared to interband and plasma frequencies, essentially similar arguments can be used to reach the same conclusions.

Application of the foregoing procedure to an applied magnetic field shows that the leading term is simply $H[\mathbf{k} + (e/\hbar c)\mathbf{A}]$, where $\mathbf{A}$ includes only the terms in $\mathbf{R}$, not those in $\mathbf{X}$. In particular, the velocity to be used in the interaction with a vector potential of long wavelength that varies slowly is $\mathbf{w}$ not $\mathbf{v}$. Kane obtained a similar result from an argument involving gage-invariance.[18]

The discussion so far shows that practically all of the previous formalism can be maintained in the self-consistent field theory. The only problem we have noticed is that the velocity $\mathbf{w}$ differs from $\mathbf{v}$. Actually, $\mathbf{w}$ can be used in calculating the current, if properly interpreted. If we consider the error thus introduced, it is the trace of $\rho(w - v)$:

$$\int \rho_k(w_k - v_k)\, dk = \int \frac{\rho(x, x')\rho(x', x)(x - x')}{|x - x'|}\, d^3x\, d^3x'$$

$$= 0. \tag{V.15}$$

Thus $\mathbf{w}$ will give the right answer, provided the exact $\mathbf{w}$, determined from the correct density matrix ρ not the unperturbed density matrix $\rho^{(0)}$, is used. This, however, is an inconvenience and it seems easier, probably, to use the velocity $\mathbf{v}$ in calculating currents.

[18a] J. Bardeen and W. Shockley, *Phys. Rev.* **80**, 72 (1950).
[19] J. Bardeen, *Phys. Rev.* **52**, 688 (1937); E. I. Blount, *ibid.* **114**, 418 (1959).

VI. The (N + 1)-Electron Problem

In this section we shall consider a case in which an exact many-body theory leads to results very similar to that of simple band theory. This is the problem of an insulator, with one extra electron, which has been treated previously by Kohn[20] and Ambegaokar[21,22] and by Klein.[23] It is beyond the scope of this paper to use the diagrammatic techniques which appear to be necessary for careful proof of some results. For these points the reader is referred to the original papers.

We shall consider the Hamiltonian

$$H = \sum_{i=0}^{N} \frac{p_i^2}{2m} + \sum_{i=0}^{N} U(\mathbf{r}_i) + \frac{e^2}{8m^2c^2} \sum_{i=0}^{N} (\boldsymbol{\sigma}_i \times \nabla U) \cdot \mathbf{p}_i + \tfrac{1}{2} \sum_{i,j} \frac{e^2}{r_{ij}} \qquad (\text{VI.1})$$

where U is the one-electron periodic potential and N is the number of electrons in the insulator. We have included the spin-orbit interactions of the electrons with the nuclei, but not the interelectronic spin-orbit or magnetic interactions. This is a reasonable procedure since the possibility of large values of nuclear charge permits the term we have kept to be important, while the small electronic charge and the exclusion principle keep the interelectronic effects small.

Since the Hamiltonian is unchanged if we add any lattice vector $\mathbf{R}$ to *all* the electronic coordinates, we can write the eigenfunction in the form

$$\psi_{n\mathbf{k}} = \exp\left(i\mathbf{k} \cdot \sum \mathbf{x}_i\right) u_{n\mathbf{k}}(\mathbf{x}_0 \cdots \mathbf{x}_N) \qquad (\text{VI.2})$$

where u is periodic in the sense just used and is antisymmetric in the electronic coordinates. If we increase each $\mathbf{x}$ by $\mathbf{R}$, $\psi_{n\mathbf{k}}$ is multiplied by $e^{(N+1)\mathbf{k}\cdot\mathbf{R}}$. When periodic boundary conditions are used, N must be a multiple of the number of cells in the lattice in each of the main directions and $\mathbf{k}$ is, therefore, restricted in such a way that $e^{iN\mathbf{k}\cdot\mathbf{R}} = 1$ for all admissible $\mathbf{k}$. Therefore we can omit the N in the previous exponential; the character of $\psi_{n\mathbf{k}}$ is $e^{i\mathbf{k}\cdot\mathbf{R}}$, as in the one-electron case. We shall assume this as a characteristic of the insulator-plus-one-electron problem in either periodic or infinite boundary condition.

[20] W. Kohn, *Phys. Rev.* **105**, 509 (1957); **110**, 857 (1958).

[21] V. Ambegaokar and W. Kohn, *Phys. Rev.* **117**, 423 (1960).

[22] V. Ambegaokar, *Phys. Rev.* **121**, 91 (1961).

[23] A. Klein, *Phys. Rev.* **115**, 1136 (1959).

We first consider the variable $\mathbf{x} = \sum \mathbf{x}_i$. Just as in the one-electron CMR,

$$\mathbf{x} = i\frac{\partial}{\partial \mathbf{k}} + \mathfrak{X}$$

(VI.3)

$$\mathfrak{X}_{nn'} = \delta(k - k') \int u_{n\mathbf{k}}^* i\frac{\partial u_{n'\mathbf{k}}'}{\partial \mathbf{k}}\, d\tau$$

where the integral is taken over all $\mathbf{x}_0 \cdots \mathbf{x}_N$, subject to the condition that $\mathbf{x}$ lie in the unit cell including the origin.

As written, we have no reason to expect that n is a discrete index. Following Kohn, we shall select the assumption that n is discrete at low energies as providing the definition of the insulator-plus-one-electron problem.

As indicated in Appendix A for the one-electron case, this representation of $\mathbf{x}$ is valid only when applied to wave functions for which $\overline{x^2}$ (not $\sum x_i^2$) is bounded.

The total velocity $\mathfrak{B}$ is given by

$$\mathfrak{B} = -\frac{i}{\hbar}\sum [\mathbf{x}_i, H]$$

$$= \sum \frac{\mathbf{p}_i}{m} + \frac{e^2}{4mc^2}\sum \boldsymbol{\delta}_i \times \nabla U_i$$

$$= -\frac{i}{\hbar}[\mathbf{x}, H]$$

$$= \frac{1}{\hbar}\frac{\partial H}{\partial \mathbf{k}} - \frac{i}{\hbar}[\mathfrak{X}, H].$$

Thus again

$$\mathbf{V}_{nn'} = +\frac{i}{\hbar}\mathbf{X}_{nn'}(H_n - H_{n'}).$$

(VI.4)

As in the one-electron case, we can define

$$H(k) = e^{-i\mathbf{k}\cdot\mathbf{x}}He^{i\mathbf{k}\cdot\mathbf{x}}$$

$$= H + \hbar\mathbf{k}\cdot\mathfrak{B} + (N + 1)\frac{k^2}{2m}.$$

(VI.5)

The only important difference lies in the factor $(N + 1)$ in the last term. Let us now attempt to find an equation analogous to (1.7) by considering the commutator

$$-\frac{i}{\hbar} [\mathbf{x}, \mathfrak{B}] = \frac{N + 1}{m}$$

$$= \frac{1}{\hbar^2} \frac{\partial \mathbf{v}}{\partial k} - \frac{i}{\hbar} [\mathbf{V}, \mathbf{X}] \qquad (\text{VI.6})$$

where we have kept only the terms which contribute to the diagonal matrix elements. These yield

$$\alpha = \frac{1}{\hbar} \frac{\partial \mathbf{v}}{\partial \mathbf{k}} = \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial \mathbf{k} \, \partial \mathbf{k}} = \frac{N + 1}{m} + \frac{i}{\hbar} [\mathbf{V}, \mathbf{X}]. \qquad (\text{VI.7})$$

This again is identical to (2.18), except for the factor $(N + 1)$ which at first sight appears rather strange.

Let us consider the perfect insulator for a moment. We could treat it in much the same way that we have proceeded above, except that in the ground state $\mathbf{v} = 0$ and $N + 1$ is replaced by N. Thus

$$0 = \frac{N}{m} + \frac{i}{\hbar} [\mathbf{V}, \mathbf{X}]_0 \qquad (\text{VI.8})$$

where the subscript means the value of the commutator in the ground state of the insulator. Thus we can rewrite (VI.7) in the form

$$\alpha = \frac{1}{m} + \frac{i}{\hbar} [\mathbf{V}, \mathbf{X}] - \frac{i}{\hbar} [\mathbf{V}, \mathbf{X}]_0 \qquad (\text{VI.9})$$

showing that we must remove the filled band contribution to the commutator or f sum in calculating the effective mass. In terms of the diagrams of Kohn and Ambegaokar, this means that we need only use graphs like the "extra-particle response" graphs.

This sum rule differs from that obtained by Ambegaokar for finite wavelengths but is not in contradiction. (Without going into detail, we will remark that (VI.9) can be obtained from his sum rule by eliminating all improper graphs. This procedure can be shown to give a null contribution at zero frequency. This result is not immediately apparent in his paper because he takes the limit $q \to 0$ before letting $\omega \to 0$. For the zero-frequency response, in which he was not primarily interested, one should reverse this order. This in no way invalidates any of his results.)

All other relations between $\mathbf{v}$'s and $\mathbf{X}$'s obtained in Part I now follow immediately. We turn next to the representation of one-electron perturbations. If the potential $\sum V(\mathbf{x}_i)$ is applied to the system, we first rewrite it as

$$\sum V(\mathbf{x}_i) = \int d^3 V(\mathbf{q}) \rho(\mathbf{q})$$

(VI.10)

$$\rho(\mathbf{q}) = \sum_{i=0}^{N} e^{i\mathbf{q}\cdot\mathbf{x}_i}.$$

Thus the matrix elements of V are

$$V_{nn'}(k, k') = \int d^3q V(\mathbf{q}) \; \delta(\mathbf{k} - \mathbf{k}' + \mathbf{q}) \rho(q)_{n\mathbf{k},n'\mathbf{k}+\mathbf{q}}. \qquad (\text{VI.11})$$

Kohn has shown that at small $\mathbf{q}$, $\rho(\mathbf{q})_{n\mathbf{k},n'\mathbf{k}+\mathbf{q}}$ approaches

$$\lim \rho_{nn'}(\mathbf{q}) = \frac{1}{\kappa} \delta_{nn'}$$

where κ is the static dielectric constant of the insulator. Thus for small $\mathbf{q}$, V can be represented as

$$V = \frac{1}{\kappa} V\left(i \frac{\partial}{\partial \mathbf{k}}\right). \qquad (\text{VI.12})$$

The term linear in q takes the form

$$d^3q V(\mathbf{q})\mathbf{q} \; \delta(\mathbf{k} - \mathbf{k}' + \mathbf{q}) \left(\frac{\partial \rho_{nn}(\mathbf{k}; \mathbf{q})}{\partial \mathbf{q}}\right)_{q=0}. \qquad (\text{VI.13})$$

Here we have a difference from the one-electron case since

$$\frac{\partial \rho_{nn'}}{\partial \mathbf{q}} \neq \frac{\mathbf{X}_{nn'}}{\kappa}. \qquad (\text{VI.14})$$

The difference between them can be interpreted as a periodic perturbation arising from the fact that the polarization of the core charge by the field is not uniform in space, but varies over a unit cell. In general, ρ replaces S in (2.20) for the representation of one-electron potentials.

We have obtained above a result for slowly varying fields. If we consider a constant field so that $V(x_i) = -eE\cdot x_i$ we obtain for V the representation

$$V = -e\mathbf{E}i\cdot\frac{\partial}{\partial \mathbf{k}} - e\mathbf{E}\cdot\mathbf{X}. \qquad (\text{VI.15})$$

This expression differs in two respects from (VI.12) and (VI.13). First, the $1/\kappa$ is missing; second, we have replaced $\partial\rho/\partial q$ by ξ. The first is not a serious problem, for in writing the potential in this way, we are implying that E is the real electrostatic field whereas in our previous treatment $\partial V/\partial x$ would correspond to D in Maxwell's theory. The second problem is more serious and shows that while the use of the dielectric constant describes the long-range effect of the field sufficiently, it fails to take into account the periodic polarization of the insulator which produced a periodic perturbation. In other words when we write $V(x_i) = -eE \cdot x_i$, we are presupposing a very special charge distribution which cancels out the effect of the core polarization. This, of course, is not what we use in practice in applying a static uniform electric field. Hence if we have occasion to work to this accuracy, as in anomalous Hall effects, we should add a term describing this atomic polarization to the potential $E \cdot x$.

Thus for the special case of an insulator plus one-electron we recover a band theory which exhibits only slight variations from the one-electron theory. The same conclusions hold in the magnetic case, as has been shown in part by Klein and Ambegaokar. We may anticipate that these results also apply for a few electrons in an insulator until the number of added electrons becomes great enough to cause important interactions among them.

Appendix A. The Representation of x

In Part I we saw that the representation of x in the CMR is not completely straightforward. This is basically related to the fact that the operator x does not have the whole Hilbert space for its domain. That is, there exist normalizable functions φ for which $\int |\varphi(x)| x^2 \, dx$ is not finite. Under these circumstances the easiest way to find the representation of x

is to let it operate on a function φ which is in its domain and see what form it takes. For any function φ we take its representation in Bloch functions:

$$\varphi(\mathbf{x}) = \frac{1}{(2\pi)^3} \int d^3k e^{i\mathbf{k}\cdot\mathbf{x}} \sum_n u_{n\mathbf{k}}(\mathbf{x}) f_n(\mathbf{k}).$$

Then for $\mathbf{x}\varphi$ we obtain

$$\mathbf{x}\varphi(\mathbf{x}) = \frac{1}{(2\pi)^3} \int d^3k e^{i\mathbf{k}\cdot\mathbf{x}}\mathbf{x} \sum_n u_{n\mathbf{k}}(\mathbf{x}) f_n(\mathbf{k}) \tag{A1}$$

$$= \frac{1}{(2\pi)^3} \int d^3k (-i) \frac{\partial}{\partial k} \sum_n f_n(\mathbf{k}) \psi_{n\mathbf{k}}(\mathbf{x})$$

$$+ \frac{1}{(2\pi)^3} \int d^3k e^{i\mathbf{k}\cdot\mathbf{x}} \sum_n i \frac{\partial}{\partial \mathbf{k}} (u_{n\mathbf{k}}(\mathbf{x}) f_n(\mathbf{k}))$$

$$= - \frac{1}{(2\pi)^3} \int d\mathbf{S} \sum f_n(k) \psi_{n\mathbf{k}}(\mathbf{x})$$

$$+ \frac{1}{(2\pi)^3} \int d^3k \sum_n \psi_{n\mathbf{k}}(x) \left(\frac{i \, \partial f_n}{\partial \mathbf{k}} + \sum_{n'} \mathfrak{X}_{nn'} f_{n'}(\mathbf{k}) \right). \tag{A2}$$

The surface integral is to be taken over a limiting surface in $\mathbf{k}$ space surrounding any points of discontinuity in $\sum f_n(\mathbf{k}) \psi_{n\mathbf{k}}(\mathbf{x})$. Purely intuitively, it seems unlikely that any physically realizable wave function could have such discontinuities. More mathematically, it can be shown that, for functions in the domain of x, there are no discontinuities. For $\sum f_n(\mathbf{k}) \psi_{n\mathbf{k}}(\mathbf{x}) = \sum_{\mathbf{K}} \varphi(\mathbf{k} + \mathbf{K}) e^{i(\mathbf{k}+\mathbf{K})\cdot\mathbf{x}}$ where $\mathbf{K}$ is a reciprocal lattice vector and $\varphi(\mathbf{k})$ is the Fourier transform of $\varphi(\mathbf{x})$. Thus if $\varphi(\mathbf{k})$ has no discontinuities, there are none in $\sum f_{n\mathbf{k}} \psi_{n\mathbf{k}}$. But any function φ, for which $\int |\varphi(\mathbf{x})|^2 \mathbf{x}^2 \, d^3x$ exists, has a uniformly continuous Fourier transform which is nearly everywhere differentiable. Thus for all φ in the domain of x, we can in fact write

$$xf_n = i \frac{\partial}{\partial \mathbf{k}} f_n + \sum \mathfrak{X}_{nn'} f_{n'}. \tag{A3}$$

In Part II, however, we discuss Wannier functions, whose representation in the CMR is given by

$$f_n(\mathbf{k}) = \delta_{nn_0} e^{-i\mathbf{k}\cdot\mathbf{R}} \sqrt{\frac{\Omega}{(2\pi)^3}}.$$

Wherever two bands are degenerate at one point in $\mathbf{k}$ space, there is a question as to how to number the bands. Generally, the only feasible

procedure is to define the nth band at each k as that whose energy is nth from the bottom. This may, however, result in a situation whereby $u_n(\mathbf{k})$ changes discontinuously on passing through such a point. This will generally be true if the energy contours are linear along the path through the degeneracy. Thus in this case it is more convenient to consider the discontinuity to reside in the u than in f. We find then for the Wannier functions for $\mathbf{R} = 0$

$$\mathbf{x}a_n = \sqrt{\frac{\Omega}{(2\pi)^3}} \int d^3k \sum \psi_{n'\mathbf{k}}(x)\xi_{n'n} - \sqrt{\frac{\Omega}{(2\pi)^3}} \int d\mathbf{S}\psi_{n\mathbf{k}}(x)$$

If there is a surface of discontinuity, the surface integral term can be written

$$-\sqrt{\frac{\Omega}{(2\pi)^3}} \int dS e^{i\mathbf{k}\cdot\mathbf{x}}(u_{n\mathbf{k}}^{(+)}(\mathbf{x}) - u_{n\mathbf{k}}^{(-)}(\mathbf{x}))$$

the surface now being considered as one-sided. If the discontinuity is a line or point, the surface shrinks to zero area. Since the $u_{n\mathbf{k}}(\mathbf{x})$ are bounded, the contribution of these surfaces vanishes.

On the other hand, plane discontinuities may occur in structures with twofold screw axes. The bands separate in energy proportionally to distance from the plane, so if the bands are labeled in order of their energy, there is a discontinuity in $u_{n\mathbf{k}}$, and the surface does not shrink in the limit. Thus the surface contribution is finite. If we calculate the expectation value of $\mathbf{x}$, we obtain

$$\int \mathbf{x}\,|\,a_n\,|^2\,d^3x = \frac{\Omega}{(2\pi)^3} \int d^3k \mathfrak{X}_{nn}$$

$$- i\frac{\Omega}{(2\pi)^3} \int dS\,d^3k\,d^3x \psi_{n\mathbf{k}}^{*}(u_{n\mathbf{k}}^{(+)}(\mathbf{x}) - u_{n\mathbf{k}}^{(-)}(\mathbf{x}))e^{i\mathbf{k}\mathbf{x}}.$$

Here the two terms in the surface integral cancel and we are left with only

$$\frac{\Omega}{(2\pi)^3} \int d^3k \mathfrak{X}_{nn} = \bar{\mathfrak{X}}_{nn}.$$

If we now consider the integral of $\mathbf{x}^2$,

$$\int \mathbf{x}^2\,|\,a\,|^2\,dV = \frac{\Omega}{(2\pi)^3} \int d^3k (\mathfrak{X}^2)_{nn}$$

$$+ \frac{\Omega}{(2\pi)^3} \int dS\cdot dS' e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}}\,|\,u_n^{+} - u_n^{-}\,|^2\,d^3x.$$

If we call z the direction perpendicular to the plane, $k_z = k_z'$ and the integration over z diverges. (In box normalization it would be proportional to the length of the box.) More particularly the integral of z^2 diverges, while the corresponding surface integrals for x and y vanish and Wannier function is, so far, well-behaved in these directions.

Near lines and points of degeneracy, as we have seen, the surface contributions to $\mathbf{x}a$ and thus to $\int |a|^2 x^2 \, dV$ vanish. This is not necessarily the case for higher power of x. Suppose that $x^n a$ can be expressed as $[i(\partial/\partial k) + \xi]^n f$, but that the integral of this quantity over a surface around a discontinuity does not vanish. There is then a surface contribution to $x^{n+1}a$ and to $\int x^{2(n+1)} |a|^2 \, dV$. However, a detailed examination shows that in such a case the contribution to $(\int x^{(n+1)} |a|^2 \, dV)$ from the volume integral diverges if n is odd, and that to $\int x^n |a|^2 \, dV$ diverges if n is even. Thus the surface contributions do not play any vital role in these cases.

Appendix B. Consequences of Inversion and Time Reversal

A large number of crystals possess inversion symmetry J, including the great majority of those which have been seriously investigated. In the absence of magnetic fields, or to the extent that such fields are neglected, all crystals have time reversal symmetry K. Since each of these operations transform a wave function of wave vector $\mathbf{k}$ into one of wave vector $-\mathbf{k}$, their product leaves the wave vector unchanged. This operator, which we shall call conjugation and denote by C, is thus a symmetry operator diagonal in $\mathbf{k}$ for all points in the BZ.

The operator C has two important properties which can be used to obtain restrictions on the matrix elements of operators:

$$\text{antilinearity: } Ca\psi = \bar{a}C\psi \tag{B1}$$

$$\text{antiunitarity } C\psi = U\bar{\psi} \tag{B2}$$

where U is a unitary operator.

Both properties follow immediately from the nature of K^{24} and the linear, unitary character of J. These are the same properties possessed by time reversal itself, which suffice to determine its consequences. If there is no spin, U has the property $U^2 = C^2 = 1$, while if spin is included $U^2 = C^2 = -1$. On the other hand, when there is spin but no spin-orbit coupling, U itself, which is spin-reversal, is a symmetry of the system and the

[24] E. P. Wigner, *Nachr. Akad. Wiss. Göttingen, Math.-physik. Kl.* p. 546 (1932); also "Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra," p. 325. Academic Press, New York, 1959.

operation of complex conjugation alone is, therefore, also a symmetry which can replace K. Thus we will adopt the point of view that $U^2 = 1$, unless there is spin-orbit coupling.

Since C is antiunitary it cannot be adequately represented by a matrix. On the other hand, its operation on a base vector can be so represented and we will define the matrix $\mathbf{C}$ in such a way that $C\varphi_i = \varphi_j C_{ji}$.

With these properties Wigner[15] showed that K produces no extra degeneracy in the absence of spin-orbit coupling, but does make every level doubly degenerate in the presence of spin-orbit coupling.

Thus, C also produces no degeneracy at $\mathbf{k}$ in the absence of spin, but makes each level at $\mathbf{k}$ doubly degenerate in its presence, as shown by Elliott.[25] C can be represented by the identity in the former case. In the latter case, the wave functions can be chosen in pairs in such a way that C has the representation $i\sigma_y$ in the two-dimensional space associated with each pair. Using (B1) it is easy to show that if a unitary transformation R is made, the matrix $\mathbf{C}$ representing C is transformed according to the rule

$$\mathbf{C}' = R\mathbf{C}R^T. \tag{B3}$$

If we restrict ourselves to generalized phase transformations, we need only consider each pair of functions separately. If $R = e^{i\varphi}e^{iA_l\sigma_l}$, we find that if $\mathbf{C} = i\sigma_y$

$$\mathbf{C}' = e^{+2i\varphi}i\sigma_y. \tag{B4}$$

In the absence of spin we consider only $R = e^{i\varphi}$. Then if $\mathbf{C} = 1$,

$$\mathbf{C}' = e^{+2i\varphi}. \tag{B5}$$

In either case $\mathbf{C}$ is a unitary matrix but does not transform as a unitary operator.

Using (B2), we obtain for the matrix elements of an operator O

$$(C\psi, OC\varphi) = (\bar{\psi}U^\dagger O U\bar{\varphi})$$
$$= (\varphi, U^\dagger \bar{O}^\dagger \bar{U}\psi)$$
$$= (\varphi, C^{-1}O^\dagger C\psi)$$
$$= \overline{(\psi, COC^{-1}\varphi)}. \tag{B6}$$

Thus

$$C^\dagger_{il}O_{lm}C_{mj} = \pm\bar{O}_{ij} \tag{B7}$$

where the $+$ sign applies if O is invariant under C, the minus sign if it changes sign.

In the absence of spin-orbit coupling, this becomes

$$e^{-2i\varphi_i}O_{ij}e^{+2i\varphi_j} = \pm\bar{O}_{ij} \tag{B8a}$$

[25] R. J. Elliott, *Phys. Rev.* **96**, 280 (1954).

where the φ's represent the phases of the two bands relative to phases for which $\mathbf{C} = 1$. In particular, if an operator is Hermitian and, like velocity, is invariant under C, its matrix elements can be made real. If O is Hermitian and changes sign under C like x, it can be made imaginary, and its diagonal elements vanish. (In the case of $\mathbf{x}$, whose singularity makes it a special case, $\mathbf{X}_{nn}$ can be made to vanish locally, but not necessarily everywhere. If there are curves of degeneracy it may not be possible to choose the phases so as to make the u's continuous in k space in such a way that $\mathbf{X}_{nn}$ vanishes everywhere. On the other hand, since $\nabla \times \mathbf{X}_{nn}$ is invariant under phase changes, it vanishes everywhere.)

In the case of spin-orbit coupling, a similar relation is obtained, but we write it only for the matrix elements between wave functions of a given pair

$$\bar{O} = \pm e^{-2i\varphi}\sigma_y O \sigma_y e^{+2i\varphi}. \tag{B8b}$$

where the σ's refer to the Pauli matrices as mathematical objects, unrelated to actual spin. If we write $O = B + \sum A_l \sigma_l$, we obtain

$$\bar{B} = \pm e^{-2i\varphi} B e^{+2i\varphi}$$

$$\bar{A}_l = \mp e^{-2i\varphi} A_l e^{+2i\varphi} \tag{B9}$$

where we preserve the complex conjugate and phase factors to take account of operators like x which are not diagonal in $\mathbf{k}$. We see that Hermitian operators which are diagonal in k have the form B if they are invariant under C, and $A \cdot \sigma$ if they change sign under C (for A and B must be real in that case). In the case of $\mathbf{x}$,

$$\mathbf{B} = i\frac{\partial}{\partial\mathbf{k}} + \mathbf{X}^{(0)}$$

$$\bar{\mathbf{B}} = -i\frac{\partial}{\partial\mathbf{k}} + \bar{\mathbf{X}}^{(0)}$$

$$= -e^{-2i\varphi}\mathbf{B}e^{2i\varphi} = -i\frac{\partial}{\partial k} - \mathbf{X}^{(0)} + 2\frac{\partial\varphi}{\partial k}.$$

Hence

$$\mathbf{X}^{(0)} = \frac{\partial\varphi}{\partial\mathbf{k}}.$$

By proper choice of φ's, $\partial\varphi/\partial k$ can be made to vanish locally and $\mathbf{X}_{nn}$ has the form $\sum A_l \sigma_l$. Therefore, $\mathbf{\Omega}$ also has this last form locally, and by its invariance, has the same form everywhere.

Appendix C. Bands in Complex k Space

We have seen that in determining the asymptotic properties of Wannier functions, the behavior of the wave functions as functions of the complex vector $\mathbf{k}$ is important. In this section we shall discuss the extension of operators and wave functions into the complex plane and their analytical properties. Kohn[12] has given such a discussion for one dimension.

In Part I we found it convenient to define the operator $H(\mathbf{k})$ which is allowed to operate only on periodic functions. If we permit $\mathbf{k}$ in Eq. (1.7) to take on complex values, $H(\mathbf{k})$ is no longer Hermitian except for real $\mathbf{k}$. It does, however, have the property

$$H^\dagger(\mathbf{k}) = H(\bar{\mathbf{k}}). \tag{C1}$$

As a consequence of the non-Hermitian character of $H(\mathbf{k})$, the eigenvectors of $H(\mathbf{k})$ are not orthogonal and $H(\mathbf{k})$ can be diagonal only when the eigenvalues are distinct. In case there are multiple eigenvalues, the matrix can be reduced to the form of diagonal blocks, one corresponding to each eigenvalue. These eigenvalues are the diagonal elements in their respective blocks; there are no matrix elements between states in different blocks.

There are, of course, two eigenvalues problems associated with $H(\mathbf{k})$

$$H(\mathbf{k})\psi_{n\mathbf{k}} = E_n\psi_{n\mathbf{k}}$$
$$\bar{\psi}_n H(\mathbf{k}) = E_n\bar{\psi}_{n\mathbf{k}}. \tag{C2}$$

Since $H(\mathbf{k})$ is not Hermitian, $\bar{\psi}$ is not in general equal to $\psi\dagger$. Instead of the orthogonormality relation

$$(\psi_n\dagger, \psi_m) = \delta_{nm}$$

we have

$$(\bar{\psi}_n, \psi_m) = \delta_{nm} \tag{C3}$$

and the additional relation

$$\bar{\psi}_{n\bar{\mathbf{k}}} = \psi_{n\mathbf{k}}^\dagger \tag{C4}$$

resulting from (C1).

The relation (C1) has, as a special case, the condition that H is Hermitian for real k, and leads to real eigenvalues in real k space. It also implies that $E_n(\bar{\mathbf{k}}) = \bar{E}_n(\mathbf{k})$. Another important property is time-reversal which states that for any solution ψ of the Schrödinger equation $H\psi = E\psi$, with a real eigenvalue E, $\bar{\psi}$ is also a solution since H is real. (In the case of spin-orbit coupling the spin must also be reversed.) In complex $\mathbf{k}$ space, we see that if $H\psi = E\psi$ for complex E, $\bar{\psi}$ is a solution of $\bar{H}$ with eigenvalue $\bar{E}$. Since the complex conjugate of $e^{i\mathbf{k}\cdot\mathbf{x}}u_{n\mathbf{k}}(\mathbf{x})$ is $e^{i(-\bar{\mathbf{k}})\cdot\mathbf{x}}\bar{u}_{n\mathbf{k}}(\mathbf{x})$, the eigenvalues $E_n(-\bar{\mathbf{k}})$ are the complex conjugates of the $E_n(\mathbf{k})$. Com-

bining this result with our previous one, we see that the sets $E_n(-\mathbf{k})$ and $E_n(\mathbf{k})$ are equal throughout complex k space. Furthermore, if $\mathbf{k}$ is pure imaginary $-\bar{\mathbf{k}} = \mathbf{k}$. Thus for each n, there is an n' such that

$$E_n(\mathbf{k}) = \bar{E}_{n'}(\mathbf{k}) \qquad k \text{ imaginary.} \qquad (C5)$$

Thus, either $n = n'$ and the energy is real, or they are conjugate pairs. In the latter case, as $\mathbf{k}$ approaches zero in imaginary $\mathbf{k}$ space, the two bands must approach the same real energy. Therefore, if a given level at $\mathbf{k} = 0$ is nondegenerate, its energy for imaginary $\mathbf{k}$ near the real axis must be real. For $\mathbf{k} = (\mathbf{K}/2) + i\mathbf{m}$, where $\mathbf{K}$ is a reciprocal lattice vector, $-\bar{\mathbf{k}}$ is equivalent to $\mathbf{k}$ and the same results hold.

Another important property possessed by H is the fact that it is analytic. This means that each matrix element of H is an analytic function of k. In the form

$$H(\mathbf{k}) = H + \hbar\mathbf{k}\cdot\mathbf{V} + \frac{\hbar^2 k^2}{2m} \qquad (C6)$$

each matrix element is in fact a simple quadratic function. The most concise statement of the analytic quality is provided by the equation

$$[R, H] = 0 \qquad (C7)$$

where

$$R = \frac{\partial}{\partial \mathbf{l}} + i\frac{\partial}{\partial \mathbf{m}} = \frac{\partial}{\partial \bar{\mathbf{k}}}$$

and

$$\mathbf{k} = \mathbf{l} + i\mathbf{m}$$
$$\bar{\mathbf{k}} = \mathbf{l} - i\mathbf{m}.$$

Wave functions are analytic if they satisfy the corresponding equation

$$R\psi = 0. \qquad (C8)$$

Evidently, if U is an analytic operator with an inverse UHU^{-1} is also analytic. In particular if U satisfies the equation

$$HU = UE \qquad (C9)$$

where E is diagonal,

$$E = UHU^{-1}$$

is analytic, whenever U^{-1} exists, provided U is analytic.

With certain exceptions, the analytic nature of U is not difficult to prove, although it requires some labor. We suppose that at some point k_0, the eigenvalue problem has been solved, and the eigenfunction found.

Then at $k_0 + \mathbf{q}$, we have

$$H_{nn'}(\mathbf{k}_0 + \mathbf{q}) = H_n \delta_{nn'} + \hbar \mathbf{q} \cdot \mathbf{V}_{nn'}(\mathbf{k}_0). \qquad \text{(C10)}$$

Here we have neglected the term $\hbar^2 q^2/2m$, which evidently does not affect the analytic nature of E or U. At $\mathbf{k}_0$, the eigenfunctions U_l have the form

$$U_{nl} = \delta_{ln}. \qquad \text{(C11)}$$

We want the eigenfunction at $\mathbf{q}$. We further suppose, for the time being, that the eigenvalues are nondegenerate. In this case, the question we are considering is essentially the convergence of perturbation theory for the present problem. In general, perturbation theory is believed to be convergent only asymptotically. In the present case, however, we shall prove that the U's can in fact be differentiated. We rewrite (C9)

$$-(H_n - E_l) U_{nl} = \sum_{n'} \hbar \mathbf{q} \cdot \mathbf{\mathfrak{B}}_{nn'}^{(k0)} U_{n'l}. \qquad \text{(C12)}$$

Taking the absolute square, we find with the Schwartz inequality

$$\sum_n |H_n - E_l|^2 |U_{nl}|^2 < |\hbar q|^2 \sum_{nn'} U_{ln}(\mathbf{\mathfrak{B}}^2)_{nn'} U_{n'l}.$$

$$< \hbar^2 |q|^2 (\mathbf{\mathfrak{B}}^2)_{ll} \qquad \text{(C13)}$$

At this point the simplest, if not most elegant, procedure is to invoke relativity theory, thus guaranteeing that $V_{ll}^2 < C^2$. Then as $\mathbf{q}$ becomes very small, this relation can be satisfied only if U_{nl} vanishes like $\mathbf{q}$ for all but one n, and if E_l approaches the corresponding H_n. If there is a degeneracy at $\mathbf{k}_0$, the only change is that U_{nl} vanishes for all n not belonging to one energy level at k_0.

We now return to (C12) and consider the level 0, labeling E_0 as the level which approaches H_0 at small q. Then

$$(H_n - E_0(\mathbf{q})) U_{n0}(\mathbf{q}) = -\hbar \mathbf{q} \cdot \mathbf{V}_{nn'} U_{n'0}(\mathbf{q}) \qquad \text{(C14)}$$

and for $n \neq 0$

$$U_{n0}(q) = \frac{\hbar \mathbf{q} \cdot \mathbf{V}_{nn'}}{H_n - E_0(\mathbf{q})} U_{n'0}(q)$$

$$\frac{\partial U_{n0}}{\partial \mathbf{k}} = \lim_{q \to 0} \frac{\mathbf{V}_{nn'}}{H_n - E_0(\mathbf{q})} U_{n'0}.$$

By the previous paragraph this limit is defined so as to yield

$$\frac{\partial U_{n0}}{\partial \mathbf{k}} = \frac{\mathbf{V}_{n0}}{H_n - H_0} \qquad \text{(C15)}$$

as would be obtained by assuming the convergence of perturbation theory. Since this limit does not depend on the realness of $\mathbf{q}$, the Cauchy-Riemann conditions are satisfied. We observe, however, that U_{00} is not determined by the foregoing procedure but by the condition (C3). The equation for $n = 0$ yields instead

$$E_0 = H_0 + h\mathbf{q} \cdot \frac{\mathbf{V}_{nn'}U_{n'0}}{U_{00}}$$

or

$$\frac{\partial E_0}{\partial \mathbf{k}} = \mathbf{V}_{00}. \tag{C16}$$

Thus, the Cauchy-Riemann conditions are satisfied at $\mathbf{k}_0$ for both E and U. Furthermore, since H is analytic and U and E are continuous, the derivatives are continuous in a neighborhood around $\mathbf{k}_0$. These conditions are sufficient to establish that E and U are analytic.

At $\mathbf{k}_0$, a point of degeneracy between two or more bands, the Hamiltonian can be brought into a form in which the nonzero elements are all along the diagonal, or above it; furthermore, $H_{nn'} = 0$ unless $H_{nn} = H_{n'n'}$. For real $\mathbf{k}$, of course, all nondiagonal elements can be removed. At neighboring points, $\mathbf{k}_0 + \mathbf{q}$ interband elements proportional to $\mathbf{q}$ will appear. By arguments essentially like those already used, the elements connecting different eigenvalues can be removed analytically to give an analytic matrix for each group of degenerate levels. We can now fix our attention on just one of these groups. For simplicity we shall discuss only double degeneracies. More complicated cases are equally straightforward, if longer. Setting the diagonal elements at $\mathbf{k}_0$ equal to zero, we then have the matrix

$$\mathbf{q} \cdot \mathbf{V}_0 + \mathbf{q} \cdot \mathbf{V}_1 + \cdots \quad U + \mathbf{q} \cdot \mathbf{V}_2 + \cdots$$
$$\mathbf{q} \cdot \mathbf{V}_3 + \cdots \quad \mathbf{q} \cdot \mathbf{V}_0 - \mathbf{q} \cdot \mathbf{V}_1 + \cdots.$$

The resulting secular equation is

$$(E - \mathbf{q} \cdot \mathbf{V}_0)^2 = U\mathbf{q} \cdot \bar{\mathbf{V}}_3 + (\mathbf{q} \cdot \mathbf{V}_1)^2 + \mathbf{q} \cdot \mathbf{V}_2 \bar{\mathbf{V}}_3 \cdot \mathbf{q}.$$

Let us consider one complex line. (We shall refer in this way to a plane in our six-dimensional space, for which the two independent directions can be chosen as the real and imaginary parts of one component of $\mathbf{k}$.) Then $U\mathbf{q} \cdot \mathbf{V}_3$ does not vanish generally and E is not analytic as a function of k_x, say:

$$E^{\pm} = \mathbf{q} \cdot \mathbf{V}_0 \pm q_x^{1/2}(U\bar{V}_{3x} + q_x\bar{V}_{3x}V_{2x} + q_xV_{1x}^2 + \cdots)^{1/2}.$$

Indeed it has a branch point of order 1. Similarly the wave functions have branch points, as we shall now see in detail.

The wave functions $C^{\pm}$ are determined by

$$(\mathbf{q}\cdot\mathbf{V}_1 + \mathbf{q}\cdot\mathbf{V}_0 - E^{\pm})\,C_1^{\pm} + (U + \mathbf{q}\cdot\mathbf{V}_2)\,C_2^{\pm} = 0$$

$$\mathbf{q}\cdot\mathbf{V}_3 C_1^{\pm} - (\mathbf{q}\cdot\mathbf{V}_1 + \mathbf{q}\cdot\mathbf{V}_0 + E^{\pm})\,C_2^{\pm} = 0$$

$$\frac{C_2^{\pm}}{C_1^{\pm}} = \frac{E^{\pm} - \mathbf{q}\cdot\mathbf{V}_1 - \mathbf{q}\cdot\mathbf{V}_0}{U + \mathbf{q}\cdot\mathbf{V}_2}.$$

Similarly for the solutions of the transposed equation, $D^{\pm}$

$$\frac{D_2^{\pm}}{D_1^{\pm}} = \frac{U + \mathbf{q}\cdot\mathbf{V}_2}{E^{\pm} + \mathbf{q}\cdot\mathbf{V}_1 + \mathbf{q}\cdot\mathbf{V}_0}.$$

Thus, the branch points appear, through E, in C and D.

It is worthy of note that as $q \to 0$, $C_2 \to 0$ and $D_1 \to 0$ for both E^{+} and E^{-}. Thus both eigenvectors approach the same vector at the degeneracy where there is only one eigenvector. For if $k = 0$, the vector $C_1 = 1$, $C_2 = 0$ is an eigenvector, but the vector $C_2 = 1$, $C_1 = 0$ is not, and conversely for the D's. This is a consequence of the non-Hermitian character of $H(\mathbf{k})$ at complex $\mathbf{k}$. Near degeneracies, not only are two eigenvectors not orthogonal, they are nearly identical except, of course, for real k.

So far we have considered the general case in which $U\mathbf{q}\cdot\bar{\mathbf{V}}_3 \neq 0$ for the complex line being considered. If $U q\cdot\bar{V}_3 = 0$, $E^{\pm}$ is analytic on the line k. For some line of $\mathbf{q}$ through any point, $\mathbf{q}\cdot\mathbf{V}_3$ is in fact zero. Along this line, the bands diverge linearly and analytically in k rather than as $|\,k\,|^{1/2}$, and the C's and D's behave similarly, although again, of course, they converge to one single eigenvector rather than two.

If U is zero, however, the behavior is much more like what we find for real $\mathbf{k}$. There are two eigenvectors at $\mathbf{q} = 0$, and the behavior is analytic on any complex line through $\mathbf{k}$. Before passing to the general case, we mention some important special cases.

(a) In crystals with inversion symmetry, the bands are. doubly degenerate throughout real $\mathbf{k}$ space when spin is included. Analyticness now requires that the degeneracy extend throughout complex $\mathbf{k}$ space, and that $U = 0$ everywhere. (This can also be proven by group theory, but the proof is more difficult than for real $\mathbf{k}$ space.)

(b) In some crystals there are planes of degeneracy in real $\mathbf{k}$ space. Again analyticness requires the extension of these degeneracies into complex k space; U also vanishes on these complex planes (defined analogously to complex lines). Thus the energies and wave functions are analytic on and near such degeneracies. This is a general feature of $(n - 1)$-dimensional degeneracies in n-dimensional space.

(c) In one-dimensional problems, if two bands are degenerate at a real **k**, they are analytic in the neighborhood of such a degeneracy. Thus Kohn's analysis[12] extends to this case also, although such degeneracies are vanishingly improbable in Herring's[26] sense, and thus of no real importance.

Returning to the general case, we expect that the conditions $U = 0$ and $\Delta E = 0$ can be satisfied only on a two-dimensional curve in our six-dimensional space. This suggests that such curves are the intersections of two surfaces of degeneracy, both involving the same pair of bands. When two branch surfaces of order 1 intersect, their branch characters cancel so far as any complex line through the intersection is concerned, for any closed path on such a line must surround an even number of branch points. Thus U must be zero at such an intersection. Conversely, if we consider the surfaces of degeneracy near a point where $U = 0$, the lowest order of k appearing in the equations defining the surface is the second. Thus there are two surfaces of degeneracy of the two bands near any such point. Although these points are not branch points for any one complex line, we may come back to a different band from the original on a closed contour which does not stay in one complex line. This will now be considered in more detail in regard to another special case.

(d) *Symmetry lines of degeneracy.* Many crystals have lines of symmetry along which degeneracies of bands occur. These degeneracies, like those on planes, extend over the complex line by arguments involving symmetry or analyticness. Likewise U is zero on the whole complex line. Thus on any intersecting complex line, the behavior of the energies and wave functions is analytic. Nevertheless, the analytical behavior in three dimensions is somewhat complicated.

(1) First let us consider the more usual case taking the two-band Hamiltonian near the degeneracy to be of the following form

$$\begin{pmatrix} +k_x V_1 & +k_y V_2 \\ k_y \bar{V}_3 & -k_x V_1 \end{pmatrix}.$$

Then the energy is given by

$$E^2 = k_x^2 V_1^2 + k_y^2 V_2 \bar{V}_3.$$

On the complex k_x line we have

$$E^{\pm} = \pm V_1 k_x.$$

In real k space, on the other hand,

$$E = \pm\sqrt{k_x^2 V_1^2 + k_y^2 V_2^2}$$

[26] C. Herring, *Phys. Rev.* **52**, 365 (1937).

where $V_3 = V_2$ and V_1 and V_2 are taken as real. Then there is no degeneracy except at $k = 0$ real k space. If we follow a path from a point k_x in the band with energy $V_1 k_x$ to a point $-k_x$, avoiding the origin and keeping the energy continuous, the energy at $-k_x$ is also $V_1 k_x$. On the other hand, had we followed the analytically defined band on the complex k line we should have ended with energy $-V_1 k_x$. Thus, although there is no branch point on the k_x line or in real k space, there are two branch surfaces defined by the equation

$$k_x = \pm k_y \frac{V_2 \bar{V}_3}{V_1^2}.$$

Any closed path, either in real space or on the complex line passes around both of these or neither of them. A closed path composed of our two previously discussed paths, however, circulates just one, and typical branch behavior results.

(2) If the energies diverge quadratically from the line of symmetry, a typical Hamiltonian is

$$\begin{pmatrix} (k_x^2 - k_y^2)\alpha_1 & k_x k_y \alpha_2 \\ k_x k_y \bar{\alpha}_3 & -(k_x^2 - k_y^2)\alpha_1 \end{pmatrix}$$

$$E^2 = \alpha_1^2 (k_x^2 - k_y^2)^2 + \alpha_2 \bar{\alpha}_3 k_x^2 k_y^2.$$

If we follow the paths defined above, in this case we wind up with the same energy we started with in both cases. The explanation is easily seen to be that we now have twice as many branch surfaces as before, since for given k_y we now have four values of k_x which make E^2 vanish. Our composite path through real space and the k_x line now surrounds an even number of such surfaces. This situation is somewhat simpler in that both types of path likely to be of most interest give the same end results, but the branch behavior still lurks close behind the surface. In particular it means that the bands have singularities at the real axis so that the perturbation expansion at this point does not converge and Wannier functions for the bands do not fall off exponentially at large distance.

On the other hand, for bands which have no degeneracies in real k space, we can now extend the integrals defining the Wannier functions as described in the text out as far as the nearest approach of a degeneracy surface.

In addition to showing the consequences of branch points in complex **k** space, Kohn's analysis demonstrated their existence. This can also be done in the present case. To this end, let us consider $H(\mathbf{k})$ at large values

of the imaginary part **m** of **k**, whose real part is 1

$$H(1 + i\mathbf{m}) = H(1) + i\mathbf{m} \cdot \mathbf{V}(1) - \frac{m^2}{2}.$$

If **m** = 0, we know that if we go continuously from 1 to 1 + **L**, where **L** is a reciprocal lattice vector, the eigenvalues are periodic and analytic. If **m** is large, however, the behavior will be dominated by the eigenvalues of $\mathfrak{B}(\mathbf{k})$. These are independent of the periodic potential U and increase continuously and analytically by **L** if we go from 1 + $i\mathbf{m}$ to 1 + **L** + $i\mathbf{m}$. Thus the eigenvalues obtained by analytic continuation are periodic for **m** = 0 but not if m is very large, though of course the set of eigenvalues at 1 is the same as the set at 1 + **L**. This situation clearly requires branch points, for by following a to a closed circuit from 1 through 1 + $i\mathbf{m}$, 1 + $i\mathbf{m}$ + **L**, and 1 + **L** back to 1 we arrive at a different energy from that with which we started. We may say that the bands at large m are tilted relative to those for **m** = 0. We can also estimate the value of m at which the branch points should occur, for, roughly speaking, they will separate the region where $H(1)$ is dominant from the region where **m** · **V**(1) prevails. Thus, **m** should be of the order of $[V_{nn'}/(E_n - E_{n'})]^{-1}$ or $X_{nn'}$. In simple cases where there are two band edges with nearly equal energies at the same point in **k** with all other bands much further away, this is easily verified.

We can also determine something about the location of the branch surfaces. Since the eigenvalues at large **m** are determined by **m** · **V**, they will be essentially the same as for free electrons. If 1 = 0, **V**(0) is just **p**/m for the Schrödinger equation, and the energies will be $\frac{1}{2}m(\hbar\mathbf{K} + i\mathbf{m})^2$. At large **m**, these levels cannot be seriously perturbed by the potential, so for general directions of **m**, the energies are complex. On the other hand, we established that near **k** = 0, the energies of bands which are nondegenerate at **k** = 0 are real for imaginary **k**. A real analytic function can become complex in this way only by having a branch point in the imaginary **k** space in every direction. That is, the branch points in imaginary **k** space form a two-dimensional manifold. Similarly, there is a surface of branch points in each three-dimensional manifold **k** = (**K**/2) + $i\mathbf{m}$, where **K** is a reciprocal lattice vector. Since the branch surfaces are four-dimensional and imaginary k space three-dimensional, one would have anticipated only a one-dimensional intersection. In the special case of one-dimensional bands, this result has the form that each band must have branch points in imaginary k space and on the lines (**K**/2) + $i\mathbf{m}$. In this case, the branch points might have been anticipated to be completely unlikely to lie on the imaginary axis (or any prechosen one-dimensional manifold), but in fact they must lie on it.

Appendix D. Derivation of Multiplication Rule

Given two functions N and 0 where product is P, we wish to find the product formula in the mixed representation. We have then

$$P(\mathbf{k}, \mathbf{R}) = \int e^{-i\mathbf{q}\cdot\mathbf{R}} P(\mathbf{k}; \mathbf{q})\, d^3q \tag{D1}$$

$$= \int e^{-i\mathbf{q}\cdot\mathbf{R}} P\left(\mathbf{k} + \frac{\mathbf{q}}{2}, \mathbf{k} - \frac{\mathbf{q}}{2}\right) d^3q$$

$$= \int e^{-i\mathbf{q}\cdot\mathbf{R}} N\left(\mathbf{k} + \frac{\mathbf{q}}{2}, \mathbf{k} + \boldsymbol{\epsilon}\right) O\left(\mathbf{k} + \boldsymbol{\epsilon}, \mathbf{k} - \frac{\mathbf{q}}{2}\right) d^3q\, d^3\epsilon \tag{D2}$$

where $\mathbf{q}$ is limited to its principal BZ, and $\boldsymbol{\epsilon}$ can be taken over any BZ.

We now wish to obtain $N[\mathbf{k} + (\mathbf{q}/2), \mathbf{k} + \boldsymbol{\epsilon}]$ in terms of $N(\mathbf{k}, \mathbf{R})$, but we run into difficulty because

$$N\left(\mathbf{k} + \frac{\boldsymbol{\epsilon}}{2} + \frac{\mathbf{q}}{4}; \frac{\mathbf{q}}{2} - \boldsymbol{\epsilon}\right) = \sum_{R_1} N\left(\mathbf{k} + \frac{\boldsymbol{\epsilon}}{2} + \frac{\mathbf{q}}{4}; \mathbf{R} + \mathbf{R}_1\right) e^{i\mathbf{R}_1\cdot[(\mathbf{q}/2) - \boldsymbol{\epsilon}]}$$

is periodic in $(\mathbf{q}/2) - \epsilon$ when $(\boldsymbol{\epsilon}/2) + (\mathbf{q}/4)$ is held constant, while $N[\mathbf{k} + (\mathbf{q}/2), \mathbf{k} + \boldsymbol{\epsilon}]$ is not. We must, therefore, always choose $\boldsymbol{\epsilon}$ so that $|(q^\mu/2) - \epsilon^\mu| < |K^\mu/2|$. Consider Fig. 1, in which we have changed
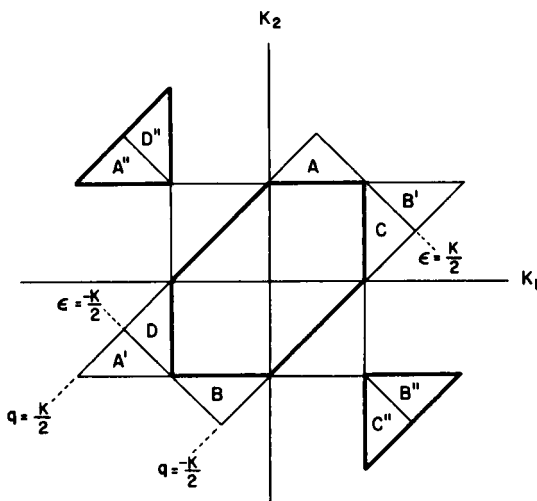

FIG. 1.

variables to κ_1, κ_2:

$$\kappa_1 = \varepsilon + \frac{q}{2}; \; \kappa_2 = \varepsilon - \frac{q}{2}. \tag{D3}$$

The oblique rectangle is the original region of integration, but in regions A and B, $N[k + (\kappa_1/2); -\kappa_2]$ is not equal to $N[\kappa + (q/2), \kappa + \varepsilon]$ since $|\kappa_2{}^{\mu}| > |\kappa^{\mu}/2|$. The latter is in this region equal to

$$N(\mathbf{k} + (\kappa_1/2) + (\mathbf{K}/2); -\kappa_2 + \mathbf{K})$$

(obtained by adding $\mathbf{K}$ to ε). Thus so far as N is concerned, we can replace A by A' and B by B' but O requires correction in these regions. Fortunately, the periodicity of $N(\mathbf{k} + (\kappa_1/2); -\kappa_2)$ in κ_2 now comes to our aid by enabling us to replace A' and D by A'' and D'' which by exactly similar reasoning turns out to give correct results for O also. Likewise B and C are moved to B'' and C''. Our result is

$$P(\mathbf{k}, \mathbf{R}) = \int_I N\left(\mathbf{k} + \frac{\kappa_1}{2}; -\kappa_2\right) O\left(\mathbf{k} + \frac{\kappa_2}{2}; \kappa_1\right) e^{iR \cdot (\mathbf{K}_2 - \mathbf{K}_1)} \, d^3\kappa_1 \, d^3\kappa_2 \tag{D4}$$

where the region of integration for each pair of components $\kappa_1{}^{\mu}$, $\kappa_2{}^{\mu}$ is the heavily outlined region of Fig. 1. We can now rewrite this as

$$P(\mathbf{k}, \mathbf{R}) = \sum_{R_1 \cdot R_2} \int_I N\left(\mathbf{k} + \frac{\kappa_1}{2}, \mathbf{R} + \mathbf{R}_1\right) O\left(\mathbf{k} + \frac{\kappa_2}{2}, \mathbf{R} + \mathbf{R}_2\right)$$

$$e^{i(\mathbf{K}_1 \cdot \mathbf{R}_2 - \mathbf{K}_2 \cdot \mathbf{R}_1)} \, d^3\kappa_1 \, d^3\kappa_2. \tag{D5}$$

To obtain a more useful form, we expand N and O in power series in R_1 and R_2, obtaining

$$P = \int_I \sum_{R_1 R_2 j, l} \frac{\partial^j N}{\partial R^j} \frac{\partial^l O}{\partial R^l} R_1{}^j R_2{}^l e^{i(\kappa_1 R_2 - \kappa_2 R_1)} \, d^3\kappa_1 \, d^3\kappa_2$$

where we have written as if for one dimension, the generalization being easy, but messy in appearance. The summations over $\mathbf{R}_1$ and R_2 now give derivatives of delta functions, which we integrate by parts, obtaining

$$P(k, R) = \sum_{il} \frac{(-1)^j (-i)^{j+l}}{2^{j+l} j! \, l!} \frac{\partial^{j+l} N}{\partial k^l \, \partial R^j} \frac{\partial^{j+l} O}{\partial k^j \, \partial R^l}. \tag{D6}$$

In this derivation we have assumed the existence of continuous functions of R with an infinite radius of convergence which are equal to N and O at all lattice points. Such functions always exist and are not even unique.

The most suitable function is

$$N(k, x) = \sum_R \int_{BZ} d^3q N(\mathbf{k}, \mathbf{R}) e^{iq \cdot (x - R)}. \tag{D7}$$

Even granting this point however, our derivation of (D6) has been rather cavalier, involving the rearrangement of a series which is not absolutely convergent. Let us then look more carefully at (D6) and its derivation. Suppose that $O(k, R)$ has only one Fourier component $O(k; q)$. Then if q is greater than the radius of convergence of any $\partial^l N / \partial R^l$, (D6) will diverge. Thus clearly if $O(k; q)$ has finite values for any q outside the surface of convergence of N and its derivatives, divergence will result, while if neither O nor N has such $O(\mathbf{k}; \mathbf{q})$ we expect convergence. The latter should include many useful cases, but we would like to extend the usefulness of (D6) into the nonconvergent region. To this end we examine (D4) and consider the approximation of O by

$$O_n = \sum_{l=0}^{n} \frac{1}{l!} \frac{\partial^l O}{\partial k^l} \left(\frac{\kappa_2}{2} \right)^l. \tag{D8}$$

For a region inside the radius of convergence, this approximation should be very good; in fact, there exists an $A(k_1)$ such that

$$\left| O - O_n \right| < A(\kappa_1) \left| \frac{\kappa_2}{2} \right|^{n+1} \quad \text{for} \quad \left| \frac{\kappa_2}{2} \right| < k_c - \epsilon$$

where k_c is the radius of convergence of $O(k; \kappa_1)$ in k, and ϵ is a finite positive number, as is A. Furthermore, outside $k_c - \epsilon$, the error should have an upper bound M. Thus the error arising from using O_n for O in (D4) is less than

$$A \int_{BZ} \int_0^{k_c - \epsilon} \left| N \left(k + \frac{\kappa_1}{2}, -\kappa_2 \right) \right| \left| \frac{\kappa_2}{2} \right|^{n+1} d^3\kappa_1 \, d^3\kappa_2$$

$$+ \int_{BZ} \int_{BZ}' |M| \left| N \left(k + \frac{\kappa_1}{2}, -\kappa_2 \right) \right| d^3\kappa_1 \, d\kappa_2 \tag{D9}$$

where the prime indicates that the integral over κ_2 excludes the region of the first integral.

Now if we consider a family of functions $N(\mathbf{k}, \mathbf{R}; \lambda)$ such that as λ becomes small the Fourier transform $N(\mathbf{k}; \mathbf{q}; \lambda)$ becomes more concentrated at small $|q|$, we can determine whether (D9) goes to zero faster than λ^n as λ goes to zero for all n. If it does, then the expansion is asymptotically convergent in λ. (For example, suppose $N(h; q; \lambda)$ behaves like $e^{-|q|/\lambda}$.) If,

furthermore, we now reproduce this process for κ_1, we establish that the following series is asymptotically convergent.

$$P(k, R) = \sum_{jl} \frac{1}{j!l!} \int \frac{\partial^l}{\partial k^l} N(k; \kappa_2) \frac{\partial^j}{\partial k^j} O(k; \kappa_1) \left(\frac{\kappa_2}{2}\right)^j \left(\frac{\kappa_1}{2}\right)^l e^{i(\kappa_1 - \kappa_2)}$$

$$d\kappa_1 \, d\kappa_2 \qquad \text{(D10)}$$

but this series is identical to (D6). Thus the possibility of asymptotic convergence for (D6) is established. Whether it actually is asymptotically convergent depends on whether there is a useful parameter λ with the stated property.

Even in cases where the series neither converges nor is asymptotic, the use of a few terms may provide a reasonable approximation. Thus we may be able to establish that the first three terms will provide an approximation better than λ^2 for small λ, which may be useful even if we cannot proceed to higher orders similarly.