PHYSICAL REVIEW VOLUME 126, NUMBER 5 JUNE 1, 1962

Bloch Electrons in a Magnetic Field

E. I. Blount
Bell Telephone Laboratories, Murray Hill, New Jersey

(Received August 14, 1961)

A formalism recently developed by the author is used to simplify Kohn's derivation of an effective Hamiltonian. The new derivation is valid for arbitrary symmetry and with the inclusion of spin-orbit effects and for bands with degeneracies. The actual form is obtained to second order in the magnetic field. It is explicitly shown that the procedure is asymptotically convergent. This Hamiltonian is then used to obtain the susceptibility in a convenient form. The nature of the actual energy levels is discussed with particular attention to the broadening, gaps in the continuous region, and magnetic breakdown.

SINCE the early days of solid-state physics it has been customary to treat the motion of Bloch electrons in a magnetic field by a Hamiltonian $H(\hbar\mathbf{k}+(e/c)\mathbf{a})$ obtained by substituting $\hbar\mathbf{k}+(e/c)\mathbf{a}$ for $\hbar\mathbf{k}$ in the energy function for a band. This was first justified by Peierls[1] within the limitations of the tight-binding approximation. Since then a number of further efforts have been made to justify this procedure culminating in the work of Kohn,[2] who, by methods which he called "shockingly complicated," was able to establish that the Hamiltonian can indeed be written as a function of $H(\hbar\mathbf{k}+(e/c)\mathbf{a})$, where $H(\mathbf{k})$ is itself a power series in the magnetic field $\mathbf{B}$.

A major purpose of the present paper is to reduce the complication of Kohn's method. We shall also demonstrate explicitly that it is asymptotically convergent. Our procedure is applicable also to bands which do have degeneracies with other bands and to electrons obeying more general wave equations than the nonrelativistic Schrödinger equation. These results will be applied to the steady susceptibility of a crystal and to the determination of the energy levels.

I. THE EFFECTIVE HAMILTONIAN

In this section we shall derive the effective diagonal-in-bands Hamiltonian for electrons described by any of the following one-electron Hamiltonians.

(a) *Schrödinger:*

$$\mathfrak{H}=\frac{1}{2m}\left(\mathbf{p}+\frac{e}{c}\mathbf{a}\right)^2+U, \tag{I.1a}$$

(b) *Pauli:*

$$\mathfrak{H}=\frac{1}{2m}\left(\mathbf{p}+\frac{e}{c}\mathbf{a}\right)^2+\frac{e^2}{8m^2c^2}[\mathbf{p}\cdot(\boldsymbol{\sigma}\times\nabla U)]+U, \tag{I.1b}$$

(c) *Dirac:*

$$\mathfrak{H}=c\boldsymbol{\alpha}\cdot[\mathbf{p}+(e/c)\mathbf{a}]+\beta mc^2+U, \tag{I.1c}$$

where U is the periodic crystalline potential. In the absence of the vector potential $\mathbf{a}$, each of the Hamiltonians has a velocity operator $\mathfrak{V}$, such that

$$\mathfrak{V}=-(i/\hbar)[\mathbf{x},H], \tag{I.2}$$

where H is the Hamiltonian with $\mathbf{a}=0$, whose eigenvalues are $E(\mathbf{k})$. All the above Hamiltonians can be written in the form

$$\mathfrak{H}=H+(e/2c)\{\mathfrak{V}\cdot\mathbf{a}\}+e^2a^2/2mc^2, \tag{I.3}$$

except that for the Dirac case we must set $1/m=0$ and the last term disappears. We now transform this equation to the crystal momentum representation (CMR). [A detailed discussion of this representation in the form used in this paper has been given by the author in another paper[3] which will hereafter be referred to as (F).] In the CMR, $\mathbf{x}$ has the representation $\mathbf{R}+\mathfrak{X}$ where $\mathbf{R}=i\partial/\partial\mathbf{k}$ and $\mathfrak{X}$ is diagonal in the wave vector $\mathbf{k}$, but not in the band index n. Correspondingly, we shall write $\mathbf{a}=A+\mathfrak{A}$, where A depends on $\mathbf{R}$ and $\mathfrak{A}$ on $\mathfrak{X}$. H and $\mathfrak{V}$ are diagonal in $\mathbf{k}$, and H is diagonal in n, while $\mathfrak{V}$ is a matrix in n. Finally,

$$\mathfrak{H}=H+\tfrac{1}{2}\{\mathfrak{V}\cdot(A+\mathfrak{A})\}+(A+\mathfrak{A})^2/2m,$$

where we have also chosen units such that $e/c=1$, or alternatively we have absorbed e/c into $\mathbf{a}$.

We will make much use of another object from the CMR,

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

where the u's are the periodic parts of the Bloch functions $\psi_{n\mathbf{k}}(\mathbf{r})=e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})$. S is related to $\mathfrak{X}$:

$$\mathfrak{X}_{nn'}(k)=\delta(\mathbf{k}-\mathbf{k}')i(\partial/\partial\mathbf{k}')S_{nn'}(\mathbf{k},\mathbf{k}'). \tag{I.5}$$

The whole difficulty in treating the Hamiltonian (I.4) is that $\mathbf{a}$ and in particular A, can be very large. Thus normal perturbation theory cannot be expected to give good results. The author has developed a procedure for treating perturbations which may be large, if they vary slowly from cell to cell [(F), Sec. III]. In this formalism one transforms from the CMR to a mixed representation to which we shall refer as pseudoclassical (PCR), in which an operator is written as a function of two variables $\mathbf{R}$ and $\mathbf{k}$, which can be treated as numbers. The quantum mechanics then rests in the multiplication rules, which are nonlocal. A typical operator then becomes $O_{nn'}(\mathbf{k},\mathbf{R})$. $\mathbf{x}$ is still written as $\mathbf{R}+\mathfrak{X}$, but $\mathbf{R}$ can now be considered a number not an operator, while $\mathfrak{X}$ is still a matrix and depends only on $\mathbf{k}$.

Alternatively, one can continue to regard $\mathbf{k}$ and $\mathbf{R}$ as operators, and our formalism provides a recipe for

---

[1] R. E. Peierls, Z. Physik **80**, 763 (1933).

[2] W. Kohn, Phys. Rev. **115**, 1460 (1959). Additional references will be found in this paper.

[3] E. I. Blount, in *Solid-State Physics*, edited by F. Seitz and D. Turnbull (Academic Press Inc., New York, 1961), Vol. 13.

obtaining other operators in the form of matrices each of whose elements is an operator function of **k** and **R**. Our multiplication rules then are equivalent to treating these functions as fully symmetrized Taylor series in **k** and **R** and multiplying them accordingly. Our formalism is valid asymptotically, however, in cases where the Taylor series diverge, as shown in (F).

We write

$$\mathfrak{H}_{nn'}(\mathbf{k},\mathbf{R})=H(\mathbf{k})+\tfrac{1}{2}\{\mathfrak{B}(\mathbf{k})\cdot[A(R)+\mathfrak{A}(\mathbf{k})]\}+[A(\mathbf{R})+\mathfrak{A}(\mathbf{k})]^2/2m. \tag{I.6}$$

(F) gives a procedure for removing interband matrix elements to arbitrary order, and we could proceed immediately in this direction, but we also wish to assure that our final expression is a function only of $\mathfrak{k}=(\mathbf{k}+A)$ in analogy with the free-electron problem. To this end we make the transformation $S_{nn'}(0,\mathbf{k})$. This has the effect of expressing the Hamiltonian in terms of functions whose periodic parts are the Bloch functions at $\mathbf{k}=0$ or Kohn-Luttinger functions. H, $\mathfrak{B}$, and $\mathbf{x}$ are altered as follows:

$$H(k)\rightarrow H(0)+\mathfrak{B}(0)\cdot\mathbf{k}+k^2/2m. \tag{I.7}$$

(We now choose $\hbar=1$.)

$$\mathfrak{B}\rightarrow\mathfrak{B}(0)+\mathbf{k}/m, \tag{I.8}$$

$$\mathbf{x}\rightarrow\mathbf{R}. \tag{I.9}$$

As a consequence of the last of these, $A+\mathfrak{A}$ is replaced by A and $\mathfrak{H}$ becomes

$$\begin{aligned}\mathfrak{H}'(k,R)&=H(0)+\mathfrak{B}(0)\cdot(\mathbf{k}+A)+(\mathbf{k}+A)^2/2m\\&=H(\mathbf{k}+A).\end{aligned} \tag{I.10}$$

Thus we already have $\mathfrak{H}'$ as a function only of $\mathfrak{k}$, and our procedure as discussed in (F) will be to diagonalize this by successive transformations, each of which will depend only on $\mathfrak{k}$; thus our final result will be of this form. This follows from the fact that commutators of the components of $\mathfrak{k}$ obey the following rule:

$$[\mathfrak{k}_i,\mathfrak{k}_j]_0=i\epsilon_{ijk}B_k,$$

and are thus constant.

We now observe that (I.10) is of the same form as (I.7) with **k** replaced by $\mathfrak{k}$. Thus we can diagonalize it to lowest order, in the sense of (F), by applying the transformation $S_{nn'}(\mathfrak{k},0)$. This transformation itself is not unitary for it depends on both **k** and **R**, and the Poisson bracket $\langle S,S^\dagger\rangle$ does not vanish. It will thus be necessary to use the perturbation procedure of (F).

Before doing so, however, it will be desirable to obtain some simplification of certain formulas in (F) which arise from the fact that we will be using functions only of ($\mathfrak{k}$). In (F), we used the Poisson bracket of two operators N and O,

$$\langle N,O\rangle=\frac{\partial N}{\partial R_i}\frac{\partial O}{\partial k_i}-\frac{\partial N}{\partial k_i}\frac{\partial O}{\partial R_i}, \tag{I.11}$$

where we use the summation convention. If N and O are functions of $\mathfrak{k}$, we obtain

$$\begin{aligned}\langle N,O\rangle&=\left(\frac{\partial N}{\partial\mathfrak{k}_j}\frac{\partial O}{\partial\mathfrak{k}_i}-\frac{\partial N}{\partial\mathfrak{k}_i}\frac{\partial O}{\partial\mathfrak{k}_j}\right)\frac{\partial\mathfrak{a}_j}{\partial k_i}\\&=-\mathbf{B}\cdot\frac{\partial N}{\partial\mathfrak{k}}\times\frac{\partial O}{\partial\mathfrak{k}},\end{aligned} \tag{I.12}$$

where $\mathbf{B}=\nabla\times\mathbf{a}$ is the magnetic field.

This also makes it clear that in higher-order terms, where we also have such terms as

$$\left(\frac{\partial N}{\partial R_i}O\frac{\partial P}{\partial k_i}-\frac{\partial N}{\partial k_i}O\frac{\partial P}{\partial R_i}\right), \tag{I.13}$$

where N, O, P can themselves be derivatives, we can replace this sum by

$$\sum_i\left(\frac{\partial N}{\partial R_i}O\frac{\partial P}{\partial k_i}-\frac{\partial N}{\partial k_i}O\frac{\partial P}{\partial R_i}\right)=-B_l\epsilon_{ijl}\frac{\partial N}{\partial k_i}O\frac{\partial P}{\partial k_j}. \tag{I.14}$$

We now write from (F) the formula for the operator product $(SHS^\dagger)_O$ to second order in **B**. [In (F), we used "order" to indicate the number of differentiations with respect to **R** in a particular term. Here, this coincides with order in **B** since **R** appears only in A.]

$$\begin{aligned}(SHS^\dagger)_O={}&SHS^\dagger+\tfrac{1}{2}i\left(\langle S,H\rangle S^\dagger+S\langle H,S^\dagger\rangle+\frac{\partial S}{\partial R_i}H\frac{\partial S^\dagger}{\partial k_i}-\frac{\partial S}{\partial k_i}H\frac{\partial S^\dagger}{\partial R_i}\right)-\frac{1}{8}\left(\frac{\partial^2 S}{\partial k_i\partial k_j}H\frac{\partial^2 S^\dagger}{\partial R_i\partial R_j}+2\frac{\partial S}{\partial k_i}\frac{\partial H}{\partial k_j}\frac{\partial^2 S^\dagger}{\partial R_i\partial R_j}\right.\\&+S\frac{\partial^2 H}{\partial k_i\partial k_j}\frac{\partial^2 S^\dagger}{\partial R_i\partial R_j}-2S\frac{\partial^2 H}{\partial k_i\partial R_j}\frac{\partial^2 S^\dagger}{\partial k_j\partial R_i}+S\frac{\partial^2 H}{\partial R_i\partial R_j}\frac{\partial^2 S^\dagger}{\partial k_i\partial k_j}+\frac{\partial^2 S}{\partial R_i\partial R_j}\frac{\partial^2 H}{\partial k_i\partial k_j}S^\dagger+\frac{\partial^2 S}{\partial R_i\partial R_j}H\frac{\partial^2 S^\dagger}{\partial k_i\partial k_j}\\&+2\frac{\partial^2 S}{\partial R_i\partial R_j}\frac{\partial H}{\partial k_i}\frac{\partial S^\dagger}{\partial k_j}-2\frac{\partial^2 S}{\partial k_i\partial R_j}\frac{\partial H}{\partial R_i}\frac{\partial S^\dagger}{\partial k_j}-2\frac{\partial^2 S}{\partial k_i\partial R_j}\frac{\partial^2 H}{\partial R_i\partial k_j}S^\dagger+\frac{\partial^2 S}{\partial k_i\partial k_j}\frac{\partial^2 H}{\partial R_i\partial R_j}S^\dagger-2\frac{\partial^2 S}{\partial k_i\partial R_j}H\frac{\partial^2 S^\dagger}{\partial k_j\partial R_i}\\&-2\frac{\partial^2 S}{\partial k_i\partial R_j}\frac{\partial H}{\partial k_j}\frac{\partial S^\dagger}{\partial R_i}-2\frac{\partial S}{\partial k_i}\frac{\partial H}{\partial R_j}\frac{\partial^2 S^\dagger}{\partial R_i\partial k_j}+2\frac{\partial S}{\partial R_i}\frac{\partial H}{\partial R_j}\frac{\partial^2 S^\dagger}{\partial k_i\partial k_j}+2\frac{\partial S}{\partial k_j}\frac{\partial^2 H}{\partial k_i\partial R_j}\frac{\partial S^\dagger}{\partial R_i}-2\frac{\partial S}{\partial k_j}\frac{\partial^2 H}{\partial R_i\partial R_j}\frac{\partial S^\dagger}{\partial k_i}\\&\left.-2\frac{\partial S}{\partial R_j}\frac{\partial^2 H}{\partial k_i\partial k_j}\frac{\partial S^\dagger}{\partial R_i}-2\frac{\partial S}{\partial R_j}\frac{\partial H}{\partial k_i}\frac{\partial^2 S^\dagger}{\partial k_j\partial R_i}+2\frac{\partial^2 S}{\partial k_i\partial k_j}\frac{\partial H}{\partial R_i}\frac{\partial S^\dagger}{\partial k_j}+2\frac{\partial S}{\partial R_j}\frac{\partial^2 H}{\partial k_j\partial R_i}\frac{\partial S^\dagger}{\partial k_i}\right),\end{aligned} \tag{I.15}$$

where all products on the right are matrix products. Using (I.14), this equation can be considerably simplified:

$$(SHS^\dagger)_O = SHS^\dagger - \tfrac{1}{2}iB_l\epsilon_{lij}(S_iH_jS^\dagger + SH_iS_j^\dagger + S_iHS_j^\dagger) \\ -\tfrac{1}{8}B_lB_m\epsilon_{lij}\epsilon_{mnp}(S_{in}HS_{jp}^\dagger + SH_{in}S_{jp}^\dagger + S_{in}H_{jp}S^\dagger + 2S_iH_nS_{jp}^\dagger + 2S_{in}H_jS_p^\dagger + 2S_iH_{nj}S_p^\dagger), \quad (I.16)$$

where we have used subscripts to indicate differentiation with respect to $\mathfrak{k}$. To obtain $(SS^\dagger)_O$, we merely set $H=1$ in (I.16):

$$(SS^\dagger)_O = 1 - \tfrac{1}{2}iB_l\epsilon_{lij}S_iS_j^\dagger - \tfrac{1}{8}B_lB_m\epsilon_{lij}\epsilon_{mnp}S_{in}S_{jp}^\dagger. \quad (I.17)$$

We can now remove the first-order term in $(SS^\dagger)_O$ by multiplying S on the left by $(1+g^{(1)})$, where $g^{(1)} = \tfrac{1}{4}iB_l\epsilon_{lij}S_iS_j^\dagger$, which is Hermitian. Applying this to (I.16), we obtain to second order

$$\mathfrak{H} = E(\mathfrak{k}) + \tfrac{1}{4}iB_l\epsilon_{lij}\{S_i(S_j^\dagger SHS^\dagger - H_jS^\dagger - HS_j^\dagger) + (SHS^\dagger S_i - S_iH - SH_i)S_j^\dagger - S_iH_jS^\dagger - SH_iS_j^\dagger\} \\ -\tfrac{1}{8}B_lB_m\epsilon_{lij}\epsilon_{mnp}\{(S_{in}HS_{jp}^\dagger + SH_{in}S_{jp}^\dagger + S_{in}H_{jp}S^\dagger + 2S_iH_nS_{jp}^\dagger + 2S_{in}H_jS_p^\dagger + 2S_iH_{jn}S_p^\dagger) \\ -[S_iS_j^\dagger, (S_nH_pS^\dagger + S_nHS_p^\dagger + SH_nS_p^\dagger)]_{(+)} - [(S_{in}S_j^\dagger + S_iS_{jn}^\dagger), (SHS^\dagger)_p]_{(-)} \\ + \tfrac{1}{2}S_iS_j^\dagger SHS^\dagger S_nS_p^\dagger - \tfrac{1}{2}[S_{in}S_{jp}^\dagger, SHS^\dagger]_{(+)} + \tfrac{3}{4}[S_iS_j^\dagger S_nS_p^\dagger, SHS^\dagger]_{(+)}\}, \quad (I.18)$$

where the subscripts $+$ and $-$ indicate anticommutators and commutators, respectively. Here the first group of terms in second order comes from (I.16), the second group from $[g^{(1)},\langle SHS^\dagger\rangle]_{(+)}$, the third from $\langle g^{(1)},SHS^\dagger\rangle + \langle SHS^\dagger,g^{(1)}\rangle$, the fourth from $g^{(1)}SHS^\dagger g^{(1)}$, and the last two from $[g^{(2)},SHS^\dagger]_{(+)}$, where $g^{(2)}$ is chosen so that if $H=1$, the second-order terms vanish. That is, $g^{(2)}$ is chosen to make $(SS^\dagger)_O$ unitary to second order:

$$g^{(2)} = \tfrac{1}{16}B_lB_m\epsilon_{lij}\epsilon_{mnp}(S_{in}S_{jp}^\dagger - \tfrac{3}{32}S_iS_j^\dagger S_nS_p^\dagger).$$

The addition of $g^{(1)}$ and $g^{(2)}$ to S does not remove interband terms to second order, but before proceeding we rewrite (I.18) in terms of the $\mathfrak{X}$'s and $\mathfrak{B}$'s, using (I.2) and (I.5).

$$\mathfrak{H} = E(\mathfrak{k}) + \tfrac{1}{4}B_l\epsilon_{lij}\{[\mathfrak{X}_i,v_j]_+ + [\mathfrak{X}_i,\mathfrak{B}_j]_+\} + B_lB_m\epsilon_{lij}\epsilon_{mnp}\{\tfrac{1}{32}[\mathfrak{X}_i[\mathfrak{X}_n,\alpha_{jp}]_+]_+ + \tfrac{1}{16}[v_j[\mathfrak{X}_n,\partial\mathfrak{X}_i/\partial\mathfrak{k}_p]_+]_+ \\ + \tfrac{1}{16}[\mathfrak{X}_i,(\partial/\partial\mathfrak{k}_j)[\mathfrak{X}_n,\mathfrak{B}_p]_+]_+ + (\delta_{jn}/8m)\mathfrak{X}_i\mathfrak{X}_p - \tfrac{1}{32}[\partial\mathfrak{X}_i/\partial\mathfrak{k}_n,[\partial\mathfrak{X}_p/\partial\mathfrak{k}_j,E]_-]_- \\ + \tfrac{1}{16}i[((\partial\mathfrak{X}_i/\partial\mathfrak{k}_n)\mathfrak{X}_j\mathfrak{X}_p + \mathfrak{X}_i\mathfrak{X}_n\partial\mathfrak{X}_p/\partial\mathfrak{k}_j), SHS^\dagger]_-\}, \quad (I.19)$$

where v_j is the intraband velocity operator $v_j = \partial E/\partial\mathfrak{k}_j$. The first-order term is obtained straightforwardly from (I.18), while the second-order term requires complicated rearrangements. The last term has only interband elements and can be removed immediately.

Equation (I.19) can now be put in a very convenient form if we introduce the following notation:

$$\tfrac{1}{2}B_l\epsilon_{lij}\mathfrak{X}_i = \mathfrak{A}_j^s, \quad (I.20)$$

where the superscript indicates that this is the part of the vector potential associated with $\mathfrak{X}$ in the symmetric gauge, regardless of the gauge we started with;

$$\tfrac{1}{4}B_l\epsilon_{lij}[\mathfrak{X}_i,\mathfrak{B}_j]_+ = -\mathbf{B}\cdot\mathfrak{M}', \quad (I.21)$$

where $\mathfrak{M}'$ is the nonsingular part of the magnetic moment operator. Equation (I.19) becomes to second order

$$\mathfrak{H} = E(\mathfrak{k} + \mathfrak{A}^s(\mathfrak{k}+\mathfrak{A}^s)) - \mathbf{B}\cdot\mathfrak{M}'(\mathfrak{k}+\mathfrak{A}^s) + (\mathfrak{A}^s)^2/2m \\ + \tfrac{1}{8}[(\partial\mathfrak{A}_j^s/\partial k_n)[\partial\mathfrak{A}_n^s/\partial k_j,E]_-]_-, \quad (I.22)$$

where $E(\mathfrak{k}+\mathfrak{A}^s)$ and $\mathfrak{M}'(\mathfrak{k}+\mathfrak{A}^s)$ are defined by the power-series expansion of E and $\mathfrak{M}'$ about $\mathfrak{k}$. We have left out the last term of (I.19), which can be easily removed by a transformation. This expression except for the last term is exactly what the Hamiltonian of an ion in an external magnetic field looks like, the first term corresponding to $(1/2m)[\mathbf{p}+(e/c)\mathbf{a}]^2$ for the center-of-charge-and-mass motion, the second term being the magnetic moment, and the third the molecular diamagnetism. The fourth term is a specifically crystalline effect as is the appearance of $\mathbf{A}^s$ in the first two.

While (I.22) is a particularly concise expression it is not very convenient for computation, and we now show a different rewriting of (I.19) using a different magnetic moment operator $\mathfrak{M}$.

$$\mathfrak{M}_l = -\epsilon_{lij}\tfrac{1}{4}[\mathfrak{X}_i,V_j]_+, \quad (I.23)$$

where V_j has only interband matrix elements. This leads us to the form

$$\mathfrak{H} = E(\mathbf{k}) - \tfrac{1}{2}[(\mathbf{B}\times\mathbf{v})\cdot\boldsymbol{\mathfrak{X}}]_+ - \mathbf{B}\cdot\mathfrak{M} + \tfrac{3}{8}[\mathfrak{A}_i^s,[\mathfrak{A}_j^s,\alpha_{ij}]_+]_+ + (\mathfrak{A}^s)^2/2m + \tfrac{1}{2}[v_j,[\mathfrak{A}_p^s,\partial\mathfrak{A}_j^s/\partial k_p]_+]_+ \\ + \tfrac{1}{4}[[\mathfrak{A}_j,v_p]_-,\partial\mathfrak{A}_p/\partial k_j]_- + \tfrac{1}{8}[\partial\mathfrak{A}_j/\partial k_p,[\partial\mathfrak{A}_p^s/\partial k_j,E]_-]_- - \tfrac{1}{2}[\mathfrak{A}_j^s,(\partial/\partial\mathfrak{k}_j)\mathbf{B}\cdot\mathfrak{M}]_+. \quad (I.24)$$

In this form the various orders of $\mathbf{B}$ are kept separate and terms involving $\mathbf{v}$ are separated from those involving V, which occurs only in $\mathfrak{M}$.

It is interesting to note that the second term can be written

$$-\tfrac{1}{2}[(\mathbf{B}\times\mathbf{v})\cdot\boldsymbol{\mathfrak{X}}]_+ = -\tfrac{1}{2}[\mathbf{F}_L\cdot\boldsymbol{\mathfrak{X}}]_+. \quad (I.25)$$

This is the energy of a dipole $\boldsymbol{\mathfrak{X}}$ in the Lorentz force $\mathbf{F}_L$, without the factor $\frac{1}{2}$ which occurs in the magnetic moment term. It means that if we add an electric field to the problem, the Hamiltonian to first order in both fields is

$$E(\mathfrak{k})-\mathbf{E}\cdot\mathbf{R}-\tfrac{1}{2}[(\mathbf{E}+\mathbf{F}_L),\cdot\mathfrak{X}]_+-\mathbf{B}\cdot\mathfrak{M}, \qquad (\text{I}.26)$$

that is, the electric and Lorentz forces appear on an equal footing. When the procedure we have used is applied to the Dirac equation, treated as a band theory, it is found that the intraband part of the first-order term is

$$-\tfrac{1}{2}[\mathbf{v}\times\mathfrak{X}]_+-\mathfrak{M}=\hbar\boldsymbol{\sigma}/m^*c,$$

where m^* is the transverse mass, this equation being valid for all $\mathbf{k}$. Thus, here it is the rather peculiar combination $\frac{1}{2}[\mathbf{v}\times\mathfrak{X}]_++\frac{1}{4}[\mathbf{V}\times\mathfrak{X}]_+$ which is simply related to $\boldsymbol{\sigma}$ the spin.

It is now appropriate to remove the first-order interband matrix elements, and for this purpose we simply use second-order perturbation theory, obtaining

$$\begin{aligned}\mathfrak{H}=E(\mathfrak{k})-(\mathbf{B}\times\mathbf{v})\cdot\mathfrak{x}-\mathbf{B}\cdot\mathbf{m}+\tfrac{3}{8}[\mathfrak{A}_i{}^s[\mathfrak{A}_j{}^s,\alpha_{ij}]_+]_++(\mathfrak{A}^s)^2/2m+\tfrac{1}{2}[v_j,[\mathfrak{A}_p{}^s,\partial\mathfrak{A}_j{}^s/\partial k_p]_+]_+\\ +E_{VD}+\tfrac{1}{4}[[A_i{}^s,v_p]_-,\partial A_p{}^s/\partial k_j]_--\tfrac{1}{2}[\mathfrak{A}_i{}^s,(\partial/\partial k_i)\mathbf{B}\cdot\mathbf{M}]_++\tfrac{1}{8}[\partial A_j{}^s/\partial\mathfrak{k}_p,[\partial A_p{}^s/\partial\mathfrak{k}_j,E]_-]_-,\end{aligned} \qquad (\text{I}.27)$$

where the lower case $\mathfrak{x}$ and $\mathbf{m}$ designate intraband matrix elements of $\mathfrak{X}$ and $\mathfrak{M}$ (which for bands degenerate throughout $\mathbf{k}$ space is taken to mean a two×two matrix in the degenerate bands[3]), and E_{VD} is defined as follows:

$$E_{VD,nn}=-\tfrac{1}{2}\sum_{n'\neq n}\frac{|\mathbf{B}\cdot\mathbf{M}_{nn'}+[(\mathbf{B}\times\mathbf{v});\mathbf{X}_{nn'}]_+|^2}{E_n-E_{n'}}. \qquad (\text{I}.28)$$

The term in E_{VD} containing $|B\cdot M_{nn'}|^2$ is the Van Vleck paramagnetism, the term in $|X^2|$ is like a dielectric effect of the Lorentz force, and the cross term is an interference between these effects. The symbols $\mathbf{M}$, $\mathbf{X}$, $\mathbf{A}$ refer to *interband* matrix elements. We have left out some terms involving intraband matrix elements of $\mathfrak{A}$ because they vanish in the diagonal terms.

This is still not the most useful form for the second-order terms. As a first step, we eliminate all factors of the form $\partial A_\mu/\partial k_\nu$ for $\mu\neq\nu$. Then the second-order terms have the form

$$\begin{aligned}\mathfrak{h}_2=-\tfrac{1}{8}[\alpha_{\mu\nu}[A_\mu{}^sA_\nu{}^s]_+]_++\frac{(\mathfrak{A}^s)^2}{2m}-\tfrac{1}{2}[\nabla_{\mathbf{k}}\cdot\mathbf{A}^s[\mathbf{A}^s\cdot\mathbf{v}]_+]_++E_{VD}+\tfrac{1}{2}[\nabla_{\mathbf{k}}\cdot\mathbf{A}^s,\mathbf{B}\cdot\mathbf{M}]_++\tfrac{1}{8}[\nabla_{\mathbf{k}}\cdot\mathbf{A}^s[\nabla_{\mathbf{k}}\cdot\mathbf{A}^s,E]]\\ +\tfrac{3}{8}[\alpha_{\mu\nu}[\mathfrak{a}_\mu{}^s\mathfrak{a}_\nu{}^s]_+]_++\tfrac{1}{2}[v_\mu,[\mathfrak{a}_\nu{}^s\partial\mathfrak{a}_\mu{}^s/\partial k_\nu]_+]_+-\tfrac{1}{2}[\mathfrak{a}_\mu{}^s,\partial\mathbf{B}\cdot\mathbf{m}/\partial k_\mu]_+\\ +\frac{1}{2}\frac{\partial}{\partial k_\nu}([A_\nu{}^s,[A_\mu{}^s,v_\mu]_+]_+-\tfrac{1}{2}[A_\nu{}^s,[\nabla_{\mathbf{k}}\cdot\mathbf{A}^s,E]_-]_--[A_\nu{}^s,\mathbf{B}\cdot\mathbf{M}]_+),\end{aligned} \qquad (\text{I}.29)$$

where $\mathfrak{a}^s$ is the intraband part of $\mathfrak{A}^s$.

$\nabla_{\mathbf{k}}\cdot\mathbf{A}^s$ and $\mathbf{B}\cdot\mathbf{M}$ have very similar structure which can be exploited to obtain a very simple form for the total Hamiltonian.

$$\begin{aligned}\mathfrak{H}=H(\mathfrak{k}+2\mathfrak{a}^s(\mathfrak{k}+2\mathfrak{a}^s))-\mathbf{B}\cdot\mathbf{m}(\mathfrak{k}+2\mathfrak{a}^s)+\mathfrak{a}^s\cdot\mathbf{v}_A+E_{VD}'+A^2/2m-\tfrac{1}{4}\alpha_{\mu\nu}[A_\mu{}^sA_\nu{}^s]_+\\ -\tfrac{1}{2}i[(R_\mu+\mathfrak{x}_\mu),(A_\mu{}^s\mathfrak{B}_\nu A_\nu{}^s+A_\nu{}^s\mathfrak{B}_\nu A_\mu{}^s+v_\nu[A_\nu{}^sA_\mu{}^s]_+)]_-.\end{aligned} \qquad (\text{I}.30)$$

In the first term, the argument is to be interpreted as $\mathfrak{k}$, plus $2\mathfrak{a}$ evaluated at $\mathfrak{k}+2\mathfrak{a}$. Furthermore, E_{VD}' is defined as follows:

$$E_{VD,n}'=\sum_{n'=n}\frac{|(\mathbf{A}^s\cdot\mathbf{V}+\{\mathbf{v}\cdot\mathbf{A}^s\})_{nn'}|^2}{E_n-E_{n'}}. \qquad (\text{I}.31)$$

We note that the operator $\mathbf{A}^s\cdot\mathbf{V}$ is *not* Hermitian. $\mathbf{v}_A$ is defined as

$$\mathbf{v}_A=-(i/\hbar)[\mathfrak{x},H-\mathbf{B}\cdot\mathbf{m}]_0,$$

the operator commutator of $\mathfrak{x}$ with $\mathfrak{H}$ evaluated to first order.

$$\mathbf{v}_A=+(i/\hbar)[\mathfrak{x},\mathbf{B}\cdot\mathbf{m}]+(i/\hbar)[\mathfrak{x},(\mathbf{B}\times\mathbf{v})\cdot(\mathfrak{x})].$$

We have used the form $\mathfrak{a}\cdot\mathbf{v}_A$ to emphasize that $\mathbf{v}_A$ is a part of the total velocity operator and this term thus has a natural place in the Hamiltonian. It was pointed out at the beginning that $1/m$ must be set equal to zero for Dirac electrons. We, nevertheless, expect a term like $\mathbf{A}^2/2m$ in the Hamiltonian. We now see that it comes from E_{VD}'. For the sum over n' must include negative energy states, for which $E_n-E_{n'}\sim 2mc^2$. The sum then becomes approximately $(\mathbf{A}\cdot\mathbf{VV}\cdot\mathbf{A})/2mc^2$. For the Dirac electron $\mathbf{VV}\sim 1c^2$ and the total contribution of the sum over negative energy states is approximately $A^2/2m$.

In (I.30), the first two terms are the only ones which exist for free electrons, while $\mathbf{A}^2/2m$ and E_{VD}' are the only ones for molecules and atoms. A different categorization is that the first three terms are determined by simple one-band parameters discussed in (F). These parameters have their analogs in the Dirac theory (F). The next two terms are explicitly interband or "molecular" in origin. The sixth and seventh terms are specifically connected with the interaction between intraband and interband effects.

The form (I.30) is valid for any choice of phases of the Bloch functions, and is indeed insensitive to the generalized phase transformation discussed in (F). The establishment of this fact serves only as a check on our result and will not be performed here.

One final rearrangement of (I.30) will be given which will be useful in the next section.

$$\begin{aligned}\mathfrak{H}=H(\mathfrak{k})-\mathbf{B}\cdot\mathbf{m}-\mathbf{B}\times\mathbf{v}\cdot\mathfrak{a}+(A^s)^2/2m+E_{VD}'-\tfrac{1}{4}\alpha_{\mu\nu}[A_\mu{}^sA_\nu{}^s]_+ \\ -\tfrac{1}{2}i[(R_\mu+\mathfrak{x}_\mu),\,(A_\mu{}^s\mathfrak{V}_\nu A_\nu{}^s+A_\nu{}^s\mathfrak{V}_\nu A_\mu{}^s+v_\nu[A_\nu{}^sA_\mu{}^s]_+)]_-+(\partial/\partial k_\nu)(v_\mu[\mathfrak{a}_\mu{}^s\mathfrak{a}_\nu{}^s]) \\ -\tfrac{1}{2}iv_\mu[\mathfrak{a}_\mu{}^s,[\mathfrak{x}_\nu,\mathfrak{a}_\nu]_-]_++\tfrac{1}{2}i[2R_\mu+\mathfrak{x}_\mu,\,[\mathfrak{a}^s,\mathbf{B}\cdot\mathbf{m}]_+]_--\tfrac{3}{8}[\mathbf{B}\cdot\mathbf{\Omega},\mathbf{B}\cdot\mathbf{m}]_+,\end{aligned}\qquad\text{(I.32)}$$

where

$$\mathbf{\Omega}=-i(\mathbf{R}+\mathfrak{x})\times(\mathbf{R}+\mathfrak{x})$$

and has been discussed in (F).

At this point we have accomplished our first objective, to show that an effective Hamiltonian can be found without interband matrix elements, to any desired order, and we have carried it out to second order.

Some final remarks are in order concerning the first-order terms in (I.27). In a crystal with inversion symmetry and no spin-orbit coupling, **m** vanishes and $\mathfrak{x}$ is a gradient $\nabla_{\mathbf{k}}\varphi$. For such a band, which in addition intersects no other bands, φ and $\mathfrak{x}$ can be made zero. In a band which has intersections, this cannot usually be done, though $\nabla_k\cdot\mathfrak{x}$ can be made zero. In crystals with inversion symmetry and spin-orbit coupling, all bands occur in degenerate pairs throughout **k** space, and we consider such pairs of bands to be one band of two-component wave functions. Then, in general **m** will not vanish nor will $\mathfrak{x}$, but the trace of **m** for each band vanishes everywhere, while the trace of $\mathfrak{x}$ behaves like $\mathfrak{x}$ in the absence of spin-orbit coupling. (In such bands **m** and $\mathfrak{x}$, like all other operators, are treated as 2×2 matrices, just as in the Pauli theory of spin.)

So far we have simply taken over the procedure of (F) without concerning ourselves with matters of convergence, either of the multiplication procedure (I.14), or the perturbation procedure. We now consider these in order.

In (F) it is pointed out that the multiplication formula (I.14) cannot be expected to converge if one factor has Fourier components $O(\mathbf{k};\mathbf{q})$ for **q** outside the range of convergence of another factor $N(\mathbf{k};\mathbf{q})$ as a function of **k**. In our present problem let us consider the first step when we take the product $S(\mathbf{k}+A,0)H(\mathbf{k}+A)S^\dagger(\mathbf{k}+A,0)$. $H(\mathbf{k}+A)$ is quadratic in **k** and A and thus gives no trouble, but $S(\mathbf{k}+A,0)$ and its adjoint have Fourier components over the whole Brillouin zone (BZ), whereas as a function of **k** it does not in general have such a large range of convergence as a power series. We can, however, explicitly see that it is asymptotically convergent at small field. This is done in Appendix A.

Each step of the perturbation procedure involves the introduction of a new pair of operators, which must be multiplied by the expressions already obtained. These new operators are all functionals of S and the polynomials H and $\mathfrak{V}$. Consequently any finite number of them will not disturb the asymptotic character of the expressions obtained. Likewise, the removal of interband elements is asymptotic in that the error in the nth order is of order B^{n+1} or higher at small B. This establishes the asymptotic convergence of our total procedure.

One exception must be made to this statement. At degeneracies we have established the asymptotic convergence of the multiplication rules, but of course the perturbation theory will break down. In such cases, we usually use degenerate perturbation theory. Thus in the present problem, we might try to use degenerate perturbation theory at degeneracies. Unfortunately it does us no particular good, for in such neighborhoods the derivatives with respect to $\mathfrak{k}$ of the transformations become very large and there is no possibility of a development in power series. Indeed, the interband matrix elements have an important meaning in such places as we shall see. Of course, (I.24), which gives a Hamiltonian depending only on $\mathfrak{k}$, is valid and useful near degeneracies.

Finally, it is desirable to show what the basis function in the transformed representation looks like. We shall content ourselves with the zeroth order approximation since this contains the basic physical idea. From $S(\mathbf{k},\mathbf{R})=S(\mathbf{k}+A,\mathbf{k})$ we must calculate the matrix elements $S(\mathbf{k}',\mathbf{k}'')$ and apply them to the Bloch function. We find

$$\begin{aligned}\psi_{n\mathbf{k}}&=\sum_{n'}\int d^3k'\,S_{nn'}(\mathbf{k},\mathbf{k}')\psi_{n'\mathbf{k}'}(\mathbf{r})\\ &=\sum_{n',R}\int d^3k'\,S_{nn'}\left(\frac{\mathbf{k}+\mathbf{k}'}{2}+A,\,\frac{\mathbf{k}+\mathbf{k}'}{2}\right)e^{i\mathbf{R}\cdot(\mathbf{k}-\mathbf{k}')}e^{i\mathbf{k}'\cdot\mathbf{r}}u_{n'\mathbf{k}'}(\mathbf{r})\\ &=e^{ik\cdot r}\sum_{n',R}\int d^3k'\,S_{nn'}\left(\mathbf{k}+A+\frac{(\mathbf{k}'-\mathbf{k})}{2}\right)e^{i(\mathbf{k}'-\mathbf{k})\cdot(\mathbf{r}-\mathbf{R})}u_{n\mathbf{k}'}(\mathbf{r}).\end{aligned}\qquad\text{(I.33)}$$

This can be expanded as a series in B just as the multiplication rules can, and the result is

$$\psi_{n\mathbf{k}}'=e^{i\mathbf{k}\cdot\mathbf{r}}u_{n,\mathbf{k}+A}(\mathbf{r})+\cdots. \qquad \text{(I.34)}$$

This first term shows that the basic transformation is to change the periodic function $u_{n\mathbf{k}}$ to the nearly periodic function $u_{n,\mathbf{k}+A}$ characteristic of the crystal kinetic momentum $\mathbf{k}+A$. This is entirely in accord with one's intuitive feelings, and some authors have used this as a basis for attacking the problem of this paper.

Note added in proof. In the case of a band whose energy range does not overlap that of any other band, it is easy to see that the matrix elements with other bands can in principle be removed exactly. This follows from the fact that for fields below a certain B_c, the exact energy levels will be so arranged that a number of states equal to the number of states in the band will have energies well separated from the remaining states. Linear combinations of these wave functions can then be taken which will be infinite in extent and similar to our modified Bloch functions. Since these functions have no strict requirements to obey such as being of the form $e^{i\mathbf{k}\cdot\mathbf{r}}\times$periodic function, there is no difficulty. For this case, then, our proof gives an asymptotic form of the Hamiltonian, but we know that the removal of interband elements is exact.

If the bands do overlap, on the other hand, we cannot guarantee the exact removal of interband elements and this part too is only asymptotic. This distinction is important in connection with magnetic breakdown (see Sec. III).

II. THE STEADY SUSCEPTIBILITY

The first use we shall make of (I.30) is to calculate the susceptibility. This is determined from the free energy by the relation

$$\chi=\frac{M}{B}=-\frac{1}{B}\frac{\partial F}{\partial B}\bigg|_{\zeta}, \qquad \text{(II.1)}$$

while F is given by

$$F=N\zeta+\mathrm{Tr}f(\mathfrak{H}), \qquad \text{(II.2)}$$

$$f(\mathfrak{H})=-kT\ln\{1+\exp[(\zeta-\mathfrak{H})/kT]\}. \qquad \text{(II.3)}$$

To obtain f correct to order B^2, we would like to expand it as a power series in $\mathfrak{H}$ and use our multiplication rule [(I.16), with $\mathfrak{H}$ substituted for S and $S^{\dagger}$]. f has only a small finite radius of convergence however, which indeed vanishes at $T=0$. Therefore, we make use of the Laplace transform of f, since $\exp(s\mathfrak{H})$ has an infinite radius of convergence. Thus we write

$$f=\int_{c-i\infty}^{c+i\infty}\varphi(s)\exp(s\mathfrak{H})ds, \qquad \text{(II.4)}$$

$$\exp(s\mathfrak{H})=\sum_n\frac{s^n}{n!}\mathfrak{H}^n. \qquad \text{(II.5)}$$

As shown in (F), the operator product $\mathfrak{H}^n$ evaluated at $(\mathbf{k},\mathbf{R})$ is not equal to the nth power of $\mathfrak{H}(\mathbf{k},\mathbf{R})$. Treating each term by the expansion procedure of (F), we obtain

$$\begin{aligned}\exp(s\mathfrak{H})=&\sum_n\frac{s^n}{(n-2)!2}[\mathfrak{H}(k,R)]^{n-2}\{\tfrac{1}{8}i\langle\mathfrak{H},\mathfrak{H}\rangle-\tfrac{1}{8}\langle\langle\mathfrak{H},\mathfrak{H}\rangle\rangle\}\\&+\sum_n\frac{s^n}{(n-3)!}\frac{[\mathfrak{H}(k,R)]^{n-3}}{6}(-\tfrac{1}{8})\langle\langle\mathfrak{H}\mathfrak{H}\mathfrak{H}\rangle\rangle'\\&+\sum_n\frac{s^n}{(n-4)!}\frac{[\mathfrak{H}(k,R)]^{n-4}}{4!}(-\tfrac{1}{4})\langle\mathfrak{H},\mathfrak{H}\rangle^2+\cdots, \qquad \text{(II.6)}\end{aligned}$$

where the prime indicates that only terms in which each $\mathfrak{H}$ is differentiated at least once are to be included, the others being in the previous term. We note that $\langle\mathfrak{H},\mathfrak{H}\rangle=0$; summing over n, and inserting the result in (II.4), we find

$$\begin{aligned}f(\mathbf{k},\mathbf{R})=&f(\mathfrak{H}(k,R))\\&-\tfrac{1}{8}B_{\lambda}B_{\mu}\epsilon_{\lambda\iota\kappa}\epsilon_{\mu\nu\rho}(\tfrac{1}{2}f''\alpha^{\iota\nu}\alpha^{\kappa\rho}+\tfrac{1}{3}f'''\alpha^{\iota\nu}v^{\kappa}v^{\rho}). \qquad \text{(II.7)}\end{aligned}$$

The integration over $\mathbf{k}$ and sum over bands yields

$$F(R)=F_0-\frac{1}{48}B_{\lambda}B_{\mu}\epsilon_{\lambda\iota\kappa}\epsilon_{\mu\nu\rho}\left(\frac{1}{2\pi}\right)^3\int d^3k\,\mathrm{Tr}f_0'\alpha^{\iota\nu}\alpha^{\kappa\rho},$$

where Tr signifies the trace of the matrix at fixed $\mathbf{k}$, $\mathbf{R}$; f_0 is f' the Fermi function, and F_0 is the trace of the first term in (II.7):

$$F_0=F(B=0)+\left(\frac{1}{2\pi}\right)^3\int d^3k\,\mathrm{Tr}\left\{f_0\mathfrak{h}^{(2)}+f_0'\frac{(\mathfrak{h}^{(1)})^2}{2}\right\}, \qquad \text{(II.9)}$$

$\mathfrak{h}^{(1)}$ and $\mathfrak{h}^{(2)}$ are, respectively, the first- and second-order terms in (I.30). The first-order contribution of $\mathfrak{h}^{(1)}$ vanishes by time-reversal for any nonferromagnetic material.

The procedure outlined could in principle be extended to higher orders, but would remain asymptotic at best. In particular it would not be able to pick up the de Haas-van Alphen effect. This is most readily found by making use of the actual spectrum and has been done by Lifshitz and Kosevch[4] for arbitrary $\mathfrak{H}(\mathfrak{k})$. There is no need to reproduce it here.

The asymptotic character of our power series in B differs from that in the free electron case, however, for in that problem $\exp(-\beta\mathfrak{H})$ has a convergent power series in $\mathfrak{H}$ and the asymptotic character arises only on taking the inverse Laplace transform. In the present case $\exp(-\beta\mathfrak{H})$ itself is given only asymptotically by (II.6). The reason for this difference is that for free electrons each term in (II.5) is a polynomial and is thus exactly evaluable by the series (II.6) which terminates for each order n.

---

[4] I. M. Lifshitz and A. M. Kosevch, Doklady Akad. Nauk S.S.S.R. **96**, 963 (1954).

The total susceptibility can now be written in detail:

(1) The first term is χ_{LP}, the Landau diamagnetism with an effective mass tensor, given by the second term in (II.8). As a term in the free energy, it is written

$$F_{LP}=-\frac{1}{48}B_\lambda B_\mu \epsilon_{\lambda\iota\kappa}\epsilon_{\mu\nu\rho}\left(\frac{1}{2\pi}\right)^3\int \mathrm{Tr}f_0'\alpha_{\iota\nu}\alpha_{\kappa\rho}d^3k. \quad \text{(II.10)}$$

(2) The contribution of the last term in (II.9) is partially cancelled by some of the second term, namely, that part coming from the eighth through the eleventh terms of (I.30). (This requires the nonobvious result that $\int f_0 v_\mu \,\mathrm{Tr}\{\mathfrak{a}_\mu,[\mathfrak{x}_\nu,\mathfrak{a}_\nu]\}d^3k=0$.) The remainder will be called the Pauli paramagnetism, and it involves an effective magnetic moment

$$F_p=\left(\frac{1}{2\pi}\right)^3\int \mathrm{Tr}f_0'\frac{(\mathbf{B}\cdot\mathbf{m})^2}{2}d^3k. \quad \text{(II.11)}$$

(3) The Van Vleck-dielectric paramagnetism comes from the fifth term in (I.30):

$$F_{VD}=\left(\frac{1}{2\pi}\right)^3\int \mathrm{Tr}f_0 E_{VD}'d^3k.$$

(4) The atomic diamagnetism, from the fourth term of (I.30), strictly speaking from the Dirac point of view, arises from the previous term, as discussed following (I.30). The split is nevertheless useful since in solids, electrons are far from the relativistic region.

$$F_{at}=\left(\frac{1}{2\pi}\right)^3\int \mathrm{Tr}f_0\frac{(\mathbf{A}^s)^2}{2m}d^3k.$$

(5) The next term is similar in appearance to the last but involves an effective mass so that its sign is not definite, a characteristic common to the remaining terms, which distinguishes them from the previous terms. We shall call this contribution pseudoatomic:

$$F_{pa}=\left(\frac{1}{2\pi}\right)^3\int \mathrm{Tr}f_0(-\alpha_{\mu\nu})\left\{\frac{A_\mu{}^s A_\nu{}^s}{2}\right\}d^3k.$$

(6) The last term in (I.30) has no simple description. We shall denote its contribution by F_Ω:

$$F_\Omega=\left(\frac{1}{2\pi}\right)^3\int \mathrm{Tr}f_0(-\tfrac{3}{8})[\mathbf{B}\cdot\mathbf{\Omega},\mathbf{B}\cdot\mathbf{m}]_+d^3k.$$

(7) Finally we have another term whose existence depends on that of a Fermi surface for we can integrate by parts, obtaining

$$F_7=\left(\frac{1}{2\pi}\right)^3\int \mathrm{Tr}f_0'\left(\frac{v_\mu}{2}\right)(A_\mu{}^s\mathfrak{B}_\nu A_\nu{}^s$$
$$+A_\nu{}^s\mathfrak{B}_\nu A_\mu{}^s+v_\nu\{A_\nu{}^s A_\mu{}^s\}d^3k.)$$

The relative importance of the various terms will differ considerably from one type of substance to another. In particular, of course, in insulators and intrinsic semiconductors the terms (1), (2), (7) are zero. When the band gaps are large in such materials, the atomic diamagnetism (4) will predominate. Narrow band gaps will increase the sizes of (3) and (5), which will tend to cancel, however, as (3) is definitely paramagnetic while the largest contribution to (5) will come from regions of large negative mass. Term (6) will be large only if, in addition to a small band gap, the material has strong spin-orbit coupling.[5] Otherwise, $\mathbf{m}$ will be about the free electron value and $\mathbf{\Omega}$ will be small. The contribution when large should be predominantly diamagnetic.

In a metal on the other hand the terms (1) and (2) may be comparable to or larger than the atomic diamagnetism. A discussion of the comparison between calculations and theory has been given by Wilson.[6] In the good metals the remaining terms are expected to be smaller, as the energy difference between energy levels at given $\mathbf{k}$ is generally large, while these terms, having effectively three such factors in denominators are increased by small energy differences.

In the case of doped semiconductors one has the opportunity to study the contribution of small numbers of added carriers, unmasked by the atomic and Van Vleck terms. This has been done in particular by Yafet and Bowers.[7] Any of the terms can be large in this case, though at small concentrations, the first two are especially enhanced by a factor of approximately the energy gap over the Fermi level.

In the semimetals one has effectively a doped semiconductor in which the intrinsic susceptibility is unknown. In Bi, for instance, it has long been supposed that the susceptibility is predominantly due to the carriers, but since it is now known that the electrons have a large magnetic moment,[5] these would appear to give a net paramagnetic contribution. On the other hand, the term (5) above could give rise to a very large diamagnetism from electrons in the filled bands near the band edges. But this is not the only possibility, since the atomic term should also be increased by small energy gaps, this contribution possibly coming from a larger region of $\mathbf{k}$ space. Term (6) F_Ω will also be diamagnetic in this case and comparable in size to (5). Whichever of these terms is the primary source of the diamagnetism, it should remain large if the electrons and holes are removed. This can apparently be achieved by making Bi-Sb alloys,[8] and the experiments of Shoenberg and Zaki Uddin[9] on such alloys provide some

---

[5] M. H. Cohen and E. I. Blount, Phil. Mag. **5**, 115 (1960); G. E. Smith, J. K. Galt, and F. R. Merritt, Phys. Rev. Letters **4**, 276 (1960).

[6] A. H. Wilson, *The Theory of Metals* (Cambridge University Press, New York, 1953), 2nd ed.

[7] Y. Yafet and R. Bowers, Phys. Rev. **115**, 1165 (1959).

[8] A. L. Jain, Phys. Rev. **114**, 1518 (1959).

[9] D. Shoenberg and Zaki Uddin, Proc. Roy. Soc. (London) **A156**, 687 (1936).

confirmation of this prediction, though they did not examine specimens in the range which later proved to have well-defined semiconducting properties. Further experiments on such alloys are strongly indicated. In the semimetals in general, more information can be obtained about specific group of carriers from more specialized experiments such as de Haas-van Alphen effect, cyclotron resonance,[10] ultrasonic attenuation,[11] and anomalous skin effect,[12] but the steady susceptibility then serves as a check on the completeness of this information.

III. EIGENFUNCTIONS AND EIGENVALUES

Having now determined the effective one-band Hamiltonian, it is of interest to look at the eigenfunctions and eigenvalues as well. As a first step in this process, it is convenient to consider what constants of the motion can be determined in advance. At this point the simplicity of the problem depends on the choice of gauge. In particular, if we choose a doubly-singular gauge (see Appendix A), the effective Hamiltonian depends only on one component of $\mathbf{R}$. This is not quite the same in general, as to say that we have two constants of the motion as we shall now see.

Consider first the case where the magnetic field is along a lattice vector of the crystal which we shall call the z axis. We choose the gauge so that $a_x=0$, $a_y=Bx$, $a_z=0$. Then our effective Hamiltonian is $H(k_x, k_y+Bx, k_z)$. k_y and k_z are constants of the motion, and the CMR wave equation is one dimensional. We could have come to this conclusion by considering the original Schrödinger Hamiltonian, so it does not depend on any of our approximations. If we now suppose that $\mathbf{B}$ is not along a short lattice vector but along a long one, so that its direction cosines have large denominators, nothing is changed in principal, but the simple statements can be made only in reference to a different unit cell. For we can obtain crystal momentum components as constants of the motion only if the Hamiltonian is periodic in the corresponding space directions. Thus we would proceed as follows for such a $\mathbf{B}$ as we have just described. We choose the z axis parallel to $\mathbf{B}$ and the shortest lattice vector in that direction as one edge of the unit cell. This determines a plane in reciprocal space which is perpendicular to $\mathbf{B}$. If the lattice vector is large, the unit cell in this plane will also be large and the spacing of the planes small. Thus the BZ will be thin in the z direction and of great area, becoming thinner and more extensive as the z-lattice vector increases in length, that is, as the denominators of the direction cosines of $\mathbf{B}$ become larger. Finally, if the direction cosines are irrational, the zone has no finite thickness, and the BZ becomes in effect an infinite plane. An intermediate case exists if $\mathbf{B}$ is perpendicular to one reciprocal lattice vector, but has some irrational direction cosines. In this case, the BZ becomes a strip of infinite length but finite width. If the BZ has a finite thickness and a_z is chosen zero, and the other components of $\mathbf{a}$ are independent of z, $\mathbf{k}\cdot\mathbf{B}/|\mathbf{B}|$ is a constant of the motion. As the BZ becomes thinner, the range of values of $\mathbf{k}\cdot\mathbf{B}/|\mathbf{B}|$ decreases and the variation of wave functions and energies with $\mathbf{k}\cdot\mathbf{B}/|\mathbf{B}|$ becomes unimportant, until in the case of irrational orientations, it can be neglected. The eigenfunctions in all cases do not fall off at large z. For rational orientation their z dependence is like Bloch functions, albeit with very large periods. For irrational orientations, the eigenfunctions remain large at large z, even though not periodic. This reflects the fact that as the BZ becomes a pancake, the unit cell becomes a long thin rod, so that the periodic part of the wave function becomes a function only of z, assuming reasonably decent behavior; conversely the z dependence relates only to the periodic part. Thus, the Wannier wave function $f(\mathbf{R})$ should show no z dependence, but only depend on x and y. The preceding discussion will probably require some getting used to and is most easily seen by imagining the irrational orientations as a limiting case of rational orientations.

We have then that $\mathfrak{H}$ can be written as $\mathfrak{H}(\mathfrak{k}_x,\mathfrak{k}_y,``k_z")$, where the quotation marks are intended to mean that there is a k_z dependence only for rational field directions. At this point, we have reduced the problem to two dimensions, without in general, however, having any constants of the motion. We can now choose our gauge in such a way as to reduce the equation to one dimension, namely by picking the vector potential to have only one component and to vary in only one direction. We will choose $a_z=a_x=0$, $a_y=Bx$, so that k_y is a constant of the motion. In general k_y will have an infinite range, but when the BZ is a strip or a finite rectangle, the range of k_y is finite.

Thus, the situation is not unlike that for free electrons. In the latter case, there is only one other quantum number left n, which is discrete, and on which alone the energy depends. n labels the "harmonic oscillator states." In our present problem, we have yet no guarantee that n is a discrete quantum number, and it must clearly do the work of both the harmonic oscillator number and the band index. To see that n cannot be so simple, suppose we have a state $\psi_1=\psi(n,k_y,k_z;\mathbf{x})$. If we now make a gauge transformation, adding Bx_0 to the vector potential, the original Hamiltonian $\mathfrak{H}(p_x,p_y,p_z;\mathbf{x})$ becomes $\mathfrak{H}(p_x, p_y+Bx_0, p_z;\mathbf{x})$ and the wave function $\psi(n, k_y+Bx_0, k_z;\mathbf{x})$ has the same energy as ψ_1. If we now translate the system by a lattice vector $\mathbf{R}$, the Hamiltonian is $\mathfrak{H}(p_x, p_y+B(x_0-R_x), p_y;\mathbf{x})$ and the wave function is $\psi(n, k_y+Bx_0, k_z;\mathbf{x}-\mathbf{R})$. Thus, if we choose x_0 a lattice vector $\mathbf{R}$, we obtain from $\psi(n,k_y,k_z;\mathbf{x})$ a different eigenfunction $\psi(n, k_y+BR_x, k_z;\mathbf{x}-\mathbf{R})$ with the same energy. The spectrum considered as a function of k_y is then periodic

---

[10] J. K. Galt, W. A. Yager, F. R. Merritt, and B. B. Cetlin, Phys. Rev. **114**, 1396 (1959).

[11] D. H. Reneker, Phys. Rev. **115**, 303 (1959).

[12] G. E. Smith, Phys. Rev. **115**, 1561 (1959).

with a periodicity Br_x where r_x is the minimum R_x (which is zero for irrational orientations). The spectrum also has the periodicity K_y, where K_y is the minimum reciprocal lattice vector in the y direction. We now distinguish three cases:

(1) K_y and K_x are infinite; r_x and r_y are zero. The spectrum is independent of k_y (or of k_x had we chosen the "opposite" gauge $a_x=-By$, $a_y=0$).

(2) K_y and K_x are finite; r_x and r_y are also finite. Then the occurrence of two periodicities K_y and Br_x means that unless they are commensurate the spectrum is independent of k_y, and likewise of k_x in the "opposite" gauge. The periods are commensurate if Br_x/K_y is rational, that is, if the flux through a unit cell in the plane perpendicular to $\mathbf{B}$ is a rational multiple of hc/e. The spectrum is not independent of k_z.

(3) K_y is finite, K_x infinite; r_x is zero, r_y finite. Br_x is zero so the spectrum is independent of k_y, but in the opposite gauge; Br_y is finite and the spectrum is not constant, but may have a periodicity Br_y as a function of k_x.

Since we have only been able to show that a level at k_y is related to one at k_y+Br_x by translation symmetry and gauge invariance, we must suppose that for rational orientations, the corresponding levels at other (k_y+Bx_0) have different energies, though for small x_0 they must be arbitrarily close. This will soon be verified. We have seen, however, that except for particular values of B, the spectrum is independent of k_y. This leads to the conclusion that the spectrum for given k_y and k_z is not discrete.

The reader may have gotten the impression that there is a remarkable difference between rational and irrational orientation which seems physically unreasonable. This difference is only apparent, and is one of mathematical simplicity alone. If we did not attempt, at rational orientations, to make the most convenient choice of gauge, we would not notice the discontinuity. As we have seen, however, it is possible at rational orientations to make a choice of gauges, such that we greatly simplify the mathematics. In both cases, there is a continuous spectrum when all quantum numbers are considered. Because of the degeneracy of eigenvalues, it is possible to use very different-looking sets of eigenfunctions. A somewhat analogous situation would arise in a crystal, which we subject to a constant sinusoidal perturbation, whose repetition vector is $\mathbf{T}$. If $\mathbf{T}$ is a lattice vector, the problem is still periodic, and the eigenfunctions can be written as Bloch functions. For other values of $\mathbf{T}$ this is not so, but we expect no discontinuities at the preferred values.

The above reasoning has not used the effective Hamiltonian, but rather has proceeded from the Hamiltonian (I.1). It will soon appear that the results from the effective Hamiltonian are in agreement. They will also give a more detailed picture of the spectrum and eigenfunctions than can be deduced from invariance arguments alone.

We have seen that the effective Hamiltonian can be written in the form $\mathfrak{H}(k_x, k_y+Bx, "k_z")$. For purposes of solving the equation of motion we can now forget about k_z immediately. We also wish to get rid of k_y. This can be done by making the transformation $e^{-ik_yk_x/B}$ which replaces the above Hamiltonian with $H(k_x,Bx)$. This is generally permissible, but not when K_x is finite, for in that case the wave function must be periodic in k_x. When k_y is removed from the Hamiltonian, we require that this boundary condition be replaced with one, that the wave function change by the factor $e^{-ik_yK_x/B}$ when k_x is increased by K_x.

Thus, we have replaced a set of different Hamiltonians with a set of different boundary conditions. When K_x is infinite no such action is required. The effect is that in either case we consider all the energy levels of $H(k_x,B_x)$ not merely those corresponding to eigenfunction periodic in k_x.

We now have a Hamiltonian $H(k_x,Bx)$, which is defined over a plane or set of planes, and which may be periodic in one or both arguments. Since the Hamiltonian is not simple, but involves only one pair of conjugate variables, the WKB procedure is indicated as a method of solution.

It is well known that the straightforward application of the WKB method leads to the familiar semiclassical quantization rules for the energy levels, and thus to the de Haas-van Alphen and related effects. It is also well known that some orbits called open extend indefinitely in k space and that the spectrum is continuous in the region of these orbits. It has been pointed out by Zil'berman and others[13] that the quantized levels are in fact subject to a slight broadening, while Kohn[2] has found that there are gaps in the open orbit region. Finally, it has recently been noted by Cohen and Falicov[14] that in strong fields the energy levels and wave functions can change in a qualitative way. These subjects are closely related and a unified discussion is desirable and appropriate at this point, because they are also closely related to the problem of the effective Hamiltonian.

We start with our two-dimensional Hamiltonian $\mathfrak{H}(k,Bx)$ which we rewrite as $\mathfrak{H}(p,q)$ where $q=k$ and $p=Bx$ and p, q satisfy the commutation relation $[p,q]=iB$.

In applying the WKB method, we suppose in effect that we have a relief map of energy as a function of p and q. We then make a two-dimensional plot of the contours for a particular energy E. The contour will consist in general of a number of curves which we shall call orbits, some closed, some open, that is infinite

---

[13] G. E. Zil'berman, Soviet Phys. (JETP), **3** (**30**), 835 (1957); **5** (**32**), 208 (1957); **6** (**33**), 299 (1958). P. G. Harper, Proc. Phys. Soc. (London) **A68**, 874, 879 (1955); A. D. Brailsford, *ibid.* **A70**, 275 (1957).

[14] M. H. Cohen and L. M. Falicov, Phys. Rev. Letters **7**, 231 (1961).

in extent. For the moment, we consider only closed orbits. If any orbit encloses an area of $(n+\frac{1}{2})B$ where n is an integer it corresponds to an eigenfunction. By going through this procedure for all energies, we obtain the quasi-classical energy levels. The procedure is only approximate, of course. The Hamiltonian has off-diagonal matrix elements; there are also overlap integrals since the wave functions are not exactly orthogonal. If we consider two neighboring orbits with n's differing by δn, then as $B\rightarrow 0$, the overlap and off-diagonal Hamiltonian matrix elements between them decrease as B^2 in the usual WKB approximation; by proceeding to the indicated higher approximations, they can be made to decrease more rapidly than any given power of B. For orbits differing by a given energy ϵ, or for orbits about different centers, on the other hand, these matrix elements fall off as $e^{-c/B}$, where c depends on the contours, but not on B. Its nature is discussed in Appendix B. This cannot be significantly improved by going to higher approximations in the WKB procedure.

On the other hand, the situation we have described is very similar to one encountered in certain idealized problems in solid-state physics, which are discussed by the tight-binding approximation. We have a number of approximate wave functions on different centers, which overlap very slightly, the matrix elements connecting them being small compared to the energy splittings between states on a given center. Two situations must be distinguished. In the first, the field is in an irrational direction and there is no periodicity. Neighboring orbits will have different energies in the WKB approximation and the spectrum is a continuum in that approximation. Then each level will be slightly shifted by interaction with its neighbors, somewhat as a nuclear spin may be thought of as having a slightly different local field from that of its neighbors. Since the spectrum is already continuous, no change in this characteristic will occur. When the field is in a rational direction, the WKB approximation will make orbits displaced from one another by a reciprocal lattice vector degenerate and although the over-all spectrum is continuous, the spectrum for a given k_z is discrete. In this case, discussed by Zil'berman, Harper, and Brailsford[13] interaction between neighboring orbits will produce a broadening into a band just as in the tight-binding treatment of bands. The width of the band will essentially be determined by the matrix element between neighboring orbits. In complex cases where not all the orbitals are identical, the band structure will also be more complicated, but can easily be determined in any given case. The physics of this broadening can be described as follows. The WKB levels do not take account of the atomic structure, but levels corresponding to different values of k_y have their centers at different positions relative to the unit cell (this point is meaningful only when the field is perpendicular to at least one reciprocal lattice vector), and thus have different energies. An additional broadening arises because levels whose k_y's differ by K_y have finite matrix elements, even though they are far apart. There will then be an interaction between them, and they will be somewhat mixed. These two sources of broadening are not actually different and appear so only in a given gauge. If the opposite gauge is chosen, the two types of broadening are interchanged. Our analysis in particular makes clear that at small field the bands are sharply defined with regions densely filled with states, alternating with energy gaps. The width of the levels is evidently of the order of $e^{-c/B}$, like the interorbit matrix elements.

We now turn to the case of open orbits. These are of two types: (1) Periodic open orbits which have periodic orbits in $\mathbf{k}$ space; these can occur only for field orientations such that the BZ in the plane is finite or a strip. (2) Nonperiodic open orbits, which can occur for irrational field directions. The latter are in no sense quantized. The former, with a proper choice of gauge, have k_y as a good quantum number; periodic boundary conditions on k_x result in a discrete spectrum for fixed k_y, though in the WKB approximation there is a continuous spectrum when all k_y's are considered.[15] As in the case of closed orbits, the open orbits will have matrix elements with each other, and with any closed orbits which may coexist with them, which decrease like $e^{-c/B}$. For the nonperiodic orbits, this results in shifts but does not change the character of the spectrum. For the periodic orbits on the other hand, the interactions introduce gaps in the spectrum.[2] This can be seen easily as follows. Just as we pictured the closed orbits as analogous to tightly bound electrons, we can imagine the open orbits to be analogous to nearly free electrons. Their orbits in (p,q) space are essentially identical to the orbits in phase space of nearly-free electrons in a weak one-dimensional periodic potential. The open orbits can be considered to belong to groups. These groups are separated by regions of closed orbits, and alternate groups represent electrons whose net velocities have opposite directions. We may suppose that the interactions between orbits within a group have been diagonalized to any desired degree and consider only interactions between orbits in different groups. So far there will be no energy gaps. But matrix elements will exist only between orbits for which k_y differs by BR_x. They are so small that we need only consider the interaction of orbits of practically the same energy. Two such orbits will now be split in energy by the amount of the matrix element between them. Thus just as in the one-dimensional nearly-free electron case, the energy levels will be practically unaffected, except for those orbits which have matrix elements with degenerate levels. At these points gaps will be introduced into the spectrum. For this purpose only neighboring groups of orbits need be considered.

---

[15] E. I. Blount, Phys. Rev. Letters 4, 114 (1960).

So far we have seen that for closed orbits the energy levels are broadened by an amount which decreases exponentially in $1/H$, and that similar gaps are opened in the open orbit spectrum for fixed k_z. Actually of course, when k_z is taken into account—or the infinite extent of the BZ in the plane when there is no k_z—the over-all spectrum is continuous. Finally, we must consider what happens in the transition region between closed and open orbits. This corresponds to the region which neither the tight-binding nor nearly-free electron approximations are very good, and requires more careful analysis. This can be done by an extension of the WKB method to cover the case where the energy is nearly equal to the height of a barrier. The necessary connection formulas can be found in Landau and Lifshitz.[16] Essentially this method has been used by Kohn, who found that there is a symmetry between energies above and below the barrier height, such that if an energy ϵ measured from the barrier height is permitted, the energy $-\epsilon$ is forbidden and conversely. This applies to periodic open orbits and the corresponding closed orbits, and is strictly true only for energies so close to the saddle point that the classical period is proportional to $\ln|\epsilon|$. It also considers interaction between only two adjacent groups of open orbits and the closed orbits between them. (The derivation of this result has not been published by Kohn, and the author is indebted to him for a discussion of it.) For low fields, the regions of validity of this approximation and our tight-binding and nearly-free-electron procedures overlap and they give equivalent results.

Since the broadening is, at low fields, negligible compared to the Landau splittings, it can for most purposes be neglected and the orbits treated as isolated. It is then much easier to map them back into a compact Brillouin zone and reintroduce k_z, for purposes of calculating densities of states. Only for open or very extended orbits is the mapping into a plane or flat zone necessary or useful in this approximation.

The final topic of this section is magnetic breakdown, which also represents the breakdown of the effective diagonal-in-bands Hamiltonian. For this discussion we revert to the Hamiltonian of (I.24) which includes interband matrix elements of order B, and consider the case where orbits of nearly the same energy in different bands also approach very closely in $\mathbf{k}$ space. To first order the Hamiltonian is

$$\mathfrak{H}(\mathfrak{k})-\tfrac{1}{2}\{\mathbf{B}\times\mathbf{v}\cdot\mathbf{X}\}-\mathbf{B}\cdot\mathbf{M}, \tag{III.1}$$

where we have absorbed the diagonal first-order terms of (I.24) into the zero-order term. The last two terms give rise to matrix elements between two orbits. We now choose directions so that the line of closest approach between the orbits is in the y direction and set $k_y=0$ as before. Then the normalized WKB wave function for one of the orbits will be

$$\psi_1(k)=\left(\frac{1}{m_1v_{y1}}\right)^{\frac{1}{2}}\exp\left[\frac{i}{B}\int_0^k R_1(k)dk\right], \tag{III.2}$$

where k stands for k_x, R for R_x a function of k_x for a given orbit, m_1 is the effective mass of the orbit, $m_1=\oint dk/v_{1y}$, and v_{y1} the y component of velocity.

The matrix element between two such orbits in different bands is then

$$Q_{12}=\frac{B}{(m_1m_2)^{\frac{1}{2}}}\int\frac{\frac{1}{2}(v_{1y}+v_{2y})X-M}{(v_{1y}v_{2y})^{\frac{1}{2}}}\times\exp\left[\frac{i}{B}\int_0^k(R_1-R_2)dk'\right]dk, \tag{III.3}$$

where X is the x component of $\mathbf{X}_{12}$ and M is the z component of $\mathbf{M}_{12}$. It is shown in Appendix B that if the orbits do not overlap Q_{12} decreases at small field as $e^{-c'/B}$. Here we consider the important special case of two bands, which are nearly degenerate. In fact we further suppose that they can be obtained from two bands which "cross" each other, by the addition of a small perturbation Δ. To a sufficient approximation, the Hamiltonian then looks like this in the unperturbed representation:

$$\begin{pmatrix}\mathbf{k}\cdot\mathbf{v}_1 & \Delta\\ \Delta & \mathbf{k}\cdot\mathbf{v}_2\end{pmatrix}, \tag{III.4}$$

with eigenvalues

$$E_\pm=\tfrac{1}{2}\mathbf{k}\cdot(\mathbf{v}_1+\mathbf{v}_2)\pm\{\Delta^2+\tfrac{1}{4}[\mathbf{k}\cdot(\mathbf{v}_1-\mathbf{v}_2)]^2\}^{\frac{1}{2}}. \tag{III.5}$$

Then for any field direction not in the direction of $\mathbf{v}_1-\mathbf{v}_2$ the orbits are hyperbolas and the bands have equal values of v_y. Furthermore the X term in the Hamiltonian is much larger than the M term. Thus for values of R near the turning point, our Hamiltonian becomes

$$H(k_x)-F_LR-F_LX, \tag{III.6}$$

where $F_L=Bv_y$. This is identical in form to the Hamiltonian which produces Zener breakdown in insulators. Thus magnetic breakdown is essentially Zener breakdown, where the electric force is supplanted by the Lorentz force. We can then obtain a good approximation for the asymptotic behavior at low field by substituting in the formula for Zener breakdown as given by Kane[17]; the important factor is the transmission amplitude which we write

$$\alpha=\exp\left(-\frac{\pi}{4}\frac{|\Delta|}{F_L|X|}\right). \tag{III.7}$$

We further have $X=v_x/2\Delta$ in our simple model, where $|v_x|$ is $|[\frac{1}{2}(\mathbf{v}_1-\mathbf{v}_2)\times\mathbf{B}/|\mathbf{B}|]|$ and $F_L=Bv_y$, so

$$\alpha=\exp\left(-\frac{\pi}{2}\frac{\Delta^2}{B|v_xv_y|}\right). \tag{III.8}$$

---

[16] L. Landau and E. Lifshitz, *Quantum Mechanics—Nonrelativistic Theory* (Addison-Wesley Publishing Company, Reading, Massachusetts, 1959).

[17] E. O. Kane, J. Phys. Chem. Solids **12**, 181 (1960).

This result is obtained by approaching from low fields. Alternatively, we may approach from high fields and consider the effect of Δ on the orbits which are present when it is zero. We can indeed calculate the matrix elements for any perturbation $\Delta_{mn'}(\mathbf{k}',\mathbf{k}'')$ if we show how to do it for the case when $\mathbf{k}'$, $\mathbf{k}''$ differ by $\boldsymbol{\kappa}$. We then write $\Delta_{nn'}(\mathbf{k}',\mathbf{k}'')=\Delta_{nn'}(\mathbf{k})$ where $\mathbf{k}=(\mathbf{k}'+\mathbf{k}'')/2$. We also choose our gauge so that the x direction is parallel to $\mathbf{k}'-\mathbf{k}''=\boldsymbol{\kappa}$. Then between two wave functions $f_1(q)$ and $f_2(q)$, the matrix element will be

$$Q_{12}=\int f_{1m}(q-\tfrac{1}{2}\kappa)\Delta_{nn'}(q)f_{2n'}(q+\tfrac{1}{2}\kappa)dq. \qquad \text{(III.9)}$$

If we suppose that the orbits for $\Delta=0$ belong to a particular band we have $n=n'$ and drop the index.

Now assuming that $\Delta(q)$ is slowly varying, we use WKB wave functions and the saddle point method to find that the matrix element consists of a contribution from each intersection of the orbits, with the orbit 2 displaced by κ. The absolute value of each contribution is

$$\pi^{\frac{1}{2}}\Delta B^{\frac{1}{2}}/[m^*|\tfrac{1}{2}(v_{1x}v_{2y}-v_{1y}v_{2x})|^{\frac{1}{2}}], \qquad \text{(III.10)}$$

where m^* is the geometric mean of the cyclotron masses of the orbits and the v's are evaluated at the points of intersection. (This formula breaks down, of course, if one of the velocities gets very small.) This quantity in particular is quite insensitive to whether 1 and 2 are the same orbit. Provided 2 is a neighboring concentric orbit, the matrix element is nearly constant and m^* is simply the cyclotron mass of either. The ratio of this expression and the separation between neighboring levels is then

$$\pi^{\frac{1}{2}}\Delta/[\tfrac{1}{2}B(v_{1x}v_{2y}-v_{2x}v_{1y})]^{\frac{1}{2}}. \qquad \text{(III.11)}$$

In our previous model, then this is just the square root of twice the exponent in (III.7). This corresponds to choosing κ to be a reciprocal lattice vector in the final structure.

The significance of the parameter we have found is more easily understood when we note that the matrix elements we have evaluated have had the form B/m^* multiplied by a function of the parameter, where B/m^* is the geometric mean of the frequencies of the two orbits involved. Now for degenerate orbits, the matrix element is the frequency of transfer from one orbit to the other in the simple-minded resonance picture, provided it is small compared to either orbit frequency. Thus, when we divide by either frequency, we obtain a sort of switching coefficient, the probability that the electron will jump from one orbit to the other on a given passage through the intersection or point of closest approach. When we divide, instead, by the geometric mean of the frequencies, we obtain a mean switching coefficient. Thus, in the low-field limit the switching coefficient between bands is given by (III.7), while in the high-field limit, the switching coefficient between "unperturbed" orbits is given by (III.11). We have derived the latter as applying between two "oscillator" levels of the same k_y, k_z. The same value would have been obtained had we considered the periodic potential to vary in the y direction and calculated the matrix element between equal oscillators at values of k_y differing by κ. This latter quantity is more closely related to the orbit picture as drawn in phase-space. It was the observation by Priestley of an orbit corresponding to the free-electron orbits in Mg that led Cohen and Falicov to postulate breakdown. These authors used the parameter $\Delta/\hbar\omega_c$ as a criterion. We now see that the correct parameter is much more favorable to breakdown. For some cases we can give a simpler form of our parameter. When the "unperturbed" orbit has an energy that can be written $E=m^*v^2/2$, and v_x and v_y are comparable, we have

$$Q\sim/\hbar\omega_c\sim\Delta/(E\hbar\omega_c)^{\frac{1}{2}},$$

where in experimentally attainable cases, E will usually be the Fermi energy which is much larger than $\hbar\omega_c$. Our derivation shows that this weakening applies not only to the case of breakdown, but to any interaction with long wavelength perturbations such as acoustical or optical vibrations.

From the emphasis we have put on magnetic breakdown it should not be inferred that this is the only case where the perturbation theory of (F) and Sec. I breaks down, though it is very likely unique in being a qualitative effect. Whenever $\frac{1}{2}\{\mathbf{B}\times\mathbf{v})\cdot\mathbf{X}_{nn'}\}$ or $\mathbf{B}\cdot\mathbf{m}$ is comparable to $E_n-E_{n'}$, care must be exercised. In such cases it will probably frequently be possible to devise methods better suited than that used here.

IV. SUMMARY

In this paper we have presented what we feel is a simple, even intuitive approach to the establishment of an effective one-band Hamiltonian for a Bloch electron in a magnetic field. Some of our manipulations are complicated, but we would emphasize that this complication arises only in an effort to simplify the actual expressions for the second-order terms. The actual proof that interband matrix elements can be removed to arbitrary order leaving a Hamiltonian $\mathfrak{H}(\mathfrak{k})$ is completed in the paragraph following (I.10).

Of previous attempts to accomplish this objective, we single out two for particular discussion. Zil'berman[13] has used a procedure equivalent to our zeroth-order transformation and showed that in the cases he considered, inversion symmetry and no spin-orbit coupling, the correction terms are of second order. Kohn[1] has used series of transformations in powers of $\mathfrak{k}$ and rearranged them to a series in powers of B to remove interband terms formally to all orders. There is a close relation between his method and ours. The simplification in our method arises from connecting the series in $\mathfrak{k}$ to one in B in the general formalism of (F), namely in the derivation of our basic multiplication rule. This has also made possible the systematic discussion of the validity of the rearrangement of the series, and in particular has made it possible to demonstrate the asymptotic character of its "convergence." This is of some importance both positively and negatively. For on the one hand, Zil'berman showed that for a particular model the eigenstates are broadened exponentially in $-1/B$, while working with a Hamiltonian accurate only to order B at best, whereas we have now established that the effective Hamiltonian can also be given with a similar asymptotic behavior. Also, the de Haas-van Alphen effect is associated with nonanalytic behavior at small B. On the other hand, if Kohn's procedure actually converged, the phenomenon of breakdown would be seriously modified, if not eliminated.

The difficulty in trying to treat the magnetic problem previously has been that one would like to "get one's hands" on both $\mathbf{k}$ and $\mathbf{R}$ simultaneously, because one has intuitively felt that the basis functions should be chosen differently at different positions in the lattice. With the usual techniques this has been difficult to do, whereas this is just the great merit of the PCR. The bands are defined as functions of $\mathbf{k}$ and $\mathbf{R}$ from the start, and our transformations are essentially a perturbation-theoretic approach to the diagonalization of the Hamiltonian for each point in "phase space." The noncommutativity leads to some specifically quantum-mechanical effects because one cannot completely localize in both $\mathbf{k}$ and $\mathbf{R}$, but basically one is "locally" diagonalizing the bands. We personally find it convenient to work with objects for which $\mathbf{k}$ and $\mathbf{R}$ can both be treated as c numbers, but this point is not essential and they can be thought of exclusively as operators if preferred.

Note added in proof. Recently, Roth [J. Phys. Chem. Solids (to be published)] and Wannier and Fredkin [Phys. Rev. **125**, 1910 (1962)] have written papers on the subject of Sec. I, which also greatly simplify the proof. Basically, both of these methods amount to the same thing as ours, though approached from quite different points of view; in both cases the method is applicable only to the present problem, rather than being a special case of a more general formalism.

The expression for the susceptibility in Sec. II is so far as we know, the first one given which is valid for electrons with spin-orbit coupling. It has also been possible to put the expression in what seems to be the most concise form yet given. We have also used simple one-band parameters whenever possible to facilitate comparison with experiments. It will also be noted that the PCR permits a relatively simple derivation of the Landau-Peierls formula.

Essentially equivalent formulas for the susceptibility have been given by Adams,[18] Hebborn and Sondheimer,[19] and Enz.[20]

APPENDIX A

In order to discuss the convergence of the multiplication rule (I.15), we start with a result of reference 3. Suppose we have a product $P=NM$ where N and M can be considered as belonging to families of functions $N(\mathbf{k},\mathbf{R};\lambda)$ or $N(\mathbf{k};\mathbf{q};\lambda)$ where λ is some parameter. Then consider the integrals

$$\begin{aligned} I_1&=\int_0^{k_c(M)-\epsilon}|N(\mathbf{k};2\mathbf{q};\lambda)|\,|q|^n d^3q,\\ I_2&=\int_{\mathrm{BZ}}' |N(\mathbf{k};\mathbf{q};\lambda)|\,d^3q, \end{aligned}\tag{A1}$$

when the prime on the integral in I_2 means that the region of integration of I_1 is to be excluded. This region, in turn, is to be interpreted as one which extends to within some finite distance ϵ of the range of convergence $k_c(M)$ of the power series expansion of M as a function of $\mathbf{k}$ (*sic*). If both I_1 and I_2 are $O(\lambda^n)$ for all n at small λ, and the same is true of analogous integrals of M the series expansion of the product is asymptotically convergent in λ. In particular, if $N(\mathbf{k};\mathbf{q};\lambda)\sim e^{-\epsilon|\mathbf{q}|/\lambda}$ where ϵ is a positive number, and M is similar, the series is asymptotic. We are concerned with establishing such inequalities for the $S(\mathbf{k};\mathbf{q})$ corresponding to $S(\mathbf{k},\mathbf{R})=S(\mathbf{k}+A,\,\mathbf{k})$ since (1) they are the only nonpolynomial functions of $\mathbf{k}$ and $\mathbf{R}$ used in the first steps of our derivations, and (2) each later step in the perturbation theory involves beside polynomials such as $H_0(\mathfrak{k})$ and $V_0(\mathfrak{k})$, only a finite number of S's and derivatives thereof.

From $S(\mathbf{k},\mathbf{R})=S(\mathbf{k}+A,\,\mathbf{k})$ we obtain

$$S(\mathbf{k};\mathbf{q})=\sum_{R_i} S(\mathbf{k}+A_i,\,\mathbf{k})e^{i\mathbf{R}_i\cdot\mathbf{q}},\tag{A2}$$

while $S(\mathbf{k}+A_i,\,\mathbf{k})$ can be expressed as

$$S(\mathbf{k}+A_i,\,\mathbf{k})=\int \mathcal{S}(\mathbf{k};\mathbf{Q})\exp(i\mathbf{Q}\cdot A_i)d^3Q,\tag{A3}$$

where

$$\mathcal{S}(\mathbf{k};\mathbf{Q})=\int S(\mathbf{k}+\mathbf{l},\,\mathbf{k})\exp(-i\mathbf{Q}\cdot\mathbf{l})d^3l,\tag{A4}$$

the last two integrals being carried out over all space and all $\mathbf{k}$ space, respectively. These equations yield

$$S(\mathbf{k};\mathbf{q})=\sum_{Ri}\int d^3Q\;\mathcal{S}(\mathbf{k};\mathbf{Q})\exp[i(\mathbf{Q}\cdot A_i+\mathbf{q}\cdot\mathbf{R}_i)].\tag{A5}$$

Now for a linear vector potential we can write

$$A_i=G\cdot\mathbf{R}_i,\tag{A6}$$

where G will be called the gauge tensor. We further

---

[18] E. N. Adams, Phys. Rev. **89**, 633 (1953).

[19] J. E. Hebborn and E. H. Sondheimer, J. Phys. Chem. Solids **13**, 105 (1960).

[20] C. P. Enz, Helv. Phys. Acta **33**, 89 (1960).

define $\mathbf{T}(\mathbf{Q})=\mathbf{Q}\cdot\boldsymbol{G}$ and write

$$S(\mathbf{k};\mathbf{q})=\sum_{R}\int d^3Q\,\mathcal{S}(\mathbf{k};\mathbf{Q})\exp[i(\mathbf{q}+\mathbf{T}(\mathbf{Q}))\cdot\mathbf{R}_i] \tag{A7}$$

$$=\sum_{\mathbf{K}}\int \mathcal{S}(\mathbf{k};\mathbf{Q})\delta(\mathbf{q}+\mathbf{K}+\mathbf{T}(\mathbf{Q}))d^3Q. \tag{A8}$$

If $\mathbf{T}(\mathbf{Q})=\mathbf{q}+\mathbf{K}$ had a unique solution, this would take a simple form. In fact, however, the tensor $\boldsymbol{G}$ may be and almost always is a singular matrix, with one or two zero eigenvalues. In such cases the solutions $\mathbf{Q}$ of this equation fill a one- or two-dimensional manifold, respectively. Before discussing these complications however, we want to know what $\mathcal{S}(\mathbf{k};\mathbf{Q})$ is.

Starting from its definition we have

$$\mathcal{S}_{nn'}(\mathbf{k};\mathbf{Q})=\int d^3l\,S_{nn'}(\mathbf{k}+\mathbf{l},\mathbf{k})\exp(-i\mathbf{l}\cdot\mathbf{Q})$$

$$=\int d^3l\int d\tau\,u_{n,\mathbf{k}+\mathbf{l}}{}^*u_{n'\mathbf{k}}\exp(-i\mathbf{l}\cdot\mathbf{Q}), \tag{A9}$$

where the spatial integration extends over one unit cell. This can now be rewritten as

$$\mathcal{S}_{nn'}(\mathbf{k};\mathbf{Q})$$

$$=\int d^3l\int d\tau\,\Psi_{n,\mathbf{k}+\mathbf{l}}{}^*\Psi_{n'\mathbf{k}}\exp[i\mathbf{l}\cdot(\mathbf{x}-\mathbf{Q})] \tag{A10}$$

$$=\int d^3l\int d^3x\int d^3k'\,\Psi_{n\mathbf{k}'}{}^*\Psi_{n'\mathbf{k}}\exp[i\mathbf{l}\cdot(\mathbf{x}-\mathbf{Q})], \tag{A11}$$

where the space integral is now over all space, as is the integral over l, while $\mathbf{k}'$ is integrated over one BZ. This last integral is now performed, yielding

$$\mathcal{S}_{nn'}(\mathbf{k};\mathbf{Q})=\int d^3l\int d^3x\,a_n{}^*\Psi_{n'\mathbf{k}}\exp[i\mathbf{l}\cdot(\mathbf{x}-\mathbf{Q})], \tag{A12}$$

where $a_n(\mathbf{x})$ is the Wannier function for the nth band. Finally, we obtain

$$\mathcal{S}_{nn'}(\mathbf{k};\mathbf{Q})=a_n{}^*(\mathbf{Q})\Psi_{n'\mathbf{k}}(\mathbf{Q}). \tag{A13}$$

We now have the basis for asymptotic convergence. For, as shown in reference 3, $a_n(\mathbf{Q})$ for a band which has no degeneracies with other bands, falls off exponentially in Q. Since $\boldsymbol{G}$ is linear in $|\mathbf{B}|$ we can write its inverse as $|\mathbf{B}|^{-1}\boldsymbol{g}^{-1}$, for nonsingular gauges (that is, those with nonsingular gauge tensor). Then for this case, we have

$$\begin{aligned}S(\mathbf{k};\mathbf{q})&\\&=\textstyle\sum_K \mathcal{S}(\mathbf{k};|\mathbf{B}|^{-1}\boldsymbol{g}^{-1}\cdot(\mathbf{q}+\mathbf{K}))\\&=\textstyle\sum_K a_n{}^*(|\mathbf{B}|^{-1}\boldsymbol{g}^{-1}\cdot(\mathbf{q}+\mathbf{K}))\Psi_{n'\mathbf{k}}(|\mathbf{B}|^{-1}\boldsymbol{g}^{-1}\cdot(\mathbf{q}+\mathbf{K}))\\&\le\textstyle\sum_K|a_n(|\mathbf{B}|^{-1}\boldsymbol{g}^{-1}\cdot(\mathbf{q}+\mathbf{K}))|).\end{aligned} \tag{A14}$$

Each term in this sum decreases exponentially in $|\mathbf{B}|^{-1}$. At small $\mathbf{B}$, the sum will be dominated by its largest term, and it is clear that $|\mathbf{B}|$ plays the role of λ and that the conditions of I_1 and I_2 are satisfied. If we now consider singular gauges, there are no serious differences.

If however, we consider bands with degeneracies, there are serious difficulties because for them the Wannier functions will not fall off rapidly enough.[3] At this point the singular gauges play an important role, for with them we can demonstrate the asymptotic convergence.

To see how this comes about, we substitute (A10) in (A5), obtaining

$$S(\mathbf{k};\mathbf{q})=\sum_{R_i}\int d\tau\,\Psi_{n,\mathbf{k}+A_i}{}^*\Psi_{n'\mathbf{k}}\times\exp[i(\mathbf{q}\cdot\mathbf{R}_i+\boldsymbol{A}_i\cdot\mathbf{x})]. \tag{A15}$$

As in the passage from (A10) to (A11), we now introduce $\mathbf{k}'$ and extend the spatial integration over all space.

$$S(\mathbf{k};\mathbf{q})=[\delta(0)]^{l-3}\sum_{R_i}\int d^3x\int d^lk'\times\Psi_{n\mathbf{k}'}{}^*\Psi_{n'\mathbf{k}}\exp[i(\mathbf{q}\cdot\mathbf{R}_i+\boldsymbol{A}_i\cdot\mathbf{x})]. \tag{A16}$$

In this equation, however, $\mathbf{k}'$ need not extend over a whole BZ, but only over points $\mathbf{k}+\mathbf{m}$ where $\mathbf{m}$ is the general vector of the form $\mathbf{m}=\boldsymbol{G}\cdot\mathbf{r}$ for all $\mathbf{r}$ or equivalently we simply integrate $\mathbf{m}$ over a fundamental region in extended k space. The factor $[\delta(0)]^{l-3}$ is required to assure that

$$S(\mathbf{k}+0,\mathbf{k})=\int S(\mathbf{k};\mathbf{q})d^3q=1.$$

To clarify this point, a digression is necessary. In a singular gauge, the vector potential cannot take on any value but is restricted to a line or plane in extended $\mathbf{k}$ space. In the case of a plane, when the points outside a BZ are mapped into it, they may lie on a family of discrete planes, if the direction perpendicular to the plane is parallel to any lattice vector (not reciprocal lattice vector). Otherwise, the points will be dense in the BZ. In the case of a line, three alternatives are possible; the points may lie on a discrete family of lines (this requires that the line be parallel to a reciprocal lattice vector), may be dense on a discrete family of planes (this requires that the line lie in a plane determined by two reciprocal lattice vectors), or may be dense in the BZ.

In the case of planes, the first possibility corresponds to a finite fundamental region, which is adopted because all points on the plane outside it are equivalent to a point inside. This means that the fundamental region is a section of a special BZ chosen to have faces parallel to the plane. The second possibility corresponds to the fundamental region being the whole plane. Likewise in the case of lines, the first possibility corresponds to a finite fundamental region, the second and third possibilities to the fundamental region being the whole line.

If the line or plane of $\mathbf{m}$ maps into discrete lines or planes in the BZ, we can easily see that there is a possibility of improving convergence. For if for given $\mathbf{k}$ we can choose the gauge so that the line or plane passing through $\mathbf{k}$ does not pass through a degeneracy, then the arguments used in (F) to prove the exponential fall-off of Wannier functions will work for these functions of the form

$$a_n(\mathbf{r};\mathbf{k},G)=\int d^l m\,\Psi_{n,\mathbf{k}+\mathbf{m}}(\mathbf{r}). \tag{A17}$$

These functions behave like Wannier functions in the direction corresponding to the manifold of $\mathbf{m}$ (which depends on G) and like Bloch functions perpendicular to it. This failure to fall off in the perpendicular directions does not hurt us, as the integral over that component of $\mathbf{x}$ simply kills the offensive $\delta(0)$'s.

The trouble with this procedure is that it is not generally applicable. To be sure, we can for any direction of magnetic field require the vector potential to lie in any desired plane, but it is not always possible to choose a plane which will avoid all degeneracies, still less to find a family of planes nearly all of which will avoid degeneracies. On the other hand, if we could confine $\mathbf{a}$ to a line, nearly all of the resulting family of lines would avoid degeneracies unless there were a plane of degeneracies, for which special measures are available. However, one cannot freely choose the line on which $\mathbf{a}$ must lie, for it is required to be perpendicular to the field direction. If as is true except for a set of field directions of measure zero, the ratios of the field components along crystal axes are irrational, there is no reciprocal lattice vector perpendicular to the field. Thus, if we wished to utilize the advantages of a doubly singular gauge, we would have points distributed densely in a BZ and thus passing through or arbitrarily close to degeneracies.

Fortunately, however, the doubly singular gauges have an additional advantage which offsets these difficulties. Namely, it permits us to use a one-dimensional integration, albeit over an infinite domain. Thus we have in effect a one-dimensional band structure over this domain. In reference 3 it is shown that in a one-dimensional band structure the Bloch functions are analytic functions of k even in the neighborhood of degeneracies on the real axis, though not in the neighborhood of those in the complex k space. The bands can be so labeled that this is true in our problem also. Thus the integration over $\mathbf{m}$ passes through no singularities. On the other hand it passes arbitrarily close to branch points, since the branch surfaces come right up to the real axis near degeneracies in real k space.[3] Thus, we cannot expect an exponential fall-off but we expect one which is faster than any power of x.

At this point we return to (A16) and rewrite it as

$$\begin{aligned}S_{nn'}(\mathbf{k};\mathbf{q})&=[\delta(0)]^{-2}\sum_{R_i}\int d^3x\,a_n(\mathbf{x};G)\Psi_{n'\mathbf{k}}(\mathbf{x})\exp[i(\mathbf{q}\cdot\mathbf{R}_i+\mathbf{T}(\mathbf{x})\cdot\mathbf{R}_i)]\\&=[\delta(0)]^{-2}\sum_{K}\int d^3x\,a_n(\mathbf{x};G)\Psi_{n'\mathbf{k}}(\mathbf{x})\delta(\mathbf{q}+\mathbf{K}+\mathbf{T}(\mathbf{x})).\end{aligned} \tag{A18}$$

It will now be easier if we choose a set of axes and a gauge. We pick $\mathbf{B}$ in the z direction and $\mathbf{a}$ in the y direction depending only on x. The crystal axes now have no particular relation to x, y, z. We have also specialized to the case of an irrational field orientation, so that there is no $\mathbf{k}$ dependence of $\mathbf{a}$. Then

$$S_{nn'}(\mathbf{k};\mathbf{q})=[\delta(0)]^{-2}\sum_{K}\int dxdz\,a_n\!\left(x,\frac{q_x+K_x}{B},z;G\right)\Psi_{n',\mathbf{k}}\!\left(x,\frac{q_x+K_x}{B},z\right)\delta(q_y+K_y)\delta(q_z+K_z). \tag{A19}$$

Consider now one term of this sum. By the Schwarz inequality we have

$$\begin{aligned}|S_{nn'}(\mathbf{k};\mathbf{q};K)|^2&\equiv\left|[\delta(0)]^{-2}\int dxdz\,a_n\!\left(x,\frac{q_x+K_x}{B},z;G\right)\Psi_{n'\mathbf{k}}\!\left(x,\frac{q_x+K_x}{B},z\right)\right|^2\\&\leq[\delta(0)]^{-2}\int dxdz\left|a_n\!\left(x,\frac{q_x+K_x}{B},z;G\right)\right|^2[\delta(0)]^{-2}\int\left|u_{n'\mathbf{k}}\!\left(x,\frac{q_x+K_x}{B},z\right)\right|^2dxdz.\end{aligned} \tag{A20}$$

For irrational orientations the second integral is independent of $(q_x+K_x)/B$ and simply cancels $[\delta(0)]^{-2}$. We now require the explicit form of a more general pseudo-Wannier function

$$a_n(x,y,z;G;Y)=\int_{-\infty}^{\infty}dk'\,\Psi_{nk'}e^{-ik'Y}, \tag{A21}$$

where Y is a continuous variable. These functions are orthogonal for different values of Y. Furthermore, consider the integral

$$P(y;Y)=[\delta(0)]^{-2}\int|a_n(x,y,z;G;Y)|^2dxdz$$
$$=[\delta(0)]^{-2}\int dkdk'e^{i(k'-k)(y-Y)}\left(\int u_{nk}{}^*(x,y,z)u_{nk'}(x,y,z)dxdz\right). \quad \text{(A22)}$$

As in (A20), the integral over x and z is independent of y and cancels $[\delta(0)]^{-2}$ so that $P(y;Y)$ depends only on $(y-Y)$, and $P(y)$ is simply the Fourier transform of

$$S_{nn}(\kappa_y)=\int S_{nn}(k_y+\kappa_y,\,k_y)dk_y. \quad \text{(A23)}$$

Since our one-dimensional bands are analytic for real k, this is an analytic function of κ_y for all real κ_y, which, however, cannot be continued into the complex plane, because branch points are arbitrarily close to the real axis. This $P(y)$ falls off more rapidly than y^{-n} for any n at large y, but more slowly than $e^{-|\lambda y|}$ for any λ. (Of course, if the band has no degeneracies, the branch surfaces are not close to the real axis and the fall-off is exponential.) This is sufficient to show that $S(\mathbf{k};\mathbf{q};\mathbf{K})$ and $S(\mathbf{k};\mathbf{q})$ vanish faster than $|q/B|^{-n}$ and that the integrals I_1 and I_2 with S substituted for n satisfy the conditions for an asymptotic expansion. Thus the multiplication rules are asymptotically convergent at small B for all products used in this paper and for all k and R, provided we have chosen a doubly singular gauge. This is true even for points of degeneracy.

Although our proof has used a special gauge, it is clear that after obtaining the expansion in that gauge one can make the transformation

$$U(\mathbf{R})=e^{i\chi(\mathbf{R})}, \quad \text{(A24)}$$

which will change the vector potential by $\nabla\chi$ for arbitrary χ. Thus the product expansion will be valid for any gauge; presumably a guage-invariant proof exists. However, we have not seen it and do not consider it necessary to look further.

Our proof for degenerate bands has assumed an irrational field orientation, since any other has zero probability. The proof is no more difficult for rational orientations, but we leave it to the interested reader.

Finally, we should note that our differentiation of $S(k,R)$ with respect to R has been somewhat cavalier for we have differentiated $S(\mathfrak{k})$ with respect to $\mathfrak{k}$ and then $\mathfrak{k}$ with respect to R, thus treating $S(\mathfrak{k})$ as a continuous function of $\mathfrak{k}$. In fact, however, in reference 3 it was shown that the fact that S is defined only on lattice points is vital in giving it an infinite radius of convergence. In fact, we should use instead of $i^l\partial^lS/\partial R^l$,

$$\int_{\mathrm{BZ}}\mathbf{q}^lS(k;q)e^{-i\mathbf{q}\cdot\mathbf{R}}d^3q=\sum_{R_i}\int d^3Q\int_{\mathrm{BZ}}d^3q\;q^l\mathcal{S}(k;Q)\exp[i(\mathbf{Q}\cdot\mathbf{A}_i+\mathbf{q}\cdot(\mathbf{R}_i-\mathbf{R}))] \quad \text{(A25)}$$
$$=\sum_K\int d^3Q\int d^3q\;\mathcal{S}(k;Q)\delta(T(Q)+K+q)q^le^{-i\mathbf{q}\cdot\mathbf{R}}. \quad \text{(A26)}$$

The true derivative would be given by

$$\frac{i^l\partial^lS}{\partial k^l}=\int d^3Q\,[\mathbf{T}(\mathbf{Q})]^l\exp(i\mathbf{T}\cdot\mathbf{R})\mathcal{S}(\mathbf{k};\mathbf{Q})=\sum_K\int d^3q\;\mathcal{S}(\mathbf{k};\mathbf{Q})\delta(\mathbf{T}(\mathbf{Q})+\mathbf{K}+\mathbf{q})e^{-i\mathbf{q}\cdot\mathbf{R}}(\mathbf{q}+\mathbf{K})^l. \quad \text{(A27)}$$

These expressions differ by the presence of K in the term with $K\neq0$. Since all such terms get small very rapidly, however, the error is less than any power of B, and thus quite acceptable to us.

APPENDIX B

In this Appendix, we are concerned with the matrix elements between states represented by single orbits in the (p,q) plane. It is convenient to work in the PCR which also yields an intuitive appreciation of what is going on. We define the object

$$P_{nn'}(q,q')=\psi_n(q)\psi_{n'}{}^*(q'), \quad \text{(B1)}$$

and transform it into a PCR operator $P_{nn'}(p,q)$ as in Sec. III of (F). The matrix element $O_{nn'}$ of an operator O is simply $\mathrm{Tr}\{P_{nn'}O\}$ and it is easily shown that

$$O_{nn'}=\int\int dpdq\;P_{nn'}(p,q)O(p,q), \quad \text{(B2)}$$

a result obtained by Moyal[21] for the continuum case. Since most O of interest are slowly varying functions of p and q, one is encouraged to attempt the integration by the method of steepest descents, since $P_{nn'}$ may be

---

[21] J. E. Moyal, Proc. Cambridge Phil. Soc. 45, 99 (1949).

expected to vary rapidly with p and q at small B. Thus in the WKB approximation a wave function for an orbit n is written

$$\psi_n(q)=\frac{1}{m_n^{*\frac{1}{2}}}\sum_\alpha \frac{e^{iS_{n\alpha}(q)/B}}{[v_{n\alpha}(q)]^{\frac{1}{2}}}, \tag{B3}$$

where $S_{n\alpha}=\int p_{n\alpha}(q)dq$, $p_{n\alpha}$ is a branch of the function $p(q)$ defined by $H(p,q)=E_n$, and m_n^* is given by

$$m_n^*=\sum_\alpha \int \frac{dq}{|v_{n\alpha}|}, \tag{B4}$$

where the sum is taken over all real branches and the integral is limited to the values of q for which p_α is real. This can be converted to

$$m_n^*=\oint \frac{dq}{v_n(q)}, \tag{B5}$$

v is $\partial H/\partial p$. Equation (B3) as it stands, is valid only for values of q such that all $p_\alpha(q)$ are real. When some are complex, more care is required in order to assure that the wave function can be said to belong to only one orbit, but Eq. (B3) will give valid results for our purposes. We then obtain for $P_{nn'}(p,q)$

$$P_{nn'}(p,q)=\sum_{\alpha\beta}\int dr \times\frac{\exp[(i/B)(S_{n\alpha}(q+\frac{1}{2}r)-S_{n'\beta}(q-\frac{1}{2}r)-pr)]}{(v_{n\alpha}v_{n'\beta})^{\frac{1}{2}}}. \tag{B6}$$

We will restrict ourselves to pairs of orbits such that a straight line can be drawn between them, and we choose our axes so that p is perpendicular to such a line. The integral over r can be performed by steepest descents, yielding

$$P_{nn'}(p,q)=\sqrt{\pi}\sum_{\alpha\beta}\left.\frac{\exp[(i/B)(S_{n\alpha}(q+\frac{1}{2}r)-S_{n\beta}(q-\frac{1}{2}r)-pr)]}{(v_{n\alpha}v_{n'\beta})^{\frac{1}{2}}(|p_{n\alpha}'-p_{n'\beta}'|)^{\frac{1}{2}}}\right|_{r=r_{\alpha\beta}(p,q)}, \tag{B7}$$

where $r_{\alpha\beta}$ satisfies the equation.

$$p_{n\alpha}(q+\tfrac{1}{2}r)+p_{n'\beta}(q-\tfrac{1}{2}r)-2p=0. \tag{B8}$$

In an integral involving $P_{nn'}$ the integration over p can also be performed by steepest descents and the major contribution will come from the region where

$$r_{\alpha\beta}=0,$$

and is given by

$$\bar{P}_{nn'}(q)=\frac{\pi}{2}\sum_{\alpha\beta}\frac{\exp[(i/B)(S_{n\alpha}(q)-S_{n\beta}(q))]}{(v_{n\alpha}v_{n'\beta})^{\frac{1}{2}}}. \tag{B9}$$

In this form it is clear that only the two branches of each orbit which are closest to each other will be important; for the exponential factor in the contributions from other pairs of branches will be much more rapidly varying and will give negligible integrals over q at small B.

Finally the main contribution of (B9) on integrating q can be made to come, on deforming the contour, from the point q_0 where

$$P_{n\alpha}(q_0)=P_{n'\beta}(q_0). \tag{B10}$$

We distinguish two cases.

(1) Orbits n and n' are in the same band. Then (B10) can be satisfied only if they have the same energy. In this case, the important factor is

$$\exp[(i/B)(S_n(q_0)-S_{n'}(q_0))]. \tag{B11}$$

If the orbits have different energies but are kept within a few level splittings of each other as $B\to 0$, the error in using this factor is small. If their energy difference is large compared to level splittings, more detailed treatment is required, but the matrix elements will then be small compared to the energy difference, so that such matrix elements will act only as small perturbations, shifting levels, but not capable of producing banding.

If the orbits have essentially the same energy the factor $\exp[(i/B)(S_n(q_0)-S_{n'}(q_0))]$ has a simple geometrical interpretation, for

$$S_n(q_0)=\int_a^{q_0} p\,dq,$$

where a is some point on orbit n. $S_n(q_0)-S_{n'}(q_0)$ is then $\int_a^b p\,dq$, where b is a point on n', and the path is a surface of constant energy. If we choose the path in imaginary q space, the integral is the area of the projection of the path on the real p—imaginary q plane. The imaginary part which determines the exponential fall-off of $O_{nn'}$ is invariant to our choice of end points on the orbits, but can be most easily visualized if we choose the points with the shortest distance between them.

It should be noted the main contribution to the Hamiltonian matrix elements will be $H(p_0,q_0)=E\times$the integral of (B9) which is just the overlap integral between the two orbits. Additional terms will be of order B smaller. Thus the principal effect of the matrix element is removed by orthogonalizing the orbitals without changing their energy.

(2) If the orbits are in different bands it should usually be possible to find solutions of (B10) even if their energies are different. In this case then, the

saddle point method is immediately applicable, and if the energies are nearly equal gives the same result (B11), with the same interpretation. In the special case considered in Sec. III, it gives the same exponential factor as was found there. For the M term in (III.3) this analysis is sufficient. The X term is somewhat more complicated, however, because X has the energy difference between the bands in its denominator, and the integral thus has a pole where the band energies are equal namely at q_0. This complicates the evaluation procedure, but only modifies the factor by which (B11) is multiplied. The actual evaluation for our model (III.4) is identical to that of Kane for Zener breakdown. If the orbit energies are unequal, the orbits may even overlap so that the integrals become larger, but this has no serious effects, because the energy difference required is so much greater than the matrix element that the orbits cannot be appreciably mixed.