Magnetic Breakdown: Effective Hamiltonian and de Haas–van Alphen Effect*

W. G. CHAMBERS
Rice University, Houston, Texas
(Received 10 March 1966)

The influence of magnetic breakdown on the usual effective-Hamiltonian theory for Bloch electrons in a magnetic field is discussed using a simple two-dimensional rectangular model. The theory is based upon an expansion of the wave function in Wannier functions, but these have to be replaced by generalized Wannier functions (suggested by Blount) to handle breakdown. It is shown how to construct a linear network essentially equivalent to the network derived previously by the author for a nearly-free-electron model. The de Haas–van Alphen effect is discussed by expressing the energy density of states in terms of a time-independent Green's function on the network. It is shown how to construct a wave function lying on a two-dimensional network as used by Pippard.

1. INTRODUCTION

THE theory of the motion of conduction electrons in metals in high magnetic fields is important in connection with experimental studies of the Fermi surface. The theory is usually based on the concept of an effective Hamiltonian and much work has been done on the derivation of such an effective Hamiltonian.[1–4] The theory is complicated by the fact that the magnetic field causes some mixing-in of wave functions of other bands into the wave function in the band of interest, somewhat as a field acting on an atom causes some admixture of higher states into the ground state, and this alters the energy or the effective Hamiltonian. In very rough terms, a perturbation expansion is derived for the effective Hamiltonian in powers of the magnetic field. Under certain circumstances, when the band gaps are very small, magnetic breakdown[5] can take place, wherein the electron tunnels from the Fermi surface for one band on to the Fermi surface for another band. The usual effective Hamiltonian theory fails in this case, for it is no longer a problem of a virtual admixture of other bands into the band of interest but a real transition between bands.

Such transitions were discussed qualitatively by Blount[2] and later by the author.[6] Pippard[7,8] set up some intuitive network models to study the effects of breakdown, and such models have been used in considering the effect on magnetoresistance.[9] The author[10] (in a paper henceforth referred to as LN) derived an equivalent network model with the assumption that the periodic potential was weak enough to be treated by the Born approximation. Such models are a departure from the usual theory of the effective Hamiltonian, and it is the aim of this paper to connect the two approaches.

This paper is, in effect, an extension of LN, and the same model will be used again, a spinless electron moving in a two-dimensional rectangular lattice at right angles to the applied field. To discuss a general problem in a special model may be a serious limitation, but the following advantages can be claimed: First, the reader has a definite model to visualize. Second, the range of possible situations that can arise with magnetic breakdown is so large that a general discussion would be excessively complicated. And third, the discussion relates to the work in LN so that an immediate comparison is possible. In fact, one of the main aims of this paper is to show that the effective Hamiltonian theory can lead to a situation very similar to that discussed in LN, and thus provides a justification for using the theory in LN outside the limitations of having a very weak lattice potential.

The magnetic field is in some sense a slowly varying perturbation on the electronic motion, suggesting that a suitable set of basis functions is provided by the Wannier functions.[11] This approach also has the advantage that the wave functions are set up in real space (rather than momentum space), giving a direct comparison with LN. Most authors have tried to avoid the use of Wannier functions, probably because their properties are not readily studied. In fact it is known that in the situation where the band gaps are very narrow, which is also the situation for obtaining breakdown, the Wannier functions are not well localized. Usually these functions are constructed by integrating over the Bloch waves in a single band, but Blount[11] has suggested that, by constructing a more general type of Wannier function from the Bloch waves of several bands, it is possible to obtain well localized functions.

* Research supported by the National Science Foundation and the National Aeronautics and Space Administration.
[1] L. M. Roth, J. Phys. Chem. Solids **23**, 433 (1962).
[2] E. I. Blount, Phys. Rev. **126**, 1636 (1962).
[3] W. Kohn, Phys. Rev. **115**, 1460 (1959). See this paper for other references.
[4] G. H. Wannier and D. R. Fredkin, Phys. Rev. **125**, 1910 (1962).
[5] M. H. Cohen and L. M. Falicov, Phys. Rev. Letters **7**, 231 (1961).
[6] W. G. Chambers, Proc. Phys. Soc. (London) **84**, 181 (1964).
[7] A. B. Pippard, Proc. Roy. Soc. (London) **A270**, 1 (1962).
[8] A. B. Pippard, Phil. Trans. Roy. Soc. (London) **256**, 317 (1964).
[9] L. M. Falicov and P. R. Sievert, Phys. Rev. **138**, A88 (1965).
[10] W. G. Chambers, Phys. Rev. **140**, A135 (1965).
[11] E. I. Blount, in *Solid-State Physics*, edited by F. Seitz and D. Turnbull (Academic Press Inc., New York, 1961), Vol. 13. See in particular pp. 309–335.

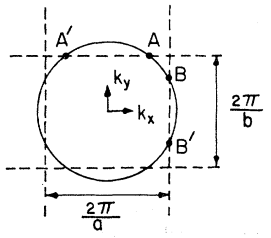

FIG. 1. The Fermi "surface" in relation to the Brillouin zone boundaries represented by dashed lines.

The subject matter of this paper covers three topics. The first is a discussion of the effective Hamiltonian theory in terms of Wannier functions. The second is the theory of the de Haas–van Alphen effect with magnetic breakdown. The effect is a very useful tool in the study of Fermi surfaces and may give additional experimental information on magnetic breakdown. Indeed, magnetic breakdown was first postulated to explain some de Haas–van Alphen results in magnesium.[9] The third topic is to show how a wave function can be localized on a Pippard network.[7,8] This should provide some further justification for this approach. The second and third topics could have been treated by the theory of LN and are therefore to some extent independent of the theory of the effective Hamiltonian.

The paper is divided into six sections, besides the introductory and concluding remarks. Section 2 describes the model. Section 3 discusses the derivation of the effective-Hamiltonian theory. Section 4 is concerned with the failure of the theory in situations where magnetic breakdown takes place and the necessary modifications based on generalized Wannier functions suggested by Blount. Section 5 discusses the treatment needed at those places on the Fermi surface where breakdown takes place, linking up with the network model in LN. All these sections are concerned with the first topic, the effective Hamiltonian. The second main topic, the de Haas–van Alphen effect, is treated in Sec. 6, and the third topic, the construction of a wave function on a network, is treated in Sec. 7.

The discussion contains assumptions that are made without proof. These assumptions perhaps seem more questionable in the Wannier representation, though some of them have been made by other authors. The Landau gauge is employed, as in LN, because it makes the system one dimensional, but it also introduces artificial turning points. These give mathematical problems which are usually ignored. There is not much discussion about the higher order terms in the expansion of the effective Hamiltonian as a power series in the field because they seem to be irrelevant, at least when there is no spin-orbit coupling. Magnetic breakdown is better treated by other means.

2. THE MODEL

The model consists of a spinless electron moving in a two-dimensional rectangular lattice in the Oxy plane with periods $\mathbf{a}=(a,0,0)$ and $\mathbf{b}=(0,b,0)$. A uniform magnetic field $\mathfrak{H}$ is applied parallel to Oz. We have in effect left out the z axis because most of the interesting problems concern the motion at right angles to the field. We shall assume that the Fermi "surface" is related to the Brillouin zone boundaries as is shown in Fig. 1. (We shall use the word "surface" by force of habit, even in a two-dimensional problem.) The energy is such that the surface has broken into the second band. We shall assume that the surface is almost circular and that the gaps which occur at points like A and B are very small. Figure 2 shows the Pippard network construction in real space. It is simply a repeated version of Fig. 1, rotated by 90° and scaled by a factor $1/h$, where h is given by

$$\mathbf{h}=e\mathfrak{H}/hc. \tag{1}$$

Here e is the electronic charge, h is Planck's constant divided by 2π, and c is the velocity of light. For later convenience the abscissa (horizontal axis) in Fig. 2 is used as the y axis, and the ordinate (vertical axis) as the x axis. We have a four-sided hole surface S in the first band and two lenses, L_1 and L_2, in the second.

The same sort of model was discussed in LN with the assumption that the periodic potential was very weak so that the Born approximation could be used. This time no such assumption is made and the problem will be treated by an effective Hamiltonian. We may imagine that the smallness of the band gaps is due to some peculiarity in the periodic potential, such as having two almost identical atoms in each unit cell.

3. THE EFFECTIVE HAMILTONIAN

The problem of deriving an effective Hamiltonian has been considered by many authors using many different methods, but the methods seem to be more or less equivalent to the following derivation based on Wannier functions.[12,13] The units are chosen so that $\hbar=2m=1$, where m is the electronic mass. The Schrödinger equation is then

$$(H-E)\psi\equiv[(\mathbf{p}-\mathbf{A})^2+V(\mathbf{x})-E]\psi(\mathbf{x})=0, \tag{2}$$

where curl $\mathbf{A}=\mathbf{h}=(0,0,h)$, V is the periodic potential satisfying

$$V(\mathbf{x}+\mathbf{R})=V(\mathbf{x}),$$

with $\mathbf{R}$ any lattice translation, and $\mathbf{p}$ is represented by

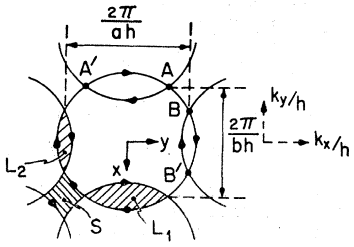

FIG. 2. Pippard network construction in real space. Note unusual choice of axes. This figure can also represent the Fermi surface in a repeated scheme, with axes as shown on right. Shaded areas represent hole orbit S and two lenses L_1, L_2.

[12] J. M. Luttinger, Phys. Rev. **84**, 814 (1951).
[13] E. Brown, Phys. Rev. **133**, A1038 (1964).

$-i\nabla_x$. In the symmetric gauge we set $\mathbf{A}=\frac{1}{2}\mathbf{h}\times\mathbf{x}$. We introduce the operator[14]

$$\mathbf{\Pi}=\mathbf{p}-\mathbf{A}+\mathbf{h}\times\mathbf{x}, \tag{3}$$

which commutes with $(\mathbf{p}-\mathbf{A})$, while the magnetic translation operators[13]

$$\tau(\mathbf{R})=\exp i\mathbf{\Pi}\cdot\mathbf{R} \tag{4}$$

commute with the Hamiltonian H and obey the relations

$$\tau(\mathbf{R})\tau(\mathbf{S})=[\exp(i\tfrac{1}{2}\mathbf{h}\cdot\mathbf{R}\times\mathbf{S})]\tau(\mathbf{R}+\mathbf{S}) \tag{5}$$

where $\mathbf{R}$ and $\mathbf{S}$ are lattice translations. This follows from the operator relation[1]

$$\exp(A+B)=(\exp B)(\exp A)\exp(\tfrac{1}{2}[A,B]) \tag{6}$$

(where $[A,B]$ is the commutator of A and B) which holds provided that $[A,B]$ commutes with both A and B. Let $a_n(\mathbf{x})$ be the Wannier function centered on the origin for the nth band of the zero-field Hamiltonian H_0. To solve (2) we expand ψ in a set of modified Wannier functions[12]

$$\begin{aligned}a_n(\mathbf{x},\mathbf{R})&\equiv\tau(-\mathbf{R})a_n(\mathbf{x})\\&=[\exp(-i\tfrac{1}{2}\mathbf{h}\times\mathbf{x}\cdot\mathbf{R})]a_n(\mathbf{x}-\mathbf{R})\end{aligned} \tag{7}$$

for the symmetric gauge. It is assumed without proof that these functions form a complete set. The assumption is discussed towards the end of Sec. 4. We shall use the Dirac notation $|m\mathbf{R}\rangle$ for $a_m(\mathbf{x},\mathbf{R})$, and then by (5) it follows that

$$\langle m\mathbf{R}|(H-E)|n\mathbf{S}\rangle=[\exp(-i\tfrac{1}{2}\mathbf{h}\cdot\mathbf{R}\times\mathbf{S})]\times[W_{mn}(\mathbf{R}-\mathbf{S})-EM_{mn}(\mathbf{R}-\mathbf{S})],$$

where

$$\begin{aligned}W_{mn}(\mathbf{R})&=\langle m\mathbf{R}|H|n0\rangle\\&=\int a_m^*(\mathbf{x}-\mathbf{R})\times[\exp(i\tfrac{1}{2}\mathbf{h}\times\mathbf{x}\cdot\mathbf{R})]Ha_n(\mathbf{x})d\mathbf{x}\end{aligned} \tag{8a}$$

and

$$\begin{aligned}M_{mn}(\mathbf{R})&=\langle m\mathbf{R}|n0\rangle\\&=\int a_m^*(\mathbf{x}-\mathbf{R})\times[\exp(i\tfrac{1}{2}\mathbf{h}\times\mathbf{x}\cdot\mathbf{R})]a_n(\mathbf{x})d\mathbf{x}.\end{aligned} \tag{8b}$$

If we substitute

$$\psi(\mathbf{x})=\textstyle\sum_{\mathbf{R},n}c_n(\mathbf{R})a_n(\mathbf{x},\mathbf{R}) \tag{9}$$

into (2) we obtain

$$\textstyle\sum_{\mathbf{S}}[\exp(-i\frac{1}{2}\mathbf{h}\cdot\mathbf{R}\times\mathbf{S})]\times[W(\mathbf{R}-\mathbf{S})-EM(\mathbf{R}-\mathbf{S})]c(\mathbf{S})=0, \tag{10}$$

where the band indices have been left out so that W and M are matrices and c is a column vector. Then if we set $\mathbf{T}=\mathbf{R}-\mathbf{S}$ and sum over $\mathbf{T}$ instead we obtain the difference equation

$$\textstyle\sum_{\mathbf{T}}[\exp(i\frac{1}{2}\mathbf{h}\cdot\mathbf{R}\times\mathbf{T})]\times[W(\mathbf{T})-EM(\mathbf{T})]c(\mathbf{R}-\mathbf{T})=0. \tag{11}$$

The coefficients $c(R)$ are of course only defined at the lattice sites, but we shall assume that it is possible to define a smooth function $c(\mathbf{r})$ of a continuous variable $\mathbf{r}$, equal to $c(\mathbf{R})$ when $\mathbf{r}=\mathbf{R}$.[11] Then (11) may be written by Taylor's theorem and Eq. (6) as

$$\textstyle\sum_{\mathbf{T}}[\exp-i(\mathbf{p}-\frac{1}{2}\mathbf{h}\times\mathbf{r})\cdot\mathbf{T}]\times[W(\mathbf{T})-EM(\mathbf{T})]c(\mathbf{r})=0, \tag{12}$$

where $\mathbf{p}=-i\nabla_r$. Then if we define the Fourier transforms

$$H(\mathbf{k})=\textstyle\sum_{\mathbf{S}}[\exp(-i\mathbf{k}\cdot\mathbf{S})]W(\mathbf{S}) \tag{13a}$$

and

$$N(\mathbf{k})=\textstyle\sum_{\mathbf{S}}[\exp(-i\mathbf{k}\cdot\mathbf{S})]M(\mathbf{S}), \tag{13b}$$

which are, of course, periodic in $\mathbf{k}$, we may write (12) as

$$[H(\mathbf{P})-EN(\mathbf{P})]c(\mathbf{r})=0, \tag{14}$$

where $\mathbf{P}=\mathbf{p}-\frac{1}{2}\mathbf{h}\times\mathbf{r}$. The matrix operator $H(\mathbf{P})$, derived from the matrix function $H(\mathbf{k})$ by replacing $\mathbf{k}$ by $\mathbf{P}$ in (13a), is called the undiagonalized effective Hamiltonian. At times the term "effective Hamiltonian" will also be used for the matrix function $H(\mathbf{k})$ and the context will indicate whether the operator or the function is implied. Equation (14) is really no more than the difference Eq. (11). We note the following points:

1) A similar equation was derived by Roth[1] in much the same manner, though the basis functions used are Fourier transforms of the Wannier functions

$$\varphi(\mathbf{k},\mathbf{x})=\textstyle\sum_{\mathbf{R}}(\exp i\mathbf{k}\cdot\mathbf{R})a(\mathbf{x},\mathbf{R})=b(\mathbf{k}-\frac{1}{2}\mathbf{h}\times\mathbf{x},\mathbf{x}), \tag{15}$$

where $b(\mathbf{k},\mathbf{x})$ is a Bloch wave of wave number $\mathbf{k}$, a solution of the zero-field Hamiltonian. Such functions have also been used by other authors.[2-4] In this case the operator $\mathbf{P}$ in (14) is replaced by

$$\boldsymbol{\kappa}=\mathbf{k}-\tfrac{1}{2}\mathbf{h}\times i\nabla_k, \tag{16}$$

and $c(\mathbf{r})$ becomes a momentum-space envelope function $c(\mathbf{k})$.

2) The function $c(\mathbf{r})$ is really only defined at the lattice sites $\mathbf{r}=\mathbf{R}$ and so of course it may be multiplied by any function periodic in the lattice without changing the problem. This fact will be used later.

3) When $h=0$ the functions $a(\mathbf{x},\mathbf{R})$ become the ordinary Wannier functions $a(\mathbf{x}-\mathbf{R})$ and in this case we find that

$$H_{mn}(\mathbf{k})=E_{mn}(\mathbf{k})\equiv E_m(\mathbf{k})\delta_{mn}, \tag{17a}$$

$$N_{mn}(\mathbf{k})=\delta_{mn}, \tag{17b}$$

where $E_m(\mathbf{k})$ is the zero-field band structure.

[14] M. H. Johnson and B. A. Lippman, Phys. Rev. **76**, 828 (1949).

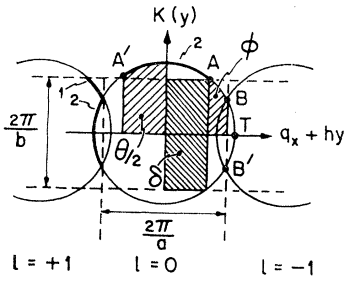

FIG. 3. Convention for $K(y)$ as a function of q_x+hy (Eq. 23). Numbers 1 and 2 label the bands. Also shown are the areas for the reflection phase δ and propagation phases θ and φ. The quantity l labels the periods.

4) We may perform gauge transformations on (14). If we set $c'(\mathbf{r})=e^{-i\chi(\mathbf{r})}c(\mathbf{r})$, where $\chi(\mathbf{r})$ is an arbitrary real function, then $c'(\mathbf{r})$ satisfies an equation like (14), but with $\mathbf{P}$ replaced by $\mathbf{P}+\nabla\chi$. We use $\chi=\frac{1}{2}hxy$ to go over to the Landau gauge when $\mathbf{P}$ becomes $(-i\partial/\partial x+hy,-i\partial/\partial y)$ in two dimensions. Then we may set $c'(\mathbf{r})=e^{iq_xx}c(y)$, and (14) becomes

$$[H(q_x+hy,-id/dy) -EN(q_x+hy,-id/dy)]c(y)=0. \quad (18)$$

This is a matrix difference equation with one independent variable y. It may be written out explicitly using (13), (6), and Taylor's theorem. It replaces the system of coupled ordinary differential equations in LN. Subsequently we shall use another gauge transform with $\chi=k_0y$ to shift the origin in $\mathbf{k}$ space.

The rest of the development is the same as that given by Roth,[1] and we only sketch it here. The work of Roth can be used without modification because the equations are gauge-invariant and may be applied to (18) as well as (14) and can be used for the operator $\mathbf{P}$ as well as the operator $\boldsymbol{\kappa}$ in (16). The second stage after the derivation of (14) or (18) is to expand $H(\mathbf{k})$ and $N(\mathbf{k})$ as a power series in h. We may use (8) and (13), but the expansion is equivalent to that given by Roth in terms of matrix elements of Bloch functions. The zero-order terms are diagonal, given by (17). The off-diagonal terms are of order h, and the third stage is to transform the matrix Hamiltonian so that these are removed. We set

$$H'(\mathbf{P})=S^\dagger(\mathbf{P})H(\mathbf{P})S(\mathbf{P}) \quad (19a)$$

and

$$N'(\mathbf{P})=S^\dagger(\mathbf{P})N(\mathbf{P})S(\mathbf{P}) \quad (19b)$$

and choose S so that $H'(\mathbf{P})$ is diagonal and $N'(\mathbf{P})$ is the unit matrix. All the quantities are periodic and so they are difference operators. The product of two difference operators $A(\mathbf{P})$ and $B(\mathbf{P})$, derived from periodic functions $A(\mathbf{k})$ and $B(\mathbf{k})$, is a difference operator $C(\mathbf{P})$, derived from a periodic function $C(\mathbf{k})$, given by

$$C(\mathbf{k})=(\exp i\tfrac{1}{2}h\cdot\nabla_k\times\nabla_l)A(\mathbf{k})B(\mathbf{l})\,|_{l=k} =A(\mathbf{k})B(\mathbf{k})+i\tfrac{1}{2}\mathbf{h}\cdot\nabla_kA\times\nabla_kB+\cdots \quad (20)$$

to first order in h. The second term on the right in (20) arises from the fact that the components of $\mathbf{P}$ do not commute. Such terms make the diagonalization in (19) more than an algebraic diagonalization and it has to be carried out step-by-step in powers of h.

Once we have carried out this process on (18) to a sufficiently high order, we may obtain Wentzel-Kramers-Brillouin (WKB) solutions for a scalar equation

$$[E_n'(q_x+hy,-id/dy)-E]u_n(y)=0, \quad (21)$$

where $E_n'(\mathbf{k})$ is the diagonal matrix element of $H'(\mathbf{k})$ for the nth band. Such a solution (to first order in h) is[15]

$$u_\pm(y)=[v(y)]^{-1/2}\exp\left(\pm i\int^y K(\eta)d\eta\right), \quad (22)$$

where $K(y)$ is the solution of

$$E'(q_x+hy,K)=E \quad (23)$$

and

$$v(y)=\partial E'(\mathbf{k})/\partial k_y \quad \text{for} \quad k_x=q_x+hy, \quad k_y=K(y). \quad (24)$$

From $u(y)$ we may construct the solution $c(y)$ of (18) by the difference relation

$$c_m(y)=S_{mn}(q_x+hy,-id/dy)u_n(y). \quad (25)$$

The calculation of the correction terms of order h and h^2 which change $E_n(\mathbf{k})$ to $E_n'(\mathbf{k})$ are important for calculating the magnetic susceptibility of Bloch electrons at low fields.[1] For our purposes, however, these corrections will be regarded as small quantitative corrections which do not affect the theory qualitatively, just as in WKB theory we do not usually take account of higher order corrections. So from now on we shall use the zero-field band structure $E_n(\mathbf{k})$ instead of $E_n'(\mathbf{k})$. For the model we are using we plot $K(y)$ as in Fig. 3. This figure is essentially a portion of Fig. 2. Since $c(y)$ is arbitrary up to multiplication by a periodic function, $K(y)$ in (22) is not defined to within $2\pi/b$. We shall therefore set up a convention giving it the values as shown in Fig. 3. The convention has the advantage that the sign of K is the same as the sign of the velocity v of (24) and it links up naturally with the method in LN. The numbers 1 and 2 in Fig. 3 are the band labels. The phase integrals in (22) can be interpreted as areas lying between the curve in Fig. 3 and the abscissa. Thus the area φ shows the phase acquired in propagation from A to B.[16] The quantity l labels the periods in the graph which is periodic in (q_x+hy) with a period $2\pi/a$.

4. MODIFICATION OF THE THEORY

For those values of y corresponding to A and A' in Fig. 3 there is no real value for $K(y)$ because of the gap between the sheets of the Fermi surface. Figure 4 shows the part of the Fermi surface around A highly enlarged.

[15] G. E. Zil'berman, Zh. Eksperim. i Teor. Fiz. 32, 296 (1957) [English transl.: Soviet Phys.—JETP 5, 208 (1957)].

[16] The phases are dimensionless. If we say that a phase corresponds to a certain area in $\mathbf{k}$ space or in real space, that area must be multiplied by h^{-1} or h to give the phase.

The figure can also represent $K(y)$ as a function of y, as the axes indicate. A simple argument[6] suggests that tunneling (breakdown) is possible with the amplitude given by e^{-U} where

$$U=\int \mathrm{Im}K(y)dy$$

integrated across the gap. Let us suppose that the curves of Fig. 4 are approximately hyperbolas; then we construct an ellipse as shown in the gap A_2A_1, with a curvature at A_2 and A_1 equal to that of the hyperbolas. Then we may write $U=h_B/h$ where h_B is, in our units, half the area of the ellipse in $\mathbf{k}$ space. We call h_B the breakdown field. Thus the gap A_2A_1 in $\mathbf{k}$ space is of the magnitude $h_B^{\frac{1}{2}}$, and for a metal with lattice spacings of a few angstroms and a breakdown field of less than 10^5 G, the gap is a fraction of a percent of the dimensions of the Brillouin zone. We shall assume this for our model.

The order of magnitude of h_B can also be written[2] as $h_B\sim\Delta^2/E_F$, where Δ is the energy band gap and E_F is the Fermi energy. Thus we may write $\Delta\sim(h_BE_F)^{\frac{1}{2}}$. A similar situation obtains at the gaps BB' (Fig. 1), and for simplicity we shall assume that the breakdown field here is about the same as at AA'. It is also useful to define a field h_Q equal to the area of the Brillouin zone in our units. It would be of the order of 10^9 G and we may write $E_F\sim h_Q$ in units of energy.

When the gaps are so small the Wannier functions are not well localized[11] and have an extent of $h_B^{-\frac{1}{2}}$, or perhaps hundreds of lattice spacings. We can show why this happens as follows: Let us consider the band-structure function $E(\mathbf{k})$ around the point A in Fig. 1 as a function of k_y. It has a branch point just off the real axis at a distance q of about $h_B^{\frac{1}{2}}$. It then follows that the Fourier transform $W(\mathbf{R})$ of the band structure cannot fall off faster than $\exp(-q|y|)$ as $\mathbf{R}$ goes along the y axis. But $W(\mathbf{R})$ is the matrix element of the zero-field Hamiltonian between $a(\mathbf{x}-\mathbf{R})$ and $a(\mathbf{x})$ and hence the Wannier functions cannot decrease faster than $\exp(-\frac{1}{2}q|y|)$. We have a similar result about the fall-off of the Wannier functions as a function of x by considering the region around B in $\mathbf{k}$-space.

This slow fall-off has serious consequences when $h\sim h_B$. First, we cannot approximate the matrix elements (8) by their zero-field values, and secondly the difference equation (11) has very many large coefficients and becomes unmanageable in practice. Other methods of deriving the effective Hamiltonian by using functions like (15) cannot avoid the problem, since they must lead to an equation like (14), which is equivalent to (11).

Blount has suggested a way out. A generalized Wannier function may be defined by

$$\alpha_n(\mathbf{x})=\frac{1}{(2\pi)^3}\sum_m\int_{\mathrm{BZ}} d\mathbf{k}\, T_{nm}^*(\mathbf{k})b_m(\mathbf{k},\mathbf{x}), \quad (26)$$

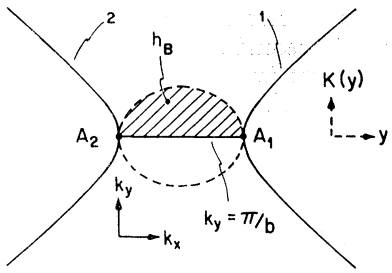

FIG. 4. Portion of Fermi surface in repeated scheme around A, highly enlarged. The breakdown field is measured by half the area of the dashed ellipse, which has the same curvature at A_1 and A_2 as the hyperbola. Numbers 1 and 2 are the band labels.

where $b_m(\mathbf{k},\mathbf{x})$ is a normalized Bloch wave for the mth band of the zero-field Hamiltonian and T is a unitary matrix. The integral is over a Brillouin zone so that $T(\mathbf{k})$ can be regarded as a periodic function. These new functions form an orthonormal complete set like the ordinary Wannier functions and we may carry out the analysis of the last section through to (18) just as before, but using these new functions. This time the zero-field effective Hamiltonian is the nondiagonal matrix

$$H_0(\mathbf{k})=T(\mathbf{k})E(\mathbf{k})T^{-1}(\mathbf{k}), \quad (27)$$

where $E(\mathbf{k})$ is the matrix defined in (17a). By an appropriate choice of T it should be possible to arrange that the matrix elements of $H_0(\mathbf{k})$ are much smoother functions of $\mathbf{k}$ than those of $E(\mathbf{k})$. This is because the branch points in $E(\mathbf{k})$ very near the real axis can be generated by the algebraic process of diagonalizing $H_0(\mathbf{k})$ and need not be in its matrix elements. Hence the Fourier transform $W_0(\mathbf{R})$ of $H_0(\mathbf{k})$ would fall off very quickly for large R. It should therefore be possible to arrange that $\alpha(\mathbf{x})$ is quite localized, to, say, within a few lattice spacings. Then we can approximate $H(\mathbf{k})$ by $H_0(\mathbf{k})$ and $N(\mathbf{k})$ by the unit matrix. The error in $(H-EN)$ would be of order h, which is small in comparison with the band gaps provided $h\ll(h_Bh_Q)^{\frac{1}{2}}$. (In a model with spin the bands may suffer from spin-degeneracy and terms of order h cannot then be neglected.)

If we wish to use the basis functions (15) instead,[1] we will modify them by replacing $b_n(\mathbf{k},\mathbf{x})$ in (15) by

$$b_n'(\mathbf{k},\mathbf{x})=\sum_m T_{nm}^*(\mathbf{k})b_m(\mathbf{k},\mathbf{x}). \quad (28)$$

The problem is now transferred to the third stage, the diagonalization procedure. We expect that the main effect is to undo the transform in (27) and we shall finish up with the Hamiltonian $E(\mathbf{P})$ to zero order in h, just as before. However, we do not have to diagonalize for all bands, but only for the first two bands, and moreover only for those values of $\mathbf{k}$ near the Fermi surface (Fig. 1). Let us suppose we are interested in a region on the Fermi surface in the first band, well away from points like A or B. Then we would have to arrange the transform (19) so that the off-diagonal elements $H'_{1m}(\mathbf{k})$ and $H'_{m1}(\mathbf{k})$ are negligibly small just in that region. The diagonal element $H'_{11}(\mathbf{k})$ would be just

$E_1(\mathbf{k})$ to zero order in h. The same idea would be used for the second band.

Except for points near A or B, the elements of the diagonalizing matrix would be expected to be smoothly varying functions of $\mathbf{k}$ in the sense that the derivative with respect to $\mathbf{k}$ would be of order h_Q^{-1}. Thus the correction terms in (20) will be of order h/h_Q and so the diagonal Hamiltonian can be written as $E(\mathbf{P})$ with an error of order h. It is evident that we cannot use a diagonalized Hamiltonian for $\mathbf{k}$ in the vicinity of the breakdown points. Thus we use a transformation which leaves the terms H'_{12} and H'_{21} large in such regions, but removes all other interband terms. In this way the effective Hamiltonian is a 2×2 matrix with smoothly varying elements. (In contrast the undiagonalized Hamiltonian derived by the theory of Sec. 3 has elements which are almost singular near the points A or B in $\mathbf{k}$ space.)

Such a Hamiltonian is not unique. Suppose we have two Hamiltonians $H_1(\mathbf{k})$ and $H_2(\mathbf{k})$ related by a unitary transform so that they give the same band structure and whose matrix elements are smoothly varying functions of $\mathbf{k}$ in the region of the Fermi surface. Then we can relate $H_1(\mathbf{P})$ to $H_2(\mathbf{P})$ as in (19), but there will be correction terms as in (20) which would be of order h. If we ignore these, the Hamiltonians are equivalent. In other words, any Hamiltonian with smoothly varying matrix elements that gives the correct band structure will be adequate for our purposes.

The usual theory of the effective Hamiltonian also suffers from nonuniqueness, but not in the zero-order terms which are given by (17). This nonuniqueness is associated with an arbitrariness in the phases of the Bloch functions or with changing the Wannier functions by using a diagonal matrix $T_{mn}=\delta_{mn}\exp[-i\theta_n(\mathbf{k})]$ in (26), where $\theta_n(\mathbf{k})$ is some real function of $\mathbf{k}$. The equivalence of such Hamiltonians does not appear to have been proved. This brings into question the assumption that the set of Wannier functions (7) or of their Fourier transforms (15) is complete. The incompleteness would not necessarily be disastrous, for (11) can be derived by a variational argument, but it would mean that changing the Wannier functions would give a slightly different effective Hamiltonian. If the set is incomplete, then it would seem that the best Wannier functions to use are the most localized, since the ordinary Wannier functions $a_n(\mathbf{x}-\mathbf{R})$ or $\alpha_n(\mathbf{x}-\mathbf{R})$ do form a complete set, and the localization would mitigate the effect of the exponential function in (7).

If we are interested in the effects of spin-orbit coupling on magnetic breakdown, then we need the term of order h in the effective Hamiltonian. We would have to consider a region in $\mathbf{k}$ space where breakdown takes place, say, centered on $\mathbf{k}_0$. We could employ Luttinger-Kohn theory[17] in this case by choosing the matrix $T(\mathbf{k})$ in (26) for $\mathbf{k}$ near $\mathbf{k}_0$ by

$$T_{mn}(\mathbf{k})=\int_\Omega d\mathbf{x}\,\chi_m^*(\mathbf{k},\mathbf{x})b_n(\mathbf{k},\mathbf{x}),$$

where the integral is over a unit cell Ω and χ is a Luttinger-Kohn function defined by

$$\chi_n(\mathbf{k},\mathbf{x})=[\exp i(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{x}]b_n(\mathbf{k}_0,\mathbf{x}).$$

The method of expanding the effective Hamiltonian as a power series in $(\mathbf{k}-\mathbf{k}_0)$, described in the next section, will give the nondiagonal Hamiltonian of $\mathbf{k}\cdot\mathbf{p}$ theory which can be diagonalized in the usual way.[17] Unfortunately, we cannot define T in this way all over the Brillouin zone, for if we regarded it as a periodic function, it would have discontinuities at the zone boundaries and the corresponding Wannier functions (26) would not be well localized.

5. TREATMENT AT SPECIAL POINTS

We shall assume that the WKB solution (22) is valid except for the vicinity of the points A, B, and the turning point T (Fig. 3). It is necessary to examine the difference equations more carefully in these regions. The usual method is to approximate them by differential equations, but first it is necessary to perform a gauge transform so that the solution is slowly varying. Suppose we are interested in a region near the point $y=y_0$, where the solution has a wave number $K_0=K(y_0)$. Then instead of the solution $u(y)$ we use the modified solution

$$\psi(y)=e^{-iK_0y}u(y) \tag{29}$$

and for convenience we shift the origin to y_0. This is equivalent to replacing the operator $\mathbf{P}=(q_x+hy, -id/dy)$ in (18) or (21) by

$$\mathbf{P}\rightarrow\mathbf{P}'=\mathbf{P}_0+\mathbf{P} \tag{30}$$

with

$$\mathbf{P}_0=(q_x+hy_0,K_0). \tag{31}$$

Around the new origin the expectation value of $\mathbf{P}$ is small and we may expand the effective Hamiltonian in powers of $\mathbf{P}$. Thus, by (13a) we would write

$$H(\mathbf{P}')=\textstyle\sum_S W(\mathbf{S})[\exp(-i\mathbf{P}_0\cdot\mathbf{S})]\times[1-i\mathbf{P}\cdot\mathbf{S}-\frac{1}{2}(\mathbf{P}\cdot\mathbf{S})^2+\cdots].$$

We ignore the higher powers of $\mathbf{P}$. This is hard to justify mathematically, though it seems physically reasonable. The assumption is that the missing terms are not important until y becomes large enough to use the WKB solution (22) and that then they can be accounted for by using the WKB solution for the true Hamiltonian rather than the truncated Hamiltonian. We are here replacing a difference equation by an approximating differential equation, the converse of the methods for solving differential equations numerically. Thus around the point T we would expect that the Hamiltonian

[17] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955); J. M. Luttinger, *ibid.* **102**, 1030 (1956).

$E_2(\mathbf{k})$ could be expanded as

$$E_2(\mathbf{k})=E_0-\alpha k_x-\beta^2 k_y^2,$$

where we have taken the origin at T, E_0 is the Fermi energy, and α and β are certain coefficients. Then (21) would become (with q_x set equal to zero)

$$(\beta^2 d^2/dy^2-\alpha h y)\psi(y)=0.$$

This is the equation for the treatment of linear turning points in WKB theory[18] and it indicates that a phase shift of $\pi/2$ must be put in at the turning point. A more careful statement goes as follows: The coefficient of the reflected wave is equal to $e^{i\pi/2}$ times the coefficient of the incoming wave if the lower limit of the phase integrals in (22) is put at the turning point. This convention for the phase of the wave function will be used because it gives a natural way of distinguishing phase shifts acquired in propagation from those due to reflection. It does not imply that (22) is a valid solution close to the turning point.

The situation is considerably more complicated at a point like A. Here we set $K_0=\pi/b$ in (29). A Hamiltonian which should describe the bands at such a point might be to first order in $\mathbf{k}$[2,19]

$$H(\mathbf{k})=\begin{pmatrix}\alpha k_x+\beta k_y+E_0 & \frac{1}{2}\Delta\\ \frac{1}{2}\Delta & \alpha k_x-\beta k_y+E_0\end{pmatrix}, \quad (32)$$

where E_0 is the Fermi energy, Δ is the band gap, and α and β are certain coefficients. This Hamiltonian leads to a second-order differential equation similar to the Weber equation and may be solved in terms of confluent hypergeometric functions,[19,20] whose asymptotic forms for large y may be calculated. These are solutions for the nondiagonal Hamiltonian (32), but it should be possible to relate them to the solutions of the diagonalized Hamiltonians $E_2(\mathbf{P}')$ for $(y-y_2)$ sufficiently negative and $E_1(\mathbf{P}')$ for $(y-y_1)$ sufficiently positive, where y_2 and y_1 are the positions of A_2 and A_1 (Figs. 4 and 5). The solutions of the diagonalized Hamiltonians $E_j(\mathbf{P}')$ $(j=1,2)$ we shall call ψ_j. These will have the form (22), but their phase integrals are to be measured from y_j, as indicated in Fig. 5, and because of the gauge transform (29) the areas are measured from a base line

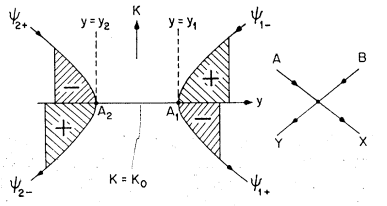

FIG. 5. (a) Signed areas measuring the phase-integrals for the solutions $\psi_{i\pm}$ of the diagonalized Hamiltonians. The subscripts $\pm$ give the direction of propagation along Oy. (b) The amplitudes of the solutions on the four arms.

$K=K_0$ (Fig. 5) instead of from the axis $K=0$, as in Fig. 3. Suppose then that a solution $(A'\psi_{2+}+Y'\psi_{2-})$ for $(y-y_2)$ sufficiently negative joins on to a solution $(B'\psi_{1-}+X'\psi_{1+})$ for $(y-y_1)$ sufficiently positive (Fig. 5b). Here the subscripts $\pm$ refer to the direction of propagation along the y axis. Then these coefficients will be related by some unitary matrix

$$\begin{pmatrix}X'\\ Y'\end{pmatrix}=\begin{pmatrix}e^{i\gamma_1}(1-Q^2)^{1/2} & e^{i\delta_1}Q\\ e^{i\delta_2}Q & e^{i\gamma_2}(1-Q^2)^{1/2}\end{pmatrix}\begin{pmatrix}A'\\ B'\end{pmatrix}, \quad (33)$$

where $\gamma_1+\gamma_2=\delta_1+\delta_2\pm\pi$ for unitarity. Here $(1-Q^2)^{\frac{1}{2}}$ is the tunneling amplitude and is given by Blount's criterion[2] or the formula at the beginning of the fourth section. Q is taken as real and represents the reflection amplitude.

Now the solutions $\psi_{j\pm}$ are solutions of $E_j(\mathbf{P}')$ and we must relate them to the solutions $u_{j\pm}$ of the $E_j(\mathbf{P})$ as in (21). This is, of course, just undoing the gauge transform (29) with $K_0=\pi/b$. But we would like to choose the origin of the phase integrals in (22) at y_j, and we would like to use the convention of Fig. 3 for $K(y)$. Then the transforms are

$$u_{j\pm}(y)=\exp[\pm i\pi(y-y_j)/b]\psi_{j\pm}(y). \quad (34)$$

Thus in effect ψ_+ has been multiplied by $e^{i\pi y/b}$ to undo the gauge transform, and ψ_- has been multiplied by $e^{i\pi y/b}$ and $e^{-2\pi i y/b}$. This last multiplication by a function periodic in the lattice does not matter, since really we are solving difference equations. Now if a solution $(Au_{2+}+Yu_{2-})$ of $E_2(\mathbf{P})$ for $(y-y_2)$ sufficiently negative joins on to a solution $(Bu_{1-}+Xu_{1+})$ of $E_1(\mathbf{P})$ for $(y-y_1)$ sufficiently positive, then we shall have instead of (33)

$$\begin{pmatrix}X\\ Y\end{pmatrix}=\begin{pmatrix}(1-Q^2)^{1/2}e^{i\gamma_1}e^{i\pi(y_1-y_2)/b} & Qe^{i\delta_1}e^{2\pi i y_1/b}\\ Qe^{i\delta_2}e^{-2\pi i y_2/b} & (1-Q^2)^{1/2}e^{i\gamma_2}e^{i\pi(y_1-y_2)/b}\end{pmatrix}\begin{pmatrix}A\\ B\end{pmatrix}. \quad (35)$$

The additional phases we shall call "extrinsic" phases, as opposed to the phases γ_i, δ_i which we shall call "intrinsic." The phase $\pi(y_1-y_2)/b$ may be interpreted as the phase acquired by the wave as it propagates through the gap while still oscillating with a wave number π/b. The phases $2\pi y_1/b$ and $2\pi y_2/b$ are the phases of positions of the turning points with respect to the periodic lattice, and these phases are really the key to the theory in LN, because they can vary from one cell in Fig. 3 ($l=0$, say) to the next. This is because changing y_1 and y_2 by $2\pi/ah$ changes the phases by

[18] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), pp. 1095–1099.

[19] U. N. Upadhyaya, University of Oregon, Eugene, Oregon (unpublished).

[20] Reference 18, p. 1403.

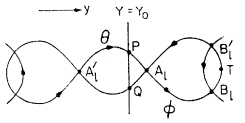

FIG. 6. Network to represent propagation of waves along the y direction. Figure shows lth link corresponding to lth period in Fig. 3. The points A_l, A_l', B_l, B_l' are junctions, as in Fig. 5(b). Waves starting from a source at $Y=Y_0$ and travelling to the right are represented as starting at P, while those travelling to the left at Q.

$(2\pi)^2/hab$, which is not necessarily an integral multiple of 2π.

The situation at the point B where $q_x+hy=\pi/b$ is physically very similar, but this time there are no "extrinsic" phase shifts. There is another difference. The roles of k_x and k_y are reversed in comparison with (32) and this leads to a slightly different equation. The effect of this can be illustrated in the classical limit $h\ll h_B$. There will then be almost total reflection at the junctions. But the points A_1 and A_2 (Fig. 5) will act like the turning points T, and thus the phases δ_1 and δ_2 in (35) will be equal to $\frac{1}{2}\pi$. In the problem for the region around B there are no turning points and so the corresponding phases are zero. This shows that the lens L_1 (Fig. 2) has its turning points at AA', whereas the lens L_2 has them at points like T. This is a mathematical rather than a physical problem and will not be treated further.

This argument is in effect a justification of the method used in LN, for it should now be evident how we may set up a network (Fig. 6). The propagation phase shifts like θ and φ are obtained by calculating the appropriate phase integrals and are represented by areas[16] as shown in Fig. 3. For the purposes of the next section we shall simplify the phase shifts at the junctions as in (35) by ignoring the difference (y_1-y_2). The "extrinsic" phase for reflection to the left $-2\pi y_2/b$ can be written, as in LN,

$$\begin{aligned} &\text{at } A_l: \quad \alpha-\delta+\epsilon l, \\ &\text{at } A_l': \quad \alpha+\delta+\epsilon l, \end{aligned} \tag{36}$$

where $\alpha=2\pi q_x/bh$, δ corresponds to the area as shown in Fig. 3, and $\epsilon=(2\pi)^2/hab$ corresponds to the area of the unit cell in Fig. 3 or of the Brillouin zone.[16] The reflections to the right have an equal and opposite value. We shall also ignore the phase $\frac{1}{2}\pi$ at the turning points T. The argument with all the phases included will be similar but somewhat more complicated.

6. THE de HAAS—van ALPHEN EFFECT

We need to calculate the terms in the energy density of states at the Fermi level E which are oscillatory functions of h. These are then put into the usual formula for the free energy,[21] and from this the oscillatory terms in the magnetic susceptibility can be calculated. The oscillatory part of the density of states can be obtained from a Green's function.[22]

The method is to find the density of states for the one-dimensional difference equation (18) and then integrate the result over q_x from 0 to $2\pi/a$ to obtain the energy density of states for the two-dimensional problem. Later we shall introduce a third coordinate z to give a model for a three-dimensional system.

We assume that (18) has been derived using the modified Wannier functions of Sec. 4, so that $N(\mathbf{k})$ may be approximated by the unit matrix. It is also convenient (as in LN) to assume that the field h has a "rational" value given by

$$h=(2\pi/ab)(\lambda/N), \tag{37}$$

where λ and N are integers, although this time it will not be assumed that λ/N has been reduced to its lowest terms. We may impose periodic boundary conditions at the ends of a long line of N lattice sites in the y direction. We shall use the notation $Y=nb$ to denote the y coordinates of these sites. Then (18) is essentially just a matrix equation with a very large matrix. If we lump the band index and position index into a Greek suffix (λ, μ, or ν) and use the summation convention, (18) may be written

$$(E\delta_{\lambda\mu}-H_{\lambda\mu})c_\mu=0. \tag{38}$$

Let us call the eigenvalues E_j. The Green's function (or rather matrix) $G_{\lambda\mu}(E)$ is defined by

$$[(E+i\epsilon)\delta_{\lambda\mu}-H_{\lambda\mu}]G_{\mu\nu}=\delta_{\lambda\nu}, \tag{39}$$

where ϵ is a positive infinitesimal. Then the energy density of states $n(E)$ is just

$$n(E)=\textstyle\sum_j \delta(E-E_j)=-\pi^{-1}\,\mathrm{Im}G_{\lambda\lambda}(E), \tag{40}$$

which can be proved by diagonalizing $H_{\lambda\mu}$ by a unitary transform. If we alter H by a unitary transform, then it is easy to show that the trace in (40) is unaltered, so that we may carry out the procedure of diagonalizing (18) in bands. This can be done almost everywhere, except for lattice sites Y near the y coordinates of A and B. But these are small regions and will not contribute much to the trace. We thus have to find the Green's function $G_n(Y,Y')$ for the diagonalized Hamiltonian $E_n(\mathbf{P})$. Then it follows that

$$n(E)\approx-\pi^{-1}\textstyle\sum_{n,Y}\mathrm{Im}G_n(Y,Y). \tag{41}$$

To calculate the Green's function approximately for a source at Y' we first replace (q_x+hy) in (21) by a constant (q_x+hY'). Then (21) becomes a difference equation with constant coefficients. For a given value of Y' we have solutions e^{ik_yY} where k_y is given by $E_n(q_x+hY',k_y)=E$ [compare (23)]. To obtain the Green's function we replace the right-hand side of (21)

[21] J. Callaway, *Energy Band Theory* (Academic Press Inc., New York, 1964), p. 258.

[22] L. M. Falicov and H. Stachowiak, Phys. Rev. **147**, 505 (1946).

by a source term $-\delta_{YY'}$ which may be written

$$-\delta_{YY'}=-\frac{b}{2\pi}\int_{-\pi/b}^{\pi/b}\exp[i\kappa(Y-Y')]d\kappa,$$

since Y and Y' are lattice translations. Then it follows that

$$G_n(Y,Y')=-\frac{b}{2\pi}\int_{-\pi/b}^{\pi/b}\frac{\exp[i\kappa(Y-Y')]}{E+i\epsilon-E_n(k_x,\kappa)}d\kappa, \quad (42)$$

where $k_x=q_x+hY'$. Since the integrand is periodic, the path of integration may be displaced upwards (for $Y>Y'$) or downwards (for $Y<Y'$) in the complex κ plane, leaving behind the residues at poles very close to the real axis. It is found that for each real positive value k_y of κ satisfying $E(k_x,k_y)=E$ there is a residue

$$-ibv^{-1}(\exp ik_yY_>)[\exp(-ik_yY_<)], \quad (43)$$

where $Y_>$ and $Y_<$ are the greater and lesser, respectively, of Y and Y' and v is the velocity $\partial E/\partial k_y$ at (k_x,k_y). This residue (43) simply represents waves propagating outwards from the source. The contribution to the integral from the displaced path falls off for $|Y-Y'|$ large and represents "evanescent" waves which are of no further interest.

The second stage is to allow the term (q_x+hy) in (21) to vary again. This will have two effects, one "quantitative" and the other "qualitative." The Green's function will be modified locally because the coefficients in the difference equation are now slowly varying, but this would be a small effect of order h. This might lead to a small change in the energy density of states which will not give an oscillatory contribution to the susceptibility. The contribution (43) to the Green's function *presumably* will now be replaced by

$$-ibu_+(Y_>)u_-(Y_<), \quad (44)$$

where $u_\pm$ are the solutions of (21), given by (22). The "qualitative" effect arises from the reflection of the outward-travelling waves (44) from distant junctions and turning points. These contribute to (41) and lead to the de Haas–van Alphen oscillations. They can be calculated by multiple-scattering theory. Suppose we have a source at $Y=Y_0$ (Fig. 6). Then we must consider all paths p in the one-dimensional system by which the wave returns to Y_0. It is better to represent the paths on a diagram like Fig. 7, similar to Fig. 2. A wave starting to travel to the right is presumed to commence at P, while one to the left, at Q. It always travels clockwise round a circle. A reflection to the left by a junction like A_l or A_l' transfers the wave on to the next circle up and a reflection to the right on to the next circle down. By (36) these reflections introduce phases dependent on q_x, a phase $2\pi q_x/bh$ for a reflection to the left, and $-2\pi q_x/bh$ for a reflection to the right. Let us consider a wave starting out at Y_0 to the right (represented by starting at P in Fig. 7), being reflected by the turning point T, and returning to Y_0 (at Q). The phase

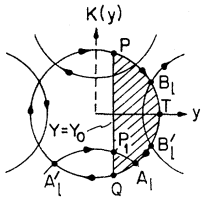

FIG. 7. Another network to represent propagation of waves along the y direction. For measuring phase-integrals vertical axis should be taken as $K(y)$.

shift due to the propagation is represented by the shaded area. The wave will also be phase shifted and attenuated at the junctions. But when we perform the sum in (41) this particular contribution will cancel out because evidently the area varies with Y_0 and cancellation will occur by "random phase." To avoid this the wave starting at P must return again from the left. Such a path in Fig. 7 might go from P to T to Q to A_l' and be reflected to P_1. However, the phase shift upon reflection at A_l' depends on q_x according to (36), and so when we later perform the integration over q_x for the two-dimensional density of states, this contribution will also cancel out. To avoid this the path must return again to its starting point on the same circle, and so in fact it must be a closed orbit.

We need therefore only consider closed orbits. It can be shown, as we might expect, that for such a path the sum of the propagation phase shifts and the "extrinsic" phase shifts at the junctions is just h^{-1} times the **k**-space area of the orbit S_p. We may therefore express the total phase shift and attenuation for a given orbit p as

$$t_p=c_p\exp(ih^{-1}S_p), \quad (45)$$

where c_p gives the attenuation and the "intrinsic" phases at the junctions and S_p is the **k**-space area of a closed path p. If the path goes n times round a given area, that area must contribute n times, and hole orbits contribute negatively. In (44) we may phase $u(Y)$ to be real at $Y=Y_0$. It is then easy to see that the contribution to $G(Y_0,Y_0)$ from a given orbit p is simply

$$-ibt_p/v(Y_0). \quad (46)$$

The integration over q_x is simple, since all the terms are independent of q_x. Let us suppose that the sample has dimension A in the x direction over which we impose periodic boundary conditions. This just introduces a factor $A/2\pi$ in the two-dimensional density of states $\nu(E)$. The sum over Y_0 on N sites is a little more difficult. To go over to a scheme in $\boldsymbol{k}$ space, we replace Y by k_x/h and the sum over Y_0 by an integral, since the rapidly varying terms cancel out by "random phase." Then we have

$$\sum_{Y_0}=\frac{1}{b}\int_0^{Nb}dY=\frac{1}{hb}\int_0^{Nbh}dk_x.$$

This [by (37)] is equal to

$$\frac{aN}{2\pi\lambda}\int_0^{2\pi\lambda/a}dk_x=\frac{aN}{2\pi}\int_0^{2\pi/a}dk_x,$$

since the terms we are keeping in the Green's function are periodic in k_x. This is valid since a displacement of the source Y_0 by $2\pi/ah$ only alters q_x, and we are keeping only terms independent of q_x. We may also write $dk_x/v(Y)$ as $ds_k/v_\perp$, where ds_k is an infinitesimal length of arc on the Fermi surface and $v_\perp$ is the magnitude of the velocity normal to the Fermi surface.

Then finally, for the oscillatory terms in the two-dimensional density of states we have

$$\nu_{\rm osc}=\frac{NAb}{(2\pi)^2}\int\frac{ds_k}{v_\perp}\sum_p (t_p+t_p^*)\,, \tag{47}$$

where at any point on the Fermi surface we have to sum over all paths p passing through the point. We integrate only within one Brillouin zone for each band.

As a simple example let us take the free-electron model. The Fermi "surface" is a circle of area πE (in our system of units). The only orbits possible are once, twice, thrice, etc., around this circle. Thus we may set $S_p=p\pi E$, which gives

$$\nu_{\rm osc}=(NAb/4\pi)\sum_{p=1}^{\infty}(e^{ip\pi E/h}+e^{-ip\pi E/h})\,.$$

The steady-state term is just $NAb/4\pi$, so we find that altogether

$$\nu=(NAb/4\pi)\sum_{n=-\infty}^{\infty}\delta(E/2h-n)\,.$$

Thus we find that the possible energy levels are multiples of $2h$. They should be odd multiples of h, but an error occurs because we have ignored a phase shift of $p\pi$ for the pth orbit due to the turning points. The formula also suggests negative energy states, but the above theory only applies for $E\gg h$, when there are propagating waves and the WKB theory can be used.

The rest of the analysis for the de Haas–van Alphen effect is straightforward and only a rough sketch is necessary. We process (47) term by term and, as we might expect, the oscillatory nature of each term comes out in the free energy. The first step is to introduce a third coordinate z and integrate over k_z to obtain the three-dimensional density of states $N(E)$. This will, in the usual way, pick out the extremal orbits by "stationary phase." Then we take a particular term of $N(E)$, $\chi_p\exp[ih^{-1}S_p(E)]$, where χ_p is some coefficient which varies slowly with h and E, and $S_p(E)$ is the area of an extremal orbit at the energy E. The expression for the free energy is[21]

$$F=N_0\mu-\int_0^{\infty}N(E)g(E)dE\,, \tag{48}$$

where N_0 is the number of particles, μ is the chemical potential, and g is given by

$$g=kT\ln(1+e^{(\mu-E)/kT})\,,$$

with T the temperature and k Boltzmann's constant. Next we define $\zeta(E)$ by

$$d^2\zeta/dE^2=N(E)\,,\quad \zeta(0)=\zeta'(0)=0\,,$$

that is, the double integral of $N(E)$ from $E=0$. Then by integrating by parts twice we may write (48) as

$$F=N_0\mu-\int_0^{\infty}\zeta(E)(d^2g/dE^2)dE\,, \tag{49}$$

where d^2g/dE^2 is the derivative of the Fermi distribution function. It is very sharply peaked at the Fermi level over a range a few kT in width. Within this range we may assume that $S_p(E)$ varies linearly with E, and so the double integration of a particular term

$$\chi_p\exp[ih^{-1}S_p(E)]$$

in $N(E)$ may be written as

$$\chi_p(ih^{-1}dS_p/dE)^{-2}\exp[ih^{-1}S_p(E)]+Ef_1+f_2\,, \tag{50}$$

where f_1 and f_2 are constants of integration. After the final integration in (49), the first term in (50) gives an oscillatory contribution to the free energy, with a period in h^{-1} corresponding to the extremal area $S_p(E_F)$, where E_F is the Fermi energy.

It is conceivable that f_1 and f_2 also give oscillatory contributions. We can see this as follows in the two-dimensional model: At some energy well below the Fermi level the constant-energy contour in Fig. 1 will be just breaking through in to the second zone. We shall thus have a rather singular situation over a narrow range of energy in this transition region. This could lead to the constants of integration in (50) being periodic in h^{-1} with periods corresponding to areas in this energy range, since they are derived by integrating across this region. Similarly, the integration over k_z in the three-dimensional situation might pick up non-extremal areas at those values of k_z where the Fermi-surface cross-section is in a transition region. This has been discussed in more detail by Roth.[23]

At any rate we have the simple result that all possible orbits in a network may contribute to the de Haas–van Alphen effect in the usual way, except that the contribution of a given orbit will be attenuated by imperfect transmission and reflection at the junctions on the orbit. The transmission and reflection amplitudes to be used are, of course, the quantities $(1-Q^2)^{\frac{1}{2}}$ and Q in (35), and not the squares which give the transmission and reflection probabilities. The nth harmonic of a fundamental period corresponding to some area should be regarded as being due to an orbit going n times around that area. In particular, we do not expect to find any periodicities corresponding to the area of the Brillouin zone (as suggested in LN) because it is not possible to find orbits with this area.

[23] L. M. Roth, Phys. Rev. **145**, 434 (1966).

7. THE NETWORK WAVE FUNCTION

So far the wave functions we have used are Bloch waves in the x direction with a wave number q_x and do not resemble wave functions lying on a network in real space, as suggested by Pippard.[7,8] Such wave functions are readily constructed, however, as eigenfunctions of the Hamiltonian (2) by the following method. Let us first consider the free-electron problem. In this case the components of the operator $\boldsymbol{\Pi}$ (Eq. 3) measure the position of the center (x_0,y_0) of a circular orbit according to,[14]

$$\Pi_x=-hy_0, \quad \Pi_y=hx_0.$$

The operators satisfy the commutation rule,

$$[\Pi_x,\Pi_y]=-ih$$

and thus the uncertainty in measuring them both obeys a relation

$$\Delta\Pi_x\cdot\Delta\Pi_y\sim h.$$

The wave function obtained on separating the Schrödinger equation in the Landau gauge as in Sec. 3 is

$$\psi=e^{iq_xx}w(y),$$

where $w(y)$ satisfies the simple-harmonic-oscillator equation

$$[-d^2/dy^2+(q_x+hy)^2-E]w(y)=0.$$

The wave function ψ is an eigenfunction of Π_x (with eigenvalue q_x), so that $\Delta\Pi_x=0,\Delta\Pi_y=\infty$. To construct a wave function where $\Delta\Pi_x=\Delta\Pi_y\sim h^{1/2}$, we use the operator $e^{i\Pi_y\lambda}$ (with $\Pi_y=p_y+hx$), which commutes with the Hamiltonian, and set[8]

$$\Psi(x,y)=\int\alpha(\lambda)e^{i\Pi_y\lambda}\psi(x,y)d\lambda, \tag{51}$$

with $\alpha(\lambda)=\exp(-\frac{1}{2}\lambda^2h)$. This wave function is circular, with its center at $(0,-q_x/h)$. The function $\alpha(\lambda)$ is determined by the kinematics of the uncertainty principle rather than the dynamical features of the Hamiltonian. So we shall try a similar construction on the envelope functions of Sec. 3 and show that this does lead to a function localized on a network like that in Fig. 2.

We start with the function

$$\psi(\mathbf{r})=e^{iq_xx}u(y),$$

where $u(y)$ satisfies (21). (Strictly speaking this function is only defined on the lattice sites.) Instead of $e^{i\Pi_y\lambda}$ we use an operator similar to $\tau(n\mathbf{b})$ [Eq. (4)] with n integral, which acts on envelope functions and is equal to $\exp inb(hx-id/dy)$ in the Landau gauge. This operator commutes with the effective Hamiltonian in (18). The integration in (51) is replaced by a discrete sum over n, and so we set up the wave function

$$\Psi(\mathbf{r})=\sum_n[\exp(-\tfrac{1}{2}n^2b^2h)]e^{i(q_x+nhb)x}u(y+nb). \tag{52}$$

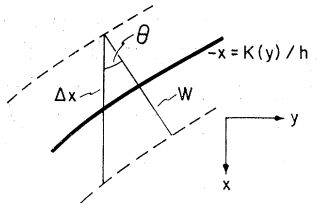

FIG. 8. Heavy line represents center of track $-x=K(y)/h$ in real space. Δx is spread of wave function in x direction; W is width of track.

We use (22) for $u(y)$, taking the positive sign. Let us write

$$J(y)=\int^y K(\eta)d\eta$$

for the phase in (22). Then by Taylor's theorem we may write

$$J(y+nb)=J(y)+nbK+\tfrac{1}{2}n^2b^2K' \tag{53}$$

to second order. Here $K=K(y)$ and $K'=dK(y)/dy$. We leave out the higher terms, assuming that the exponential in (52) cuts off the sum for large values of n. Then we have

$$\Psi(\mathbf{r})\approx[v(y)]^{-\frac{1}{2}}e^{iq_xx}[e^{iJ(y)}][\chi(x,y)] \tag{54a}$$

with

$$\chi(x,y)=\sum_n[\exp(-\tfrac{1}{2}n^2b^2h)]\times\exp i(Knb+nbhx+\tfrac{1}{2}K'n^2b^2). \tag{54b}$$

We expect χ to peak for $x=-K/h$, for then the phase is stationary as a function of n at $n=0$. It will also peak for $x=-K/h-2\pi m/bh$ for m any integer. Thus it is easy to see that the wave function will be localized on a network as in Fig. 2. The argument fails at points like A and B, because the effective Hamiltonian cannot be diagonalized and at T, because the WKB solution (22) is invalid, but it is valid over the major part of the network.

We may also calculate the way the wave function falls off as x deviates from the value $-K/h$. For this purpose we replace the sum in (54b) by an integral. This should not cause a serious error since $b^2h\sim h/h_Q\ll1$. We find that the wave function is like a Gaussian with a width $\Delta x=h^{-1}[(h^2+K'^2)/h]^{\frac{1}{2}}$. Figure 8 shows a portion of the track $x=-K(y)/h$ (note that the axes are chosen unconventionally). The track is shown as a continuous line. The tangent of the angle θ is just the slope $h^{-1}dK(y)/dy$. The line Δx shows the spread of the wave function about the track. Resolving Δx on to the direction at right angles to the track, we find that the wave function has a width $W=h^{-\frac{1}{2}}$ which can be interpreted as the uncertainty in the position of the electron owing to its zero-point motion in the magnetic field. This width is of the order of $(h_Q/h)^{\frac{1}{2}}$ lattice spacings and is constant along the track.

Thus the function $\Psi(\mathbf{r})$ in (52) is localized on a network as in Fig. 2. From (52) we find that if we compare

the wave function at $(x+2\pi/hb,y)$ with its value at (x,y), the magnitude is the same but the phase is changed by $2\pi q_x/hb$, because of the phase-factor $e^{iq_x x}$ in (54a). There is no such periodicity over a distance $2\pi/ha$ in the y direction because the wave-function amplitudes are not periodic in the network in Fig. 6 since the phase shifts at A_l and A_l' depend on l according to (36). This is what we would expect by setting up the wave function as a Pippard network[8] in the Landau gauge used in this paper.

8. CONCLUDING REMARKS

We have set up the theory of the effective Hamiltonian in terms of Wannier functions and difference equations. This method can be used to derive the wave functions in configuration space, and it is hoped that this approach gives a clear picture of the situation as well as of the problems involved. It is certainly one of the more elementary methods, since all the matrix elements are finite and do not involve δ functions. When breakdown is liable to occur, the Wannier functions have to be modified, and than the effective Hamiltonian must be treated, at least in the regions of breakdown, as a matrix.

The use of such a theory has enabled us to justify the network model used by the author,[10] which can readily be shown to be essentially equivalent to those used by Pippard.[7,8] Then we have considered the problem of the de Haas–van Alphen effect with breakdown. The results are very much as we might expect. Finally, we have shown how to construct a wave function lying on a network as suggested by Pippard.[7,8]

All this has been done in the framework of a simple rectangular model. For the sake of completeness the theory should be extended to deal with more complicated models, but the author does not expect that any surprising new results would be found. The problem of spin-orbit coupling has not been treated and it may be important, particularly when the band gaps across which breakdown takes place are produced primarily by spin-orbit coupling. In this case we would have to consider the effect of the first-order term in the effective Hamiltonian expanded as a power series in h.

ACKNOWLEDGMENTS

The author would like to thank Professor E. Brown, Professor L. M. Falicov, and Professor G. Trammell for their help in discussing some of the topics in this article.

PHYSICAL REVIEW VOLUME 149, NUMBER 2 16 SEPTEMBER 1966

Band Structure and Fermi Surface of White Tin*†

GIDEON WEISZ‡
Department of Physics and Institute for the Study of Metals, University of Chicago, Chicago, Illinois
(Received 4 February 1966)

The band structure and Fermi surface of metallic white tin are successfully calculated, using a local-pseudopotential approximation. A new, simple, and accurate model is found for including the spin-orbit coupling within the framework of a pseudopotential Hamiltonian. The results are compared with experimental data, and good agreement is found throughout.

I. INTRODUCTION

THIS paper contains a theoretical calculation of the band structure and Fermi surface of white tin, and a comparison of the results with experimental data. It also introduces a new method for dealing with the spin-orbit interaction. The study of tin is motivated by many circumstances and considerations. These are discussed in the following paragraphs, beginning with the experimental situation and proceeding to the problems and possibilities to which the tin lattice structure gives rise.

In recent years, the Fermi surface of tin metal has been the subject of a number of experimental investigations, but the complexity of the surface has thus far prevented it from being fully understood. The present calculation by a pseudopotential method is intended to furnish a guide and a stimulus to further experimentation. The following experiments have shed some light on the electronic structure of Sn: de Haas–van Alphen effect,[1] galvanomagnetic properties,[2–4] cyclotron size

* Supported in part by the National Science Foundation.
† Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Chicago, Chicago, Illinois.
‡ Present address: Division of Engineering and Applied Physics, Harvard University, Cambridge, Massachusetts.

[1] A. V. Gold and M. G. Priestley, Phil. Mag. **5**, 1089 (1960).
[2] N. E. Alekseevskii, Yu. P. Gaidukov, I. M. Lifshitz, and V. G. Peschanskii, Zh. Eksperim. i Teor. Fiz. **39**, 1201 (1960) [English transl.: Soviet Phys.—JETP **12**, 837 (1961)].
[3] N. E. Alekseevskii and Yu. P. Gaidukov, Zh. Eksperim. i Teor. Fiz. **41**, 1079 (1961) [English transl.: Soviet Phys.—JETP **14**, 770 (1962)].
[4] R. C. Young, Phys. Rev. Letters **15**, 262 (1965).