Quantization of Coupled Orbits in Metals II. The Two-Dimensional Network, with Special Reference to the Properties of Zinc

A. B. Pippard

[317]

QUANTIZATION OF COUPLED ORBITS IN METALS

II. THE TWO-DIMENSIONAL NETWORK, WITH SPECIAL REFERENCE TO THE PROPERTIES OF ZINC

By A. B. PIPPARD, F.R.S.

Cavendish Laboratory, University of Cambridge

CONTENTS

In an earlier paper the wave functions and eigenvalues for an electron moving in a magnetic field, and interacting with one component of lattice potential, were analysed in terms of a model of coupled localized orbits. The model is now examined in more detail and shown to be a reasonable approximation to one possible representation of the true wave function. It is then extended to cover the case of a two-dimensional metal, the model now consisting of a network of interlocking orbits, on which an electron can move with specified probability amplitude for making a transition between orbits at any junction point. The problem of periodicity of the structure is discussed carefully, and it is found that the phase changes accompanying gauge transformations assume great importance. It is shown that the magnetic field imposes a periodicity on the network which is not in general compatible with that of the lattice potential, and the consequences are briefly investigated with the conclusion that they are probably observable only with difficulty. A special case, the hexagonal network, is then solved exactly, the magnetic field being chosen to avoid the above-mentioned difficulty of incompatible periodicities. From the solution an energy level diagram is constructed, showing how the free-electron levels are broadened by the lattice potential and, as this is made stronger, reconstruct themselves into the sharp level system predicted by Onsager's semi-classical method. In the intermediate stages of the process the bands contact each other frequently and other types of singularity appear. It is claimed that the structure revealed by this simple model is more elaborate than anything that could be readily derived by a perturbation treatment of the magnetic field. The electrons are able to move as quasi-particles in straight lines in any direction through the lattice, the velocity being derived from the energy level structure by the standard formula $\hbar^{-1}\nabla_k E$. When the bands are at their broadest the velocity is comparable with that of a free electron near the corners of the Brillouin zone. The contribution of the quasi-particles to the conductivity of the metal is evaluated on the assumption that the width of individual bands is rather less than kT, so that much of the rapid variation of conductivity with Fermi energy is smoothed out. The variations that are left are still considerable and have a periodicity determined by the smallest quantized orbits.

The results of the theory are applied to the fairly extensive, though not always consistent, observations of oscillatory behaviour in zinc. The anomalous variation with field strength of the de Haas–van Alphen amplitude can be satisfactorily explained if it is assumed that the energy gap across the sides of the Brillouin zone is about 0·027 eV. The vigorous resistance oscillations, attributed by Stark to magnetic breakdown changing some of the hole orbits into electron orbits, are shown to require more than this, though this effect is certainly important and is implied by the theory. It is suggested that the quasi-particles provide the necessary extra mechanism to account for resistance and Hall-effect data, but quantitative comparison is far from satisfactory, and it is concluded that more data and further analysis are probably needed. Stark's proposal that the fine structure of the oscillations are due to spin, with a g-factor of 34, is disputed since it appears that the quasi-particle conductivity possesses the right sort of fine structure to account for the observations.

INTRODUCTION

The phenomenon of magnetic breakdown was discovered by Cohen & Falicov (1961) who invoked it to explain certain peculiarities of the de Haas–van Alphen effect in magnesium (Priestley 1963). Since then Stark (1962) has argued that the remarkable oscillatory magneto-resistance of zinc can be accounted for by breakdown, but he has not given a complete theory. The present work develops further a model devised to represent the behaviour of electrons subject to breakdown, to see how far it can account for the observations, particularly those of Stark and of Reed & Brennert (1963) on transport effects, and of Dhillon & Shoenberg (1955) on the de Haas–van Alphen effect.

The original model (Pippard 1962 a, referred to as I) treated a gas of nearly free electrons in a magnetic field, only one Fourier component of lattice potential being supposed to exist. It was argued that the wave function could be represented by a linear chain of free-electron orbits coupled by Bragg reflexions, but the validity of the model was not scrutinized very closely. It is worth going more deeply into this matter before attempting to analyse a two-dimensional network, such as more nearly represents the situation in magnesium and zinc, for it is easier to justify with a one-dimensional model the belief that the network of coupled orbits is a very close approximation to the true wave function. There is another reason for returning to the linear chain; as soon as one tackles the two-dimensional network, problems of gauge transformation arise which were sufficiently unobtrusive in the earlier work to be inadvertently ignored. None of the quoted results was incorrect, but the arguments leading to them need emendation. The way will then be clear to the extension of the model and a derivation of the energy level structure of the two-dimensional network. This will enable the de Haas–van Alphen effect to be discussed, but not the magneto-resistance, which requires a theory of the transport properties of the model. At this stage once more the linear chain helps to clarify the assumptions and make the results more plausible. Finally the theory will be compared with experimental observations, with a measure of success but far from perfect agreement.

Some of the ideas presented here are to be found embedded in far more sophisticated and general treatments of the problem of electrons in a lattice under the influence of a magnetic field (Harper 1955; Kohn 1959; Blount 1962). Indeed it is only just to point out that the conception of magnetic breakdown is clearly formulated in Harper's work, though not remarked on as implying any interesting experimental consequences. It has been our aim to eschew generality, and to concentrate a limited armoury of elementary mathematics on the reduction of a model simple enough to yield exact solutions. The complexity of the

solutions so obtained is in itself an apologia for the attempt, since by its very example it provides a convincing explanation for the far greater, truly appalling, complexity of the general treatments.

Gauge transformations

To represent an electron moving in a circular orbit it is convenient to choose for the vector potential the form $\mathbf{A} = \frac{1}{2}\mathbf{H} \times \mathbf{r}$. If the centre of the orbit coincides with the origin of $\mathbf{r}$ (gauge centre), the wave function may then be confined to the vicinity of the classical orbit and have a phase varying evenly round the orbit. For example, if $\mathbf{H}$ lies along the z axis and we consider only the variation of ψ in an (x, y)-plane, we may choose one of the solutions given by Dingle (1952)

$$\psi = e^{il\phi} r^l e^{-\frac{1}{4}\alpha r^2}, \tag{1}$$

in which (r, ϕ) are polar co-ordinates in the (x, y)-plane, l is the azimuthal quantum number and $\alpha = e\mathbf{H}/\hbar$ (note that e is the electronic charge, so that α is oppositely directed to $\mathbf{H}$; by α we mean the (positive) magnitude of α). The corresponding energy

$$E = (l + \tfrac{1}{2}) |e| \hbar H/m. \tag{2}$$

The maximum value of $\psi\psi^*$ occurs at a radius r_0 equal to $(2l/\alpha)^{\frac{1}{2}}$, which is as near as can be expected to the classical orbit radius $[(2l+1)/\alpha]^{\frac{1}{2}}$ for an electron whose energy is given by (2). Near the maximum $\psi\psi^*$ varies roughly as $e^{-\alpha(r-r_0)^2}$ when l is large, so that the radial width of the wave function is about $2/\sqrt{\alpha}$, or the relative width $(2/l)^{\frac{1}{2}}$. We shall be mainly concerned with orbits for which l is large, 10^3 or more, and the electron is confined to a circular track whose width is less than 4 % of its radius.

If (1) is a solution for the free electron in a uniform field, an equally good solution can be constructed representing an electron on an orbit centred otherwise than on the origin, say at $\mathbf{R}$ $(= (X, Y, 0))$. If the gauge centre were at $\mathbf{R}$, i.e. if $\mathbf{A} = \frac{1}{2}\mathbf{H} \times (\mathbf{r} - \mathbf{R})$, the wave function would have the form

$$\psi = e^{il\phi} |\mathbf{r} - \mathbf{R}|^l e^{-\frac{1}{4}\alpha |\mathbf{r} - \mathbf{R}|^2}, \tag{3}$$

and $\psi\psi^*$ would be confined to a narrow ring centred on $\mathbf{R}$, the phase of ψ varying uniformly round the ring. The phase variation may conveniently be described by a local wave vector $\mathbf{k}'$ whose value is easily seen to be $r_0^2[(\mathbf{r} - \mathbf{R}) \times \alpha]/2|\mathbf{r} - \mathbf{R}|^2$. In particular, round the classical orbit radius, where $|\mathbf{r} - \mathbf{R}| = r_0$,

$$\mathbf{k}' = \tfrac{1}{2}(\mathbf{r} - \mathbf{R}) \times \alpha = \tfrac{1}{2}\alpha(y - Y, X - x, 0). \tag{4}$$

Now in the network model which we shall develop we shall have occasion to change the gauge of $\mathbf{A}$ from time to time. For example, when discussing simultaneously a number of orbits centred on different points it is sometimes convenient to describe each separately by a wave function such as (3), which implies a different gauge centre for each, and sometimes to describe all in the same gauge. To change from one representation to the other we need to know what transformation ψ suffers when the gauge centre is moved. As is well known, the amplitude of ψ remains the same but the phase is altered; of course the absolute phase of ψ is meaningless, but the phase relationships between ψ at different points are significant, and it is the latter which are changed by a gauge transformation. If the electron is confined to a ring, as in (3), before the transformation, it is afterwards confined to the same ring, but the phase of ψ no longer varies uniformly round the ring. The rule for gauge

transformations is that if $\mathbf{A}$ is changed by the addition of $\nabla\beta$, ψ must be multiplied by $e^{ie\beta/\hbar}$.

Suppose then that at a point $\mathbf{r}$ the phase of ψ is ϵ when $\mathbf{R}$ is the centre of gauge. Then when the gauge centre is moved to the origin of co-ordinates, ϵ must suffer a change $\Delta\epsilon$ given by the expression $\Delta\epsilon = e\beta/\hbar$, where $\nabla\beta = \frac{1}{2}\mathbf{H}\times\mathbf{R}$, or $\beta = \frac{1}{2}(\mathbf{H}\times\mathbf{R}).(\mathbf{r}+\mathbf{c})$, $\mathbf{c}$ being an arbitrary constant. Hence

$$\Delta\epsilon = \frac{1}{2}(\boldsymbol{\alpha}\times\mathbf{R}).(\mathbf{r}+\mathbf{c}) = \frac{1}{2}\boldsymbol{\alpha}.[\mathbf{R}\times(\mathbf{r}+\mathbf{c})]. \tag{5}$$

More generally let us suppose the gauge centre to be moved from $\mathbf{R}_1$ to $\mathbf{R}_2$. By moving it first to the origin and then to $\mathbf{R}_2$ we see that (5) takes the form

$$\Delta\epsilon = \frac{1}{2}\boldsymbol{\alpha}.[(\mathbf{r}+\mathbf{c})\times\mathbf{S}], \quad \text{where} \quad \mathbf{S} = \mathbf{R}_2 - \mathbf{R}_1. \tag{6}$$

It will be observed from (6) that $\Delta\epsilon$ depends on the choice of origin and, what is the same thing, on an arbitrary constant $\mathbf{c}$. This is no more than another affirmation that the absolute phase is without significance, for when we consider how the phase relationship between two points is altered by the gauge transformation we find that the choice of origin is immaterial and $\mathbf{c}$ vanishes. Thus if ϵ_1 is the phase at $\mathbf{r}_1$, ϵ_2 at $\mathbf{r}_2$, it follows from (6) that

$$\Delta(\epsilon_2 - \epsilon_1) = \frac{1}{2}\boldsymbol{\alpha}.[(\mathbf{r}_2 - \mathbf{r}_1)\times\mathbf{S}]. \tag{7}$$

A particular consequence of (7), of use later, is that the phase difference between two wave functions at the same point ($\mathbf{r}_2 = \mathbf{r}_1$) is independent of the choice of gauge centre. We may also note at this point another expression for the gauge transformation, involving the local wave vector $\mathbf{k}'$, defined as $\nabla\epsilon$. When a gauge transformation is made by adding $\Delta\mathbf{A}$ ($=\nabla\beta$) to $\mathbf{A}$, $\mathbf{k}'$ changes by $\Delta\mathbf{k}'$, where

$$\Delta\mathbf{k}' = e\nabla\beta/\hbar = e\Delta\mathbf{A}/\hbar. \tag{8}$$

This result, from which (7) may be derived by integration, expresses the invariance of $\hbar\mathbf{k}' - e\mathbf{A}$, which determines the local density of kinetic energy and must be independent of choice of gauge.

It is clear from (7) that so long as we describe all parts of a wave function in the same gauge there is no ambiguity about the phase relationships. When we come to discuss the problem of an electron jumping from one orbit to another we shall find the phase relationships assuming great importance, and in any calculation it will be imperative to employ the same gauge in describing the wave functions of the two orbits. On the other hand, the periodicity (which is an important property of the wave functions) is easier to appreciate if each orbit is gauged with respect to its own centre, as in (3). Now in changing the representation from the former (common gauge) to the latter (mixed gauge) we carry out a different gauge transformation on each orbit, and according to (6) re-introduce the ambiguity associated with choice of origin and the integration constant $\mathbf{c}$. This ambiguity may, however, be resolved by adopting a definite convention; having chosen the origin of co-ordinates (and this choice is arbitrary) we stick to our choice, and we choose $\mathbf{c}$ to be zero, so that (6) takes the form

$$\Delta\epsilon = \frac{1}{2}\boldsymbol{\alpha}.(\mathbf{r}\times\mathbf{S}). \tag{9}$$

Within this convention we may shuttle to and fro between common and mixed gauges and never run into trouble provided we always return to common gauge for any calculation involving exact phases.

The foregoing argument is concerned with the type of gauge transformation which will be most useful in developing the network model, but there is another type helpful in discussing the validity of the model. When $\mathbf{A}$ is written as $\frac{1}{2}\mathbf{H}\times\mathbf{r}$, i.e. $\frac{1}{2}H(-y,x,0)$ for a field along z, the solutions of Schrödinger's equation for a free electron may be obtained readily in polar co-ordinates, and (1) is simply one of the complete set of orthogonal solutions in this system. As is well known the solutions have a high degree of degeneracy, so that many different representations of the same wave function are possible by linear combinations of degenerate solutions. Thus we may represent the ring-like solution in terms of the set of rather simple Cartesian functions obtained by writing $\mathbf{A}$ as $H(0,x,0)$, which yields Schrödinger's equation in the form

$$\nabla^2\psi+2\mathrm{i}\alpha x\,\partial\psi/\partial y-\alpha^2x^2\psi+2mE\psi/\hbar^2 = 0. \tag{10}$$

Solutions of (10) may be found of the form

$$\psi = \mathrm{e}^{\mathrm{i}k_y y}X(x-x_0), \tag{11}$$

$X(x-x_0)$ being a solution of the harmonic oscillator equation,

$$X'' - \{(\alpha x+k_y)^2 - 2mE/\hbar^2\}X = 0, \tag{12}$$

centred on the point $x_0 = -k_y/\alpha$. Suppose now we have a ring centred on the point $(0,Y)$ and $\mathbf{A} = \frac{1}{2}H(Y-y,x,0)$, so that the phase varies uniformly round the ring and the wave fronts are radial. If we now transfer to a gauge in which $\mathbf{A} = H(0,x,0)$, we have that

$$\Delta\mathbf{A} = \tfrac{1}{2}H(y-Y,x,0).$$

Hence, from (8), $\Delta\mathbf{k}' = \tfrac{1}{2}\alpha(y-Y,x,0),$

and from the original value of $\mathbf{k}'$ given by (4) it follows that in the new gauge the phase variation round the classical orbit is described by a wave vector $\alpha(y-Y,0,0)$; the wave fronts are no longer radial but at every point round the classical orbit are normal to the x axis. Moreover, the spacing of the wave fronts is just the same as in the oscillator function $X(x)$. In fact the ring orbit is very much like an annulus sliced out of the particular function of the set (11) which has $k_y = 0$. Clearly the ring can be built up by Fourier synthesis of the functions (11) with a range of k_y centred on zero; the wave function can indeed be cast in the form

$$\int \exp\left\{-\frac{k_y^2}{2\alpha}+\mathrm{i}k_y y\right\}X(x+k_y/\alpha)\,\mathrm{d}k_y,$$

to show that the spectrum has a half-width $(2\alpha)^{\frac{1}{2}}$. According to (11), therefore, the functions involved have x variations which are oscillator functions centred in a range along the x axis of roughly $\pm(2/\alpha)^{\frac{1}{2}}$. This is in fact just about the radial width of the ring-like wave function and is much less than the orbit radius. The narrow range of oscillator centres is important for the next stage of the argument.

THE LINEAR CHAIN

Physical basis of the model

In I the two-dimensional motion of an electron, in a periodic potential $V_g\cos gx$ and a magnetic field normal to the plane of motion, was represented by a chain of linked circular orbits. This expressed the intuitive conception that if the lattice potential was weak the

electron could spend most of its time in a hardly perturbed free-electron orbit, but at certain points in the orbit the wave number was such that Bragg reflexion could switch it to a similar, but displaced, orbit. It is only for a short segment of the orbit that the Bragg condition is satisfied, and if V_g is small there is a good chance that the electron will penetrate the reflexion region and remain on its original orbit. This is the phenomenon of magnetic breakdown. When only one lattice component $V_g \cos gx$ is of importance the orbits which are accessible from a given initial orbit are strictly limited and form a linear chain in the y direction of evenly spaced intersecting orbits. Consider two such orbits centred at $(0, 0)$ and $(0, Y)$, and intersecting at $(x, \frac{1}{2}Y)$. If we choose the centre of each orbit to be its own gauge centre we see from (4) that the orbit centred on the origin has wave number $\alpha(\frac{1}{4}Y, -\frac{1}{2}x)$ at the intersection and the other $\alpha(-\frac{1}{4}Y, -\frac{1}{2}x)$. Now to write the condition for Bragg reflexion we must derive the local wave number in both orbits with respect to the same gauge centre, and when we transfer the gauge centre of the second orbit to the origin, making use of (8), its wave number at the intersection becomes $\alpha(-\frac{3}{4}Y, -\frac{1}{2}x)$. Thus a switch from the first to the second orbit involves addition of a wave vector $\alpha(-Y, 0)$, such as is contributed by the component e^{-igx} of the lattice if $g = \alpha Y$. Hence the linked orbits are evenly spaced at a separation of g/α. A more general statement, of use later, is that a switch of orbit centres through a vector $\mathbf{S}$ is accomplished by virtue of a lattice component of wave number $\mathbf{\dot{g}}$, where

$$\mathbf{\dot{g}} = \mathbf{\alpha} \times \mathbf{S}. \tag{13}$$

Let us now reconsider the problem raised in the last section, of a Fourier synthesis of a single free electron orbit in terms of the functions (11), but instead let us synthesize a chain of independent free electron orbits centred at $(0, nY)$, n being any integer. We shall impose phases on the individual orbits such that when each is gauged with respect to its own centre there is a steady phase variation as represented by a wave vector κ_y, i.e. the phase difference between corresponding points on neighbours is $\kappa_y Y$. On transforming everything into the same gauge, $\mathbf{A} = H(0, x, 0)$, all corresponding points suffer the same phase change, so that the pattern still has a strict periodicity described by κ_y, and must be synthesized from a discrete spectrum chosen from (11), only those terms being used for which

$$k_y = \kappa_y + 2n\pi/Y. \tag{14}$$

The oscillator functions are centred on $-k_y/\alpha$, so that their centres form an evenly spaced set of points on the x axis with spacing $2\pi/(\alpha Y)$. Now we have seen that if the orbit spacing Y were determined by a lattice component of wave number g, we should have $\alpha Y = g$, and the oscillator centres would be spaced $2\pi/g$ apart, which is just one lattice distance, a. This neat result enables us to proceed to discuss in a simple manner the situation when a real lattice is present.

So long as the lattice potential V is a function of x alone, (10) may be supplemented by the addition of $-V\psi$, and separation of variables accomplished as in (11), but with X obeying a rather more complicated equation than (12):

$$X'' - \{(\alpha x + k_y)^2 - (2m/\hbar^2)(E - V_g \cos gx)\} X = 0. \tag{15}$$

If $V_g \ll E$ the general character of X is clear. There are three ranges of x separated by the points x_1 and x_2 at which Bragg reflexion can occur. Except in the vicinity of these points the lattice perturbs the equation only to the extent that Bloch waves, in the absence of

a magnetic field, are affected by the lattice when their wave number is not close to a Brillouin zone boundary; that is, if V_g is very small there are three sections, $-\infty < x < x_1$, $x_1 < x < x_2$ and $x_2 < x < \infty$, in which the solutions are free-electron like. But there are narrow regions around x_1 and x_2 where the lattice must be explicitly taken into account, and it is the solution in these regions which supplies the relations between the amplitudes and phases in the three sections. There is no need in the present context to find these relations, nor to determine the eigenvalues of (15), which are of course fixed by the vanishing of X at $\pm\infty$. If we now attempt to synthesize a linear chain of ring-like orbits out of these solutions we note that to obtain the required spacing Y we need only solutions of (15) which are identical in energy and in the form of X, except for being shifted by integral numbers of lattice spacings. If we combine these solutions in the same proportion and phases as we did the free-electron solutions, we shall construct the required chain having the same narrow tracks for the wave function almost everywhere. The exceptions are the regions round x_1 and x_2 where the tracks overlap, and we can say nothing about them without detailed solutions of (15). Of course the amplitude is now no longer constant all round each orbit, but the orbit is divided into four sections in each of which it is constant, corresponding to whether x is less than x_1, greater than x_2, or between x_1 and x_2, the last applying to two sections on each orbit. We noted in connexion with the free electron that the range of oscillator centres was little more than the radial width of the track. Thus the confused regions in the synthesis, in which functions taken from different regions overlap, is also little wider than the track itself, and confined to the regions in space where two tracks are superimposed.

This argument shows how the network model may be regarded as a good approximation to one possible representation of the wave functions. We imagine the wave function confined to its track and moving from one junction to another as a free-electron wave. All that we ask of (15) is that it provide the connexion formulae to determine the amplitudes and phases of the waves leaving a junction. In fact the amplitudes have been in effect calculated by Blount (1962), as discussed in I, the formula being quoted in (11) of I. We shall continue to use the concepts of I, that a wave of unit amplitude arriving at a junction continues past the junction along the same orbit with amplitude p, and suffers partial Bragg reflexion into another orbit with amplitude q. We defer further discussion of the magnitudes and phases of p and q.

The foregoing justification of the network model has relied on V_g being small, so that the energy contours (when $\mathbf{H} = 0$) in k-space are circular except near the zone boundaries, and the wave functions free-electron-like nearly everywhere. If V_g is not so small, so that the energy contours are distorted, the argument is harder and will not be attempted. But we may hazard the guess that a representation is still possible in which the electrons are confined to narrow orbits. The solutions of (15) will not be free-electron-like anywhere and we cannot be sure, as above, that they will now combine so as to limit the width of the wave function; it seems not unlikely however that such is what will happen, with the orbit no longer circular but of the same form as the energy contour and turned through $\frac{1}{2}\pi$, as in Onsager's (1952) semi-classical analysis. We shall not take the argument to this point in what follows; there is enough to discuss in the almost-free case for which the model has so much sounder a basis.

Transport properties of the linear chain

Equation (11) for the free electron has even symmetry in $x - k_y/\alpha$, and the solutions also have such symmetry as precludes any overall transport of charge by any stationary state, as might be expected for electrons moving in independent closed orbits. The same is not so for (15), which is even only when $gk_y = n\pi\alpha$; the solutions under these conditions are centred either on a lattice point or midway between two, and from (14) it may be seen that successive orbits in the chain are either in phase or in antiphase. These solutions have equal amplitudes on both sides of the chain and carry no resultant current, but all other solutions have unequal amplitudes on the two sides, so that more current runs up one side than down the other. This point was noted in I but not discussed. From the amplitudes evaluated in I it is straightforward, if tedious, to compute the current which may be conveniently expressed in terms of the effective velocity v_y of an electron along the chain. The result of this computation may be stated most simply in the form well known to apply to the Bloch functions describing electrons in the absence of a field

$$v_y = \hbar^{-1} \mathrm{d}E/\mathrm{d}\kappa_y. \tag{16}$$

This result, first discovered in this context by Harper (1955), may be proved directly from the structure of the equations but we shall defer the proof until later, when it is convenient to discuss the problem more generally.

The simplicity of (16) encourages us to treat the electron on the linear chain as a quasi-particle moving in one dimension and having an energy level structure determined by the solutions of (15), in other words a series of bands as in figure 9 of I. In this way we eliminate the magnetic field from further participation, but it is important to discover whether the real system and the quasi-particle respond to an electric field in the same way. Let us first consider the response in the representation which leads to (15). If $\mathbf{A} = H(0, x, 0)$ the time-dependent Schrödinger equation takes the form, when an electric field $\mathscr{E}$ is applied,

$$\nabla^2\Psi + 2\mathrm{i}\alpha x\,\partial\Psi/\partial y - \alpha^2 x^2\Psi + 2m/\hbar^2\{e(\mathscr{E}_\kappa x + \mathscr{E}_y y)\Psi - V\Psi + \mathrm{i}\hbar\,\partial\Psi/\partial t\} = 0, \tag{17}$$

and solutions can be found of the type $\Psi = \mathrm{e}^{\mathrm{i}k_y y} F(x, t)$, in which

$$\dot{k}_y = e\mathscr{E}_y/\hbar, \tag{18}$$

and

$$\partial^2 F/\partial x^2 + 2\mathrm{i}m/\hbar\,\partial F/\partial t - \{(\alpha x + k_y)^2 + 2m(e\mathscr{E}_\kappa x - v)/\hbar^2\}F = 0. \tag{19}$$

If then we start with the system in a stationary state defined by E and k_y, and apply a weak electric field for a time t, removing it before the energy has changed by anything comparable with the band width, then according to (18) k_y changes in a well-defined way, while we know that the energy supplied by the source of the field is $e\mathscr{E}_y v_y t$, so that E also changes in a well-defined way,

$$\dot{E} = e\mathscr{E}_y v_y. \tag{20}$$

Equations (18) and (20) are consistent with (16) and it is clear from this that in response to an electric field the solutions of the type (11), $X(x)\,\mathrm{e}^{\mathrm{i}k_y y}$, behave like Bloch functions, the solutions remaining pure provided the field is weak enough.

The requirement that the field be weak is analogous to the condition that must be applied when the same argument is used with Bloch functions. Strictly (20) gives the rate of change of the expectation value of E, and any state with the correct k_y may in principle

be produced. However, these states of given k_y form a discrete set, one to each band, and if $\mathscr{E}$ is so weak as to produce only a small change of state in a time such that $\hbar/\tau$ is much smaller than the energy separation of the states, the probability of interband transitions is negligible. That is to say, (18) describes correctly the linear response of the system to small fields. We shall revert to the interband transitions presently, since their role in the transport properties is not to be neglected.

To complete the argument we must consider the response of the linear chain to a weak electric field. We have seen how this can be synthesized from functions like (10), and it is readily proved that all solutions of the same E but different k_y (i.e. shifted relative to one another by integral numbers of lattice spacings) carry the same current. Moreover, these solutions are orthogonal and a linear combination also carries the same current.† From this it is readily seen that, since each component in the synthesis of the linear chain remains pure and evolves according to (18), so too the linear chain retains its form and evolves in accordance with the equations applicable to a quasi-particle of specified $E(\kappa)$

$$\hbar\dot{\kappa} = e\mathscr{E}; \quad \dot{E} = e\mathscr{E}.\mathbf{v}; \quad \hbar\mathbf{v} = \nabla_\kappa E. \tag{21}$$

In this case $E = E(\kappa_y)$ and all motion is confined to the y direction.

Having said this we must recollect that a free electron in crossed $\mathscr{E}$ and $\mathscr{H}$ drifts normal to both at a speed $\mathscr{E}/\mathscr{H}$, and this effect, which is the consequence of interband transitions, is not contained in (21), which describes only the accelerative effects of the field. This may be adequate if the relaxation time is long, but if it is short the sharp distinction between interband and intraband contributions to the current is not helpful. To illustrate this point we consider the response of the system to a short impulsive field $\mathscr{E}\delta t$ applied at time $t = 0$. By integrating the time-dependent Schrödinger equation it may readily be shown that if an electron is in a state ψ before the impulse, immediately afterwards it is described by $\psi e^{i\mathbf{q}.\mathbf{r}}$, where $\mathbf{q} = e\mathscr{E}\delta t/\hbar$, and the current is changed by $e^2\mathscr{E}\delta t/m$. Subsequently the current so established varies in a more or less complex fashion, the time integral of the current giving the response of the system to a steady field (Pippard 1962b). The same argument may be applied to an assembly of electrons; initially if they are in equilibrium there is no current, and any current established by the impulse ultimately decays through collisions. Once again, the time integral of the current is the response to a steady field.

Consider then an assembly of N electrons in equilibrium in states of the type (11), X being solutions of (15). Any one electron, initially described by $X_{k_y}e^{ik_y y}$, is described after an impulse $\mathscr{E}_y\delta t$ by $X_{k_y}e^{i(k_y+q_y)y}$. This is not a solution of Schrödinger's equation, but can be synthesized from the set of solutions having wave number (k_y+q_y), of which one comes from the same band as the initial wave function and has amplitude unity (to first order in q_y), while the others come from other bands and have amplitudes proportional to q_y. The contribution of the term from the same band is the intraband current which may be derived from (21). This does not normally account for the total initial current $Ne^2\mathscr{E}_y\delta t/m$, the difference being the interband current contributed by the other terms. Since the Fourier components have different energies they beat among themselves, and if

† This is not true for all orthogonal functions carrying the same current. For example $\cos kx$ and $\sin kx$ carry zero current, but the linear combination e^{ikx} does not. I am obliged to Dr J. M. Ziman for this simple example. In (10) it is the term $e^{ik_y y}$ that ensures the result quoted.

the relaxation time is long compared with all beat periods the integrated interband contribution is small compared with the intraband contribution. To give a simplified example, in the absence of collisions the intraband current might lie along y and have magnitude $a_1 \mathscr{E}_y \delta t$, while the interband current might consist of a single rotating current (as is produced by an impulse on a free-electron assembly) having components $a_2 \mathscr{E}_y \delta t \cos \omega_c t$ along y and $a_2 \mathscr{E}_y \delta t \sin \omega_c t$ along x. The initial current $(a_1 + a_2) \mathscr{E}_y \delta t$ is equal to $Ne^2 \mathscr{E}_y \delta t/m$. If we suppose the currents to decay with relaxation time τ we have for the response to the impulse

$$\delta J_y = (a_1 + a_2 \cos \omega_c t) \mathscr{E}_y \delta t \, \mathrm{e}^{-t/\tau},$$
$$\delta J_x = a_2 \sin \omega_c t \, \mathscr{E}_y \delta t \, \mathrm{e}^{-t/\tau}.$$

Thus integrating over all positive time we have for the steady response

$$J_y = a_1 \tau \mathscr{E}_y + a_2 \tau \mathscr{E}_y (1 + \omega_c^2 \tau^2)^{-1}, \tag{22}$$
$$J_x = a_2 \omega_c \tau^2 \mathscr{E}_y (1 + \omega^2 \tau^2)^{-1}. \tag{23}$$

The second (interband) term in (22) is smaller than the first by a factor which varies as $(1 + \omega_c^2 \tau^2)^{-1}$, so that when collisions are infrequent it may be neglected in the parallel component of the current. But the transverse component, due entirely to interband transitions, tends to a constant, $a_2 \mathscr{E}_y/\omega_c$, as $\tau \to \infty$. For free electrons $a_1 = 0$ and $a_2 = Ne^2/m$, while $\omega_c = eH/m$, so that $J_x \to Nev_x$, where v_x is the drift velocity $\mathscr{E}_y/H$. It should be particularly noted in (22) that if $\omega_c \tau \ll 1$, $J_y = (a_1 + a_2) \tau = Ne^2 \tau/m$, i.e. interband and intraband currents must be treated as equals. This is obvious from the derivation; if the current established by the impulse decays before it has time to oscillate, the integral is determined by the initial current and the relaxation time, and the dissection into separate components is without value. We shall have occasion to refer back to this argument later.

Two-dimensional network

Structure and general properties

We now extend the model to a two-dimensional metal in which more than one component of lattice potential is capable of causing a switch from one orbit to another, so that the whole plane is covered with linked orbits. We assume the orbits to be free-electron-like except at a switching point where two overlap. Let the ionic lattice (I-lattice) have basis vectors $\mathbf{a}$ and $\mathbf{b}$, with reciprocal lattice vectors $\dot{\mathbf{g}}_a$ and $\dot{\mathbf{g}}_b$:

$$\dot{\mathbf{g}}_a = \frac{2\pi \boldsymbol{\alpha} \times \mathbf{a}}{\boldsymbol{\alpha}.(\mathbf{a} \times \mathbf{b})}, \quad \dot{\mathbf{g}}_b = \frac{2\pi \boldsymbol{\alpha} \times \mathbf{b}}{\boldsymbol{\alpha}.(\mathbf{a} \times \mathbf{b})}, \tag{24}$$

$\boldsymbol{\alpha}$ being employed here simply as a vector normal to the plane. Associated with each reciprocal lattice vector by (13) is a switching vector $\mathbf{S}$ which is the shift of orbit centre induced by Bragg reflexion from the lattice component concerned. Comparison of (13) and (24) shows that there are two basic switching vectors,

$$\mathbf{S}_a = \frac{2\pi \mathbf{a}}{\boldsymbol{\alpha}.(\mathbf{a} \times \mathbf{b})}, \quad \mathbf{S}_b = \frac{2\pi \mathbf{b}}{\boldsymbol{\alpha}.(\mathbf{a} \times \mathbf{b})}, \tag{25}$$

i.e.
$$\mathbf{S}_a = s\mathbf{a}, \quad \mathbf{S}_b = s\mathbf{b},$$

where
$$s = 2\pi/(\alpha \Sigma), \tag{26}$$

Σ being the area of unit cell of I-lattice, $|\mathbf{a} \times \mathbf{b}|$. Thus the centres of the orbits which are coupled by the lattice potential form a lattice (O-lattice) similar to the I-lattice but s times as big, as shown in figure 1. It should be noted that s is determined by the I-lattice and the magnetic field, but is independent of the energy of the electron. It must also be remarked that if the network in figure 1 is a good representation of a possible wave function, the whole diagram may be shifted laterally to give an equally good representation. These different degenerate functions are not in general mutually orthogonal, for which reason care must be taken in using them in perturbation calculations. Nor is it easy to prove that the number of independent solutions has the right value. We shall not attempt a rigorous development of this part of the argument, but merely assume that for the accountancy of states there are s independent network systems, a result derived as follows. The degeneracy of the Landau levels of a free electron gas is $\alpha/2\pi$ for unit area of the metal; this may be taken as the number of independent free electron orbits. Since the O-lattice has a cell of area $s^2\Sigma$, there are $(s^2\Sigma)^{-1}$ orbits per unit area linked into a single network. Thus to include all independent orbits in the networks the number of independent networks must be $\alpha s^2\Sigma/2\pi$, which from (26) is seen to be s.
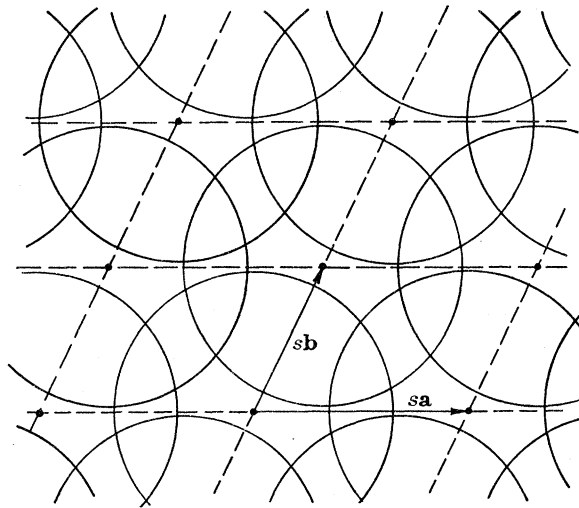

FIGURE 1. Two-dimensional network with basis vectors $s\mathbf{a}$ and $s\mathbf{b}$.

For certain values of the magnetic field strength s is integral, and all the coupled orbits have their centres identically situated on the I-lattice. If the lattice potential were strong enough to influence the electrons all round their orbits it would obviously be a great simplification of the problem to choose the special field strengths which make all orbits identical. What is not so obvious is that even in the case we treat here, where the lattice is weak enough only to be of importance at switching points, the exact value of s is still a matter of serious concern. This is because the vector potential imposes a phase variation on the wave function of a periodicity which in general is not commensurable with the periodicity of the network. The way in which the value of s enters will become clearer when we have discussed the phase changes that occur at a switch.

Consider two neighbouring orbits Ω_1 and Ω_2, as shown in figure 2, and let P be a point at which Bragg reflexion causes a switch from Ω_1 to Ω_2. If we take the same gauge centre for both, the phase change on switching is determined by the phase at P of the lattice

component responsible for the switch. To say this is to neglect an unknown phase change which can only be discovered by detailed analysis, but as it takes the same value at all corresponding switches we neglect it for the moment. Since from (13) the lattice component has wave number $\boldsymbol{\alpha} \times \mathbf{S}$, the phase at $\mathbf{r}$ is $(\boldsymbol{\alpha} \times \mathbf{S}) . \mathbf{r}$ and we may write for the phase change on switching

$$\epsilon_2 - \epsilon_1 = (\boldsymbol{\alpha} \times \mathbf{S}) . \mathbf{r} = \boldsymbol{\alpha} . (\mathbf{S} \times \mathbf{r}). \tag{27}$$

Now we have seen earlier that if all orbits are described with respect to the same gauge centre they have different phase patterns and the periodicity implied by figure 1 is obscured.
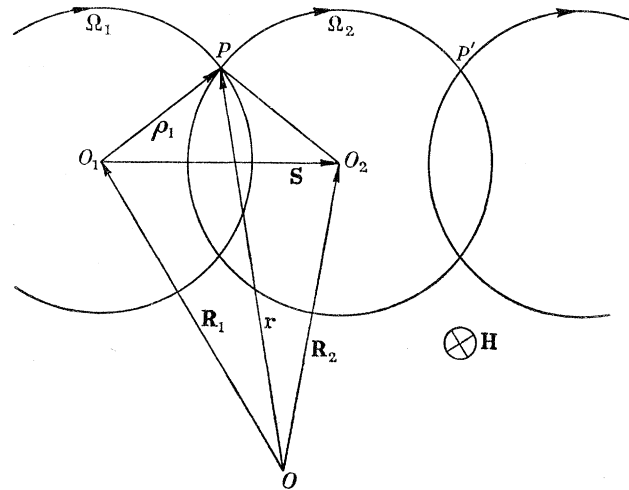

FIGURE 2. Switching from orbit Ω_1 to orbit Ω_2.

To regain the periodicity we gauge each orbit with respect to its own centre, so that around each there is a uniform phase variation. The price we pay for this enforced periodicity is that (27) must be changed by a gauge transformation. We have already noted as a consequence of (7) that (27) is independent of the choice of gauge centre, provided the same is used for both orbits. Let us therefore start with both gauged from O_1 the centre of Ω_1, and shift the gauge centre of Ω_2 to O_2, when, according to the convention incorporated in (9),

$$\Delta\epsilon_2 = -\tfrac{1}{2}\boldsymbol{\alpha} . (\mathbf{S} \times \mathbf{r}). \tag{28}$$

Thus when each orbit has its own gauge centre the phase change on switching, δ, is equal to $\epsilon_2 + \Delta\epsilon_2 - \epsilon_1$, or from (27) and (28)

$$\delta = \tfrac{1}{2}\boldsymbol{\alpha} . (\mathbf{S} \times \mathbf{r}). \tag{29}$$

The geometrical interpretation of this result is shown in figure 2, a typical switch for a negative particle whose orbit runs clockwise about $\mathbf{H}$; $\tfrac{1}{2}\mathbf{S} \times \mathbf{r}$ is directed outwards normal to the diagram and has magnitude equal to the area of the quadrilateral OO_1PO_2. Since for a negative particle $\boldsymbol{\alpha}$ is also directed outwards, δ is simply $|\boldsymbol{\alpha}|$ times the area OO_1PO_2, and is positive if the letters have the same sense of rotation as the particle in its orbit. It will be noted that by expressing $\mathbf{r}$ as $\mathbf{R}_1 + \boldsymbol{\rho}_1$, the phase change can be dissected into an external part δ_0, equal to $\tfrac{1}{2}(\mathbf{S} \times \mathbf{R}_1) . \boldsymbol{\alpha}$ and represented by the triangle OO_1O_2, and an internal part δ_i, equal to $\tfrac{1}{2}(\mathbf{S} \times \boldsymbol{\rho}_1) . \boldsymbol{\alpha}$ and represented by the triangle O_1PO_2. The external part depends only on the co-ordinates of the orbit centres involved, the internal only on the position of the switch point relative to the orbit centres. This dissection will be of value when we come to discuss the periodicity of the wave function.

Let us first, however, observe how these phase changes play a part in determining the quantization of orbits when switching is either absent or complete. In the former case the electron travels in a circular orbit, and the variation of phase along its path, when the centre of gauge is the orbit centre, is uniform and equal to α times the area swept out by the radius vector of the electron. The quantum condition is that the area shall be $2l\pi/\alpha$. In the latter case, when switching occurs at each cross-over point (29) enables us to represent the phase change around the orbit by contributions from sectors of free-electron orbits supplemented by quadrilaterals at each switch. Consider then an orbit in which the free-electron orbits have successive centres $O_1 O_2 \ldots O_n O_1$. If we dissect each quadrilateral into two triangles the sum of the phase changes due to external parts is represented by the area of the polygon $O_1 O_2 \ldots O_n O_1$, taken as positive if the letters have the same sense as the free-electron orbit; the internal parts give a contribution which depends on the precise form of the orbits. Three examples are shown in figure 3. Figure 3(a) shows a lens-shaped orbit $ABCDA$, in
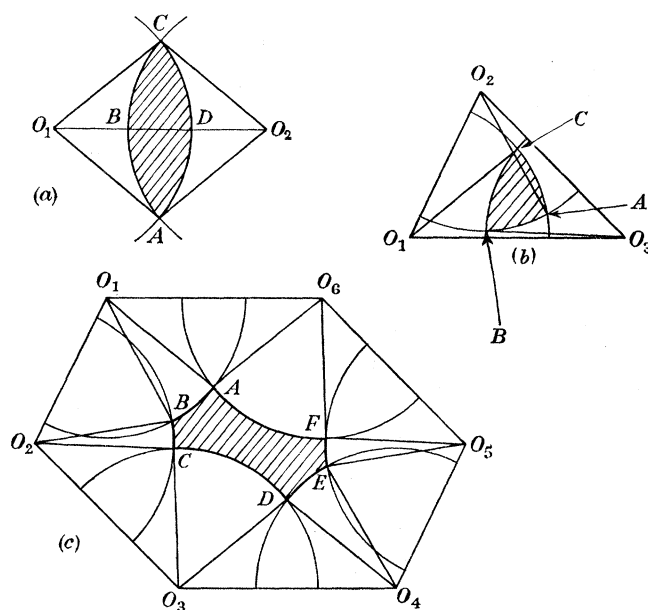

FIGURE 3. Relation between the area of an orbit and the phase changes
taking place round it.

which the motion of the electron gives contributions to the phase change represented by the sectors $O_1 CDA$ and $O_2 ABC$, both being positive, and the internal parts of the switches at A and C give triangles $O_1 AO_2$ and $O_2 CO_1$, both being negative. There is no external contribution, and the total phase change is represented by the area of the lens $ABCD$. Figure 3(b) shows a triangular orbit; the orbit from C to A contributes a sector $O_1 CA$, and the internal part of the switch at A a (negative) triangle $O_1 AO_2$, the two together yielding the negative curvilinear area $O_1 CAO_2$. The whole orbit then gives the area of the orbit ABC minus the area of the triangle $O_1 O_2 O_3$, but to this must be added the external contribution of the switches, which is the (positive) area $O_1 O_2 O_3$. The final result then is that the phase change round the orbit is represented by the area of the orbit. Finally 3(c) shows a hexagonal hole orbit, for which the sectors and triangles add to give the unshaded area

$$O_6 O_5 O_4 O_3 O_2 O_1 + ABCDEF,$$

while the external part is the negative area $O_1O_2O_3O_4O_5O_6$. The resultant is the negative orbit area *ABCDEF*. Thus the model yields the Onsager rule for quantizing orbits by their area, and distinguishes electron and hole orbits by sign.

We are now in a position to discuss intermediate cases, where, as in I, there is a certain probability of switching at each junction, and the wave functions are not confined to single orbits, but spread in the form of a network over all space, as in figure 1. We shall describe the wave function by its amplitude and phase at each point, always taking the local orbit centre as centre of gauge when denoting the phase. Thus at each switch we must supply a phase shift δ as given by (29). The notation used is shown in figure 4, where p and q

$$A \qquad\qquad pB + qAe^{i\delta}$$
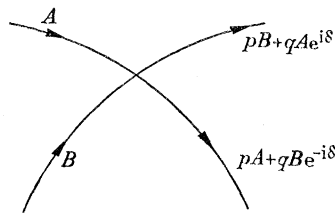$$B \qquad\qquad pA + qBe^{-i\delta}$$

FIGURE 4. Amplitude and phase relationships at a junction.

represent the amplitudes of unswitched and switched waves. It was argued in I, and the argument is not modified here by the introduction of δ, that $pp^* + qq^* = 1$ and that p and q are in phase quadrature. This result is however unjustifiably restrictive, as Dr E. I. Blount has kindly pointed out. If the coefficients p and q describe the phase relationships between the waves on different branches at the actual point of intersection, it is certainly not necessary that q be real and p imaginary (as was assumed for convenience). On the other hand, it is possible to find a reference point on each branch, near the intersection, such that the phase relationships between the waves at these points are correctly described by real q and imaginary p. We shall continue to use this convention, with $q^2 - p^2 = 1$, in what follows, since we have no detailed calculation to provide the actual phases of the switched waves. Thus the phase length of each branch of the network, as inserted into the calculations, is to be thought of as measured between these new reference points rather than between the intersections. For any given value of q the calculated energy level structure, such as that shown in figure 8, is correct in form but must be regarded as having possibly suffered a vertical shift through an incorrect assumption about the relative phase length of the different branches. Since the vertical shift may well differ for different values of q, the whole diagram of energy levels is liable to be sheared vertically, perhaps to a considerable extent. Nevertheless, the shear must be a smooth, if not monotonic, function of q, so that the topological form of the diagram will not suffer, and this is the aspect of the problem which is of greatest interest.

We may now ascribe amplitudes to all the separate paths in the network, and relate them to one another by writing down compatibility equations for each junction. In I we treated a one-dimensional network and found periodic solutions of the Bloch type, corresponding to wave motion down the network. One might expect to be able to do the same in two dimensions, but this is not possible in general, on account of the variation of δ from one cell in the network to another. The internal part of δ offers no difficulty—it is the same for switches at corresponding junctions; but the external part varies, since for a given switching

vector $\mathbf{S}$ its value depends on the position $\mathbf{R_1}$ of the orbit centre. The simplest periodicity is achieved by choosing the value of $\mathbf{H}$ so that the phase change corresponding to any polygon of the O-lattice is an integral multiple of 2π; for this we must make the triangle which is half the unit cell in the O-lattice represent an integral multiple of 2π. Since the triangle has area $\frac{1}{2}s^2\Sigma$, $\mathbf{H}$ must be such that

$$\tfrac{1}{2}s^2\Sigma\alpha = 2n\pi,$$

i.e. from (26), $s = 2n.$†   (30)

When $\mathbf{H}$ is such that (30) is satisfied, all unit cells of the O-lattice are identical, the orbits being identically situated on the I-lattice, and the phase changes at corresponding switches being the same except for integral multiples of 2π. Under these special conditions the wave function has the periodicity of the diagram in figure 1. If we choose $\mathbf{H}$ instead so that s is an odd integer, corresponding switches do not all have the same value of δ, but the problem is not much more complex in this case since, as can readily be shown, the behaviour is strictly periodic when a cell twice as big in linear dimensions is chosen as the basis. If, however, to go to an extreme, s is irrational, every corresponding switch has a different value of δ and there is no unit cell, however large, which can be taken as the basis for a periodic solution.

We shall return to this matter later, but meanwhile we discuss in some detail the behaviour when (30) is satisfied and periodic solutions can be found with only one orbit per unit cell. These solutions are similar to Bloch functions in that the wave functions on different orbits differ only in phase (provided each is gauged with respect to its own centre) and the phase changes can be characterized by a wave vector $\boldsymbol{\kappa}$. Thus two orbits whose centres are $\mathbf{S}$ apart differ in phase at corresponding points by $\boldsymbol{\kappa}.\mathbf{S}$. For a given choice of orbit centres $\boldsymbol{\kappa}$ determines the energy of the state, since only for a particular energy can the phase relationships at all junction points be satisfied. But there are similar networks differing from one another by uniform displacement, and the same relation does not hold for all, as we now prove.

We saw when analyzing the switch displayed in figure 2 that at P the phase change δ is given by (29). Now when O_2 is taken as the centre of gauge, the phase along the second orbit varies linearly, and ϕ, the phase difference between P and P', is proportional to the angle PO_2P', and independent of the magnitude of $\mathbf{R_2}$. Thus if P' and P are corresponding points, they differ in phase by $\delta+\phi$, when each point is gauged with respect to the centre of its own orbit. It is for this system of gauging that $\boldsymbol{\kappa}$ is defined, and

$$\boldsymbol{\kappa}.\mathbf{S} = \delta+\phi = \tfrac{1}{2}(\mathbf{r}\times\boldsymbol{\alpha}).\mathbf{S}+\phi.$$

If (30) is satisfied, we may neglect any part of $\mathbf{r}$ which is a vector of the O-lattice, and write $\mathbf{r}$ as $\mathbf{r_0}+\boldsymbol{\rho}_1$, $\mathbf{r_0}$ being the distance of the nearest orbit centre from the origin. It then follows that

$$\boldsymbol{\chi}.\mathbf{S} = \tfrac{1}{2}(\boldsymbol{\rho}_1\times\boldsymbol{\alpha}).\mathbf{S}+\phi,$$ (31)

where $$\boldsymbol{\chi} = \boldsymbol{\kappa}+\tfrac{1}{2}\boldsymbol{\alpha}\times\mathbf{r_0}.$$ (32)

† Since the area of unit cell of the O-lattice is $s^2\Sigma$, the flux contained in unit cell is $s^2\Sigma H$, or an even multiple of the 'flux quantum' h/e. Alternatively, the special values of H are integral submultiples of that field which produces one flux quantum in each unit triangle of the I-lattice.

The right-hand side of (31) is determined by the energy of the electron, since ρ_1 and ϕ are functions of energy and independent of the location of the orbits. Thus it is χ rather than κ that is determined by the energy, $E = E(\chi)$, so that different orbit systems may have different wave numbers but the same energy if they are differently located.

The difference in wave number for differently located degenerate states is however something of an artifact, arising from the choice of a different O-lattice as the lattice of gauge centres for each network. Let us instead arbitrarily select a standard O-lattice, e.g. that whose nearest point to the origin is $\mathbf{r}_1$, and refer a differently located orbit network to this system. For example, consider a point $P_1(\mathbf{r})$ on an orbit network defined by $\mathbf{r}_0$, the centre of the orbit containing P_1 being at $\mathbf{r}_0 + \mathbf{R}$, where $\mathbf{R}$ is an O-lattice vector; and another point $P_2(\mathbf{r} + \mathbf{S})$ on a neighbouring orbit whose centre is at $\mathbf{r}_0 + \mathbf{R} + \mathbf{S}$. If $\mathbf{r}_0 + \mathbf{R}$ and $\mathbf{r}_0 + \mathbf{R} + \mathbf{S}$ are the gauge centres, as in our previous treatment, the phase difference $\epsilon_2 - \epsilon_1$ is $\kappa.\mathbf{S}$. Now we shift the gauge centre for P_1 from $\mathbf{r}_0 + \mathbf{R}$ to $\mathbf{r}_1 + \mathbf{R}$, and for P_2 from $\mathbf{r}_0 + \mathbf{R} + \mathbf{S}$ to $\mathbf{r}_1 + \mathbf{R} + \mathbf{S}$, using the rule expressed by (9). Then

$$\epsilon_1 \to \epsilon_1 - \tfrac{1}{2}\boldsymbol{\alpha}.[\mathbf{r} \times (\mathbf{r}_0 - \mathbf{r}_1)]$$

and

$$\epsilon_2 \to \epsilon_2 - \tfrac{1}{2}\boldsymbol{\alpha}.[(\mathbf{r} + \mathbf{S}) \times (\mathbf{r}_0 - \mathbf{r}_1)],$$

so that

$$\epsilon_2 - \epsilon_1 \to \kappa.\mathbf{S} - \tfrac{1}{2}\boldsymbol{\alpha}.[\mathbf{S} \times (\mathbf{r}_0 - \mathbf{r}_1)],$$

i.e.

$$(\chi - \tfrac{1}{2}\boldsymbol{\alpha} \times \mathbf{r}_1).\mathbf{S} \quad \text{(from (32))}. \tag{33}$$

This result shows that if we dissect the plane into identical cells, one to each O-lattice point, and choose a standard point in each cell as gauge centre for all points in the cell, then so long as (30) is obeyed each network solution is strictly periodic in this representation. Moreover, from (33) we see that since the effective wave number, $\chi - \tfrac{1}{2}\boldsymbol{\alpha} \times \mathbf{r}_1$, does not involve $\mathbf{r}_0$, it is the same for all degenerate solutions, though the value depends on the point chosen as gauge centre in each cell. Thus we may construct linear superpositions of degenerate solutions which perhaps no longer exhibit an obvious network structure, yet are also strictly periodic.

To return to our original gauge convention and its consequence expressed by (32), it will be seen that of all the degenerate orbit networks only that one whose O-lattice contains the origin, so that $\mathbf{r}_0 = 0$, has χ and κ equal. For this case, subject to (30), the external part of δ vanishes for all switches, and this makes it the most elegant to analyse in detail. But obviously no lack of generality arises, since any spectrum $E(\kappa)$ derived for this network applies to any differently located network if we interpret it as $E(\chi)$. We now proceed along these lines to analyse the special case of a hexagonal network.

The hexagonal network

There is reason, apart from simplicity, for choosing this special case for detailed treatment, since the hexagonal array shown in figure 5 closely resembles the central section, normal to the c axis, of the Fermi surface in magnesium and zinc, metals which have been observed to show breakdown effects in moderate magnetic fields. If the I-lattice is specified by three vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ oriented at $120°$ to each other and each of magnitude a, the O-lattice is specified by the vectors $s\mathbf{a}_1$, $s\mathbf{a}_2$, $s\mathbf{a}_3$. All switches are identical apart from orientation, so that one value only is required for δ, and if $\mathbf{r}_0 = 0$ this is simply the internal

part, represented by one of the shaded triangles in figure 5. To look for periodic solutions
we choose one orbit and ascribe a (complex) amplitude to the wave function at the centre
point of each of its twelve segments, as shown in figure 6(a); neighbouring orbits are
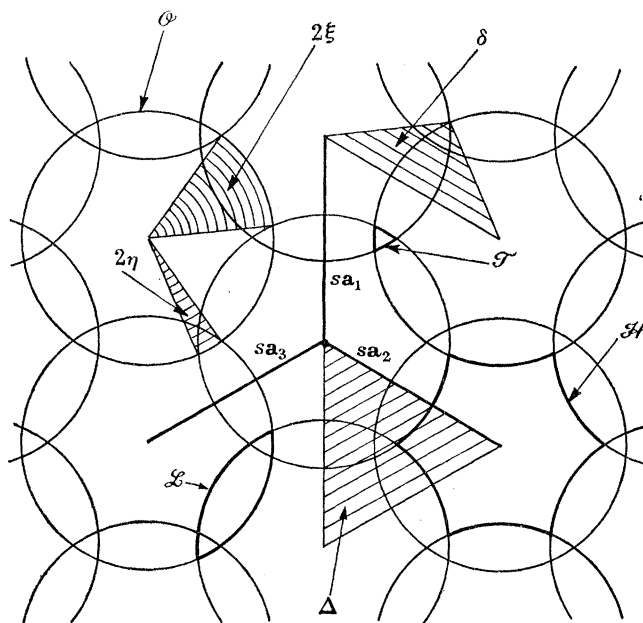

FIGURE 5. The hexagonal network, showing geometrical meaning of the
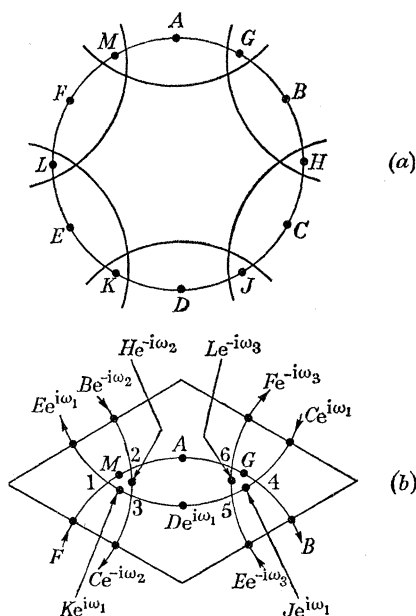phase parameters involved in the analysis.

FIGURE 6. Assumed amplitudes on network: (a) round one orbit, (b) in unit cell.

assumed to have the same amplitude distribution, phase shifted by ω_1, ω_2, or ω_3 according
to the relative positions of the neighbours, ω_i being written for $s\mathbf{a}_i.\boldsymbol{\kappa}$ (or $s\mathbf{a}_i.\boldsymbol{\chi}$ to generalize
to arbitrarily located O-lattices). In 6(b) are shown the amplitudes at the centre points of
segments in a unit cell, from which the amplitudes at the junction points are determined,

2ξ being written for the phase length of the longer paths (labelled A to F in figure 6(a)), and 2η for the shorter paths (G to M). The six junctions yield twelve relations as follows:

$$
\begin{aligned}
&(1) \quad E\,\mathrm{e}^{\mathrm{i}(\omega_1-\xi)} = pK\,\mathrm{e}^{\mathrm{i}(\omega_1+\eta)} + qF\,\mathrm{e}^{\mathrm{i}(\xi+\delta)}; && M\,\mathrm{e}^{-\mathrm{i}\eta} = qK\,\mathrm{e}^{\mathrm{i}(\omega_1+\eta-\delta)} + pF\,\mathrm{e}^{\mathrm{i}\xi}. \\
&(2) \quad H\,\mathrm{e}^{-\mathrm{i}(\omega_2+\eta)} = pB\,\mathrm{e}^{-\mathrm{i}(\omega_2-\xi)} + qM\,\mathrm{e}^{\mathrm{i}(\eta-\delta)}; && A\,\mathrm{e}^{-\mathrm{i}\xi} = qB\,\mathrm{e}^{-\mathrm{i}(\omega_2-\xi-\delta)} + pM\,\mathrm{e}^{\mathrm{i}\eta}. \\
&(3) \quad C\,\mathrm{e}^{-\mathrm{i}(\omega_2+\xi)} = pH\,\mathrm{e}^{-\mathrm{i}(\omega_2-\eta)} + qD\,\mathrm{e}^{\mathrm{i}(\omega_1+\xi+\delta)}; && K\,\mathrm{e}^{\mathrm{i}(\omega_1-\eta)} = qH\,\mathrm{e}^{-\mathrm{i}(\omega_2-\eta+\delta)} + pD\,\mathrm{e}^{\mathrm{i}(\omega_1+\xi)}. \\
&(4) \quad J\,\mathrm{e}^{\mathrm{i}(\omega_1-\eta)} = pC\,\mathrm{e}^{\mathrm{i}(\omega_1+\xi)} + qG\,\mathrm{e}^{\mathrm{i}(\eta-\delta)}; && B\,\mathrm{e}^{-\mathrm{i}\xi} = qC\,\mathrm{e}^{\mathrm{i}(\omega_1+\xi+\delta)} + pG\,\mathrm{e}^{\mathrm{i}\eta}. \\
&(5) \quad L\,\mathrm{e}^{-\mathrm{i}(\omega_3+\eta)} = pE\,\mathrm{e}^{-\mathrm{i}(\omega_3-\xi)} + qJ\,\mathrm{e}^{\mathrm{i}(\omega_1+\eta-\delta)}; && D\,\mathrm{e}^{\mathrm{i}(\omega_1-\xi)} = qE\,\mathrm{e}^{-\mathrm{i}(\omega_3-\xi-\delta)} + pJ\,\mathrm{e}^{\mathrm{i}(\omega_1+\eta)}. \\
&(6) \quad F\,\mathrm{e}^{-\mathrm{i}(\omega_3+\xi)} = pL\,\mathrm{e}^{-\mathrm{i}(\omega_3-\eta)} + qA\,\mathrm{e}^{\mathrm{i}(\xi+\delta)}; && G\,\mathrm{e}^{-\mathrm{i}\eta} = qL\,\mathrm{e}^{-\mathrm{i}(\omega_3-\eta+\delta)} + pA\,\mathrm{e}^{\mathrm{i}\xi}.
\end{aligned}
\tag{34}
$$

It is short work to eliminate six of the variables, but the other six involve considerable manipulation, entirely devoid of interest. Eventually the result is obtained:

$$
\begin{aligned}
C &\equiv \cos\omega_1 + \cos\omega_2 + \cos\omega_3 \\
&= \frac{\sin 6(\xi+\eta) - 2q^3\sin 6(\xi+\tfrac{1}{2}\delta) + q^6\sin 6(\xi-\eta+\delta) - 3q^2(1-q^2)^2\sin 2(\xi-\eta+\delta)}{2q(1-q^2)\,[q\sin 2(\xi-\eta+\delta) - \sin 2(\xi+2\eta-\tfrac{1}{2}\delta)]}.
\end{aligned}
\tag{35}
$$

To see the meaning of this result, it is convenient to remember that the phase angles ξ, η and δ all represent areas in the orbit diagram, as shown in figure 5, and that the four orbits of interest, those shown in figure 3 and the free electron circular orbit, have areas as follows:

$$
\begin{aligned}
\text{free-electron orbit:} \quad & \mathcal{O} = (12\xi + 12\eta)/\alpha; \\
\text{lens orbit:} \quad & \mathcal{L} = (4\xi + 8\eta - 2\delta)/\alpha; \\
\text{triangular orbit:} \quad & \mathcal{T} = (\Delta - 3\delta + 6\eta)/\alpha; \\
\text{hexagonal orbit:} \quad & \mathcal{H} = (6\Delta - 12\xi - 6\delta)/\alpha.
\end{aligned}
\tag{36}
$$

Hence the parameters occurring in (35) may be rewritten:

$$
\begin{aligned}
6(\xi+\eta) &= \tfrac{1}{2}\alpha\mathcal{O} = \alpha(2\Delta + \mathcal{T} - \tfrac{1}{2}\mathcal{H}), \\
6(\xi+\tfrac{1}{2}\delta) &= \alpha(3\Delta - \tfrac{1}{2}\mathcal{H}), \\
6(\xi-\eta+\delta) &= \alpha(4\Delta - \mathcal{T} - \tfrac{1}{2}\mathcal{H}), \\
2(\xi-\eta+\delta) &= \alpha(\tfrac{4}{3}\Delta - \tfrac{1}{3}\mathcal{T} - \tfrac{1}{6}\mathcal{H}), \\
2(\xi+2\eta-\tfrac{1}{2}\delta) &= \tfrac{1}{2}\alpha\mathcal{L}.
\end{aligned}
\tag{37}
$$

As the electron energy is altered, $\mathbf{H}$ being constant, the orbit areas change, and (35) defines the band structure of the system, the dependence of energy on χ. If the quantum numbers of the orbits are large, (35) is a rapidly oscillatory function and the changes in the areas (37) which occur in a narrow range of energy are conveniently expressed in terms of the perimeters of the orbits concerned. Thus the area of a free electron orbit of energy E is $2\pi mE/(eH)^2$, so that

$$
\delta\mathcal{O} = \frac{2\pi m}{(eH)^2}\,\delta E.
\tag{38}
$$

If we write γ for ξ/η, we can associate a fraction $\tfrac{1}{6}\gamma/(1+\gamma)$ of $\delta\mathcal{O}$ with each of the six sectors 2ξ, and a fraction $\tfrac{1}{6}/(1+\gamma)$ with each of the sectors 2η. In general, for an orbit of area $\mathcal{A}$ whose

perimeter is composed of n elements 2ξ and m elements 2η,† the variation of area with energy may be written

$$\delta \mathscr{A} = (n\gamma + m)\, x/\alpha \tag{39}$$

in which x is a dimensionless form of the energy increment,

$$x = \frac{\pi m\, \delta E}{3(1+\gamma)\, eH\hbar}. \tag{40}$$

In this way, using (37), we may rewrite (35) so as to exhibit the relative rates of change of the different terms in a small energy interval,

$$C = \frac{\sin 3(\gamma+1)\, x - 2q^3 \sin 3\gamma x + q^6 \sin 3(\gamma-1)\, x - 3q^2(1-q^2)^2 \sin(\gamma-1)\, x}{2q(1-q^2)\,[q \sin(\gamma-1)\, x - \sin(\gamma+2)\, x]}. \tag{41}$$
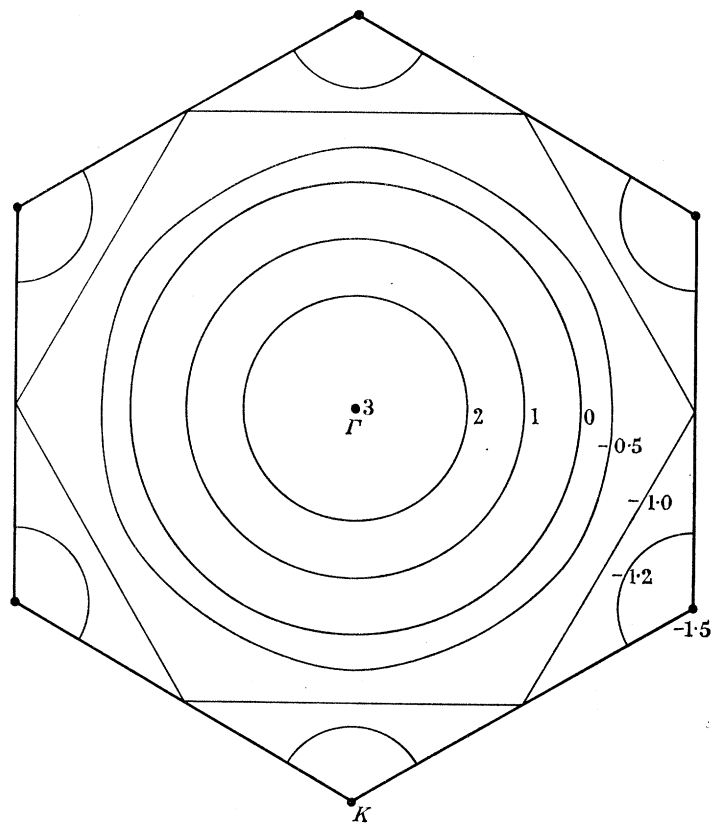

FIGURE 7. Contours of C (constant energy) on the χ-plane.

In writing (41) we have omitted all constant phase factors corresponding to the precise situation when $x = 0$; we have in fact assumed that when $x = 0$ all the areas are such that each sine function in (35) vanishes. For our purposes this is an entirely trivial specialization.

It will be observed that there is a separation in (41) of terms involving χ from all the rest. This means that contours of constant C in χ-space are contours of constant energy, and have the same form for all bands and all values of the coupling constant q. They are the same as in the tight-binding calculation for a hexagonal array of atoms with only nearest-

† The sign of m or n indicates whether a given element contributes to an increase or decrease of area when the energy is increased. Thus $\mathscr{T}$ has $m = +3$ while $\mathscr{H}$ has $n = -6$.

neighbour interaction, and are shown in figure 7. The zone defining the range of variation of χ is the Brillouin zone corresponding to the O-lattice and is thus s times smaller in linear dimensions, s^2 in area, than the zone of the I-lattice. It consequently holds $1/s^2$ states for each ion of the I-lattice, but as we have supposed there to be s independent networks we conclude that each filled zone in χ space accounts for $1/s$ states per ion. This is precisely the degeneracy of the Landau levels of free electrons, and the result merely confirms our accountancy of states, and illustrates how the highly degenerate Landau states are spread into bands by the coupling, yet not to the extent of eliminating the last s-fold degeneracy of the bands themselves.

For all its compactness $C(\gamma, x, q)$ is a very complicated function, with rather more intricate properties than might have been expected before the calculation was performed. To illustrate (41) extensive computations have been made for one particular choice of γ, which has been taken as $31/5$. For this value C is periodic in x with period 10π, the second half of the range repeating the first in reverse. Figure 8 shows a section of this range containing illustrations of the principal features of interest. Only the bounds of the physically accessible solutions are shown, each band lying between the contours $C = 3$ (centre of zone) and $C = -\frac{3}{2}$ (corners of zone). When $q = 0$ the energy levels are sharp, at the appropriate spacing for uncoupled free electron orbits; when $q = 1$ they are again sharp, and form two sets of levels spaced in inverse proportion to the areas of the hexagonal and triangular orbits. The latter are twice as numerous as the former, and their levels occur in pairs; at P, for example, where $x = \frac{10}{3}\pi$, three levels coincide. For intermediate values of q the levels are broadened by orbit coupling, but by no means uniformly. It appears from the computations, but has not been proved analytically, that all saddle-points of $C(x, q)$ occur when C is either 3 or $-\frac{3}{2}$. These saddle-points, denoted by S in figure 8, mark points at which two zones contact one another, either at the centre, Γ, or at a corner, K. A more elaborate contact is marked by T; at such points the trajectories of the zeros of numerator and denominator in (41) osculate, and there is a momentary sharpening of the level which passes through T. The detailed structure of the contours round a point such as T is shown in figure 8 (b) from which the highly singular nature of the surface is obvious. A point T is always accompanied by two saddle-points, and if these three are brought to coincidence a junction point like F results (though in fact F may only have the character of T, but compressed into a very small range of the variables).

It will be observed that all contours between 3 and $-\frac{3}{2}$ span the diagram from $q = 0$ to 1; this permits the number of eigenstates to be the same for all q, as is physically necessary. Conservation of states demands that no saddle-point shall lie in this range, but it is not obvious why all saddle-points should lie at the bounds of the range, rather than in the forbidden regions.

The revelation by explicit calculation of the nexus of saddle-points and even worse singularities may be considered one of the principal results of an extraordinarily tedious and inelegant analysis, for it makes clear the limitations of any perturbation approach to the energy-level structure, which would need delicate handling to span the full range of q. Furthermore, if the simplest model is capable of such anfractuosity it need be no occasion for marvel that such general treatments as have been essayed should be distinguished more for erudition than for usefulness.

Transport properties of the two-dimensional network

In principle we could solve (34) completely to determine the amplitudes and phases on all arms of the network. Such labour would be necessary if we wished to compute interband matrix elements for a complete treatment of transport phenomena, but fortunately some
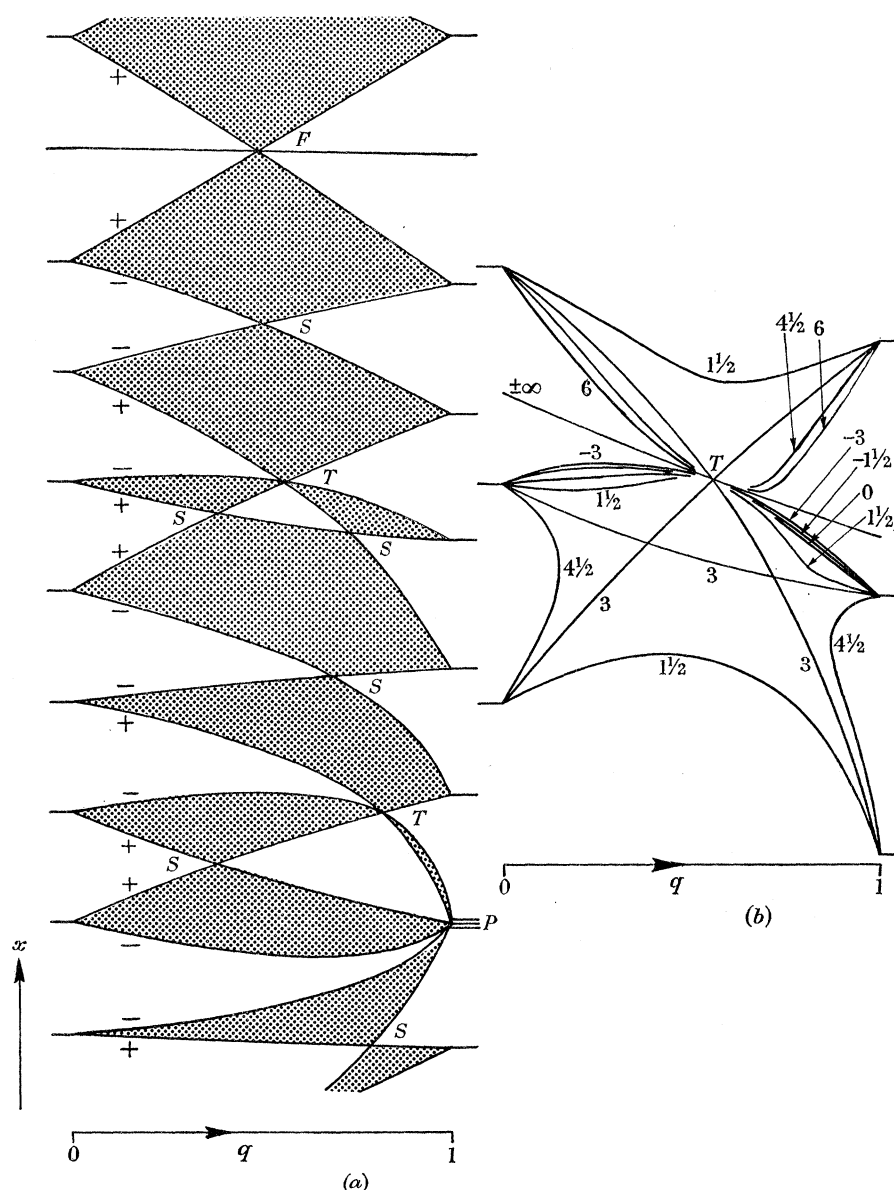

FIGURE 8. Energy band structure for hexagonal network, $\gamma = 31/5$. The shaded regions are bounded by the lines $C = 3$ $(+)$ and $C = -\frac{3}{2}$ $(-)$. The sharp levels on the left of (a) $(q = 0)$ are the free-electron levels, on the right $(q = 1)$ the levels of hexagonal and triangular orbits. The structure near T is shown enlarged in (b).

of the more interesting results for our present purpose can be achieved more simply. What we shall see is that the solutions of the network, as with the one-dimensional chain, are current-carrying states and that we can ascribe an effective velocity to each state by the formula $\hbar^{-1}\nabla_\chi E$, just as if the band structure of figure 8 applied to particles moving in the absence of a magnetic field. The proof is a straightforward extension of a standard theorem

and in developing it we ignore for the moment the s-fold degeneracy of the states characterized by E and $\boldsymbol{\chi}$, returning to this important point afterwards. We do not however confine ourselves to one particular network, but suppose that we have a superposition of states of the same E and $\boldsymbol{\chi}$, making up some function $\psi'_\chi(\mathbf{r})$ which we saw from (33) retains the periodicity of $\boldsymbol{\chi}$. That is to say, if the plane is divided into cells and $\psi'_\chi(\mathbf{r})$ in each cell is gauged with respect to a gauge centre lying in the cell (for definiteness let the gauge centres form an O-lattice containing the origin), then

$$\psi'_\chi(\mathbf{r}) = e^{i\boldsymbol{\chi}\cdot\mathbf{R}}u'_\chi(\boldsymbol{\rho}) = e^{i\boldsymbol{\chi}\cdot\mathbf{r}}u_\chi(\boldsymbol{\rho}), \tag{42}$$

where, as before, $\mathbf{r} = \mathbf{R}+\boldsymbol{\rho}$, $\mathbf{R}$ being an O-lattice vector, and $u_\chi(\boldsymbol{\rho}) = e^{-i\boldsymbol{\chi}\cdot\boldsymbol{\rho}}u'_\chi(\boldsymbol{\rho})$. Now let us transfer all gauge centres to the origin, so that the wave function takes the form, according to (5),

$$\psi_\chi(\mathbf{r}) = e^{\frac{1}{2}i(\boldsymbol{\alpha}\times\mathbf{R})\cdot\mathbf{r}}\psi'_\chi(\mathbf{r}) = e^{i(\boldsymbol{\chi}+\frac{1}{2}\boldsymbol{\alpha}\times\mathbf{R})\cdot\mathbf{r}}u_\chi(\boldsymbol{\rho}). \tag{43}$$

By using the same gauge for all parts of the wave function we destroy the appearance of strict periodicity, but not in such a way as to invalidate the effective mass theorem, which we now prove.

In the presence of a magnetic field we evaluate the mean electronic velocity from the expression

$$\mathbf{v} = -\frac{1}{m}\int\psi^*(i\hbar\nabla+e\mathbf{A})\psi\,d\mathbf{r}$$

$$= -\frac{1}{m}\int u_\chi^*(i\hbar\nabla-\hbar\boldsymbol{\chi}-\tfrac{1}{2}\hbar\boldsymbol{\alpha}\times\mathbf{R}+e\mathbf{A})u_\chi\,d\boldsymbol{\rho}, \quad \text{from (43),} \tag{44}$$

if ψ_χ is normalized in the unit cell. Although u_χ and $\mathbf{R}$ are discontinuous at the cell boundaries, there is no boundary contribution to (44) since ψ_χ itself is continuous. To manipulate $\mathbf{v}$ it is convenient to write Schrödinger's equation in full

$$\left\{-\frac{\hbar^2}{2m}\nabla^2-\frac{ie\hbar}{m}\mathbf{A}\cdot\nabla+\frac{e^2\mathbf{A}^2}{2m}+V\right\}\psi_\chi = E\psi_\chi,$$

and to substitute (43) to yield the equation

$$\mathscr{H}_\chi u_\chi = Eu_\chi,$$

in which

$$\mathscr{H}_\chi = -\frac{\hbar^2}{2m}\nabla^2-\frac{i\hbar}{m}(\hbar\boldsymbol{\chi}+\tfrac{1}{2}\hbar\boldsymbol{\alpha}\times\mathbf{R}-e\mathbf{A})\cdot\nabla$$

$$-\frac{e\hbar}{m}\mathbf{A}\cdot(\boldsymbol{\chi}+\tfrac{1}{2}\boldsymbol{\alpha}\times\mathbf{R})+\frac{\hbar^2}{2m}(\boldsymbol{\chi}+\tfrac{1}{2}\boldsymbol{\alpha}\times\mathbf{R})^2+\frac{e^2A^2}{2m}+V. \tag{45}$$

Let us now evaluate $E(\boldsymbol{\chi}+\delta\boldsymbol{\chi})$ by writing $\mathscr{H}_{\chi+\delta\chi}$ as $\mathscr{H}_\chi+\delta\boldsymbol{\chi}\cdot\nabla_\chi\mathscr{H}_\chi$ and treating the second term as a perturbation. Then to first order in $\delta\boldsymbol{\chi}$

$$\delta E = \delta\boldsymbol{\chi}\cdot\nabla_\chi E = \int u_\chi^*(\delta\boldsymbol{\chi}\cdot\nabla_\chi\mathscr{H}_\chi)u_\chi\,d\boldsymbol{\rho},$$

or

$$\nabla_\chi E = \int u_\chi^*(\nabla_\chi\mathscr{H}_\chi)u_\chi\,d\boldsymbol{\rho}. \tag{46}$$

From (45) we have that

$$\nabla_\chi\mathscr{H}_\chi = -\frac{i\hbar^2}{m}\nabla-\frac{e\hbar}{m}\mathbf{A}+\frac{\hbar^2}{m}(\boldsymbol{\chi}+\tfrac{1}{2}\boldsymbol{\alpha}\times\mathbf{R}), \tag{47}$$

so that, substituting (47) in (46) and comparing with (44), we achieve the desired result,

$$\nabla_k E = \hbar \mathbf{v}. \tag{48}$$

In deriving (48) we applied first-order perturbation treatment to one of a set of degenerate levels. This procedure is only valid if we first choose such linear combinations of our network functions as form an orthogonal set whose members do not combine under the influence of the perturbation, i.e. if ψ_{χ_1} and ψ_{χ_2} are two such members we must have that

$$\int u^*_{\chi_2}(\nabla_\chi \mathscr{H}_\chi) u_{\chi_1} \, d\boldsymbol{\rho} = 0,$$

or, what is equivalent,
$$\int \psi^*_{\chi_2} \mathbf{v} \psi_{\chi_1} \, d\mathbf{r} = 0. \tag{49}$$

Now if we assume that the network functions display in themselves enough variety to form a complete set (as is presumably true, since it is true for a network of independent orbits of free electrons), it is always possible in principle to construct the appropriate set of ψ_χ for the present purpose, and we note from (48) that all the members of the set have the same mean velocity. Hence any linear combination of these ψ_χ has the same velocity, since from (49) all cross-terms vanish. Thus (48) applies to all states characterized by χ.

This result implies that every electronic state has the property of transporting charge through the metal, in spite of the magnetic field, and all directions of transport are possible. Herein lies an important distinction between the network and a metal having open orbits, for the latter has only certain well-defined directions of transport. As we discussed in connexion with the linear chain, (48) tempts us to introduce a set of quasi-particles whose properties are defined by $E(\chi)$, and forget about the magnetic field. But we must justify this by demonstrating the appropriate response to an electric field, as may be readily achieved. It is convenient to use the impulse method, applying a field $\mathscr{E}$ for a time δt, at $t = 0$, to an electron in the stationary state ψ_χ and thereby changing the wave function into ψ', where
$$\psi' = \psi_\chi e^{i\mathbf{q}\cdot\mathbf{r}} \quad \text{and} \quad \mathbf{q} = (e/\hbar)\mathscr{E}\,\delta t.$$

As in the earlier discussion of the linear chain, ψ' is not a stationary state but may be synthesized from one or more states $\psi_{\chi+q}$ belonging to the same band as ψ_χ together with other states of wave number $\chi + \mathbf{q}$ taken from other bands. The intraband terms have a total amplitude of unity, and contribute a steady change in current, while the interband terms contribute oscillatory currents of amplitude proportional to q. The result just proved, that a superposition of states of the same χ carries the same current as each component, show that the intraband current is just that associated with a change from χ to $\chi + \mathbf{q}$, exactly as if we were concerned with a quasi-particle defined by $E(\chi)$. For a steady field
$$\hbar\dot{\chi} = e\mathscr{E}, \tag{50}$$
or, in terms of effective masses,

$$\dot{v}_j = \frac{e}{\hbar}\mathscr{E}_i m_{ij}^{-1}, \quad \text{where} \quad m_{ij}^{-1} = \frac{1}{m}\frac{\partial^2 E}{\partial \chi_i \partial \chi_j}. \tag{51}$$

We shall make no attempt to calculate the interband terms, which may be expected to be extremely complicated, to judge from the variety of orbits available on the network, any one of which being in principle available to a classical electron should make its contribution to the oscillatory response to the impulse. As discussed earlier, in connexion with (22), the interband terms are of little consequence to the current parallel to $\mathscr{E}$ if the relaxation time

is long enough, so that $\omega_c \tau \gg 1$ for any oscillations. It is only this case which we shall treat further, for in the other limit, $\omega_c \tau \ll 1$, the system behaves like a free electron assembly and in the intermediate range the complexity is daunting. For the Hall current, normal to $\mathscr{E}$, (23) shows that the interband terms are the sole contributors. Some of the extended orbits available are electron-like and some are hole-like, and indeed as q varies from 0 to 1 the behaviour changes from purely electron-like to predominantly hole-like, since the hexagonal orbits are so much larger than the triangular; but whether this transition is monotonic, or how fast it proceeds as q varies, is something we have lacked the determination to discover.
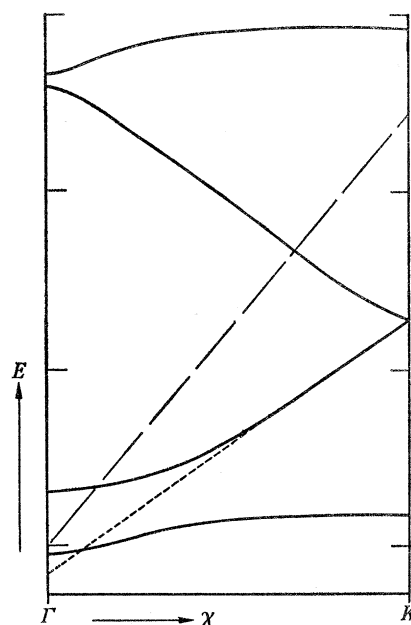

FIGURE 9. Part of the variation of E with χ along the line ΓK in figure 7. The broken line represents $dE/d\chi$ for a free electron at the Brillouin zone corner, and the markers at the sides are free-electron levels.

In view of the similarity between the network and an electron in a band with no magnetic field, it is instructive to present the solution of the network in the conventional way, displaying E as a function of χ for a particular propagation direction. In figure 9 is shown a section of such a diagram for χ lying along the line ΓK in figure 7; the value of q has been chosen so that a contact point S is included. At such a point the particle velocity given by (48) is comparable to that of a free electron of the same energy, as we shall shortly show, but in the bands on either side the effective mass is very high and the velocity low. The narrow bands are caused by zeros in the denominator of (41) and it may be noted that each zero disrupts the expected backward and forward alternation of the bands.

To estimate the order of magnitude of the velocity we note that when contact occurs the gradient $dE/d\chi$ may well be large enough to span about 1·4 times the separation of free electron levels $(eH\hbar/m)$, within the range of χ, as may be seen from the dotted line in figure 9. If k_0 is the radius from the centre to a corner of the Brillouin zone, the range of χ is k_0/s, so that the dotted line represents a velocity v given by the expression

$$v = \frac{1\cdot4 eH\hbar}{m} \frac{s}{\hbar k_0} = \frac{1\cdot4 \times 3\sqrt{3}}{4\pi} \frac{\hbar k_0}{m}.$$

That is, v is about 0·6 of the velocity of a free electron at the zone corner. It is clear that at a contact point the electron is putting its best foot forward, considering that it cannot go straight through the metal but must keep to the arms of the network.

We are now ready to consider what might actually be observed in practice if our two-dimensional metal existed, confining attention to such a pure sample that $\omega_c \tau \gg 1$. It is clear that the conductivity can in principle vary very rapidly with the position of the Fermi level on the band structure of figure 9, and hence, for a given metal, with the strength of the magnetic field (for the moment we ignore the limitation of our analysis to special values of $\mathbf{H}$, and suppose that the energy scale of the band structure expands in proportion to $\mathbf{H}$, leaving the details unaltered). It would be necessary, however, to go to a low temperature to reveal all the details in a moderate field, say 10 kG, the field strength around which interesting effects have been observed in zinc. For at this field strength $\hbar \omega_c / k$ for free electrons is a temperature of about 1 °K, and only at considerably lower temperatures would the Fermi level be sharp enough to enable the fine structure to be distinguished. We therefore discuss effects arising from the coarse structure of the bands, particularly those associated with the long periodicity due to the quantizing of the small triangular orbits. We shall suppose the temperature to be fixed and calculate the conductivity when account is taken of the Fermi distribution.

We assume a constant relaxation time τ so that the linearized Boltzmann equation has the form

$$(e\mathscr{E}/\hbar) . \nabla_\chi f_0 = (f_0 - f)/\tau,$$

in which f_0 is the Fermi function $[e^{(E-E_0)/kT} + 1]^{-1}$ and f the distribution function.

Hence

$$\Delta f \equiv f - f_0 = -e\tau\mathscr{E} . \mathbf{v} \, \mathrm{d}f_0/\mathrm{d}E.$$

If an area $\mathrm{d}\chi$ of reciprocal space in the zone of figure 7 holds $\beta \mathrm{d}\chi$ states per unit area of metal (we include in β the s-fold degeneracy of the networks), its contribution to the current may be written

$$\mathrm{d}\mathbf{J} = \beta e \mathbf{v} \Delta f \, \mathrm{d}\chi,$$

so that the component parallel to $\mathscr{E}$ has magnitude

$$\mathrm{d}J_\parallel = -\beta e^2 v^2 \tau \mathscr{E} \cos^2 \theta (\mathrm{d}f_0/\mathrm{d}E) \, \mathrm{d}\chi,$$

where θ is the angle between $\mathscr{E}$ and $\mathbf{v}$, i.e. between $\mathscr{E}$ and the normal to the energy contour passing through $\mathrm{d}\chi$. Now the conductivity σ of our hexagonal model is necessarily isotropic, so that we may just as well write the conductivity of a random polycrystal, that is, average over all values of θ, to yield the contribution of $\mathrm{d}\chi$ to σ,

$$\mathrm{d}\sigma = -\tfrac{1}{2}\beta e^2 v^2 \tau (\mathrm{d}f_0/\mathrm{d}E) \, \mathrm{d}\chi. \tag{52}$$

Now in figure 7 let us introduce curvilinear co-ordinates χ_n measured normal to the energy contours, and χ_t along the contours, so that $\mathrm{d}\chi = \mathrm{d}\chi_n \mathrm{d}\chi_t$; moreover $v = \hbar^{-1} \mathrm{d}E/\mathrm{d}\chi_n$. Substituting in (52) and integrating round an annulus between two contours ΔE apart we have that

$$\Delta\sigma = -\frac{\beta e^2 \tau}{2\hbar^2} \frac{\mathrm{d}f_0}{\mathrm{d}E} \Delta E \oint \left| \frac{\mathrm{d}E}{\mathrm{d}\chi_n} \right| \mathrm{d}\chi_t,$$

or, from (40),

$$\Delta\sigma = -\frac{3\alpha\beta e^2 \tau (1+\gamma)}{2\pi m} \frac{\mathrm{d}f_0}{\mathrm{d}E} V(x) K(C) \Delta E, \tag{53}$$

in which $V(x)$ is written for $|\mathrm{d}x/\mathrm{d}C|$, which may be computed from (41), and $K(C)$ for $|\oint(\mathrm{d}C/\mathrm{d}\chi_n)\,\mathrm{d}\chi_t|$, which is a property of the zone in figure 7 and only dependent on x in so far as $C(x)$ is defined by (41). If we now substitute $\mathrm{d}f_0/\mathrm{d}E$ and write μ for $3(1+\gamma)\hbar eH/(2\pi mkT)$, we have on integrating (53) the conductivity when the Fermi level is at x_0,

$$\sigma(x_0) = \frac{\beta e^2 \mu^2 \tau kT}{2\hbar^2}\int_{-\infty}^{\infty} V(x)\,K(C)\,\mathrm{sech}^2\left[\mu(x-x_0)\right]\mathrm{d}x. \tag{54}$$

This result can be made more readily interpretable by comparing $\sigma(x_0)$ with a standard conductivity σ_0, which we choose to be the conductivity of a gas of free electrons sufficient in number to fill the Brillouin zone, i.e. 2 per atom. Then (54) takes the form

$$\sigma(x_0)/\sigma_0 = \frac{3(1+\gamma)\mu}{8\pi^2}\int_{-\infty}^{\infty} V(x)\,K(C)\,\mathrm{sech}^2\left[\mu(x-x_0)\right]\mathrm{d}x. \tag{55}$$

Before presenting detailed calculations based on (55) it is worth remarking that if the structure of each band were the same (e.g. if x were a linear function of C as normally occurs when the band is reasonably narrow), each band, apart from the Fermi factor, would make a contribution to σ proportional to the square of its width. A coarse structure to σ may therefore be interpreted as due to a tendency for the mean band width to vary with x, and such a tendency may be discerned in figure 8, though perhaps not so clearly as might be expected from the considerable variations of σ which are revealed by computation of (55).

To show the variations of σ the same model has been used as led to (41), and the integral in (55) has been computed as a function of x_0 in the range $5 < x_0 < 10$ running through just over two periods of the oscillations due to the small triangular orbits. A value of 4 was chosen for μ, corresponding to $\hbar\omega_c = 1\cdot16kT$, e.g. 1 °K at 10 kG. Figure 10 shows how σ/σ_0 oscillates at different values of q. From these curves we may extract the mean conductivity $\bar{\sigma}$ and the fundamental oscillatory frequency σ_1 which has the periodicity of the quantized levels of the triangular orbits. The variation of $\bar{\sigma}/\sigma_0$ and σ_1/σ_0 with q are shown in figure 11. It will be observed that $\bar{\sigma}/\sigma_0$ may under favourable conditions be as high as 0·18, a conductivity equal to that of 0·36 free electron per atom. This is in line with our earlier estimate of the quasi-particle velocity. It is also worth noting the change of sign of σ_1 when q is about 0·7, and the fine structure in the individual curves of figure 10. These points are all of course consequences of the structure exhibited in figure 8. The rapid rise of σ_1 as q falls from unity results from the broadening of the levels diverging from a point such as P; the divergence is so rapid as to leave a marked deficiency of levels in this energy range when q has fallen to 0·6, and hence a dip in the curve for σ; while the reversal of the sign of σ_1 when $q \doteq 0\cdot7$ is a consequence of the tendency, visible in figure 8, for the contact points to occur at lower values of q when the energy is intermediate between the energy levels of the triangular orbits.

It is instructive to see that the same sort of behaviour occurs over a wide range of γ. For example, put γ equal to unity in (41), so that ξ and η are the same, and (41) takes the form

$$C = \frac{(q^3 - \cos 3x)\sin 3x}{q(1-q^2)\sin 3x}. \tag{56}$$

We retain the terms $\sin 3x$ in numerator and denominator to indicate the presence of sharp levels whenever $3x = n\pi$. The band structure is shown in figure 12. When q is near unity

the levels broaden rapidly at points such as P, marking the coincidence of three levels, but by the time q has fallen to 0·5 the phase of the oscillations of breadth has reversed, and maximum conductivity is found at values of x lying between the points P. The two different energy gaps become equal when q is about 0·65, at which value we expect the first Fourier
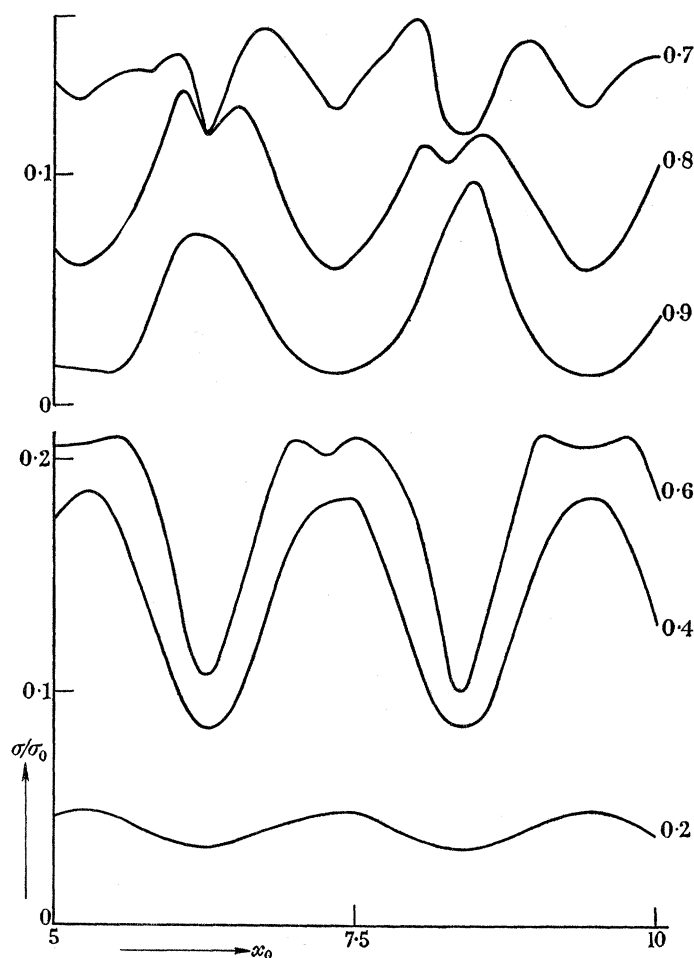

FIGURE 10. Variation of σ/σ_0 with Fermi energy, x_0, for different values of q, computed from (55). To avoid confusion the diagram is divided into two sections. Values of q are shown beside each curve.
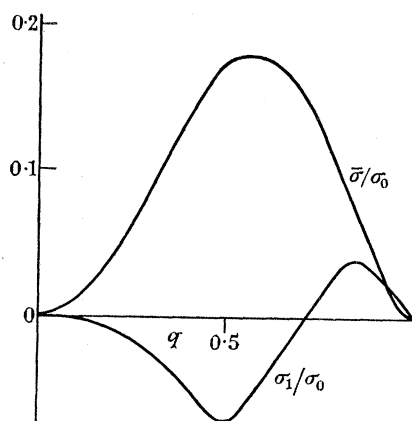
FIGURE 11. Variation with q of $\bar{\sigma}/\sigma_0$ and σ_1/σ_0, derived from figure 10.

42-2

component σ_1 to change sign. There is obviously a great similarity in behaviour for two quite different values of γ.

At the other extreme consider the behaviour of (41) when γ is very large. To compute (41), as was done in constructing figure 8, is unnecessary if we are interested only in qualitative properties, on account of the fact that if C is put equal to 3 ($\chi = 0$), (41) factorizes into 6 linear equations for q, corresponding to 6 independent sets of contours:

$$q = \frac{\cos\frac{1}{2}[(\gamma+1)x+\frac{1}{3}n\pi]}{\cos\frac{1}{2}[(\gamma-1)x-n\pi]}. \tag{57}$$

Thus the contours $C = 3$, marking one boundary of each band, are readily constructed, and the contours $C = -\frac{3}{2}$ may then be sketched in without difficulty. Sections of the band
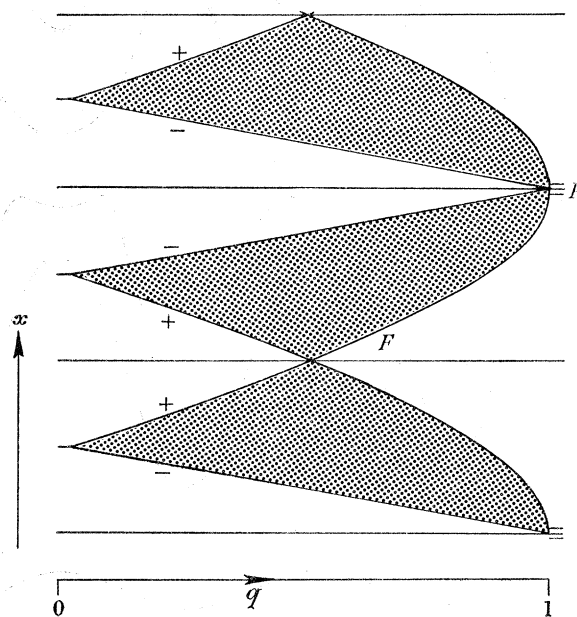

FIGURE 12. Level structure as a function of q when $\gamma = 1$.

structure so constructed are shown in figure 13, for $\gamma = 70$, demonstrating how the shape is affected by the proximity of quantized levels of the small triangular orbits, which occur when $x = \frac{2}{3}n\pi$. At such a point three levels coincide (if γ is an integer) when $q = 1$, but there is only one level at the same value of x when $q = 0$. The divergence of the bands away from P is very marked when γ is large, and is exemplified by the contour labelled V. As γ tends to ∞, V tends towards two straight lines. From every quantized level at $q = 0$ and 1 must spring one contour $C = 3$ and one $C = -\frac{3}{2}$; the excess of levels when $q = 1$ imposes a warping of the band structure which is only eliminated half-way between the points $\frac{2}{3}n\pi$, through the influence of the slightly larger separation of levels on the right-hand side. The stages in the reduction of the warping are shown in the diagram. It is the long-range influence of the little triangular orbits, extending right through the level scheme, however large γ may be, that causes even the tiniest orbits to influence the properties strongly when breakdown occurs. It will be seen that where the warping is most pronounced, near P, the levels broaden very rapidly as q falls from unity, while where there is no warping, as in figure 13 (c), the greatest breadth is reached when $q = \frac{1}{2}$. Thus the change of phase of the fundamental oscillation of conductivity, seen in figure 11, occurs for all γ. The same is true

of the crevasses in $\sigma(x_0)$ seen in figure 10. It is clear from figure 13 that when q is about 0·7, σ has its peak neither at points such as P nor half-way between, but at the quarter-way points such as 13 (b). In fact the general pattern of conductivity illustrated in figures 10 and 11 may be expected to reveal itself at all values of γ.
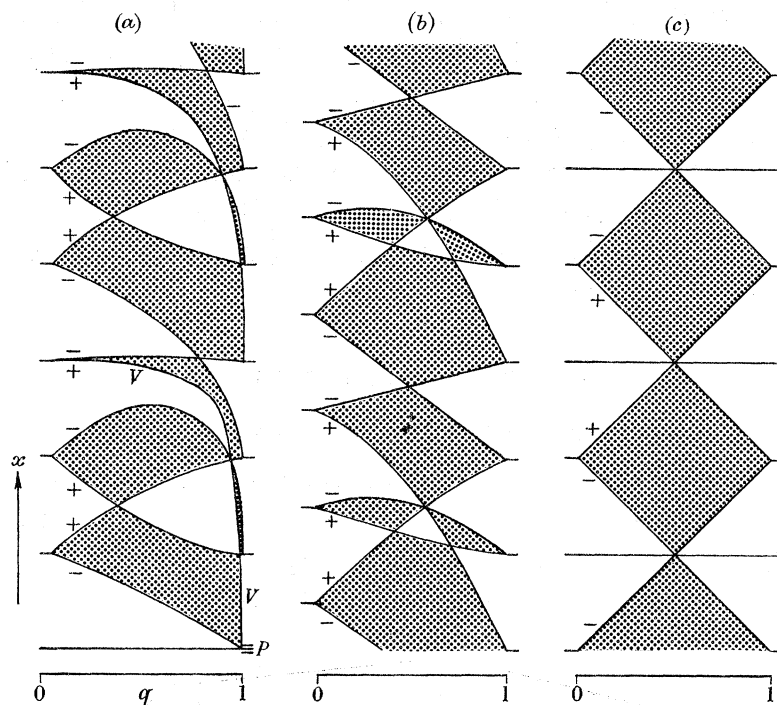

FIGURE 13. Level structure as a function of q when $\gamma = 70$. (a) near quantized level of triangular orbits, (b) quarter-way between such levels, (c) half-way between such levels.

Intermediate field strengths

The preceding discussion has been limited to such special cases as satisfy (30), the magnetic field being of such a strength that s was an even integer. We must briefly consider the problem that arises when this is not so. We have seen that at equivalent points on the network the internal part of δ is constant, but the external part δ_0 ($=\frac{1}{2}\alpha.(\mathbf{S} \times \mathbf{R})$) is determined not only by the switching vector $\mathbf{S}$ but by the co-ordinate of the orbit centre $\mathbf{R}$. Since $\mathbf{S}$ and $\mathbf{R}$ are both vectors of the O-lattice, which is s times larger than the I-lattice whose unit cell has area Σ, δ_0 may be written in the form

$$\delta_0 = \frac{1}{2}\alpha\lambda s^2 \Sigma$$

$$= \pi\lambda s \quad \text{from (26)},$$

where λ is an integer, equal to the number of cells of the O-lattice in the parallelogram defined by $\mathbf{S}$ and $\mathbf{R}$. If $\frac{1}{2}s$ is expressed as μ/ν, μ and ν being integers having no common factor except unity,

$$\delta_0 = 2\pi\lambda\mu/\nu,$$

and equivalent values of δ_0 are obtained only for such switches as have $\lambda\mu/\nu$ differing by integers. This implies that equivalent points are separated by ν vectors of the O-lattice, and the unit cell of the network is ν times as great in linear dimensions as the O-lattice cell. As $\mathbf{H}$ is steadily increased, s oscillates with infinite rapidity between rational and irrational values, and ν correspondingly indulges in infinitely fast oscillations of infinite amplitude.

If we are to solve the network problem exactly for any intermediate value of H, we must write down the compatibility conditions at $6\nu^2$ different junctions, giving $12\nu^2$ equations in $12\nu^2$ unknowns, only three of which appear in each equation. Even with this last proviso the magnitude of the problem is far from encouraging, and we shall take it no further. We may expect, however, that the solution would involve a finer subdivision of the energy bands of figure 8, each one splitting into ν sub-bands, and that the energy separation of the sub-bands would now be so small that under attainable conditions of temperature, relaxation time and field homogeneity no very striking effect could be expected. It may be noted, however, that the structure of the energy levels shows a cyclical variation, reverting to its greatest simplicity whenever H^{-1} is an integral multiple of $4\pi e/(\hbar\Sigma)$. Any observable effect would therefore resemble the de Haas–van Alphen effect due to an orbit having half the area of the Brillouin zone.

<div align="center">Comparison with experiment</div>

<div align="center">*The electronic structure of zinc*</div>

We may now look to see how closely the behaviour of a real metal reveals this sort of structure when magnetic breakdown occurs. The observations on zinc by Dhillon & Shoenberg (1955) and by Berlincourt & Steele (1954) are of value for discussing the de Haas–van Alphen effect, while the observations by Renton (1961), Stark (1962),
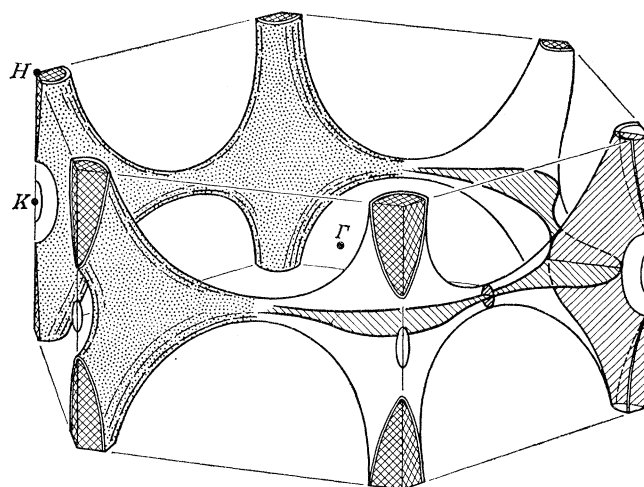

FIGURE 14. Fermi surface of zinc, from Reed & Brennert (1963).

Alekseyevski & Gaidukov (1962), Reed & Brennert (1963) enable us to discuss transport effects in zinc. We shall say little about the extensive measurements of transport effects by Reynolds and his co-workers (e.g. Grenier, Reynolds & Zebouni 1963), since they used samples of less purity than did the other workers and interpretation of the breakdown effects is far from straightforward. It is not easy in any circumstances to compare the predictions of a two-dimensional theory with the performance of a complex metal like zinc, and we need feel little surprise that the comparison is imperfect. Nevertheless, in spite of difficulties and considerable unresolved discrepancies, there are encouraging points of similarity between theory and experiment.

The Fermi surface of zinc is complicated, involving several electron and hole surfaces, the number of electrons and holes being equal if no breakdown occurs. The significant surfaces for our purpose are the second-zone 'monster' and the third-zone needles shown in figure 14. A plane normal to the c axis through the centre contains an orbit round the outside of the monster which is the smoothed-off counterpart of the hexagonal hole orbit of our model, while the needles correspond to the triangular orbits. It is convenient, in order to represent the variety of orbits, to section the surfaces by a plane which is tilted slightly from the basal plane, and in figure 15 is shown schematically such a section, the axis of tilt
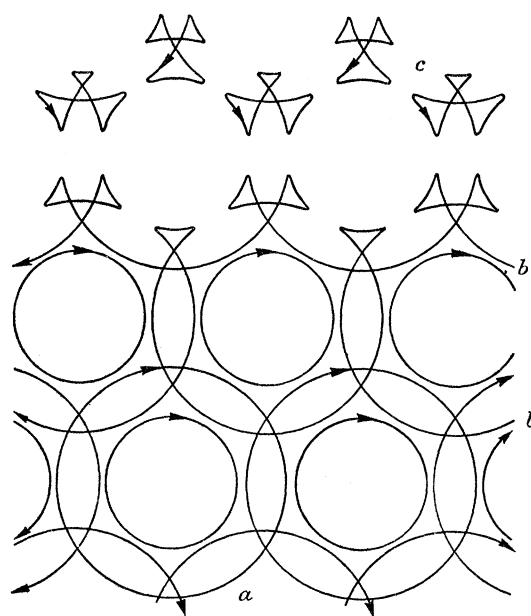

FIGURE 15. Tilted central section of Fermi surface of zinc, to show different types of orbit.

being (10$\bar{1}$0). When the plane passes through the centre of the zone the orbits (a) are the same as in figure 5, but neighbouring sections are able to cut the needles while missing the thin necks of the monster, thus giving rise, if breakdown is complete, to open orbits (b), and further out to trefoils (c); if there is no breakdown each loop of a trefoil is a separate orbit. The general validity of this picture has been established by the experiments of Joseph & Gordon (1962), Stark (1962) and Reed & Brennert (1963), and the theoretical work of Harrison (1962). The numerical magnitudes that concern us may conveniently be tabulated:

Length of side of Brillouin zone $= 1 \cdot 58 \times 10^8 \, \text{cm}^{-1}$.

Height of Brillouin zone $= 1 \cdot 29 \times 10^8 \, \text{cm}^{-1}$.

The zone contains one state per atom, i.e. $6 \cdot 5 \times 10^{22}$ states per cm^3.

Volume of monster $\doteqdot \frac{1}{12}$ volume of Brillouin zone.

Cross-sectional area of necks of monster $= 4 \cdot 3 \times 10^{13} \, \text{cm}^{-2}$.

Diameter of neck (assumed circular) $= 7 \cdot 4 \times 10^6 \, \text{cm}^{-1} = \frac{1}{17}$ height of Brillouin zone.

Cross-sectional area of needles $= 1 \cdot 5 \times 10^{12} \, \text{cm}^{-2}$.

Length of needles $= 2 \cdot 3 \times 10^7 \, \text{cm}^{-1} = \frac{2}{11}$ height of Brillouin zone.

Cyclotron mass of needles for orbits in basal plane $= \frac{1}{140}$ free electron mass (Dhillon & Shoenberg 1955).

This last fact determines the value of γ in our model, for the quantized levels of the needles are 140 times further apart than the free electron levels, and this requires γ to be 70, rather than 6·2 as in the detailed calculations presented earlier. We have seen, however, that the essential features of the level structure are not highly dependent on γ, and can infer what is needed without further elaborate computation.

The length of the needles is such that one might reasonably expect the energy gap across the Brillouin zone face to vary along their length, so that the value of q in a given field probably also varies, with a stationary value at the middle. Since, however, the oscillatory effects are dominated by an extremal cross-section, provided the field is not so high that the Fermi level lies near the lowest quantized state, we might hope to concern ourselves only with the central region of a needle. But unfortunately in zinc the needles are so thin that when the field is only 30 kG the lowest quantized state roughly coincides with the Fermi surface at the central section. It is therefore too much to expect that at the highest fields for which data exist (about 23 kG) there can be exact agreement with a simplified theory that assumes q to take the same value for all sections of the needles. On the other hand, the angular variation of the de Haas–van Alphen period is so close to that expected for an ellipsoid that we have no reason to suppose the energy gap to vary widely, such as, for example, vanishing at some point on the needle where the two zones might have an accidental degeneracy. It may further be noted that the network model was developed on the assumption that the regions of confusion at junctions did not overlap, but the needles in zinc are too small for this to be true; one cannot therefore expect the energy level structure to follow the theory exactly, though one may hope for the deviation to be small.

The de Haas–van Alphen effect

The low-frequency de Haas–van Alphen oscillations are due to the needles, and the effect of breakdown is primarily to diminish the amplitude of the oscillations, since fewer electrons are able to complete several circuits without moving on to different orbits. More precisely the effect is seen in the rearrangement and broadening of the energy levels as q falls below unity, so that the widely spaced levels (P in figure 8) due to the needles are spread out among the rest. On account of temperature, scattering and other disturbances the levels are not observed as sharp changes in magnetization as the field is increased, causing them to cross the Fermi surface, but as roughly sinusoidal variations as if only the fundamental Fourier components of the level density were being observed. In order, therefore, to estimate the expected reduction of amplitude due to breakdown it is convenient to compute the fundamental amplitude as a function of q, and we have done this for the two cases $\gamma = 31/5$ and 1, whose level schemes are shown in figures 8 and 12. It is necessary of course to take account of the way in which C varies across each band, for it is clear from figure 9 that different bands differ in detail; one must also give appropriate weight to each part of the range of C according to the area it defines in the χ zone of figure 7. The results of the computations are shown in figure 16, from which it is clear that the behaviour is not highly sensitive to the value of γ. When $\gamma = 1$ the reduction factor ζ is very close to q^3, while when $\gamma = 31/5$ it is nearer $q^{3·6}$. It is also worth noting that when the triangular orbits are not part of a network, but simply linked to three much larger orbits as shown in figure 15 for non-central sections, ζ is again q^3, as proved in the Appendix. For the purpose of

comparison with experiment we shall take ζ to be q^3, so that by use of Blount's (1962; see also I) expression for q we have

$$\zeta = (1 - e^{-H_0/H})^{\frac{3}{2}},\tag{58}$$

where

$$H_0 = \pi\epsilon^2/(4\hbar e v_x v_y),\tag{59}$$

ϵ being the energy gap at the zone boundary, v_x and v_y the normal and tangential components of the free-electron velocity at the boundary.
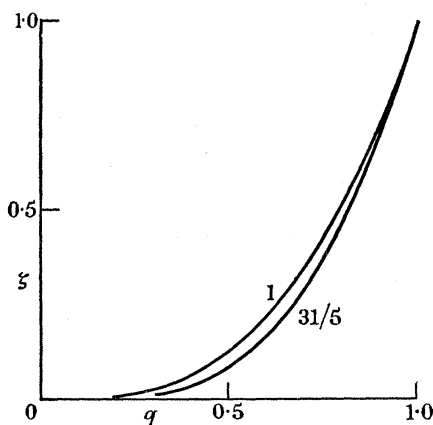

FIGURE 16. Reduction factor ζ for de Haas–van Alphen amplitude, for $\gamma = 1$ and $31/5$.
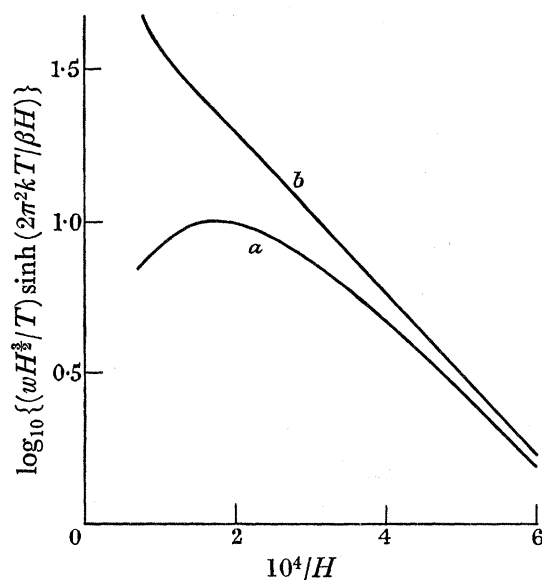

FIGURE 17. Field variation of de Haas–van Alphen amplitude in zinc: (a) as observed by Dhillon & Shoenberg (1955), (b) corrected for breakdown. See Dhillon & Shoenberg for an explanation of the symbols.

As primary data we choose the observations of Dhillon & Shoenberg (1955), presented in their figure 7. In fact the magnetic field was inclined at 20° to the c axis for these measurements, but Berlincourt & Steel (1954) state that the only effect of inclination is on the magnetic field scale, not on the shape, so we may take these data to apply to the present case if the period of the oscillations is multiplied by 1·064. This observation by Berlincourt & Steele is clearly explained by the insensitivity of ζ to anything in the model except the value

of q. Now if breakdown did not occur, and the needles were perfect ellipsoids, the amplitude of the oscillations should vary with field in such a way that the experimental curve drawn in figure 17 (taken from Dhillon & Shoenberg's figure 7) was straight, as such curves usually are for other metals. The question is whether we can find such a value of H_0 as will account for the considerable deviation from linearity at the highest fields. As can be seen from figure 17, with $H_0 = 6\,\mathrm{kG}$ the curve is rectified almost to the highest fields, and we need not be too concerned with the remaining discrepancy for the reasons explained above, though it may not be too fanciful to suggest that part at least might be explained by an increase of the gap with distance from the central section. From this comparison we estimate the energy gap at the centre of the zone boundary to be $0 \cdot 027\,\mathrm{eV}$. There are no valid theoretical estimates with which this estimate may be compared.

Magneto-resistance

The comparison is much more tenuous in this case, partly because of the incompleteness of the theory and partly because of the inconsistency of the data. All workers agree on the large oscillations of the transverse magneto-resistance when $\mathbf{H}$ lies along the c axis, but the precise magnitude of the resistance seems subject to wide variations. Thus three groups have used samples whose resistance ratio S (resistance at room temperature divided by residual resistance) was about $15\,000$, and their published curves indicate that in a field of $22\,\mathrm{kG}$ the resistance is increased by a factor of 200 (Reed & Brennert 1963), 300 (Stark 1962) or 1800 (Alekseyevski & Gaidukov 1962). The last is so wildly different as to suggest either a miscalculation or an error in orientation. But the first two values are the result of obviously careful work, and it seems likely that the behaviour is even more highly sensitive to the orientation of the sample and the field than was suspected by the authors. At all events we must not expect exact agreement between theory and experiment. We shall concentrate on the results of Reed & Brennert, not because we believe them to be necessarily the best but because they alone give details of the Hall field as well as the magneto-resistance and so allow us to calculate the components of the conductivity tensor to compare with theory. With $\mathbf{H}$ along the c axis the conductivity tensor is isotropic in the basal plane and determined by two components σ_{11} and σ_{12}, related to the resistivity ρ_{11} and the Hall constant ρ_{12}/H by the equations

$$\sigma_{11} = \frac{\rho_{11}}{\rho_{11}^2 + \rho_{12}^2}, \quad \sigma_{12} = \frac{\rho_{12}}{\rho_{11}^2 + \rho_{12}^2}.$$

Figure 18 shows part of the field variation of the conductivity deduced from Reed & Brennert's curves. It has been assumed that for their specimen S was $15\,000$.

The analysis of these curves is enormously aided by the large value of S, which ensures that in the interesting range of field strengths $\omega_c \tau \gg 1$. Of course, the complexity of the Fermi surface implies a considerable range of ω_c and probably also of τ, but if for the moment we assume constant values we may estimate them readily. For when $\omega_c \tau \gg 1$, and the number of electrons and holes are equal, the Hall field vanishes and the resistance increases as H^2, being approximately doubled when $\omega_c \tau = 1$. In a field of $1\,\mathrm{kG}$ there is no Hall field and the resistance of Reed & Brennert's sample is increased about 25 times, so that we may judge $\omega_c \tau$ to be about 5. This is not inconsistent with what is known of the electronic parameters of zinc. In the range above $4\,\mathrm{kG}$, then, we have $\omega_c \tau > 20$.

Consider now a plane section, normal to **H**, of one sheet of the Fermi surface. If this were a section of the free electron sphere it would contribute to the current according to the equations

$$\sigma_{11} = \frac{\sigma_0}{1+\omega_c^2\tau^2}, \quad \sigma_{12} = \frac{\omega_c\tau\sigma_0}{1+\omega_c^2\tau^2},$$

or, when $\omega_c\tau \gg 1$, $\qquad \sigma_{11} \approx \sigma_0/(\omega_c^2\tau^2), \quad \sigma_{12} \approx \sigma_0/(\omega_c\tau),$ \hfill (60)
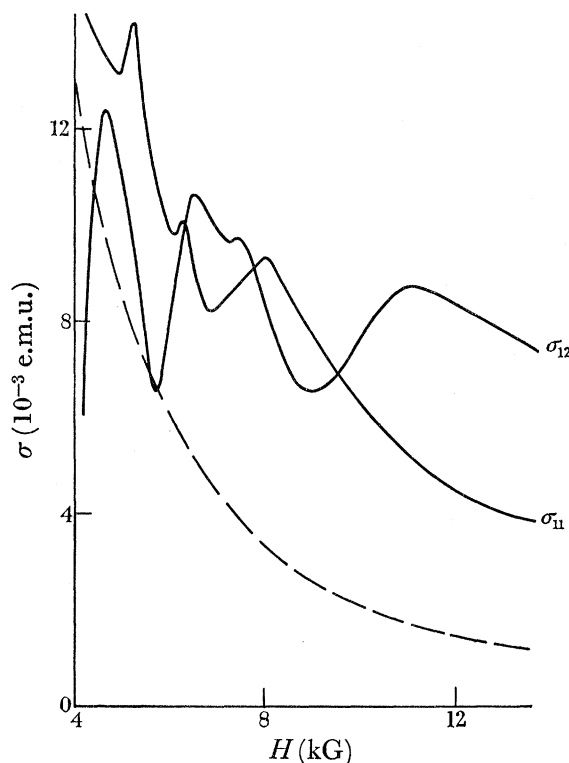
σ_0 being the zero-field conductivity.

FIGURE 18. Variation with H of σ_{11} and σ_{12} in basal plane of zinc, deduced from Reed & Brennert's data. The broken curve is the estimated contribution to σ_{11} of the majority of the Fermi surface.

In practice, of course, the sections are not circular, but it is easy to show that the form (60) still holds in the high field limit, though the meaning of σ_0 may be different (see, for example, Pippard 1962b, p. 94). We may now add separately the contributions of all parts of the Fermi surface which are electron-like and hole-like, and obtain conductivity components for each in the form

$$\sigma_{11} \approx \alpha N^2 e^2/(\sigma_0 H^2), \quad \sigma_{12} \approx Ne/H. \qquad (61)$$

Here we have made use of the well-known dependence of σ_{12} in the high-field limit only on the number of carriers, and we have introduced a numerical factor α to allow for the departure of the orbits from circular form. From (61) it is clear that as $H \to \infty$ the resistivity ρ_{11} tends to α/σ_0, and α is thus unity for the free-electron metal which shows no magneto-resistance. Single-carrier metals normally have only a small magneto-resistance, and even some which have electrons and holes, such as Al and In, saturate with a resistance only 2 or 3 times the zero-field value. It is therefore likely that a value of α no greater than 3 is needed to allow for the shape of the Fermi surface.

We may now write for the high-field conductivity components in a metal containing electrons and holes, and no open orbits,

$$\sigma_{11} = \frac{e^2}{H^2}\left[\left(\frac{\alpha N^2}{\sigma_0}\right)_+ + \left(\frac{\alpha N^2}{\sigma_0}\right)_-\right], \qquad \sigma_{12} \approx (N_- - N_+)e/H$$

$$= \frac{2e^2\overline{(\alpha N^2)}}{H^2\sigma_0}, \qquad\qquad\qquad = e\Delta N/H, \qquad (62)$$

in which σ_0 must now be interpreted as the zero-field conductivity of the complete metal. If we examine figure 18 in the light of these expressions we note that above $4\,\mathrm{kG}$ ΔN does not vanish, but oscillates about a mean value that implies the presence of more electrons than holes. It is the essence of the explanations put forward by Stark and by Reed & Brennert that breakdown turns some of the hole orbits (the hexagons of our network model) into free-electron orbits, and that the probability of breakdown is oscillatory with the field. This is certainly very likely, but cannot be satisfactorily checked in the network model without calculating the matrix elements for interband transitions. It is, however, not the whole story, for σ_{11} not only falls too slowly as H increases (it drops by a factor 3 instead of 9 as H goes from 4 to $12\,\mathrm{kG}$), but at the higher fields is clearly larger than can easily be accounted for from (62). At $12\,\mathrm{kG}$, if we take σ_0 to be $2\cdot 6\,\mathrm{e.m.u.}$ (corresponding to $S = 15\,000$), $(\overline{\alpha N^2})^{\frac{1}{2}}$ deduced from σ_{11} is $5\cdot 8 \times 10^{22}$. Thus if we take α as 3, we need about $3\cdot 4 \times 10^{22}$ each of electrons and holes per cm^3, slightly over $\frac{1}{2}$ an electron and $\frac{1}{2}$ a hole per atom. This is much more than can be accounted for on the basis of present beliefs about the Fermi surface in Zn, which suggest about $\frac{1}{8}$ of each per atom as a maximum.[†] The broken curve in figure 18 indicates about how much of σ_{11} might be due to the major parts of the Fermi surface acting in accordance with (62). What is left over, the greater part at the higher fields, we suggest is due to the quasi-particles which are able to travel in straight lines through the metal and contribute to σ_{11} only. It may be noted that if σ_{11} followed the broken curve of figure 16 the resistance at $12\,\mathrm{kG}$ would be only 50 times the zero-field resistance instead of 130 times as observed by Reed & Brennert. It is therefore necessary to postulate some process additional to the conversion of hole orbits into electron orbits by breakdown.

When we consider the matter quantitatively, however, it is clear that the matter is not satisfactorily settled. The discrepancy lies in the number of quasi-particles needed to explain the experimental results, which is several times less than the theory predicts should be present. For the section of the zone which is capable of producing quasi-particles is determined by the necks of the monster, and is about $\frac{1}{17}$ of the height of the zone. According to the network model the maximum value of $\overline{\sigma}/\sigma_0$ should then allow the quasi-particles to achieve the conductivity of a free-electron gas of such density as could fill $1\,\%$ of the zone, i.e. $0\cdot 01$ electron per atom. With the same relaxation time as in the real metal, containing 2 electrons per atom but somewhat diminished in conductivity by zone boundary effects, it

[†] In the accountancy of electrons and holes for this purpose one must not count the monster as a hole surface alone, for the orbits round the inside of the monster are electron-like. To give a simple example, a toroidal hole surface must be allowed as many hole states as are contained both within the torus and within the space enclosed by planes closing the ends, and as many electron states as are contained within the latter space, the volume of the torus itself representing the excess number of hole states.

is unlikely that the quasi-particles should have a conductivity much worse than $0.01\sigma_0$, or 26×10^{-3} e.m.u. This is some five times larger than is needed to account for the experimental results, and it is hard to justify the hope that the calculations might be inexact to this degree.

It is not worth pursuing the question further at this stage. We have had to take experimental data from published curves rather than the primary information, and have noted considerable discrepancies between different workers. It is unlikely that this is the sole reason for the disagreement between theory and experiment, but until new data have been obtained under very carefully controlled conditions, so that the conductivity components are reliably known, it is probably futile to attempt to gauge the success or failure of the theory. It is encouraging, however, to note its considerable success in the simpler question of accounting for de Haas–van Alphen amplitudes.

One final point may be made, concerning the fine structure of the resistance oscillations. It was suggested by Stark that the de Haas–van Alphen levels of the needle orbits are split by reason of the electronic spin, and that a large g-factor of 34 causes this splitting to be resolvable. If one looks at the observed line shapes it is hard to maintain this contention, for they cannot be analyzed into two simple sets of lines shifted relative to one another, as would be expected if spin were the cause. But the calculations exhibited in figure 10 show a structure in $\sigma(x_0)$ just like what is needed to produce the observed linesplitting, and it seems more reasonable to look for the explanation here, and to point to the fine structure as further evidence that the general nature of the results for the network model are in some sort of correspondence with the truth.

APPENDIX

We consider a small orbit which is coupled by breakdown to n identical large orbits, as indicated in figure 19 for the case $n = 3$. A typical normal mode of the system will have the same amplitude on each large orbit and a constant amplitude round the small orbit, but with a regular phase progression θ from one large orbit to the next. Thus if the amplitude of the small orbit is taken as unity, and on the large orbits as A, and if the phase length of the small orbit is $n\phi$ and of the large orbits $N\phi$ (ϕ depends on the electronic energy), the amplitudes at a junction are as shown in the figure, and the compatibility equations are as follows:

$$e^{i(\theta-\phi)} = pA\,e^{iN\phi} + q,$$

$$A = p + qA\,e^{iN\phi}.$$

Hence
$$\omega = \tan^{-1}\left\{\frac{\cos\epsilon - q}{\sin\epsilon}\right\}, \tag{A 1}$$

where $\epsilon = \phi - \theta$ and $\omega = \frac{1}{2}(N\phi - \epsilon)$.

Not surprisingly, this result implies a spectrum of sharp levels for all values of q, and one can trace a given level through from $q = 0$ to $q = 1$, the value of θ being constant on each level. Each value of θ gives an identical set of levels, shifted relative to one another so that when $q = 1$ the quantized levels of the small orbit ($\epsilon = 0$) occur when $\phi = \theta$. There are thus n such levels in one cycle of ϕ, and to determine the strength of the de Haas–van Alphen oscillations of the small orbit we must find the nth Fourier component of the level density considered as a function of ϕ. Since each value of θ gives the same structure we may as well consider only the case $\theta = 0$, for which (A 1) may be written

$$\tfrac{1}{2}(N-1)\phi = \tan^{-1}\left\{\frac{\cos\phi - q}{\sin\phi}\right\}. \tag{A 2}$$
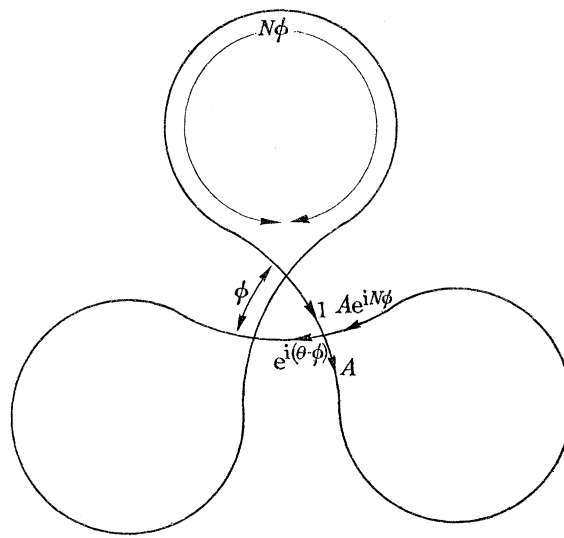

FIGURE 19. Small orbit coupled by breakdown to three identical large orbits.

Now when N is very large the levels are very closely spaced; when $q = 0$, for example, they are evenly spaced at such values of ϕ that $(N+1)\phi/\pi$ is an odd integer. To see how much the lines are displaced when ϕ is around ϕ_0 and $q \neq 0$ it is sufficient to substitute ϕ_0 for ϕ in the right-hand side of (A 2), and hence arrive at an expression for the displacement

$$\delta\phi(\phi_0) = \frac{2}{N-1}\tan^{-1}\left\{\frac{\cos\phi_0 - q}{\sin\phi_0}\right\} - \tfrac{1}{2}\pi, \tag{A 3}$$

always taking the same branch of the inverse tangent. From (A 3) we derive the separation of neighbouring lines

$$\Delta\phi = \pi[1 + \mathrm{d}(\delta\phi)/\mathrm{d}\phi_0],$$

and hence the level density $\rho(\phi_0)$, which is inversely proportional to $\Delta\phi$:

$$\rho(\phi_0) \propto 1 - \mathrm{d}(\delta\phi)/\mathrm{d}\phi_0 \quad \text{if} \quad N \gg 1$$

$$\propto 1 + \frac{2}{N-1}\frac{1 - q\cos\phi_0}{1 + q^2 - 2q\cos\phi_0}.$$

For the amplitude of the de Haas–van Alphen oscillation, I, we form

$$\int_{-\pi}^{\pi}\rho(\phi_0)\cos n\phi_0\,\mathrm{d}\phi_0,$$

i.e.
$$I \propto \frac{1-q^2}{q} \int_{-\pi}^{\pi} \frac{\cos n\phi_0 \, d\phi_0}{c - \cos \phi_0}, \quad \text{where} \quad c = \frac{1+q^2}{2q},$$

$$\propto q^n.$$

When no breakdown occurs ($q = 1$) we may take I as unity, so that q^n is the factor ζ by which breakdown reduces the de Haas–van Alphen amplitude of the small orbit.

References

Alekseyevski, N. E. & Gaidukov, Yu. P. 1962 *J.E.T.P.* **43**, 2094.

Berlincourt, T. G. & Steele, M. C. 1954 *Phys. Rev.* **95**, 1421.

Blount, E. I. 1962 *Phys. Rev.* **126**, 1636.

Cohen, M. H. & Falicov, L. 1961 *Phys. Rev. Lett.* **7**, 231.

Dhillon, J. S. & Shoenberg, D. 1955 *Phil. Trans.* A, **248**, 1.

Dingle, R. B. 1952 *Proc. Roy. Soc.* A, **211**, 500.

Grenier, C. G., Reynolds, J. M. & Zebouni, N. H. 1963 *Phys. Rev.* **129**, 1088.

Harper, P. G. 1955 *Proc. Phys. Soc.* A, **68**, 879.

Harrison, W. A. 1962 *Phys. Rev.* **126**, 497.

Joseph, A. S. & Gordon, W. L. 1962 *Phys. Rev.* **126**, 489.

Kohn, W. 1959 *Phys. Rev.* **115**, 1460.

Onsager, L. 1952 *Phil. Mag.* **43**, 1006.

Pippard, A. B. 1962*a* *Proc. Roy. Soc.* A, **270**, 1 (Paper I).

Pippard, A. B. 1962*b* The dynamics of conduction electrons. In *Low temperature physics* (ed. De Witt, Dreyfus and de Gennes). New York: Gordon and Breach.

Priestley, M. G. 1963 *Proc. Roy. Soc.* A, **276**, 258.

Reed, W. A. & Brennert, G. F. 1963 *Phys. Rev.* **130**, 565.

Renton, C. A. 1961 *Proc. 7th Int. Conf. Low Temp. Phys.*, p. 216 (ed. Graham and Hollis Hallett), University of Toronto Press.

Stark, R. W. 1962 *Phys. Rev. Lett.* **9**, 482.