Tunneling from an Independent-Particle Point of View

WALTER A. HARRISON
General Electric Research Laboratory, Schenectady, New York
(Received February 20, 1961)

A method is developed for calculating wave functions through regions of varying band structure. This method is applied to tunneling problems using the transition-probability approach of Bardeen. It is found that the experiments of Giaever involving tunneling into superconductors cannot be understood strictly in terms of an independent quasi-particle model of the superconductor. The observed proportionality of the tunneling probability to the density of states depends upon the matrix elements being constant which, in turn, depends upon a many-particle feature of the problem. This feature does not carry over to fluctuations in the density of states arising from band structure, and contributions to the current are not expected to be proportional to the density of states in that case. Instead, a projection in wave-number space of the appropriate constant-energy surface enters. Tunneling systems are discussed which involve semiconductors, semimetals, and transition metals as well as simple metals. Finally, alterations in the properties arising from alterations in the nature of the boundary regions are discussed.

I. INTRODUCTION

BARDEEN[1] has discussed tunneling from a many-particle point of view. He did not, however, discuss systems in which the band structure varies with position. In the following treatment we restrict ourselves to an independent-particle approximation, but extend the work to allow for variations in the local band structure.

We proceed by first developing a method for construction of one-particle wave functions, taking particular care that these conserve current locally. Matrix elements are then calculated in terms of these functions and the tunneling current determined.

Within this framework we consider the Giaever experiment[2] (tunneling into superconductors) using an independent-particle model of the superconductor. The failure of this approach sheds some further light on the treatment by Bardeen.

The same formulas are applied to systems involving tunneling into semiconductors, semimetals, and transition metals as well as into simple metals. Finally, we discuss the dependence of the tunneling properties on the nature of the boundary between different regions of the system.

II. DETERMINATION OF WAVE FUNCTIONS

We consider first a simple system which is divided by the plane $x=0$ into two regions of known band structure, and an electron energy which lies within the allowed bands in both regions. If suitable boundary conditions were applied to either one of these regions alone, we could construct a complete set of Bloch functions corresponding to the appropriate band structure. We will assume that wave functions for the composite system may be given, to a good approximation, by a linear combination of an incoming and an outgoing Bloch function on each side of the boundary, with each of these waves having the same component of wave number parallel to the boundary. This latter condition corresponds to an assumption of specular reflection and transmission by the boundary and will be relaxed in Sec. VI. This cannot be an exact eigenfunction for general band structure since we are required to match the wave functions on the entire plane $x=0$, but have only the coefficients of four waves at our disposal.

The problem of constructing wave functions is now reduced to the problem of obtaining two matching conditions upon the wave function; these, in conjunction with normalization and external boundary conditions, will uniquely determine the eigenstates. It will be a great mathematical simplification to assume reflection symmetry in the band structure at all points; this would correspond to time-reversal symmetry in a one-dimensional problem.

It will be convenient to represent the *Bloch* waves by *plane* waves of the same wave number, and to represent linear combinations of Bloch waves by linear combinations, ϕ, of the corresponding plane waves. Since it is the *Bloch* functions which are to be matched smoothly at the boundary, the function ϕ may not be smooth at the boundary. We write generalized boundary conditions:

$$\beta\phi \text{ continuous}, \quad \alpha\partial\phi/\partial x \text{ continuous}, \tag{1}$$

across the boundary, where α and β may depend upon energy and transverse component of wave number and are different for different band structures. These parameters may be related to the current-density operator by noting that we may associate

$$J_x=\frac{\hbar}{2mi}\left[\phi^*\alpha\frac{\partial}{\partial x}\beta\phi-\phi\alpha\frac{\partial}{\partial x}\beta\phi^*\right], \tag{2}$$

with the x component of current density. This is permissible since with $\alpha=\beta=1$, this J_x is the current density for free electrons and J_x is conserved across all boundaries between differing band structures. The particular ordering of terms will be appropriate when the equations are generalized to continuously varying band structures and will guarantee the conservation of current in the volume of the material.

We may identify the velocity associated with a state of wave number k_x as $\hbar\alpha\beta k_x/m$ by noting the form of the current density, (2). This, in turn, must equal $\partial H/\partial p_x$

[1] J. Bardeen, Phys. Rev. Letters **6**, 57 (1961).
[2] I. Giaever, Phys. Rev. Letters **5**, 147, 464 (1960).

$=(1/\hbar)\partial E/\partial k_x$ in a one-particle approximation. Thus the product $\alpha\beta k_x$ may be evaluated in terms of the band structure as

$$\alpha\beta k_x=(m/\hbar^2)\partial E/\partial k_x. \tag{3}$$

Equation (3) will provide sufficient information about the matching parameters for most of our treatment, but further knowledge will be needed when we treat specialized boundary regions in Sec. VI.

In terms of these parameters the problem of obtaining wave functions for a system in which the band structure changes discontinuously at planes of constant x is completely determined. Within regions of constant band structure, the x variation of ϕ is obtained from the equation

$$\partial^2\phi/\partial x^2+k_x^2\phi=0$$

(noting the assumed reflection symmetry in the band structure). At planes of discontinuity we apply the conditions, (1).

It may be noted that the matching conditions introduce discontinuities in the slope and value of ϕ, and that these are required in order to conserve electron flux at the boundaries. The method strongly resembles the earlier treatment by the author[3] of wave functions in perturbed monovalent metals utilizing a cellular method. In that case variations of band structure in three dimensions were allowed.

The method is readily generalized to continuous variations of the band structure by taking the planes of discontinuity closer and closer together and, in the limit, obtaining a differential equation for the x variation of ϕ;

$$\frac{\partial}{\partial x}\frac{\alpha}{\beta}\frac{\partial}{\partial x}\phi+\beta\alpha k_x^2\phi=0. \tag{4}$$

Here ϕ, α, β, and k_x are functions of position which are assumed to be slowly varying over a single atomic cell. The latter three need not be slowly varying over a wavelength of ϕ; that is, we have not yet made a WKB approximation.

In order to treat problems involving tunneling, we must extend the above treatment to electron energies which lie in the forbidden band. We do this simply by analytically continuing the band structure ($E[\mathbf{k}]$) and the parameters α and β into the space of complex $\mathbf{k}$. This is the procedure used by James[4] in treating one-dimensional problems and by Peterson[5] in developing a crystalline WKB approximation.

III. CALCULATION OF THE TUNNELING CURRENT

We now have a well-defined procedure for constructing wave functions, and we may proceed to calculate eigenstates and the tunneling current. We will use the very convenient procedure developed by Bardeen.[1]

Bardeen has written the probability per unit time of the transition of an electron in a state a on one side of the tunneling region to a state b on the other side:

$$P_{ab}=(2\pi/\hbar)|M_{ab}|^2\rho_b f_a(1-f_b),$$

where M_{ab} is the matrix element for the transition, ρ_b is the density of states at b, and f_a and f_b are the probabilities of occupation of the states a and b, respectively. Further, he shows that we may write the matrix element in terms of the matrix element of the current operator between states a and b which are constructed to continue to drop exponentially beyond the tunneling region.

$$M_{ab}=-i\hbar J_{ab}, \tag{5}$$

where the x component of the current density operator J_{ab} is to be evaluated in the tunneling region between the states a and b which decay in opposite directions.

M_{ab} vanishes unless the transverse wave number k_t is the same for the initial and final states (specular transmission); thus ρ_b is a density of states for fixed k_t. We sum over all states a of fixed k_t, sum over k_t, multiply by 2 for spin and multiply by the electronic charge e to obtain the total current to the right. Subtracting the current to the left, we have finally

$$j=\frac{4\pi e}{\hbar}\sum_{k_t}\int_{-\infty}^{\infty}|M_{ab}|^2\rho_a\rho_b(f_a-f_b)dE. \tag{6}$$

The integral over energy is taken at fixed transverse wave number k_t.

In evaluating the matrix elements, we construct states which are sinusoidal in a positive-energy region and drop exponentially in an adjacent negative-energy region. We assume that, except near the transition region, the band structure is uniform. We assume for the time that the band structure is slowly varying in the transition region and make a WKB approximation. In Sec. VI we consider the consequences of making an abrupt approximation instead.

The WKB approximation is readily applied to Eq. (4) by assuming α, β, and k_x vary slowly in the x direction. The wave functions are found to be of the form

$$(\alpha\beta k_x)^{-\frac{1}{2}}\exp\left(i\int^x k_x dx\right),$$

except near turning points. The occurrence of the factor $(\alpha\beta k_x)^{-\frac{1}{2}}$ could have been forseen from the form of the current operator and the requirement of current conservation. Connection formulas across regions where $\alpha\beta k_x$ is small are the same as the usual ones. If x_a is the point on the left of the tunneling region at which $\alpha\beta k_x=0$, the x variation of a left-hand state is given by

[3] W. A. Harrison, Phys. Rev. **110**, 14 (1958).
[4] H. M. James, Phys. Rev. **76**, 1602 (1949).
[5] G. A. Peterson, Bull. Am. Phys. Soc. **5**, 161 (1960).

$$C_a(\alpha\beta k_x)^{-\frac{1}{2}}\cos\left(\int_x^{x_a} k_x dx+\gamma_a\right), \quad x<x_a,$$
$$(\tfrac{1}{2}C_a)(\alpha\beta k_x)^{-\frac{1}{2}}\exp\left(-\int_{x_a}^{x}|k_x|dx\right), \quad x>x_a, \tag{7}$$

where k_x is chosen to make the local energy independent of position. C_a is to be determined by normalization; if L_a is the extension of the crystal to the left of x_a, then

$$C_a^2=2(\alpha\beta k_x)_a/L_a.$$

Similarly, right-hand states may be constructed by replacing the subscript a by b.

We may now evaluate the matrix element for transition using its relation to the current-density matrix element [Eq. (5)], the form of the current density operator [Eq. (2)], and the wave function [Eq. (7)].

$$|M_{ab}|^2=\left(\frac{\hbar^2}{2m}\right)^2\frac{(\alpha\beta k_x)_a}{L_a}\times\frac{(\alpha\beta k_x)_b}{L_b}\exp\left(-2\int_{x_a}^{x_b}|k_x|dx\right). \tag{8}$$

This is to be substituted into Eq. (6) for the current density;

$$j=\frac{2e}{h}\sum_{k_t}\int_{-\infty}^{\infty}\exp\left(-2\int_{x_a}^{x_b}|k_x|dx\right)(f_a-f_b)dE. \tag{9}$$

In the determination of (9) the density of states was written as

$$\rho=(L/\pi)(\partial E/\partial k_x)^{-1}. \tag{10}$$

We notice the conspicuous absence of the density of states factor in Eq. (9). This is a direct consequence of our independent-particle model and the resultant reciprocal relation between the particle velocity [Eq. (3)] and the density of states [Eq. (10)]. We will see, in Sec. VI, that the exact form of the integrand in the expression for the current depends upon the nature of the barrier region, but only a quite artificial model would restore the simple proportionality to the density of states.

IV. TUNNELING INTO SUPERCONDUCTORS

The absence of the densities of states in the integrand in Eq. (9) is contrary to the experimental results of Giaever and to the result given by Bardeen. The inconsistency results from a failure of the independent-particle model. In order to obtain agreement with experiment, we may (following Bardeen) assert that the density of states is that for quasi-particles, but the current entering the matrix element is given by the value for normal metals. From the independent-particle point of view, this violates charge conservation at the boundary; from the many-particle point of view this is the assertion that back-flow around the quasi-particles should be neglected. In either framework, the essential problem is gauge invariance, and its resolution is beyond the realm of the independent-particle model. It seems remarkable that the simple experimental result depends so directly upon the subtleties of the many-particle system.

In any case, we do not expect this essential breakdown of the independent-particle model to occur in tunneling involving normal metals, semimetals, and semiconductors. Neither, then, do we expect to find the simple proportionality of the ac conductance to the density of states.

V. TUNNELING INTO NONSUPERCONDUCTING MATERIALS

We may obtain a more informative description of Eq. (9) by replacing the sum over k_t by an integral over the projection S of a constant energy surface onto the plane of the barrier; i.e., $d^2k_t=dS$. After rearranging factors, we obtain for the current per unit area,

$$j=\frac{e}{2\pi^2\hbar}\int_{-\infty}^{\infty}dE(f_a-f_b)\int dS\, e^{-\eta}, \tag{11}$$

with

$$\eta=2\int_{x_a}^{x_b}|k_x|dx.$$

η depends on energy and on the transverse wave number associated with dS. The second integral is over regions of k_t corresponding to positive-energy states on both sides. Thus if we define the "shadow" of a constant-energy surface to be its projection in wave-number space on a plane parallel to the barrier, the integral over S is over the overlap of the shadows from the two sides for the energy E.

The current between corresponding energy shells on the two sides of the barrier is proportional to the difference in the probability of occupation of the states in the two shells and to a surface integral over the overlap of their shadows. If one or both of the energy surfaces is small, as in a semiconductor or a semimetal, η becomes essentially constant and the integral over S is simply the exponential times the area of the shadow overlap. If, on the other hand, both of the surfaces are large, as when both sides are simple metals, the exponential becomes small for high angles of incidence. Only electrons moving approximately normal to the surface tunnel and the integral is equal to the value of the exponential at normal incidence times that area of the constant-energy surface for which appreciable numbers of electrons tunnel. This area depends only upon the properties of the tunneling region and may be readily evaluated for simple models by expanding $|k_x|$ for small k_t and integrating over S.

A. Simple Metals

The integral over S in Eq. (9) is a constant and the integral over energy is equal simply to the difference in

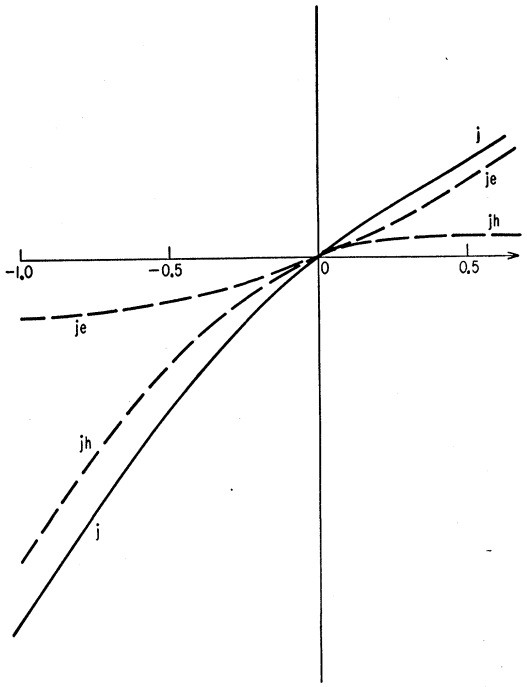

FIG. 1. The current-voltage characteristic for tunneling from a metal into an idealized semimetal. The hole mass was taken equal to three times the electron mass. Voltages are measured in units of the electron Fermi energy; the scale of the ordinate is arbitrary. The currents associated with the electron and the hole band are shown separately as well as the total current.

Fermi energy (or to the applied voltage). The behavior is Ohmic as observed by Giaever for normal metals.

B. Semimetals

We consider the tunneling between a simple metal and a semimetal. There is Fermi surface in both the electron band and the hole band and two contributions must be added. We calculate a sample characteristic neglecting the dispersal of the Fermi surface over the Brillouin zone and take η the same for both holes and electrons; we assume spherical surfaces and an effective-mass approximation for both signs of carriers with the result shown in Fig. 1.

We note that the irregularities are somewhat "washed out" and would be further reduced by the inclusion of other bands. The success of the nearly free-electron approximation for treating bismuth[6] suggests that there will be many other band edges near the Fermi energy and that the collection of all bands resembles a single free-electron band. This would suggest a generally linear characteristic with minor fluctuations.

C. Transition Metals

The very large fluctuations in the density of states in this case would have provided spectacular results if the tunneling were proportional to the density of states. However, the Fermi surfaces associated with both d and s bands are expected usually to be large. The current into each is determined by an area of surface which depends only on the properties of the insulating barrier, so the contributions from the two bands should be comparable. Only near band edges should there be fluctuations in the characteristic and these should be more pronounced near a d-band edge since the energy surface and its shadow area change more rapidly with energy in that case. Such irregularities have not been observed.

D. Esaki Diode

The application to the tunnel diode is straightforward. The band structure is continued into the plane of complex $\mathbf{k}$ in the evaluation of η. In this case, with a reasonable model of the tunneling region, η will depend upon the applied voltage.

An interesting feature of the analysis is the failure of the one-dimensional density of states to enter the integral. The same feature occurs in the treatment of tunneling as a Zener current in a uniform field.[7] In that analysis, an artificial density of states enters which depends only on the applied field and the magnitude of a reciprocal lattice vector.

VI. NATURE OF THE TRANSITION REGION

It has been necessary in this treatment to make two specific assumptions about the boundary between different regions of the system. We have assumed that this region is sufficiently planar that the transmission of the barrier is specular. We have also assumed that the band structure varies sufficiently slowly that we may make a WKB approximation.

The experiments on superconductors shed no light on either of these questions. The one-dimensional density of states which enters Eq. (6) is proportional to the three-dimensional density of states which would enter if the transmission were diffuse. Furthermore, the assertion that the matrix element is the same for the superconducting and normal metals is sufficient to assure its relative constancy whether the boundary is gradual or sharp.

In the tunnel diode it is reasonable to expect that both assumptions are valid and the question remains open with respect to the boundary region with deposited or oxidized films. The observed diffuse reflection at metallic surfaces[8] would suggest diffuse transmission also. This is not conclusive, however, since macroscopically diffuse reflection is not inconsistent with microscopic specular reflection if the transition region is irregular.

We consider briefly some modifications expected if

[6] W. A. Harrison, J. Phys. Chem. Solids **17**, 171 (1960).

[7] E. O. Kane, J. Phys. Chem. Solids **12**, 181 (1959).

[8] A. H. Wilson, *The Theory of Metals* (Cambridge University Press, New York, 1954), p. 248.

either of these assumptions as to the nature of the boundary is not valid.

A. Sharpness of the Boundary

If the boundary region were sharp, we should proceed with the calculation of the wave functions by matching rather than with the WKB approximation. The boundary conditions of Eq. (1) are directly applicable. We assume sharp boundaries and uniform band structures in each of the three regions, numbered 1, 2 and 3 (the tunneling region being numbered 2). We obtain

$$|M_{ab}|^2=\left(\frac{\hbar^2}{2m}\right)^2\frac{1}{L_1}\frac{1}{L_3}\frac{16\beta_1^2\nu_2^2\beta_3^2}{(\nu_1^2+\nu_2^2)(\nu_2^2+\nu_3^2)}e^{-\eta},$$

where

$$\nu_i=\beta_i/\alpha_i k_{xi}. \tag{12}$$

When multiplied by the density-of-states factors, this becomes

$$|M_{ab}|^2\rho_a\rho_b=\frac{4}{\pi^2}\frac{\nu_1\nu_2^2\nu_3}{(\nu_1^2+\nu_2^2)(\nu_2^2+\nu_3^2)}e^{-\eta}. \tag{13}$$

In contrast to the result of the WKB approximation, this is not independent of energy. Furthermore, the dependence upon the matching parameters is not simply through the product $\alpha\beta k$, which we were able to evaluate in terms of the band structure. We must consider these parameters in more detail.

The complete Bloch function may be written as $u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$. If $u_{\mathbf{k}}$ is normalized such that its average value is unity, the factor $e^{i\mathbf{k}\cdot\mathbf{r}}$ may be associated with our function ϕ. When the band structure changes, we must match the *Bloch functions* at the cell boundaries, whereas our matching condition is that $\beta\phi$ be matched at the boundary. We may therefore associate β with the value of the normalized $u_{\mathbf{k}}$ at the cell boundary.

$$\beta^2=\frac{|u_k(\text{cell})|^2\Omega}{\int_{\text{cell}}|u_{\mathbf{k}}(\mathbf{r})|^2d\tau}.$$

Here Ω is the cell volume, β is of order unity at the edge of an s band and zero at the edge of a p band, and β^2 is equal to the parameter γ introduced by Bardeen[9] in his extension of the Wigner-Seitz cellular method. Our matching conditions are equivalent to those used in the earlier extension of the cellular method to nonperiodic structures.[3]

From Eqs. (3), (10), and (12) we see that ν is proportional to β^2 times a density of states. Thus, Eq. (13) indicates that the density of states has reappeared in the integrand for the tunneling current. However when the density of states becomes sufficiently large, it dominates in the denominator and the result is *inversely* proportional to the density of states.

If we were to attempt to use the form of Eq. (13) to obtain agreement with the superconducting experiments within an independent-particle approximation, it would be necessary to take β in the superconductor (and in the same metal when it is normal) to be negligible compared to that in the tunneling region. This seems quite inappropriate and we maintain our conclusion that the independent-particle model fails in that case.

The dependence upon the density of states given in Eq. (13) offers some hope for interesting behavior with semimetals and transition metals, but the hope is weak. This dependence on the details of the wave functions seems unrealistic when the boundary is not ideal. Further, a sample comparison between an exact calculation and a WKB approximation for free-electron tunneling indicates that the WKB approximation becomes quite good as the thickness of the transition approaches and exceeds a single electron wavelength.

B. Diffusion of Transmission

There is no unique way to introduce a diffuse transmission. One might, however, expect three qualitative conclusions to hold in any case.

First, if the boundary is gradual we might still expect the cancellation of the density of states by the velocity in the matrix element even in the diffuse case since this arose from current conservation. In particular, we should not expect the simple proportionality to the density of states to reappear.

Second, we expect no appreciable modification of the results for simple metals (and superconductors) which have isotropic electronic properties.

Third, we expect significant differences in tunneling into semimetals and degenerate semiconductors depending upon the nature of the transmission since their Fermi surfaces lie in limited regions of wave-number space. However, the general smoothing of fluctuations because of the cancellation of the density of states and the contribution of many bands should remain.

[9] J. Bardeen, J. Chem. Phys. **6**, 367 (1938).