Tunneling from a Many-Particle Point of View*

RICHARD E. PRANGE
University of Maryland, Department of Physics and Astronomy, College Park, Maryland
(Received 22 March 1963)

The Hamiltonian for a system of interacting electrons in the presence of a barrier is transformed into one which can be separated into three parts. Two of the parts describe the electrons on the "right" and "left" sides of the barrier, while the third is a transition term which allows tunneling through the barrier. It is shown that it is not reasonable to make the "right"- and "left"-hand Hamiltonians commute, but the effect of this failure is easily accounted for in perturbation theory. In addition, corrections to the transition operator arising from interelectronic interaction are written down and discussed briefly. They may be both quantitatively and qualitatively important.

I. THE TUNNELING HAMILTONIAN

EXPERIMENTS in which electrons are observed to tunnel through insulating layers have played an increasingly important role in recent times.[1] The tunneling process has, of course, been well understood since the discovery of quantum mechanics. Bardeen[2] has very recently pointed out, however, that in the case of a many-electron system, the tunneling process is very conveniently regarded as in time-dependent perturbation theory. In other words, it is seen as a transition process from one set of nearly stationary states to another.

Let us keep in mind, to be definite, a system of interacting electrons enclosed in cubical boxes separated by a high, narrow barrier. It will be a convenient simplification in the later development if we suppose that the boxes and the barrier are symmetrically disposed, as in Fig. 1.

The most convenient formulation of the problem will involve showing that the complete Hamiltonian of the many electron system, including the insulating barrier, can be split into the form,

$$H=H_R+H_L+T. \tag{1}$$

Here H_R is the Hamiltonian of the electrons to the "right" of the barrier, H_L the Hamiltonian for the left-hand electrons, and T is the operator which makes possible the transition of electrons from left to right. It is perhaps clear intuitively from the work of Bardeen that such a formulation of the problem exists, and indeed, several papers have appeared in which the Hamiltonian (1) was taken for granted.[3]

It is the purpose of this note to show in what sense (1) is actually correct. Perhaps this is not an entirely empty exercise since an instructive difficulty has to be overcome. Furthermore, it is possible to discuss in detail the effect of interactions on the above formulations, which is always difficult to do in an intuitive fashion.

* This research was supported by the United States Air Force through the Air Force Office of Scientific Research.
[1] I. Giaever, Phys. Rev. Letters **5**, 147, 464 (1960).
[2] J. Bardeen, Phys. Rev. Letters **6**, 57 (1961).
[3] M. H. Cohen, L. M. Falicov, and J. C. Phillips, Phys. Rev. Letters **8**, 316 (1962); J. Bardeen, *ibid.* **9**, 147 (1962).

II. LOCALIZED STATES

The dilemma which arises is as follows: In defining the "right-hand" and "left-hand" Hamiltonians, one naturally needs to define right- and left-hand states. Now, first of all, these wave functions have to be in the domain of the original Hamiltonian operator, H. Otherwise, the new states will have to be expanded in old states of arbitrarily high energy. In the concrete model we have adopted this simply means that they must vanish at the sides of the boxes $|x|=a$, as well as being twice differentiable. Secondly, we would like the left-hand states to be complete on the left-hand side. This requirement insists that should an electron flow from a wire into the box to the left of $-b$, its wave function could be expressed in terms of left-hand states only. Thirdly, one should like, if possible, that all left-hand states be orthogonal to all right-hand states, in order that H_R should commute with H_L. Finally, the states have to form a complete set when taken all together.

If one could produce such a set of (single-electron) wave functions $\phi_k{}^r(x)$ and $\phi_k{}^l(x)$, it would be a simple matter to expand the annihilation operator as

$$\psi(x)=\sum a_k\phi_k{}^r(x)+b_k\phi_k{}^l(x). \tag{2}$$

One would then replace the creation and annihilation operators in the second-quantized form of H,

$$H=\int dx\,\psi^\dagger(x)[P^2/2m+V(x)]\psi(x) + \frac{1}{2}\int\int\psi^\dagger\psi^\dagger W\psi\psi \tag{3}$$

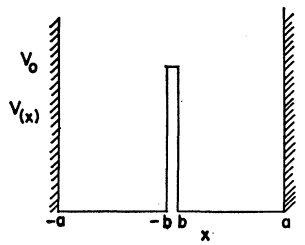

FIG. 1. Barrier potential $V(x)$ versus x.

by the expression (2). H_R could be identified as that part of H containing only "a" operators, H_L would contain only "b" operators, and T would be the remainder.

Unfortunately, no such states appear to exist. For, should all of the $\phi_k{}^l(x)$ vanish identically for positive values of the coordinate, then either they do not form a complete set for negative x, or they are too singular at $x=0$. And if they have support in the right-hand region, they cannot be orthogonal to the $\phi_k{}^r(x)$ should these functions be complete.

We are, in consequence, forced to compromise the requirements on the wave functions. There are two approaches which can be taken. The first possibility is to maintain the requirement of mutual orthogonality, but to abandon the requirement of separate "right" and "left" completeness. Instead, the right and left states will be made "as complete as possible" in their respective regions.

If this is done, a mathematically convenient formalism results, since the operator T can be treated systematically by ordinary perturbation theory. We shall show below a simple method for achieving this result. One must pay a price, however. In practice, the model which we are studying is not really of interest by itself. We also need a method of hooking up wires and batteries connecting the right-hand electrons with those on the left. Now wires and batteries are cumbersome to treat quantum mechanically, to say the least. It is essential to have a simple and convincing method of mathematically treating the classical part of the apparatus. These methods usually involve introducing electrons into the right-hand *states* and taking them out of the left-hand ones. However, the wires introduce electrons into the right-hand *side*. If the right-hand states are not complete, we have to take into account the possibility that electrons flowing into the right *side* partially go into the left-hand *states*. Methods may be advanced to get around this difficulty, but they seem rather contrived. The approach does contain the possibility of conveniently studying higher order effects, which is becoming fashionable.[4]

Another approach is to abandon the orthogonality condition, requiring instead the states to be "as orthogonal as possible," and to keep the separate completeness of the right and left states. In that case, the commutator $[H_R,H_L]$ is an operator of the order of magnitude of T.

III. ORTHONORMAL STATES

In the symmetrical model which we are considering, a particularly simple set of states springs to mind. The exact eigenstates of the single-particle Hamiltonian may be called $e_k(x)$, $o_k(x)$. They have even and odd parities under the reflection of the x coordinate, respectively. These states occur in pairs, the energy of the even (odd) state being $\epsilon_k(\mp)T_{kk}$, to an excellent approximation, where T_{kk} happens to be the diagonal element of Bardeen's transition operator,

$$T_{kk}=[2\epsilon_k(2m(V_0-\epsilon_k))^{1/2}/mV_0a] \times\exp[-2b(2m(V_0-\epsilon_k))^{1/2}],$$

and ϵ_k is the energy level calculated as if the barrier were very wide. With appropriate choices of phase for the $e_k(x)$, $o_k(x)$, one naturally defines

$$\sqrt{2}\phi_k{}^{r,l}(x)=e_k(x)\pm o_k(x). \tag{4}$$

The $\phi_k{}^r(x)$ are large only in the right-hand region, but there is a long tail (proportional to T_{kk}) in the "wrong" region. Defining the a and b operators according to (2), the Hamiltonian becomes

$$H=\sum_k \epsilon_k a_k{}^\dagger a_k+W_R+\sum_k \epsilon_k b_k{}^\dagger b_k+W_L -\sum_k T_{kk}(a_k{}^\dagger b_k+b_k{}^\dagger a_k)+W_T. \tag{5}$$

We have not bothered to write out the contributions from the interaction term. This is taken up below.

Neglecting the interaction terms then, we can calculate the time derivative of the number of particles in right-hand *states*. We find

$$\dot{N}_R=(N_L{}^0-N_R{}^0)T\sin 2Tt. \tag{6}$$

We have assumed that T_{kk} is slowly varying with k, and has average value T. We have also supposed that N_R+N_L is constant, and that at $t=0$, the numbers of particles present in the right (left) states is $N_{R(L)}{}^0$.

Formula (6) exhibits the oscillatory behavior well known from elementary quantum mechanics. Such behavior is not to be expected in the context of real metals, where one would anticipate finding a constant transition rate for not-too-long times. If (6) does not show a constant transition rate at any time, it can only be a result of the too clever definition of the states $\phi_k{}^r$. We have put electrons into the long tail of the $\phi_k{}^{r,l}$ on the "wrong" side of the barrier, already accomplishing the greater part of that tunneling which usually proceeds at a constant rate. If we look at the number of electrons in right-hand states we will not obtain a true picture of the tunneling, at least not without further arguments. By the same token, one must handle gingerly any states which are not confined rather strictly to one side or the other.

IV. NONORTHOGONAL STATES

The other alternative is to find a separation of the Hamiltonian, as in Eq. (1), based on states which are really confined to one side or the other. A natural set of states is given by the eigenstates of H_1, where

$$H_1=-\nabla^2/2m+V_1(x), \tag{7}$$

and

$$V_1(x)=V(x), \quad x>-b; \qquad =V_0, \quad x\leq -b. \tag{8}$$

[4] B. D. Josephson, Phys. Letters, **1**, 251 (1962).

This particular choice is convenient, but not mandatory. We call the eigenstates $\phi_k(x)$ and also define

$$\chi_k(x)=\phi_k(-x). \tag{9}$$

Each of the sets ϕ_k, χ_k is complete in the entire space ($a>|x|$). However, to represent a function which has large amplitude on the left by ϕ_k, it is necessary to employ large k values. Let us also introduce c_k and d_k operators according to

$$\psi(x)=\sum_k c_k\phi_k(x)=\sum_k d_k\chi_k(x). \tag{10}$$

At this stage, we want to rewrite the Hamiltonian in terms of the c's and d's in such a way that low-energy states of the original Hamiltonian are expressed in terms of correspondingly low-energy ϕ_k and χ_k functions. To that end, we employ the representation

$$e_k(x)=\sum_m[\phi_m(x)+\chi_m(x)]\lambda_{mk}, \tag{11}$$

$$o_k(x)=\sum_m[\phi_m(x)-\chi_m(x)]\mu_{mk}. \tag{12}$$

The expansion coefficients satisfy the equations

$$\lambda_{km}=(\phi_k,e_m)-\sum_n(\phi_k,\chi_n)\lambda_{nm}, \tag{13}$$

$$\mu_{km}=(\phi_k,o_m)+\sum_n(\phi_k\chi_n)\mu_{nm}, \tag{14}$$

which can be solved by iteration. We have next to compute (ϕ_k,e_m). This matrix element is diagonal in all quantum numbers except the momentum component conjugate to x. We shall suppress these other quantum numbers, as we have up to now.

The largest matrix elements between states of practically the same energy are found to be, after a tedious calculation,

$$\begin{aligned}(\phi_k,e_m)&=2^{-1/2}\delta_{k,m}+(\phi_k,e_m-2^{-1/2}\phi_m)\\&=2^{-1/2}[\delta_{km}-T_{kk}/(\epsilon_k-\epsilon_m)+\tfrac{1}{2}\xi_{km}+\cdots],\end{aligned} \tag{15}$$

$$(\phi_k,o_m)=2^{-1/2}[\delta_{km}+T_{kk}/(\epsilon_k-\epsilon_m)-\tfrac{1}{2}\xi_{km}+\cdots], \tag{16}$$

where we have introduced

$$\xi_{km}=(\phi_k,\chi_m). \tag{17}$$

The dots represent terms of higher order in T_{kk}.

In addition, there are large matrix elements (ϕ_k,e_m) when k is so large that the wavelength on the *left* is practically given by m. In that region, however, we find $(\phi_k,e_m)\simeq 2^{-1/2}(\phi_k,\chi_m)$.

Thus, we find to sufficient approximation, that

$$\lambda_{kl}=2^{-1/2}[\delta_{kl}-T_{kl}/(\epsilon_k-\epsilon_l)-\tfrac{1}{2}\xi_{kl}], \tag{18a}$$

$$\mu_{kl}=2^{-1/2}[\delta_{kl}+T_{kl}/(\epsilon_k-\epsilon_l)+\tfrac{1}{2}\xi_{kl}], \tag{18b}$$

and there are no large matrix elements to high-energy states.

In the remainder of this section, we neglect all interelectron interaction. Substitution of expressions (10)–(12) and (18) into the Hamiltonian yields at once

$$\begin{aligned}H=&\sum_{k,l,n}(\epsilon_k-T_{kk})\lambda_{km}(c_m^\dagger+d_m^\dagger)\lambda_{kn}(c_n+d_n)\\&+(\epsilon_k+T_{kk})\mu_{km}(c_m^\dagger-d_m^\dagger)\mu_{kn}(c_m-d_m)\\=&\sum_k\epsilon_k(c_k^\dagger c_k+d_k^\dagger d_k)\\&-\sum T_{kl}(c_k^\dagger d_l+d_l^\dagger c_l)-\frac{\epsilon_k+\epsilon_l}{2}\xi_{kl}[c_k^\dagger d_l+d_l^\dagger c_k]\\&+\sum_{kmn}\frac{\epsilon_k T_{kn}T_{mk}}{(\epsilon_k-\epsilon_n)(\epsilon_k-\epsilon_m)}(c_m^\dagger c_n+d_m^\dagger d_n)+\cdots.\end{aligned} \tag{19}$$

The last term, as well as others indicated by dots, may be dropped if we wish to keep only the terms of lowest order in the tunneling exponential. The desired separation of the single-particle Hamiltonian has, thus, been achieved with

$$H_R=\sum\epsilon_k c_k^\dagger c_k, \tag{20a}$$

$$H_L=\sum\epsilon_k d_k^\dagger d_k. \tag{20b}$$

However,

$$[H_R,H_L]=\sum_{r,s}\epsilon_r\epsilon_s\xi_{rs}(c_r^\dagger d_s-d_r^\dagger c_s), \tag{21}$$

which is of the order of T_{rs}. This means that the eigenstates of H_R+H_L are not just the products of eigenstates of H_R, H_L separately.

Let us, for example, calculate the current through the barrier, J, where

$$J_{\rm op}=\frac{e}{2mi}\int dydz\left[\psi^\dagger(xyz)\frac{\partial}{\partial x}\psi(xyz)-\left(\frac{\partial}{\partial x}\psi^\dagger(xyz)\right)\psi(xyz)\right]_{x=0}. \tag{22}$$

It may readily be shown that

$$J_{\rm op}=-\frac{e}{i}\sum_{kk'}(c_k^\dagger d_{k'}-d_{k'}^\dagger c_k)T_{kk'} \tag{23}$$

by the methods of Bardeen.[2] We propose to calculate

$$J=(\Phi(t),J_{\rm op}\Phi(t)), \tag{24}$$

where $\Phi(t)$ is the state reducing at $t=0$ to the product state

$$\Phi(0)=\Phi_R\Phi_L|0\rangle, \tag{25}$$

$\Phi_R|0\rangle$ being an eigenstate of H_R.

It is easy to verify that the failure of H to commute with H_L is just compensated by the term containing ξ_{kl} in (19). Indeed,

$$\begin{aligned}&(H_R+H_L)\Phi_R\Phi_L|0\rangle\\&\quad=\Phi_L H_R\Phi_R|0\rangle+\Phi_R H_L\Phi_L|0\rangle\\&\qquad+\sum\epsilon_k\xi_{kn}(c_k^\dagger d_n+d_k^\dagger c_n)\Phi_R\Phi_L|0\rangle\end{aligned} \tag{26}$$

to the first-order tunneling. In consequence, we may drop that term, at the same time assuming $\{c_r^\dagger,d_l\}=0$. This leads to the usual formula for the current proportional to $|T_{kl}|^2$.

V. EFFECT OF INTERACTIONS

We turn next to the problem of interactions. We treat here for simplicity only short-range interactions of simple type. (The range should be small even in comparison with the width of the barrier.)

Perhaps the easiest way to write down the terms of the interaction Hamiltonian is to employ the representation, obtained from (11), (12), and (18),

$$\psi(x)\approx\sum_n \tilde{\phi}_n c_n+\sum_n \tilde{\chi}_n d_n. \tag{27}$$

We have introduced the notation

$$\tilde{\phi}_n=\phi_n-\sum_m \xi_{mn}\chi_m. \tag{28}$$

Here ξ_{mn} is given by (17) with the understanding that the matrix elements connecting states of widely different energies are to be dropped.

We may immediately write down the interaction terms

$$\begin{aligned}W=&\{\tfrac{1}{2}\sum W_{pqrs}{}^{(0)}c_p{}^\dagger c_q{}^\dagger c_r c_s+\tfrac{1}{2}\sum W_{pqrs}{}^0 d_p{}^\dagger d_q{}^\dagger d_r d_s\}\\&+\{\tfrac{1}{2}\sum W_{pqrs}{}^{(1)}[c_p{}^\dagger d_q{}^\dagger d_r d_s+d_p{}^\dagger c_q{}^\dagger c_r c_s+\text{H.c.}]\}\\&+\{\tfrac{1}{2}\sum W_{pqrs}{}^{(2)}c_p{}^\dagger c_q{}^\dagger d_r d_s+\text{H.c.}\}\\&+\sum W_{pqrs}{}^{(3)}c_p{}^\dagger d_q{}^\dagger d_r c_s.\end{aligned} \tag{29}$$

Here, we have written

$$W_{pqrs}{}^{(0)}=\int\int W(x-y)\tilde{\phi}_p(x)\tilde{\phi}_q(y)\tilde{\phi}_r(y)\tilde{\phi}_s(x), \tag{30a}$$

$$W_{pqrs}{}^{(1)}=\int\int W(x-y)\times[\tilde{\phi}_p(x)\tilde{\chi}_q(y)\tilde{\chi}_r(y)\tilde{\chi}_s(x)-(p\leftrightarrow q)], \tag{30b}$$

$$W_{pqrs}{}^{(2)}=\int\int W(x-y)[\tilde{\phi}_p(x)\tilde{\phi}_q(y)\tilde{\chi}_r(y)\tilde{\chi}_s(x)], \tag{30c}$$

$$W_{pqrs}{}^{(3)}=\int\int W(x-y)[\tilde{\phi}_p(x)\tilde{\chi}_q(y)\times(\tilde{\chi}_r(y)\tilde{\phi}_s(x)-\tilde{\chi}_r(x)\tilde{\phi}_s(y))]. \tag{30d}$$

Because of the distinction between $\tilde{\phi}$ and ϕ we see that the usual interaction terms, $W^{(0)}$, are slightly modified. However, this is of no interest. On the other hand, $W^{(1)}$ is a term leading to tunneling. It will be small of order T and could be quite important numerically for strongly interacting systems. The remaining terms are of second order in T and should, for the sake of consistency, be dropped. $W^{(2)}$ is of some interest, however, in that it shows that a direct tunneling of two particles is possible, even apart from the higher order contributions of T and $W^{(1)}$.

We still have to take into account the fact that the large interaction terms, $W^{(0)}$, on the first line (29) do not commute with one another. It is not difficult to show that this can be accounted for, at least in first order, by a change in $W^{(1)}$. The result is that if we replace the second line of (29), (containing $W^{(1)}$), by

$$\tfrac{1}{2}\sum_{pqrs}\{W_{pqrs}{}^{(1)\prime}(c_p{}^\dagger d_q{}^\dagger d_r d_s+d_p{}^\dagger c_q{}^\dagger c_r c_s)+W_{pqrs}{}^{(1)\prime\prime}(d_p{}^\dagger d_q{}^\dagger d_r c_s+c_p{}^\dagger c_q{}^\dagger c_r d_s)\}, \tag{31}$$

then we may regard the c's and d's as commuting operators. Expression (31) must be thought of as operating to the right. In (31), $W^{(1)\prime}$ is the same as expression (30b) with the tilde removed from the $\tilde{\chi}$ functions. $W^{(1)\prime\prime}$ is the same as (30b) with both the $\tilde{\chi}$, $\tilde{\phi}$ functions replaced by χ, ϕ functions, respectively.

Clearly, we could have treated electron-phonon interaction terms in the same fashion. These terms make a contribution to the tunneling operator in which electrons crossing the barrier simultaneously emit or absorb a phonon. Such "phonon-assisted" tunneling may have been observed in recently reported experiments.[5]

In summary, we have succeeded in the aim of splitting the Hamiltonian as in Eq. (1). The parts H_R and H_L do not quite commute, but in first order this can be easily accounted for. In higher orders, the situation is not quite so clear. In particular, a calculation of the second-order shift in the ground-state energy cannot be carried out without taking into account the failure of the main parts of the Hamiltonian to commute. For example, in the free electron case, the ground-state energy is rigorously

$$2\sum_{\epsilon k<\mu}\epsilon_k{}^{(+)}+\epsilon_k{}^{(-)}, \tag{32}$$

whereas naive perturbation theory on (19), would lead one to think that the ground-state energy is slightly lower. Thus, the calculation of Ref. (4), which is based on such straightforward perturbation theory, may be slightly suspect.

It may be well to remember, in addition, that the model we have employed for the barrier is exceedingly simple. It contains no impurities or any source of incoherence whatsoever. Thus, we have ruled out from the beginning any possibility that, for example, the T operator might have a rapidly varying or random phase. The phase of the transition operator is of no importance in the experiments performed thus far, but it is all important in the theory of Ref. 4.

Note added in proof. Although the minor reservations mentioned in the last two paragraphs have not been completely settled within the context of the present paper, we have no doubt that the calculation of Josephson[4] is essentially correct. In fact, the effect predicted by Josephson has probably been observed.[6]

The author would like to thank Professor Falk, Professor Ferrell, and Professor Glick for stimulating conservations.

[5] B. N. Taylor and E. Burstein, Phys. Rev. Letters **10**, 14 (1963).

[6] P. W. Anderson and J. M. Rowell, Phys. Rev. Letters **10**, 230 (1963).