Direct calculation of the tunneling current

J. Phys. C: Solid St. Phys., 1971. Vol. 4. Printed in Great Britain

Direct calculation of the tunneling current

C. CAROLI, R. COMBESCOT, P. NOZIERES, and D. SAINT-JAMES
Groupe de Physique des Solides de l'Ecole Normale Supérieure†, Faculté des Sciences de Paris, 9, Quai Saint-Bernard, Paris (5è).
MS. received 15*th July* 1970

Abstract. A simple model Hamiltonian is proposed for a metal–insulator–metal tunneling junction, which permits the direct calculation of the tunneling current without introducing any effective Hamiltonian. The model rests on the use of localized functions; this procedure avoids the difficult matching problem at the boundary between the barrier and the electrodes. Moreover the use of Kjeldysh's perturbation theory for nonequilibrium systems allows an explicit calculation of the current to all orders in the applied bias. It is found that the transfer coefficient appearing in the expression for the current is energy dependent. This model can be systematically extended to include many body effects.

1. Introduction

In the past few years, a great deal of experimental work has been performed on various types of tunneling junctions with more and more refined techniques. This now allows the observation of fine structures in the I–V curves, such as phonon effects in MIM (metal–insulator–metal) junctions (Klein and Léger 1968, Rowell *et al.* 1969), or vibration spectra of radicals located in the insulating barrier (Lambe and Jaklevic 1968). On the other hand, the theoretical approach to the problem of tunneling still remains founded essentially on the theory of Bardeen (1961), that is, on the transfer Hamiltonian model. It is well known that this model gives, in most cases, a very good description of the observed effects, but that it has not been possible yet to deduce it firmly from first principles. Moreover, as has been recently pointed out by several authors (Appelbaum and Brinkman 1969), this model is not sufficient to treat many body effects properly.

In most treatments of tunneling, one tries, starting from the initial Hamiltonian of the total MIM system, to transform it into an effective 'tunneling Hamiltonian', H_{eff}, of the form

$$H_{\text{eff}} = H_{\text{L}} + H_{\text{R}} + \mathbf{T} \tag{1}$$

where H_{L} and H_{R} respectively describe, as exactly as possible, the left and right isolated electrodes, while $\mathbf{T}$ is a transfer term which describes the probability for an electron to tunnel through the barrier. One of the problems related to this formulation is that $\mathbf{T}$ obviously corresponds, not to an ordinary potential, but to the $\mathbf{T}$-matrix for the tunneling process, that is, it is in general energy dependent. The energy dependence is neglected in the transfer Hamiltonian theory, which makes this theory unsuitable for describing many body effects; many body processes, for instance phonon emission and absorption in the barrier, must appear essentially as retardation effects on the transfer term of the effective Hamiltonian.

In this paper, starting from a very simple model Hamiltonian for a MIM junction, we calculate directly the tunneling current, without having to make use of an intermediate effective Hamiltonian of the type given by equation (1). Our method takes advantage of the following two points:

(i) We describe the system, not in terms of quasi-free electron wave functions, but of functions localized on atomic sites (analogous to Wannier functions). The decomposition

† Laboratoire associé au Centre National de la Recherche Scientifique.

of the system into three parts (left electrode, barrier, right electrode) results immediately from this picture, without giving rise to the usual mathematical difficulties.

(ii) It is usually considered that a description of the system as a whole does not permit calculation of the current, since tunneling is a nonequilibrium process. This is one of the main reasons for using an effective Hamiltonian of type (1); H_L and H_R correspond to two disconnected systems, each of which is separately at equilibrium, the transfer term $\mathbf{T}$ induces transitions between these two systems, and one calculates the transition probability by means of first order time dependent perturbation theory. Actually, the current can be calculated directly, even at finite voltage, with the help of the theory of perturbations in nonequilibrium systems (Kjeldysh 1965, Craig 1968), which generalizes the diagram techniques well known for the equilibrium case. Moreover, within this framework it is also possible to take into account many body effects systematically, without any of the difficulties encountered in the usual tunneling formalism.

In § 2, we describe and discuss our model Hamiltonian and calculate the corresponding electron Green function at zero voltage. We then compare this result with what would be obtained from an effective Hamiltonain model.

In § 3, we derive the tunneling current for such a system, and we examine briefly the possible generalizations of our model.

2. Model Hamiltonian and Green function at zero voltage

2.1. *The model Hamiltonian*

We first consider, for simplicity, a one dimensional system, made of semi-infinite metallic electrodes M and M′, and of an insulating barrier I; for the time being, we neglect all many body interactions. We assume that any one electron wave function may be described in terms of an *orthonormal* basis $\varphi_l^{(n)}$, the function $\varphi_l^{(n)}$ being localized in the vicinity of site l (n is a band index). In terms of the corresponding creation and annihilation operators C^+ and C, the Hamiltonian has the form

$$H = \sum_{\substack{l,l' \\ n,n'}} T_{ll'}^{nn'} C_l^{(n)+} C_{l'}^{(n')} \tag{2}$$

If the system were homogeneous, the $\varphi_l^{(n)}$ would be chosen as the ordinary Wannier functions. We assume that a similar set may be constructed for a non-homogeneous system, which retains the same properties, that is, each function $\varphi_l^{(n)}$ is essentially localized within the lth atomic cell, and there is one $\varphi_l^{(n)}$ corresponding to each level of the lth atom†.

One can think of these pseudo-Wannier functions as being built by the following process: one would first build localized functions by means for instance of a LCAO method, then orthogonalize these localized functions. The main effect of the orthogonalization condition is to mix the LCAO function with the corresponding functions centred on the nearest neighbours of l (or more generally on a few neighbours, the amount of admixture of $\phi_{l+j}^{(n)}$ to $\phi_l^{(n)}$ decreasing exponentially with $|j|$). The contribution of the ϕ of high energy (for instance those which belong to the continuum and therefore are not localized) does not affect seriously the localized character of the orthogonalized function $\varphi_l^{(n)}$ corresponding to $\phi_l^{(n)}$; on the one hand they lie much higher in energy than the ϕ_i we are interested in, so the corresponding amount of admixture is very small; on the other hand the sum of the contributions of all these quasi-free functions gives rise to destructive interferences so that the total contribution is essentially localized within one (or at most a few) atomic distance. It is clear that the φ_i centred on sites which are not too close to the metal–insulator boundaries (typically, farther than a few atomic distances) are practically identical to the Wannier functions of the bulk medium. Close to the boundaries, the modification of these functions is related to the surface effects due to the metal–insulator contacts.

In the calculations, we neglect all interband processes, so we retain in equation (2) only

† This amounts to saying that, in a homogeneous system there is, on a given site i, one Wannier function corresponding to each band.

the φ_l constructed from one given LCAO level: this corresponds, for the metals M and M', to a one band approximation. The insulator is also described in the equivalent of a one band model; we retain only the φ_l corresponding to the band which is closest in energy to the energy level of the tunneling electrons†.

In a homogeneous material, the matrix elements of the Hamiltonian T_{lm} between two Wannier functions decrease rapidly with $|l-m|$. This remains true in the M–I–M' junction, so that we can safely neglect the T_{lm} which would couple a site of M with a site of M'. (This assumes that the barrier is at least 3 or 4 atomic layers thick, which is usually the case in the experiments.)

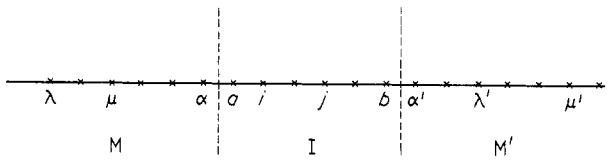

Figure 1. Schematic representation of a M–I–M' junction. The dashed lines represent the boundaries between the sites of the insulator and the two metals.

Moreover, we will assume for simplicity that the electrode M and the insulator are coupled by only one matrix element $T_{\alpha a}$, where α is the last site of M and a the first site in the barrier (see figure 1). In the same way, M' and I are coupled only through $T_{b\alpha'}$. All the calculations can be carried out with an arbitrary range for the M–I and I–M' couplings, they lead to a much more heavy algebra but to the same qualitive results as the simple nearest neighbour assumption.

The Hamiltonian can therefore be written as

$$H = \sum_{\lambda\mu} T_{\lambda\mu} C_\lambda^+ C_\mu + \sum_{\lambda'\mu'} T_{\lambda'\mu'} C_{\lambda'}^+ C_{\mu'} + \sum_{ij} T_{ij} C_i^+ C_j + (T_{\alpha a} C_\alpha^+ C_\alpha + \text{hc}) + (T_{\alpha' b} C_{\alpha'}^+ C_b + \text{hc}) \quad (3)$$

The unprimed greek indices describe the atomic sites of M(M'), the italic indices label the sites in the barrier.

The first term of equation (3) may be viewed as a 'left electrode Hamiltonian' (near the barrier, $T_{\lambda\mu}$ depends on both λ and μ, partly because of ordinary surface effects, but also because the Wannier functions are distorted by the presence of the insulator). The other terms of equation (3) respectively correspond to the right electrode, to the barrier and to the transfers between them. The separation of these various terms is here very natural; in order to establish the connection of the present approach with the usual description based on continuous wave functions throughout the system, one would need to go to the limit where the intersite distance goes to zero. At the moment, we have not been able to overcome mathematical difficulties due to the matching of wave functions at the boundaries.

2.2. *The Green functions at zero bias*

We now calculate the equilibrium Green function of the system described by the Hamiltonian of equation (3). We introduce the following notations

$$T_{\alpha a} = T \qquad T_{\alpha' b} = T' \quad (4)$$

Let $g_{\lambda\mu}(\epsilon)$ be the Green function of the left electrode alone (corresponding to the Hamiltonian $\Sigma T_{\lambda\mu} C_\lambda^+ C_\mu$). The propagator $g_{\lambda\mu}$ is given by all the diagrams that go from site λ to site μ without ever entering the barrier. Similarly, we denote by $g_{\lambda'\mu'}(\epsilon)$ and $g_{ij}(\epsilon)$ the propagators

† One can easily generalize the calculations to the case of a two band model for the insulating barrier. This will be discussed in § 3.

‡ This assumption is not crucial. The exponentially small direct coupling between M and M' gives rise simply to a correspondingly small correction to the transmission coefficient of the barrier.

for the right electrode and for the barrier alone. Since the barrier is insulating, g_{ij} decreases exponentially with $|i-j|$.

The full propagator $G_{pq}(\epsilon)$ (where the site indices p and q run over the whole system) is the solution of

$$\epsilon G_{pq}(\epsilon) - \sum_{p_1} T_{pp_1} G_{p_1 q}(\epsilon) = \delta_{pq} \tag{5}$$

In order to calculate the full propagator $G_{\lambda\mu}(\epsilon)$, we must take into account possible excursions of the electron across the (M–I) interface. If we single out the first and last paths followed in the M region, we may write

$$G_{\lambda\mu}(\epsilon) = g_{\lambda\mu}(\epsilon) + g_{\lambda\alpha}(\epsilon) T G_{aa}(\epsilon) T g_{\alpha\mu}(\epsilon)$$

Similarly, we obtain for the (M′–M′) and (M–M′) propagators

$$\begin{aligned} G_{\lambda'\mu'}(\epsilon) &= g_{\lambda'\mu'}(\epsilon) + g_{\lambda'\alpha'}(\epsilon) T' G_{bb}(\epsilon) T' g_{\alpha'\mu'}(\epsilon) \\ G_{\lambda\mu'}(\epsilon) &= g_{\lambda\alpha}(\epsilon)\, T G_{ab}(\epsilon)\, T' g_{\alpha'\mu'}(\epsilon) \end{aligned} \tag{6}$$

We must now calculate the full propagator $G_{ij}(\epsilon)$ inside the barrier. For that purpose, we note that an excursion of the electron into M or M′ may be viewed as a self-energy process superimposed on the barrier propagation. The corresponding self-energy operator has only two nonzero matrix elements

$$\begin{aligned} \Sigma_{aa} &= \Sigma = T^2 g_{\alpha\alpha} \\ \Sigma_{bb} &= \Sigma' = T'^2 g_{\alpha'\alpha'} \end{aligned} \tag{7}$$

and the Dyson equation for G_{ij} reads

$$G_{ij}(\epsilon) = g_{ij}(\epsilon) + g_{ia}(\epsilon)\Sigma(\epsilon) G_{aj}(\epsilon) + g_{ib}(\epsilon)\Sigma'(\epsilon)\, G_{bj}(\epsilon) \tag{8}$$

The solution of equation (8) is straightforward, and yields for instance

$$\begin{aligned} G_{ab} &= \frac{g_{ab}}{D} \\ G_{aa} &= \frac{g_{aa}(1 - T'^2 g_{bb} g_{\alpha'\alpha'}) + T'^2 g_{ba} g_{ab} g_{\alpha'\alpha'}}{D} \\ G_{bb} &= \frac{g_{bb}(1 - T^2 g_{aa} g_{\alpha\alpha}) + T^2 g_{ba} g_{ab} g_{\alpha\alpha}}{D} \end{aligned} \tag{9}$$

where the denominator D is given by

$$D = (1 - T^2 g_{aa} g_{\alpha\alpha})(1 - T'^2 g_{bb} g_{\alpha'\alpha'}) - T^2 T'^2 g_{ab} g_{ba} g_{\alpha\alpha} g_{\alpha'\alpha'} \tag{10}$$

the exact Green function G is thus completely determined.

We note that the M–M′ propagator may be written as

$$G_{\lambda\lambda'} = \sum_{\mu\mu'} g_{\lambda\mu} \mathscr{T}_{\mu\mu'} g_{\mu'\lambda'} \tag{11}$$

where $\mathscr{T}$ is an effective **T**-matrix given by

$$\mathscr{T}_{\mu\mu'} = \frac{T T' G_{ab}}{D} \delta_{\mu\alpha} \delta_{\mu'\alpha'} \tag{12}$$

$\mathscr{T}_{\mu\mu'}$ is equivalent to the transfer matrix element introduced by Bardeen.

Note that, as we have already noticed, and as was to be expected, since the barrier is insulating, g_{ab}, and therefore $\mathscr{T}$, decrease exponentially when $|a-b|$, that is, the barrier thickness, increases.

Besides the fact that we have an explicit expression for the **T**-matrix, our result differs from the usual one in several ways.

(i) $\mathscr{T}$ depends on the energy ϵ, a point which has been recognized to be essential in the proper treatment of inelastic tunneling. Physically, this simply means that the electron spends a finite time inside the barrier.

(ii) $\mathscr{T}$ depends on the electrode propagators $g_{\alpha\alpha}$ and $g_{\alpha'\alpha'}$ essentially because an electron inside the barrier can make excursions into the electrodes.

(iii) A simple minded calculation of the full G in terms of $\mathscr{T}$ usually gives erroneous results. For instance, the usual transfer Hamiltonian theory would yield for the M–M propagator

$$\begin{aligned} G_{\lambda\mu} &= g_{\lambda\mu} + g_{\lambda\alpha}\mathscr{T}g_{\alpha'\alpha'}\mathscr{T}g_{\alpha\mu} \\ &= g_{\lambda\mu} + g_{\lambda\alpha}\frac{T^2T'^2g_{ab}g_{ba}g_{\alpha'\alpha'}}{D^2}g_{\alpha\mu} \end{aligned} \tag{13}$$

which is obviously different from the exact expression (4). The error stems from the fact that the standard method ignores processes in which the propagator turns backward to M *inside* the barrier, before it has reached the I–M′ interface.

3. The tunneling current

3.1. *The current operator*

Let P be a point between site i and $i+1$; the current J at point P is the difference between the flow of electrons from left to right and from right to left. We thus expect an operator J of the form

$$J_{\mathrm{P}} = \sum_{\substack{l\geqslant i+1\\ m\leqslant i}} A_{lm}C_l^+C_m - \sum_{\substack{l\leqslant i\\ m\geqslant i+1}} A_{lm}C_l^+C_m \tag{14}$$

In order to calculate A_{lm}, we write the continuity equation, which in our discrete representation reads

$$J_{\mathrm{P}} - J_{\mathrm{P}'} + \frac{\partial\rho_i}{\partial t} = 0 \tag{15}$$

P′ is a point between sites $(i-1)$ and i, while $\rho_i = eC_i^+C_i$ is the electron charge at site i (or, more exactly, *around* the site i, over a region depending on the localization of the pseudo-Wannier functions). Using the Hamiltonian of equation (2), we find

$$\frac{\partial\rho_i}{\partial t} = \frac{1}{i\hbar}[\rho_i, H] = \frac{e}{i\hbar}\sum_m (T_{im}C_i^+C_m - T_{mi}C_m^+C_i) \tag{16}$$

A comparison of equations (14) and (16) yields

$$A_{lm} = \frac{e}{i\hbar}T_{lm} \tag{17}$$

In a stationary state, the current may be calculated at any point in the system. Since in our simple model the barrier is coupled to M(M′) through a single matrix element $T_{\alpha a}(T_{\alpha' b})$, it is particularly convenient to compute the current between α and a (or between b and α'). From now on, we shall choose P between sites α and a. In that case, equations (14) and (17) lead to

$$J = \langle J_{\mathrm{P}}\rangle = \frac{eT}{i\hbar}(\langle C_\alpha^+C_a\rangle - \langle C_a^+\ C_\alpha\rangle) \tag{18}$$

The calculation of J is reduced to that of the Green function $G_{\alpha a}$ for equal times.

3.2. *Calculation of the current*

We now assume that the barrier is biased at a voltage V; the chemical potentials μ and μ' in both electrodes differ by a quantity

$$\mu' - \mu = eV \tag{19}$$

If the electrodes are of a different nature, there may exist a contact potential, so that the difference in the electrostatic potentials $(\varphi' - \varphi)$ is not necessarily equal to V. We shall assume that such contact potentials have been eliminated by appropriate shifts of the band structure; the applied bias may then be described by a Hamiltonian

$$\sum_i V_i C_i^+ C_i \tag{20}$$

where V_i is the 'effective' potential at site i, going from 0 on one side to V on the other.

Ordinary perturbation theory may be used to obtain the *linear* response of the junction to the bias V; the term (20) is then treated as a small perturbation, and one only needs to calculate the appropriate correlation function for a junction in thermodynamic equilibrium (that is, such that $\mu = \mu'$). Unfortunately, such an approach does not work if we want to calculate the structure of the $(I$–$V)$ characteristics; the system is then far from equilibrium $(\mu \neq \mu')$, and the usual methods of perturbation theory do not apply. We shall use instead a generalized perturbation method due to Kjeldysh (1965) which remains valid for off equilibrium situations (5). This technique is briefly described in the Appendix.

We proceed along the same lines as in subsection 2.2, that is, we treat the I–M and I–M′ couplings as perturbations; to zeroth order, the three regions I, M and M′ are completely disconnected, and no current can flow between them. The applied bias is included in the zeroth order Hamiltonian; the voltage V_i acts to modify the one electron energies, and we assume that the latter have been redefined accordingly; the only effect of the bias is then to maintain a difference in chemical potentials between the two sides. We note that to zeroth order, the existence of such a difference $(\mu' - \mu)$ does not preclude the existence of thermodynamic equilibrium (no current can flow across the barrier). We can thus construct our perturbation theory on a well defined initial state. To be more specific, the potentials are constant in M and M′ (although different); there is no electric field, and the metals are in equilibrium. Inside the barrier, there is an electric field, but usually the Fermi level lies somewhere in a forbidden band; no current can flow, and again the insulator responds reversibly to the existing electric field. (Such a conclusion would not apply in the Fowler–Nordheim regime, when the tunneling electron can enter the conduction band of the insulator; our theory as such does not apply to this case.)

The initial state being thus unambiguously defined, we establish adiabatically the I–M and I–M′ couplings, and we calculate the current through the barrier. The Kjeldysh method allows us to perform the calculation to all orders in the coupling, thus describing a basically nonequilibrium (although stationary) situation.

One might raise a major objection to the above procedure; it amounts to establishing first the dc bias, and only later the coupling between the barrier and the electrode. Physically, it is the reverse that is true; the transfer matrix elements are always there, and the dc bias is established afterwards; it is not obvious that the corresponding limits can be interchanged. We believe that such a difficulty does not appear in the experimental situation in which the two electrodes are connected to a generator which controls the bias and serves as a source (or sink) of carriers when current flows. The relation between bias $V(t)$, current $J(t)$ and conductance $G(t)$ is instantaneous

$$J(t) = V(t)G(t)$$

and it does not matter whether the bias or the conductance is established first. The situation would be different if the junction were isolated: the flow of current would then change the

capacitance of the junction, and the bias at time t would depend on the past history of V and G†.

We now turn to the detailed application of Kjeldysh's method to our problem. We introduce three Green functions

$$\begin{aligned} G^{a}_{pq}(t,t') &= -i\theta(t-t')\langle[C_p(t), C_q^{+}(t')]_{+}\rangle \\ G^{r}_{pq}(t,t') &= i\theta(t'-t)\langle[C_p(t), C_q^{+}(t')]_{+}\rangle \\ F_{pq}(t,t') &= -i\langle[C_p(t), C_q^{+}(t')]_{-}\rangle \end{aligned} \tag{21}$$

The site indices p and q run over the whole system

$$\begin{aligned} [A,B]_{+} &= AB + BA \\ [A,B]_{-} &= AB - BA \end{aligned}$$

and $\theta(t)$ is the ordinary step function.

The current J given by equation (18) can be expressed as

$$J = \frac{eT}{\hbar}\{G^{c}_{\alpha a}(t,t_{+}) - G^{c}_{a\alpha}(t,t_{+})\} \tag{22}$$

where G^{c} is the usual causal Green function

$$\begin{aligned} G^{c}_{pq}(t,t') &= -i\langle T\{C_p(t)\,C_q^{+}(t')\}\rangle \\ &\equiv \tfrac{1}{2}\{G^{a}_{pq}(t,t') + G^{r}_{pq}(t,t') + F_{pq}(t,t')\} \end{aligned} \tag{23}$$

We also define the corresponding Green functions $g^{a}_{pq}(t,t')$, $g^{r}_{pq}(t,t')$ and $f_{pq}(t,t')$ for the zero order system (no coupling between barrier and electrodes). We note that g^{r}, g^{a} and f vanish when p and q belong to two different subparts of the junctions (for example $g^{(a)}_{\lambda i} = 0$).

The time Fourier transforms of G^{a}, G^{r} and F obey the following Dyson's equations‡ (see Appendix):

$$\begin{aligned} G^{a}_{pq}(\omega) &= g^{a}_{pq_1}(\omega)\{\delta_{q_1 q} + \Sigma_{q_1 q_2} G^{a}_{q_2 q}(\omega)\} \\ G^{r}_{pq}(\omega) &= g^{r}_{pq_1}(\omega)\{\delta_{q_1 q} + \Sigma_{q_1 q_2} G^{r}_{q_2 q}(\omega)\} \\ F_{pq}(\omega) &= f_{pq}(\omega) + g^{r}_{pq_1}(\omega)\Sigma_{q_1 q_2} F_{q_2 q}(\omega) + f_{pq_1}(\omega)\Sigma_{q_1 q_2} G^{a}_{q_2 q}(\omega) \end{aligned} \tag{24}$$

where the self-energy Σ_{pq} is defined as

$$\Sigma_{pq} = T(\delta_{p\alpha}\delta_{aq} + \delta_{pa}\delta_{\alpha q}) + T'(\delta_{pb}\delta_{\alpha' q} + \delta_{p\alpha'}\delta_{bq}) \tag{25}$$

Let us first consider the advanced and retarded functions G^{r} and G^{a}; the corresponding Dyson's equations are identical to those introduced in § 2 in our description of the equilibrium system—except that an index r (or a) has been affixed to all propagators. The previous algebra remains valid; for instance, when calculating $G^{a}_{\alpha a}$, we may start from site α, and single out the first 'crossing' vertex at which the propagator penetrates into the barrier, in $G^{a}_{a\alpha}$, we pick out the last crossing vertex before the propagator reaches site α. We thus obtain

$$\begin{aligned} G^{a}_{\alpha a} &= g^{a}_{\alpha\alpha} T G^{a}_{aa} \\ G^{a}_{a\alpha} &= G^{a}_{aa} T g^{a}_{\alpha\alpha} \end{aligned} \tag{26}$$

† In such a case, the correct limit is obtained by letting first the volume of the electrodes go to infinity (which makes the capacitance of the junction infinite), the adiabatic switching time going to infinity only later.

‡ In the general nonequilibrium situation, all the Green functions depend separately on the two time variables t and t'. However, since we are interested here in a stationary situation, they depend only on $(t - t')$.

Similar equations hold for G^r. We see that G^r and G^a do not contribute to the expression (22) for the current, which reduces to

$$J = \frac{eT}{2\hbar}\{F_{\alpha a}(t, t_+) - F_{a\alpha}(t, t_+)\} \tag{27}$$

We are then left with the calculation of $F_{\alpha a}$ and $F_{a\alpha}$.

The chain of Dyson's equations for $F_{\alpha a}$ is, according to equation (24)†

$$\begin{aligned}
F_{\alpha a} &= f_{\alpha\alpha} T G^a_{aa} + g^r_{\alpha\alpha} T F_{aa} \\
F_{aa} &= f_{aa} + f_{aa} T G^a_{\alpha a} + f_{ab} T' G^a_{\alpha' a} + g^r_{aa} T F_{\alpha a} + g^r_{ab} T' F_{\alpha' a} \\
F_{\alpha' a} &= f_{\alpha'\alpha'} T' G^a_{ba} + g^r_{\alpha'\alpha'} T' F_{ba} \\
F_{ba} &= f_{ba} + f_{bb} T' G^a_{\alpha' a} + f_{ba} T G^a_{\alpha a} + g^r_{bb} T' F_{\alpha' a} + g^r_{ba} T F_{\alpha a}.
\end{aligned} \tag{28}$$

(The various matrix elements of G^r and G^a are obtained from equation (9).) We now assume that, in the range of energy of interest‡, the density of states in the isolated barrier is strictly zero. This means that we restrict ourselves to electrons tunneling in the true gap of the insulator (for instance, we assume that there is no impurity bound state in this domain). Since the density of states is related to the imaginary part of the Green functions, in this energy range the g functions of the barrier are real, so that

$$\begin{aligned}
g^a_{ij} &= g^r_{ij} = g^c_{ij} = g_{ij} \\
f_{ij} &= 0
\end{aligned} \tag{29}$$

The resolution of equation (28) is then straightforward, and yields

$$\begin{aligned}
F_{\alpha a} = {} & TT'^2 f_{\alpha'\alpha'} G^a_{ba} G^r_{ab} g^r_{\alpha\alpha} + \frac{T f_{\alpha\alpha}}{D^a D^r}(1 - T'^2 g^a_{\alpha'\alpha'} g_{bb})(1 - T'^2 g_{bb} g^r_{\alpha'\alpha'})\, g_{aa} \\
& + TT'^2 G^a_{ab} G^r_{ba} f_{\alpha\alpha} g^a_{\alpha'\alpha'}(1 - T'^2 g^r_{\alpha'\alpha'} g_{bb})
\end{aligned} \tag{30}$$

In the same way, one obtains $F_{a\alpha}$, and finally

$$\begin{aligned}
F_{\alpha a}(\omega) - F_{a\alpha}(\omega) = {} & TT'^2 G^a_{ba}(\omega)\, G^r_{ab}(\omega)\{(g^r_{\alpha\alpha}(\omega) - g^a_{\alpha\alpha}(\omega))\, f_{\alpha'\alpha'}(\omega) \\
& + f_{\alpha\alpha}(\omega)(g^a_{\alpha'\alpha'}(\omega) - g^r_{\alpha'\alpha'}(\omega))\}
\end{aligned} \tag{31}$$

where $G^{a(r)}_{ba}(\omega) = g_{ba}(\omega)/D^{a(r)}$ is the Green function which describes the propagation of an electron from one side of the barrier to the other in the real system; the renormalizing factor $1/D^{a(r)}$ expresses the probability of an electron path, going from a to b, penetrating into the electrodes.

At this stage, it is convenient to introduce two extra Green functions

$$\begin{aligned}
G^+_{pq}(t, t') &= \mathrm{i}\langle C^+_q(t')\, C_p(t)\rangle \\
G^-_{pq}(t, t') &= -\mathrm{i}\langle C_p(t)\, C^+_q(t')\rangle
\end{aligned} \tag{32}$$

† According to Kjeldysh, the terms linear in f in equation (28) should vanish. This conclusion does not hold here. Mathematically, such a statement amounts to claiming that the product $f(1 + \Sigma^a G^a) = fg^{a-1}G^a$ vanishes because $f(g^a)^{-1} = 0$. In fact in our case the factor G^a is singular, and the product has a finite value, as can be seen by direct inspection. We suspect that the difference with Kjeldysh stems from the fact that our model does not involve any relaxation mechanism in the electrodes.

‡ We will check on the final result that, as can be expected, the energies of interest range between the Fermi levels of the two electrodes.

with similar definitions for the zeroth order limits $g^{\pm}$. We verify easily that

$$\begin{aligned} g^{+} + g^{-} &= f \\ g^{+} - g^{-} &= g^{\mathrm{r}} - g^{\mathrm{a}} \end{aligned} \tag{33}$$

The expression (31) may then be recast in the form

$$F_{\alpha a}(\omega) - F_{a\alpha}(\omega) = 2TT'^{2} G^{\mathrm{a}}_{ba}(\omega)\, G^{\mathrm{r}}_{ab}(\omega)\, \{g^{+}_{\alpha\alpha}(\omega)\, g^{-}_{\alpha'\alpha'}(\omega) - g^{-}_{\alpha\alpha}(\omega)\, g^{+}_{\alpha'\alpha'}(\omega)\} \tag{34}$$

Together with equation (27), equation (34) determines the current J.

By performing a spectral analysis of $g^{+}_{\alpha\alpha}$ and $g^{-}_{\alpha\alpha}$, we obtain, in the limit of zero temperature

$$\begin{aligned} g^{+}_{\alpha\alpha}(\omega) &= 2\mathrm{i}\theta(\mu - \omega)\, \mathrm{Im}\, g_{\alpha\alpha}(\omega) \\ &= 2\mathrm{i}\pi\theta(\mu - \omega)\, \rho_{\alpha}(\omega) \\ g^{-}_{\alpha\alpha}(\omega) &= -2\mathrm{i}\pi\theta(\omega - \mu)\, \rho_{\alpha}(\omega) \end{aligned} \tag{35}$$

where $\rho_{\alpha}(\omega)$ is the density of states on site α in the isolated electrode. A similar result holds for $g^{\pm}_{\alpha'\alpha'}$, μ being replaced by μ'.

Inserting equations (34) and (35) into the expression (27) for the current, one obtains†

$$J = -\frac{(2\pi)^2 e\, T^2\, T'^2}{\hbar} \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi}\, G^{\mathrm{a}}_{ba}(\omega)\, G^{\mathrm{r}}_{ab}(\omega)\, \rho_{\alpha}(\omega)\, \rho_{\alpha'}(\omega)$$

$$\times \{\theta(\omega - \mu)\, \theta(\mu' - \omega) - \theta(\mu - \omega)\, \theta(\omega - \mu')\} \tag{36}$$

In the case of figure 2 where $\mu' > \mu$, equation (36) reduces to

$$J = -\frac{(2\pi)^2 e\, T^2\, T'^2}{\hbar} \int_{\mu}^{\mu + eV} \frac{\mathrm{d}\omega}{2\pi}\, G^{\mathrm{a}}_{ba}(\omega)\, G^{\mathrm{r}}_{ab}(\omega)\, \rho_{\alpha}(\omega)\, \rho_{\alpha'}(\omega) \tag{37}$$

3.3. *Discussion*

Equation (37) shows that the current is, as usual, built up from the contribution of the states whose energies range from μ to $\mu' = \mu + eV$, that is, which are occupied in electrode M' and empty in electrode M. The main difference between our result and the expression obtained from the conventional tunneling Hamiltonian lies in the form of the 'transfer coefficient', which in our model is equal to

$$|\mathscr{T}|^2 = T^2\, T'^2\, G^{\mathrm{a}}_{ba}(\omega)\, G^{\mathrm{r}}_{ab}(\omega) \tag{38}$$

We note that $|\mathscr{T}|^2$ is frequency dependent. In the simple case that we have considered, this dependence is very weak as long as eV remains much smaller than the barrier energy. On the other hand, this fact becomes essential if we are to include self-energy effects, such as electron–phonon interactions. First of all, $|\mathscr{T}|^2$ depends mostly on the energy, and not on the momentum, a point which has been stressed by various authors. Moreover, $|\mathscr{T}|^2$ depends on the electrode propagators $g_{\alpha\alpha}$ and $g_{\alpha'\alpha'}$, through the renormalizing denominators in equation (9); this is a new feature of the present calculations, which may affect profoundly the shape of the self-energy structures in the I–V characteristics.

It should be stressed that some of the simplifying features of the above model can be easily relaxed, namely:

(*a*) one can extend the calculation to three dimensions, in the simple case where the component of the wave vector parallel to the M–I and I–M' boundary planes $\boldsymbol{k}_{\parallel}$ is conserved. This means that we neglect the Moiré patterns which would arise from a difference between

† Note that G_{ba} must include the effect of the finite electric field inside the barrier, which presumably accounts for the so called 'background current'. Such an effect is easily calculated if the bias is much smaller than the barrier height.

the periodicities in M, I and M′ in the yz plane (parallel to the boundaries). Moreover, we assume that each electrode is coupled to the barrier through a Hamiltonian of the form

$$H_{\mathrm{IM}} = \sum_{\alpha a} T_{\alpha a}\, C_\alpha^+\, C_a + \mathrm{hc} \tag{39}$$

where α lies in the first atomic plane of the metal, and a in the first atomic plane of the insulator (the range of $T_{\alpha a}$ in the yz plane is unimportant). With this assumption, the x

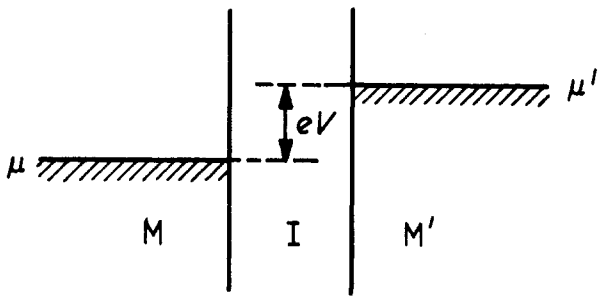

Figure 2. Schematic representation of the Fermi levels of the two electrodes in the presence of an applied voltage V.

component of the current is given by the same formal expression as in the one dimensional case

$$J = \frac{\mathrm{i}e}{\hbar} \sum_{\alpha a} (T_{\alpha a}\, C_\alpha^+\, C_a - T_{a\alpha}\, C_a\, C_\alpha) \tag{40}$$

In order to take advantage of the translational invariance along the yz plane, we Fourier transform all the quantities with respect to y and z; we define for instance

$$\begin{aligned} T(\boldsymbol{k}_\parallel) &= \sum_{\alpha a} T_{\alpha a} \exp\{\mathrm{i}\boldsymbol{k}_\parallel \cdot (\boldsymbol{R}_\alpha - \boldsymbol{R}_a)\} \\ G_{pq}(\boldsymbol{k}_\parallel, \omega) &= \sum_{\boldsymbol{p}_\parallel, \boldsymbol{q}_\parallel} G_{\boldsymbol{R}_p, \boldsymbol{R}_q}(\omega) \exp\{\mathrm{i}\boldsymbol{k}_\parallel \cdot (\boldsymbol{p}_\parallel - \boldsymbol{q}_\parallel\} \end{aligned} \tag{41}$$

(where p and $\boldsymbol{p}_\parallel$ denote respectively the x and (y, z) components of $\boldsymbol{R}_p$). For each value of $\boldsymbol{k}_\parallel$ we are left with a one dimensional problem, and hence the current is given by

$$J = \int \frac{\mathrm{d}^2\boldsymbol{k}_\parallel}{(2\pi)^2} \frac{(2\pi)\, eT^2(\boldsymbol{k}_\parallel)\, T'^2(\boldsymbol{k}_\parallel)}{\hbar} \int_\mu^{\mu + eV} \frac{\mathrm{d}\omega}{2\pi} G^{\mathrm{a}}_{ba}(\boldsymbol{k}_\parallel, \omega)\, G^{\mathrm{r}}_{ab}(\boldsymbol{k}_\parallel, \omega)\, \rho_\alpha(\boldsymbol{k}_\parallel, \omega)\, \rho_{\alpha'}(\boldsymbol{k}_\parallel, \omega) \tag{42}$$

where $\rho_\alpha(\boldsymbol{k}_\parallel, \omega)$ is the density of states in the isolated electrode on the last atomic plane and for a given value of the parallel momentum

$$\rho_\alpha(\boldsymbol{k}_\parallel, \omega) = -\frac{1}{\pi} \operatorname{sgn}(\omega - \mu)\, \mathrm{Im}\, g^{\mathrm{c}}_{\alpha\alpha}(\boldsymbol{k}_\parallel, \omega) \tag{43}$$

(*b*) We have assumed that a one band model is sufficient to describe the insulating barrier. Our calculations can be generalized without difficulty to the case of a two band model. This could be of interest if the energy range of interest lies close to the middle of the gap separating the valence and conduction band of the insulator. In that case, the transfer

coefficient appearing in J could be somewhat lowered, due to interference effects between transfer processes which take place via the two different bands.

(c) The most interesting extension of the present model consists in including many body effects, such as electron–phonon interactions in the electrodes, or interaction with localized excitations or resonant states inside the barrier. This can be done systematically with the help of Kjeldysh's generalized diagram technique. It is then possible to describe in detail the influence of these various effects on the transfer coefficient $|\mathscr{T}|^2$, as well as on the density of states of the electrodes.

These various extensions are presently under study, and will be published separately. Here we shall only stress the main features of our model. It is based on the following two ingredients:

(i) a description in terms of localized 'Wannier' functions, which avoids the difficult matching problem at the boundary between the barrier and the electrodes,

(ii) the use of Kjeldysh's perturbation theory for nonequilibrium processes, which allows an *explicit* calculation of the propagators and of the current, to all orders in the applied bias and in the electrode–barrier couplings.

The assumption of nearest neighbour coupling is not essential. Quite generally, the problem reduces to the calculation of the propagator G_{ij} inside the barrier: if the latter contains p layers, we are faced with a $p \times p$ secular equation (which becomes $2p \times 2p$ if we include two energy bands in the insulator). Possible excursions of the electron inside the electrode only appear as self-energy corrections to this barrier secular equation. The expression for the current resembles the usual result, but with a transfer coefficient $|\mathscr{T}|^2$ which we calculate explicitly, and which depends on the energy, and also on the properties of the electrodes.

Appendix

The usual diagram technique involves the calculation of chronological products of the form†

$$\langle \phi_0 \,|\, T\{A(t)\, B(t') \ldots\} \,|\, \phi_0 \rangle$$

(at zero temperature). $A(t)$, $B(t')$ are operators in the Heisenberg representation.

In the interaction representation, this expression becomes

$$\langle \phi_0 \,|\, S(-\infty, +\infty)\, T\{\tilde{A}(t)\, \tilde{B}(t') \ldots S(+\infty, -\infty)\} \,|\, \phi_0 \rangle$$

where the matrix $\mathbf{S}$ is defined by

$$S(t', t) = T\left[\exp\left\{-\mathrm{i} \int_t^{t'} H_{\mathrm{int}}(t_1)\, \mathrm{d}t_1\right\}\right]$$

The interaction Hamiltonian H_{int} is supposed to be adiabatically turned on from $t = -\infty$, and turned off at $t = +\infty$. The operators in the interaction representation are $\tilde{A}(t)$, $\tilde{B}(t')$, $H_{\mathrm{int}}(t_1)$.

In the absence of irreversible effects

$$S(+\infty, -\infty)\, \phi_0 = \mathrm{e}^{\mathrm{i}\alpha} \phi_0$$

so that

$$\langle \phi_0 \,|\, T\{A(t)\, B(t') \ldots\} \,|\, \phi_0 \rangle = \frac{\langle \phi_0 \,|\, T\{\tilde{A}(t)\, \tilde{B}(t') \ldots S(+\infty, -\infty)\} \,|\, \phi_0 \rangle}{\langle \phi_0 \,|\, S(+\infty, -\infty) \,|\, \phi_0 \rangle}$$

† Abrikosov, Gorkov, Dzialoshinski, 1965, *Quantum Field Theory in Statistical Physics*, (Oxford: Pergamon) p. 49.

This statement is not true if the evolution of the system is irreversible, since the system either absorbs or emits energy during this evolution.

But it is possible to generalize the diagram technique. Indeed

$$\langle \phi_0 \,|\, S(-\infty, +\infty)\, T\,\{\tilde{A}(t)\,\tilde{B}(t') \ldots S(+\infty, -\infty)\} \,|\, \phi_0 \rangle$$
$$= \langle \phi_0 \,|\, T_c\,\{\tilde{A}(t)\,\tilde{B}(t') \ldots S_c\} \,|\, \phi_0 \rangle \tag{A.1}$$

where T_c is an operator ordering the times from right to left, not as usual from $-\infty$ to $+\infty$, but along a contour C made of a 'positive' branch going from $-\infty$ to $+\infty$ and with a 'negative' branch going from $+\infty$ to $-\infty$. Thus, on the positive branch, the times are

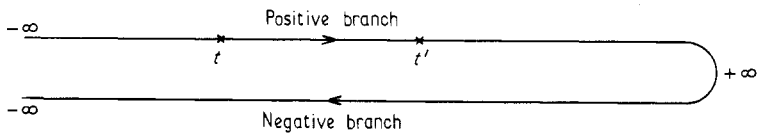

Figure 3. Representation of contour C.

ordered from $-\infty$ to $+\infty$, while they are ordered from $+\infty$ to $-\infty$ on the negative branch. Any time of the negative branch is considered as posterior to any time of the positive branch. The times $t, t' \ldots$ of the chronological product are on the positive branch. In the preceding expression

$$S_c = S(-\infty, +\infty)\, S(+\infty, -\infty)$$

where the times of $S(-\infty, +\infty)$ are on the positive branch and the times of $S(+\infty, -\infty)$ on the negative branch. It is clear that the presence of this term will introduce times on the negative branch of the contour in the Dyson equation.

It is now possible to calculate the expression (A.1) with the usual diagram technique; but since the times belonging to the negative branch (noted with subscript $-$) are posterior to the times belonging to the positive branch (noted with subscript $+$), we must use four Green functions

$$G^{+}_{ij}(t, t') = -i\langle \phi_0 \,|\, T_c\{c_i(t_+)\, c_j^{+}(t'_-)\} \,|\, \phi_0 \rangle$$
$$= i\langle \phi_0 \,|\, c_j^{+}(t')\, c_i(t) \,|\, \phi_0 \rangle$$
$$G^{-}_{ij}(t, t') = -i\langle \phi_0 \,|\, T_c\{c_i(t_-)\, c_j^{+}(t'_+)\} \,|\, \phi_0 \rangle$$
$$= -i\langle \phi_0 \,|\, c_i(t)\, c_j^{+}(t') \,|\, \phi_0 \rangle$$

$$G^{c}_{ij}(t, t') = -i\langle \phi_0 \,|\, T_c\{c_i(t_+)\, c_j^{+}(t'_+)\} \,|\, \phi_0 \rangle$$
$$= -i\langle \phi_0 \,|\, T\{c_i(t)\, c_j^{+}(t')\} \,|\, \phi_0 \rangle$$
$$\tilde{G}^{c}_{ij}(t, t') = -i\langle \phi_0 \,|\, T_c\{c_i(t_-)\, c_j^{+}(t'_+)\} \,|\, \phi_0 \rangle$$
$$= -i\langle \phi_0 \,|\, \tilde{T}\{c_i(t)\, c_j^{+}(t')\} \,|\, \phi_0 \rangle$$

where T orders the times from $-\infty$ to $+\infty$ and $\tilde{T}$ from $+\infty$ to $-\infty$.

In a diagram calculation the integration is performed along the contour C; this is equivalent to perform the integration from $-\infty$ to $+\infty$, and to sum over subscripts $+$ and $-$. Because the negative branch of the contour C goes from $+\infty$ to $-\infty$, any point of this branch corresponds to a $(-)$ sign, or equivalently any interaction on this branch corresponds to a $(-)$ sign.

Summing over subscripts $+$ and $-$ is equivalent to using 2×2 matrices. Therefore the

usual diagram technique may be used, if one defines a Green function matrix

$$\mathbf{G} = \begin{pmatrix} G^{c} & G^{-} \\ G^{+} & \tilde{G}^{c} \end{pmatrix}$$

A Dyson's equation may then be written

$$G_{ij}(t, t') = G^{0}_{ij}(t, t') + \int G^{0}_{ik}(t, t_1)\Sigma_{kl}(t_1, t_2)\, G_{lj}(t_2, t')\, \mathrm{d}t_1\, \mathrm{d}t_2 \tag{A.2}$$

where

$$\mathbf{\Sigma} = \begin{pmatrix} \Sigma^{c} & \Sigma^{-} \\ \Sigma^{+} & \tilde{\Sigma}^{c} \end{pmatrix}$$

is the self-energy matrix and $\mathbf{G}^{0}$ is the Green function matrix in the absence of the interaction.

The functions $G^{\pm}$, G^{c}, $\tilde{G}^{c}$ are connected by

$$G^{c} + \tilde{G}^{c} = G^{+} + G^{-}$$

In the same way

$$\Sigma^{c} + \tilde{\Sigma}^{c} = -(\Sigma^{+} + \Sigma^{-})$$

Thus among the four equations represented by equation (A.2), only two are independent. This becomes clearer if we make the following canonical transformation:

$$\mathbf{G} \rightarrow \frac{1 - \mathrm{i}\boldsymbol{\sigma}_y}{2} \begin{pmatrix} G^{c} & G^{-} \\ G^{+} & \tilde{G}^{c} \end{pmatrix} \frac{1 + \mathrm{i}\boldsymbol{\sigma}_y}{2} = \begin{pmatrix} 0 & G^{a} \\ G^{r} & F \end{pmatrix} \tag{A.3}$$

where†

$$\boldsymbol{\sigma}_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}$$

$$G^{a} = G^{c} - G^{+} = -\tilde{G}^{c} + G^{-}$$

$$G^{r} = G^{c} - G^{-} = -\tilde{G}^{c} + G^{+}$$

$$F = G^{c} + \tilde{G}^{c} = G^{+} + G^{-}$$

and

$$\begin{pmatrix} \Sigma^{c} & \Sigma^{-} \\ \Sigma^{+} & \tilde{\Sigma}^{c} \end{pmatrix} \text{becomes} \begin{pmatrix} \Omega & \Sigma^{r} \\ \Sigma^{a} & 0 \end{pmatrix}$$

with

$$\Sigma^{a} = \Sigma^{c} + \Sigma^{+} = -(\tilde{\Sigma}^{c} + \Sigma^{-})$$

$$\Sigma^{r} = \Sigma^{c} + \Sigma^{-} = -(\tilde{\Sigma}^{c} + \Sigma^{+})$$

$$\Omega = \Sigma^{c} + \tilde{\Sigma}^{c} = -(\Sigma^{+} + \Sigma^{-})$$

Thus we obtain three equations and the Fourier transforms of the equations for G^{a} and G^{r} are evidently complex conjugates.

† Keldysh's notation for G^{a} and G^{r}, that we use there, is not the usual one. This last is obtained by exchanging a and r.

In our case, the perturbation T is instantaneous and

$$H_{\text{int}}(t) = \{T\, C_\alpha^+(t)\, C_a(t) + T'\, C_{\alpha'}^+(t)\, C_b(t)\} + \text{hc}$$

where time t is either on the positive or on the negative branch of contour C so that H_{int} cannot connect the two branches.

Consequently $\boldsymbol{\Sigma}$ has the following form in the transformed representation of equation (A.3):

$$\sum = \begin{pmatrix} 0 & T \\ T & 0 \end{pmatrix} (\delta_{p\alpha}\, \delta_{qa} + \delta_{pa}\, \delta_{q\alpha}) + \begin{pmatrix} 0 & T' \\ T' & 0 \end{pmatrix} (\delta_{p\alpha'}\, \delta_{qb} + \delta_{pb}\, \delta_{q\alpha'})$$

References

APPELBAUM, J. A., and BRINKMAN, W. F., 1969, *Phys. Rev.*, **186**, 464–70.
BARDEEN, J., 1961, *Phys. Rev. Lett.*, **6**, 57–9.
CRAIG, R. A., 1968, *J. math. Phys.*, **9**, 605–11.
KELDYSH, L. V., 1965, *Sov. Phys.–JETP*, **20**, 1018–26.
KLEIN, J., and LEGER, A., 1968, *Phys. Lett.*, **28A**, 134–5.
LAMBE, J., and JAKLEVIC, R. C., 1968, *Phys. Rev.*, **165**, 821–32.
ROWELL, J. M., MCMILLAN, W. L., and FELDMANN, N. L., 1969, *Phys. Rev.*, **180**, 658–68.