Theory of Electron Tunneling via Real Intermediate States*

F. Mezei†
Physikalisches Institut der Universität Frankfurt am Main, Frankfurt am Main, Germany
(Received 6 May 1971)

A detailed theory of two-step electron transport through localized states is worked out. The quantitative results obtained are found to agree with existing experimental data. Analogies to the Anderson model of localized moments are discussed.

The general problem we are dealing with can be stated as follows. Let three different systems be given, say, A, B, and P, each of them having several electron states. There is an interaction between electrons on A and P, and on B and P. We are interested in electron transport between A and B due to the above indirect coupling through P.

Recently Giaever and Zeller[1,2] (GZ) have investigated such a problem in studying tunnel junctions containing small metallic particles inside the barrier. They suggested that in general there are two basic kinds of electron transition processes between A and B:

(i) Two-step transitions, when the electron first goes from, say, A to P, being localized there, and thereafter makes a subsequent transition from P to B. This type of process we will call transitions via real intermediate states.

(ii) Transitions via virtual intermediate states on P as given by higher-order perturbation theory.

GZ pointed out that process (i) may explain the experimentally observed anomalous, strongly non-Ohmic conductance behavior, because the charging up of P by the temporary localization of the electron needs an activation energy. Their treatment, however, fails to give a detailed microscopic mechanism and thus is not able to predict quantitative relations such as those between line shapes and temperature dependence.

In this paper we attempt to give a more rigorous and quantitative theory. Following GZ we will focus our attention on the Coulomb energies of localized electrons. We will apply, however, a self-consistent approach analogous to the Anderson[3] model of local moments. We feel that the original picture of activation energies is not very precise and not suitable for a microscopic treatment.

In our model we assume that the over-all relaxation time of single particle states on P is short compared to the electron transition rates between P and A or B. This has the consequence that subsequent transitions have no phase coherence, and occur as the separate quantum transitions of process (i). This situation can be formulated by using statistical mechanical transport theory in terms of localized electron states on either A, B, or P.

For temperatures above 1 °K it is reasonable to assume also that the coupling to lattice vibrations is strong enough to keep the electron system on P at the temperature of the surrounding heat bath,[4] since the transition rates in relevant tunneling experiments are as small as 10^5 sec^{-1} per particle.

The Hamiltonian under these assumptions may be given as

$$H = H_A + H_B + H_P + H_{AP} + H_{BP} \quad . \tag{1}$$

H_C describes free electron states in C, where in what follows C stands for A or B, and H_{CP} can be taken in the simplest approximation as

$$H_{CP} = \sum_{k,m} V_{km}(c_k^* d_m + d_m^* c_k) \ ,$$

where c_k and d_m are one-electron destruction operators to the states k in C and m in P, respectively. (k and m stand for all the quantum numbers.) In particular, this form of H_{CP} is just the tunneling Hamiltonian approach,[5] which seems to be well justified in this case, since no interface effects of interest are involved.[6]

H_P deserves a little more attention. We take it to be

$$H_P = \sum_m \epsilon_m d_m^* d_m + \tfrac{1}{2} \sum_{\substack{m,m' \\ m \neq m'}} U_{mm'} d_m^* d_m d_{m'}^* d_{m'} \ , \tag{2}$$

resembling the Anderson Hamiltonian.[3] H_P has to be considered as an effective Hamiltonian containing dressed quantities, especially since the matrix elements of the static Coulomb interaction $U_{mm'}$ contain all external screening (polarization of the neighborhood, image forces[7]) and internal screening effects. In the limiting case of P being a single atom, $U_{mm'}$ corresponds to the intra-atomic Coulomb interaction of the Anderson model, while for metallic particles much larger than the Thomas-Fermi screening radius it approaches e^2/C_P, where e is the electronic charge and C_P is the capacitance of the particle in the given neighborhood.[2] Since, in fact, we are concerned with eigenstates of (2) differing only in a few of the occupation numbers n_m, we may replace $U_{mm'}$ by a constant U, and for later convenience we write ϵ_m as $\epsilon_m' + V_0$, where V_0 is the electrostatic potential due to the background distribution of ions and frozen in dipoles

in and near P. Thus we are given the spectrum of (2) as

$$E(N, n_m) = \sum_m \epsilon'_m n_m + V_0 N + \tfrac{1}{2} UN(N-1) ,$$

where $n_m = 0$ or 1, and $\sum_m n_m = N$.

We adopt usual quasiequilibrium statistical mechanics of transport phenomena. Thus all of A, B, and P will be described by equilibrium density matrices corresponding to the temperature T and to the local electrochemical potentials μ_A, μ_B, and μ_P, respectively. The density matrix for P is given this way as

$$\rho_P = \sum_N P_N \rho_N ,$$

where ρ_N is the canonical density matrix for the N electron system with free-electron energies ϵ'_m. For convenience we make the assumption, applicable to the experimental situation in GZ's work, that the average spacing of the energy levels is much less than kT. Thus ρ_N corresponds[8] to the usual Fermi distribution with an internal Fermi energy ϵ'_F, which can be taken as a constant over the small range of N's we are concerned with. Thus P_N is given as ($\beta = 1/kT$):

$$P_N = \frac{\exp\{-\beta[\epsilon'_F N + V_0 N + \frac{1}{2} UN(N-1) - \mu_P N]\}}{\sum_N \exp\{-\beta[\epsilon'_F N + V_0 N + \frac{1}{2} UN(N-1) - \mu_P N]\}} .$$

In calculating electron transition rates, we determine the single-electron energies in a self-consistent manner. If P is in a state with N electrons on it, an additional electron coming to the state m will sense the average field of the N electrons on P, i.e., its final energy will be

$$\epsilon^f_m(N) = \epsilon'_m + V_0 + NU .$$

Similarly, an electron leaving the state m will have an initial energy corresponding to the average field of the other $N-1$ electrons on P:

$$\epsilon^i_m(N) = \epsilon'_m + V_0 + (N-1)U .$$

In other words $\epsilon^f_m(N)$ is found to be the energy gain of the electron gas on P if an electron is added, and $\epsilon^i_m(N)$ the energy loss if an electron is subtracted. It is just the difference between $\epsilon^f_m(N)$ and $\epsilon^i_m(N)$ which gives rise to the anomalous conductive behavior in question. This can be visualized in a simplified manner as follows. At very low temperatures, transition to the state m is possible only from the filled states below μ_C, that is only for $\mu_C > \epsilon^f_m(N)$. Similarly, transition from the state m to C can occur only for $\mu_C < \epsilon^i_m(N)$. Thus there is no transition possible in between. Since such a situation holds for all N's and m's, we expect a strong reduction of the total, averaged current in the region, where the average direction of the individual transitions is reversed. This region is centered at zero bias, of course.

Now we may solve our transport problem for given environmental parameters μ_A, μ_B, and T. We express directly the partial current between P and C in the first Born approximation as

$$I_{CP} = W_C \sum_N P_N \int_{-\infty}^{+\infty} \{ f(E-\mu_C)[1 - f(E - \epsilon'_F - V_0 - NU)] - [1 - f(E-\mu_C)] f(E - \epsilon'_F - V_0 - NU + U) \} \, dE ,$$

with

$$f(x) = (e^{\beta x} + 1)^{-1} .$$

Here W_C is proportional to an appropriate average of V_{km}'s and to the electron density of states on P and C taken to be energy independent. We have to fulfill the charge-conservation requirement

$$I_{AP} + I_{BP} = 0 \qquad (3)$$

by the proper choice of μ_P, the only variable parameter. The particular value of I_{AP} satisfying Eq. (3) is the net transported current I^* from A to B. It may be noted that I^* depends on the parameters U, T, $V_0 + \epsilon'_F - \mu_A$, W_A, W_B, and $\mu_A - \mu_B = eV$, where V is the voltage difference between A and B.

What remains is the choice of V_0. GZ have pointed out that because of frozen-in polarization effects

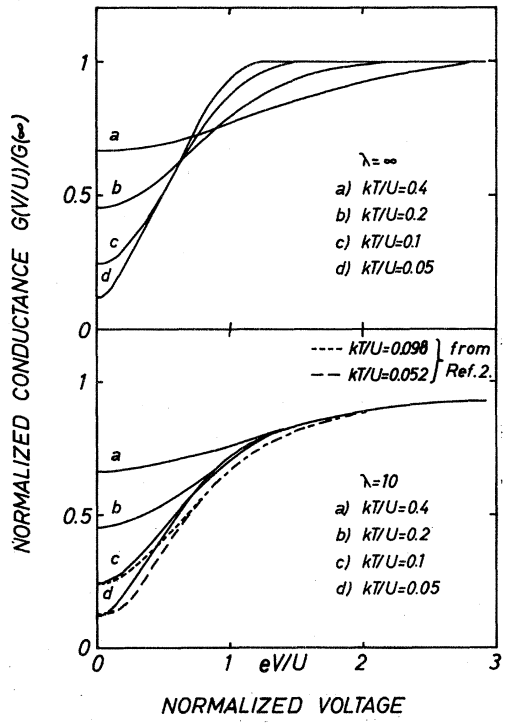

FIG. 1. Calculated conductance-vs-voltage characteristics for several values of the normalized temperature kT/U and the asymmetry parameter λ. For comparison two of the experimental characteristics taken from Fig. 3 of Ref. 2 are also shown normalized with $U = 2.7$ meV (dashed lines).

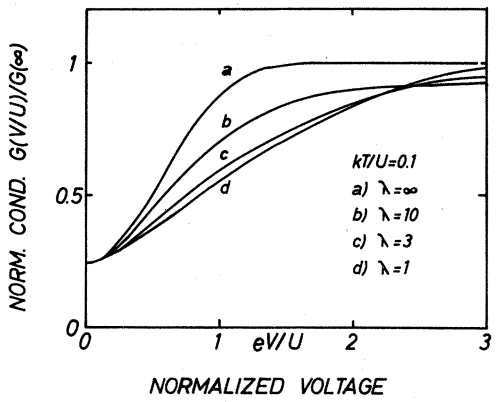

FIG. 2. The dependence of the calculated conductance-vs-voltage characteristics on the asymmetry parameter λ at a given temperature.

in the surroundings of P, V_0 is distributed continuously rather than having only discrete values corresponding to integer numbers of ionic charges. Making use of the fact that I^* is a periodic function of V_0 with a period of U, I^* can be easily averaged over the smooth distribution of V_0 by integrating over a full period. The obtained averaged net current $I(V, U, T, W_A, W_B)$ is our final result.

The outlined calculations have been performed numerically, and the results are represented in Figs. 1 and 2 as the normalized conductance $G(V)/G(\infty)$ vs voltage characteristics, where $G(V) = \partial I(V)/\partial V$ and $G(\infty)$ is the limiting value of $G(V)$ for high voltages, which is independent of V and T. The asymmetry parameter λ is defined as W_A/W_B if $W_A/W_B > 1$ and W_B/W_A otherwise. Note the remarkable difference between the characteristics with $\lambda = \infty$ and $\lambda = 10$.

In particular our results imply that for $kT < 0.2U$,

$$G(0)/G(\infty) \simeq 2.47\, kT/U \qquad (4)$$

independently of λ. The linear dependence of the zero-bias conductance on the temperature for low T's has been already pointed out by GZ; the proportionality constant, however, could not be calculated by their treatment. Eq. (4) is a rather important statement relating the temperature dependence to the width of the conductance anomaly, which is not very strongly affected by the *a priori* unknown λ. (However, for the sample fabrication procedure applied by GZ one would expect $\lambda = 10$ to be in the correct order.) This yields an immediate comparison between theory and experiment. We have performed this for the experimental curves taken from Fig. 3 of Ref. 2 by GZ. From the observed temperature dependence of $G(0)$, we have obtained an averaged value $U = 2.7$ meV from Eq. (4), a value which was used to represent two of the experimental characteristics in our Fig. 1. [Since the last data point was at $V = 5$ mV, we have actually extrapolated $G(\infty)$ to be 15% greater than $G(5\text{ mV})$ with the help of our calculated curves using an estimated U.] The good agreement obtained by this one-parameter fit seems to be convincing. A more detailed comparison with more experiments is beyond the scope of the present paper.

We may mention that our quasiequilibrium approach to determine P_N's is, in principle, justified for voltages below kT only. A somewhat closer inspection shows, however, that our procedure approximates the more rigorous result obtainable by the solution of the kinetic equations reasonably well almost in the entire region of V, T, and λ investigated; deviations beyond a few percent are likely only for $\lambda = 1$ if $V > U$.

Finally we comment on the analogy of our model and the Anderson model. As Denton, Mühlschlegel, and Scalapino[9] have pointed out recently, free metallic particles with an odd number of electrons show a Curie-type magnetic behavior at low temperatures, if the spin-orbital coupling is not too strong. This obviously holds for particles between two metals also if the mixing terms V_{km} are negligible. However, upon reducing the isolation between the particle and one of the electrodes, the mixing terms may become quite important on the energy scale of the average level spacing and U, which will be for particles of about 10-Å diameter of the order of 5 and 100 meV, respectively. Such particles may well serve as the magnetic "impurities" discussed by Anderson[7] in the derivation of Appelbaum's Hamiltonian[10] for magnetic tunneling. According to this, the Schrieffer-Wolff transformation may apply, and the magnetic behavior of the particles may correspond to the s-d model. Thus such particles may cause both the present type and the Appelbaum-type tunneling anomalies, in accordance with the experimental fact that their appearance is correlated. The consequent possibility that small metallic particles semi-isolated from a bulk metal may show Kondo-type magnetic behavior, might deserve some attention as well. These problems will be discussed elsewhere.

The author is indebted to I. Giaever and H. R. Zeller for stimulating discussions and remarks in the early stage of this work, and also to L. L. Hirst for helpful discussions.

---

*This work has been supported by the Sonderforschungsbereich 65 "Festkörperspektroskopie." Darmstadt-Frankfurt.

†On leave from the Central Research Institute for Physics, Budapest, Hungary.

[1] I. Giaever and H. R. Zeller, Phys. Rev. Letters 20, 1504 (1968).

[2] H. R. Zeller and I. Giaever, Phys. Rev. 181, 789

(1969).

[3]P. W. Anderson, Phys. Rev. 124, 41 (1961).

[4]The phonon scattering rate of conduction electrons in metals is of the order $10^{16}(T/\Theta)^3$ sec^{-1}, where T is the temperature, Θ is the Debye temperature.

[5]See C. B. Duke, *Tunneling in Solids* (Academic, New York, 1969).

[6]A. Zawadowski, Phys. Rev. 163, 341 (1967); J. A. Appelbaum and W. F. Brinkman, *ibid.* 186, 464 (1969).

[7]P. W. Anderson, Phys. Rev. Letters 17, 95 (1966).

[8]R. Kubo, J. Phys. Soc. Japan 17, 975 (1962).

[9]R. Denton, B. Mühlschlegel, and D. J. Scalapino, Phys. Rev. Letters 26, 707 (1971).

[10]J. A. Appelbaum, Phys. Rev. Letters 17, 91 (1966); Phys. Rev. 154, 633 (1967); J. A. Appelbaum, J. C. Phillips, and G. Tzouras, *ibid.* 160, 554 (1967).