Effect of Inelastic Processes on Resonant Tunneling in One Dimension

A. Douglas Stone
IBM T.J. Watson Research Center, Yorktown Heights, New York 10598

and

P. A. Lee
Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
(Received 25 September 1984)

We consider the effect of inelastic scattering on tunneling resonances in one dimension using a Breit-Wigner scattering formalism. We show that the peak transmission at resonance is decreased by the ratio of the intrinsic resonance width to the inelastic-scattering rate. For disorder-localized one-dimensional systems this predicts that resonant-tunneling conduction, in addition to variable-range-hopping conduction, will be observable at temperatures below 0.01 K in present experimental systems.

PACS numbers: 71.55.Jv, 72.10.Bg, 72.15.Cz

The existence of scattering resonances in particle collisions is a fundamental feature of quantum theory and has long been used as a probe of the internal energy-level structure of atomic and subatomic particles. Recently Azbel and co-workers[1–3] have made an exciting suggestion: that the sharp structure found in the conductance of quasi-one-dimensional systems as a function of Fermi energy represents a resonant-scattering process which probes the detailed energy-level structure of a macroscopic solid. However, in any resonant process the inverse resonance width sets the time scale for the scattering event, and for the conjectured resonant-tunneling process this time scale is very long. In a solid, inelastic processes due to thermal fluctuations typically occur on a much shorter time scale (even at low temperatures); therefore it is crucial to have a theory of the effect of inelastic processes on tunneling resonances in one dimension (1D) in order to make sensible contact with experiment. We propose such a theory below.

First, we briefly summarize the previous work on this problem. Recently Fowler, Hartstein, and Webb[4] and several other groups[5] succeeded in studying the conductance, G, of very small MOSFET devices which behave like two-dimensional strips of length 10^5 Å and width ~ 300 Å, and of variable electron density. At the lowest temperatures (50 mK) they found that the conductance of these devices varied nonmonotonically over several orders of magnitude as a result of very small changes in electron density (Fermi energy), and that the structure disappeared around 1 K, but was reproducible within a given sample. It was suggested[1–3] that this structure arose from conduction electrons resonantly tunneling through the localized states of this disordered "wire." The zero-temperature theory of the transmission resonances of noninteracting electrons in a random potential was shown to predict a reproducible structure which depended on the particular impurity configuration of the sample[1,6]; and recently[3] a phenomenological extension of that theory to finite temperature was proposed. Very recently one of the authors[7] showed that reproducible structure with many of the features found in the experiment could also be obtained by taking into account finite-size fluctuations in the Mott variable-range-hopping conduction through such small samples (and ignoring resonant processes).

The theory we propose is also phenomenological[8] but differs from that of Ref. 3 in an important feature to be described below. The physical motivation of our approach is the following. The simplest case of 1D resonant transmission occurs when a particle of very well-defined energy E, at zero temperature, is incident upon two consecutive barriers of height much greater than E, with a classically allowed region between them. At almost all energies the particle is almost totally reflected, but in exponentially small energy intervals of width Γ (the leak rate out of the interior of the scatterer) the particle transmission is greatly enhanced. Because the spread in the incident particle's energy must be less than Γ to see a strong resonance, its interaction time with the scatterer cannot be less than Γ^{-1}. During this interaction a relatively large electronic density builds up in the allowed region between the barriers, because the wave leaking through the first barrier is constructively interfering with reflections off the second barrier. The situation is exactly the same for a scatterer consisting of many small barriers of random height, except that the sequence of internal reflections leading to constructive interference is much more complicated. In this case a large electronic density builds up in a disorder-localized resonant state of a 1D resistor on a time scale $\tau_r \sim \Gamma^{-1} \sim (2\pi/\Delta\epsilon)\exp(L/L_0)$, where L is the sample length, L_0 is the localization length, and $\Delta\epsilon$ is the spacing between resonances. For the experimental system of Fowler, Hartstein, and Webb one finds $\tau_r \gtrsim 0.01$ s. However, for these systems at finite temperature there is certainly a compet-

 © 1985 The American Physical Society

ing conduction mechanism, variable-range hopping,[9] which occurs on a much shorter time scale $\tau_m \sim (1/\omega_0)\exp(T_0/T)^{1/2}$, where ω_0^{-1} is a time on the order of 10^{-13} sec. Imagine putting a field on the 1D resistor at finite T on resonance. For times less than τ_m the resonant density will build up in the scatterer exactly as it would at zero T. However, after a time τ_m the electron will scatter incoherently to another site and the electronic density in the center of the sample will never be able to build up to nearly its full resonant density. Thus, the peak transmission at resonance will be reduced by a factor τ_m/τ_r.

To make the argument quantitative we employ a formalism analogous to that used to derive the Breit-Wigner formula, familiar in nuclear physics, for the scattering cross section near resonance when several decay modes (elastic and inelastic) are possible.[10] However, our derivation differs from the usual Breit-Wigner result in two important ways. First, we are considering a 1D scatterer with very low transmission away from the resonance; thus the total scattering is always large, but mostly backscattering (reflection), and we must show that the forward scattering (transmission coefficient) behaves like the cross section in the Breit-Wigner formula. Second, our basic result cannot depend on the assumption that the potential is symmetric, which requires a significant generalization of the usual approach. Here, however, we present only the details of the symmetric case for simplicity.

We consider an arbitrary 1D scattering potential with $V(x)=V(-x)$, $V(x\rightarrow\infty)=0$ and (outermost) classical turning points at $\pm L$. Any normalized solution of the Schrödinger equation is fully determined by specifying its logarithmic derivative at a point. Therefore the logarithmic derivatives of two independent solutions in the asymptotic region $[|x| >> L, V(x)\simeq 0]$ in principle contain all the necessary information about the behavior of the solutions inside the scattering region as a function of energy. The idea is to choose a basis where the logarithmic derivatives for $|x| >> L$ are slowly varying with energy, except at resonance, when the logarithmic derivative of one of the solutions varies rapidly. Expansion of this logarithmic derivative near the resonance energy will give the behavior of the transmission coefficient near resonance. For a symmetric $V(x)$, the correct basis to choose is the symmetric and antisymmetric states at each energy. This is because the resonances are closely related to the discrete eigenstates of the Schrödinger equation with fixed boundary conditions at $\pm L$, which have a definite parity if the potential is symmetric. At most energies the parity eigenfunctions of our scattering problem have very little probability density inside the scatterer; however, at a resonance the wave function with the appropriate parity is peaked within the scattering region and has a shape similar to the corresponding eigenstate of the discrete problem.

In general, all solutions have $\psi(x)=Ie^{ikx}+Oe^{-ikx}$ for $x << -L$ and $\psi(x)=I'e^{-ikx}+O'e^{ikx}$ for $x >> L$. Define the transfer matrix by

$$\begin{pmatrix} I' \\ O' \end{pmatrix} = M \begin{pmatrix} I \\ O \end{pmatrix},$$

where

$$M=\begin{pmatrix} -r/t & 1/t \\ 1/t^* & -r^*/t^* \end{pmatrix},$$

and r and t are the reflection and transmission amplitudes for a wave incident from $-\infty$. The symmetric and antisymmetric solutions for $|x| >> L$ are obtained from the eigenvectors of this matrix with eigenvalues ± 1. These states may be chosen to be $\psi_\pm(x) = \cos[kx-\phi_\pm(E)]$ for $x << -L$, where $\exp(2i\phi_\pm) = r \pm t$. Therefore

$$t=\tfrac{1}{2}[\exp(2i\phi_+)-\exp(2i\phi_-)]. \quad (1)$$

The logarithmic derivatives $\psi'_\pm(x)/\psi_\pm(x)$ are antisymmetric so we need only consider them in the region $x\rightarrow -\infty$. We define, for $x << -L$,

$$b_\pm(x,E) \equiv -\frac{1}{k}\frac{\psi'_\pm(x)}{\psi_\pm(x)} = \tan[kx-\phi_\pm(E)]. \quad (2)$$

We can now express the transmission amplitude in terms of $b_\pm$ using (1) and (2),

$$t=\frac{e^{2ikx}}{2}\left\{\frac{(1-ib_+)^2}{1+b_+^2}-\frac{(1-ib_-)^2}{1+b_-^2}\right\}. \quad (3)$$

Far from a resonance $b_\pm$ is a very slowly varying function of energy and from (1) we see $\phi_+-\phi_- \simeq |t| << 1$, so that $b_+ \approx b_-$. Assume that E is near a symmetric resonance, so that $b_+(E)$ varies rapidly between $\pm\infty$, while b_- remains constant. Since t is independent of x we can evaluate Eq. (3) anywhere in the asymptotic region, and it is convenient to evaluate it at a point x_0 which is a node of $\psi_-(x)$, so b_- is infinite. Then maximum resonance occurs at the energy E_r when b_+ is zero. We expand b_+ around E_r, $b_+(E)\simeq(\partial b/\partial E)(E-E_r)$, and define the elastic-resonance width $\Gamma_e = 2/(\partial b/\partial E)$. Substitution into (3) yields, near resonance,

$$t=i\exp(2ikx_0)\frac{\frac{1}{2}\Gamma_e}{(E-E_r)+\frac{1}{2}i\Gamma_e} \quad (4)$$

which is the usual result when only elastic decay is possible.

We now allow for the possibility of inelastic decay modes by assuming that the scattering potential has a small constant imaginary part, $\frac{1}{2}\Gamma_i$,[10] which absorbs part of the incident flux and causes a breakdown of unitarity. Then the energy at resonance is shifted to

$E+\frac{1}{2}i\Gamma_i$ and by Eqs. (3) and (4), this gives an imaginary part to b_+, $\text{Im}(b_+)=\Gamma_i/\Gamma_e$. Substituting this into (3) yields

$$|t_e|^2=\frac{(\frac{1}{2}\Gamma_e)^2}{(E-E_r)^2+(\frac{1}{2}\Gamma)^2}, \tag{5}$$

where the total decay width $\Gamma\equiv\Gamma_e+\Gamma_i$ and we now must interpret t in Eq. (3) as the *elastic* transmission amplitude t_e. We can calculate the total *inelastic* scattering (reflection plus transmission) using

$$|t_i|^2+|r_i|^2=1-|t_e|^2-|r_e|^2$$

$$=\frac{\frac{1}{2}\Gamma_e\Gamma_i}{(E-E_r)^2+(\frac{1}{2}\Gamma)^2}, \tag{6}$$

where the elastic reflection coefficient $|r_e|^2=\frac{1}{4}\times|\exp(2i\phi_+)+\exp(2i\phi_-)|^2$, which can be evaluated near resonance in the same manner as t_e above. Assuming the probabilities of inelastic forward and backscattering are equal, (5) and (6) give the total transmission coefficient in the presence of inelastic scattering as

$$T=|t_i|^2+|t_e|^2=\frac{\frac{1}{4}\Gamma_e\Gamma}{(E-E_r)^2+(\frac{1}{2}\Gamma)^2}. \tag{7}$$

The peak transmission is reduced in the presence of inelastic scattering by Γ_e/Γ, if $\Gamma_i>>\Gamma_e$. The same reduction of the peak transmission is obtained when $V(x)$ is asymmetric; the details of this argument will be given elsewhere.[11] Note, the above argument always assumes a monoenergetic incident wave. In a real system the electrons have a Fermi-Dirac energy distribution and their energies are smeared out by KT around the Fermi level, E_F. If E_F is within KT of E_r, then the *integrated* resonant transmission (assuming Γ is much less than the resonance spacing $\Delta\epsilon$) is, by (7) and (4), independent of Γ_i and proportional to Γ_e, since inelastic scattering broadens the resonance while damping it. This is to be contrasted with the theory given in Ref. 3 where it is proposed that inelastic scattering broadens the resonance but does not damp it, so that the integrated resonant transmission is actually *increased* by inelastic scattering. In our theory the integrated resonant transmission is always proportional to Γ_e but the nonresonant (hopping) transmission is proportional to Γ_i, and will rapidly dominate the resonant transmission when $\Gamma_i>>\Gamma_e$. Therefore, either by looking at the peak transmission or the integrated transmission one arrives at the same sensible physical criterion: *Resonant tunneling is only observable when the intrinsic (elastic) resonance width is greater than or equal to the inelastic width.*

We now specialize to the case of a short, quasi-one-dimensional, disordered resistor, where the important inelastic conduction mechanism is variable-range hopping, with $\Gamma_i=\Gamma_m=\omega_0[\exp-(T_0/T)^{1/2}]$, and the elastic width, defined earlier, is the leak rate out of a state localized at the center of the sample. The criterion for seeing large resonant tunneling effects becomes

$$\frac{\Gamma_m}{\Gamma_e}=\frac{2\pi\omega_0}{\Delta\epsilon}\exp\left[-\left(\frac{t_0}{T}\right)^{1/2}+\frac{L}{L_0}\right]\leq 1. \tag{8}$$

Roughly speaking the criterion (8) means that resonant tunneling is unimportant until the Mott hopping length is half the sample length, which is reasonable since this is the distance an electron must elastically tunnel to get into a strongly resonant state. If we take the experimental parameters of Ref. 3, and assume the most favorable (longest) possible value for the localization length, the temperature T_c at which (8) is satisfied is about 10 mK. Since the lowest temperature at which experiments have been done is 50 mK, it appears unlikely that resonant tunneling is responsible for the structure in the conductance seen in present experiments.

As noted earlier, this structure may be explicable solely in terms of finite-size fluctuations in variable-range-hopping conduction. The two mechanisms may be distinguished experimentally by looking at the structure $\ln G$ vs E_F as a function of temperature. Above T_c the peaks are due to fluctuations in the critical path for Mott hopping and thus should change substantially both in location and magnitude when the temperature is changed by an amount $(T_0T)^{1/2}$, which is a typical energy barrier in Mott hopping. Very recently careful measurements of this type have been made and appear to agree better with the hopping-fluctuations theory.[12] In general, the observed rapid growth of the peaks with increasing temperature is inconsistent with our above calculations (but not with those of Ref. 3). Below T_c there are two possible behaviors depending on how far E_F is from the energy of a good resonance, E_r. Fermi energies within KT of some E_r will give rise to resonance peaks in $\ln G$ whose locations and maximum intensities are essentially temperature independent. When the Fermi energy is near a resonance, but not within KT, the primary conduction mechanism will be activated hopping to the resonance,[3] and this will give a triangular shape to the peaks of $\ln G$ with slope $1/KT$. In the valleys the primary conduction mechanism will still be Mott hopping until the temperature is so low that the hopping length becomes longer than the sample length. Then, at energies where $(E_F-E_r)/KT>L/L_0$ the activation barrier to the resonance is so high that the electrons prefer to tunnel nonresonantly through the sample. This will give rise to a flat, temperature-independent background with $\ln G\simeq-2L/L_0$ (where the conductance G is measured in units of e^2/h). Resonant tunneling behavior of this type may well be observable experimentallly if it is possible to fabricate shorter samples with characteristics similar to those already studied.

We thank Mark Azbel for many stimulating discussions on this problem, and A. Fowler and A. Hartstein for useful discussions of their experimental results. We also gratefully acknowledge the hospitality of the Aspen Center for Physics, where some of this work was done. This work was partially supported by Joint Services Electronics Program under Grant No. DAAG 29-83-K-0003.

[1]M. Ya. Azbel, Solid State Commun. **45**, 527 (1983).

[2]D. P. DiVincenzo and M. Ya. Azbel, Phys. Rev. Lett. **50**, 2102 (1983).

[3]M. Ya. Azbel, A. Hartstein, and D. P. DiVincenzo, Phys. Rev. Lett. **52**, 1641 (1984).

[4]A. B. Fowler, A. Hartstein, and R. A. Webb, Phys. Rev. Lett. **48**, 196 (1982), and Physica (Utrecht) **117&118B**, 661 (1983).

[5]W. J. Skocpol, L. D. Jackel, E. L. Hu, R. E. Howard, and L. A. Fetter, Phys. Rev. Lett. **49**, 951 (1982); R. F. Kwasnick, M. A. Kastner, J. Melngailis, and P. A. Lee, Phys. Rev. Lett. **52**, 224 (1984); R. G. Wheeler, K. K. Choi, A. Goel, R. Wisnieff, and D. E. Prober, Phys. Rev. Lett. **49**, 1674 (1982).

[6]I. M. Lifshitz and V. Ya. Kirpichenkov, Zh. Eksp. Teor. Fiz. **77**, 989 (1979) [Sov. Phys. JETP **50**, 499 (1979)].

[7]P. A. Lee, Phys. Rev. Lett. **52**, 1641 (1984).

[8]An earlier attempt to resolve this question on the basis of a microscopic model is described in A. D. Stone, M. Ya. Azbel, and P. A. Lee, Phys. Rev. B **31**, 1707 (1985).

[9]J. Kurkijärvi [Phys. Rev. B **8**, 922 (1973)] has shown that hopping conduction is activated and not variable range for a sufficiently long 1D system; however, numerical simulation (Ref. 7) and experiment (Ref. 4) indicate that Mott hopping is indeed the relevant conduction mechanism for these systems.

[10]A. Bohr and B. Mottelson, *Nuclear Structure* (Benjamin, New York, 1969), Vol. 1 pp. 430–447. See also L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Pergamon, Oxford, 1977), 3rd ed., Vol. 3, p. 603.

[11]A. D. Stone and P. A. Lee, unpublished.

[12]R. A. Webb, A. Hartstein, J. J. Wainer, and A. B. Fowler, unpublished.