ELECTRON TUNNELING in SOLIDS

A conference report by **P. J. Price**

QUANTUM tunneling of a particle through a barrier was introduced in the theory of field ionization of atoms (1927), and to account for field emission of electrons from surfaces and for alpha emission by nuclei (1928). Electron tunneling within a solid, though more complicated than tunneling of a free particle, arises in essentially the same way from a mode of wave propagation. It has recently become the basis of a new area of research. The figures below depict these varieties of tunneling process. In Fig. 1, according to classical mechanics, the particle coming from the left would be reflected at the "classical turning point", *A*, where the potential energy equals its original kinetic energy. The Schrödinger wave governing its quantum behavior penetrates into the "forbidden" region, however, and in consequence the particle may travel onward from the second turning point, *B*. The probability of transmission, instead of reflection, decreases exponentially with increasing forbidden path length, *AB*.

Tunneling between bands was first proposed by Zener[1] as an explanation of dielectric breakdown in a strong electric field. Figure 2 shows a somewhat different situation: a semiconductor junction in which the two sides are so heavily doped that the electron and hole distributions are degenerate (that is, the Fermi level lies inside the conduction and valence bands respectively). Figure 2 is the conventional potential-profile depiction of a semiconductor junction. The two "allowed" regions, to the left of *A* and to the right of *B*, are where the electron's energy places it respectively in the conduction band and in the valence band of Bloch states. The Schrödinger wave penetrates into the "forbidden" region, *AB*, where the electron energy corresponds to the exponentially varying wave functions of the "forbidden gap". When an applied emf separates the Fermi levels on the two sides, the electrons lying in the energy range between the two Fermi levels tunnel across the junction, carrying a current. In the conventional "forward bias" direction, when the left-hand (n) side is made negative as in Fig. 2, the tunneling electrons flowing to the p side appear in the valence band (whereas for the usual diffusion current they would appear as "minority carriers" in the conduction band). With reverse bias, the passage of electrons from the valence band to the conduction band is essentially the phenomenon postulated by Zener. In practice, the width of the junction is some tens of Angstroms, and the field strength is $\sim 10^6$ volts/cm. It might be an interesting question,

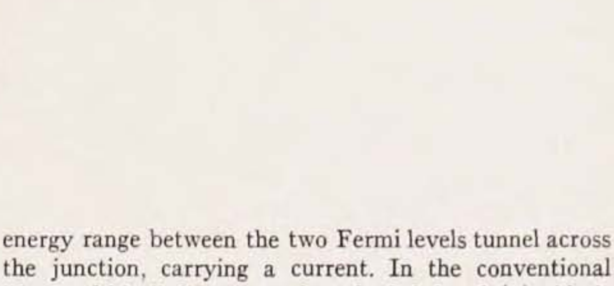

Fig. 1. Tunneling of a free particle

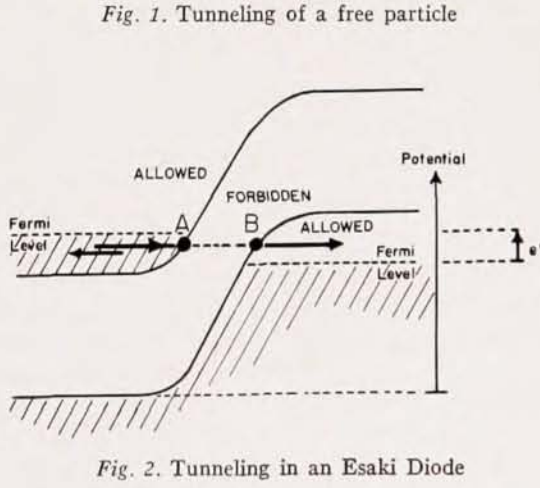

Fig. 2. Tunneling in an Esaki Diode

P. J. Price is an IBM Research staff member. His principal research interest is electronic phenomena in solids.

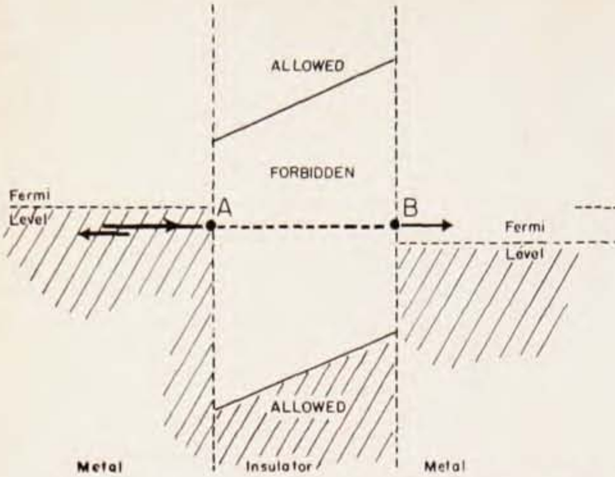

Fig. 3. Tunneling through a barrier between metals

for philosophers of research, why such tunneling junctions were made, and their properties recognized, only after a decade of intensive research on *p-n* junctions. Their discovery by Esaki,[2] and recognition of their possible value as electronic devices, initiated the present widespread research on electron tunneling in junctions.

The Esaki diode research stimulated a revival of interest in the kind of electron tunneling depicted in Fig. 3. The insulator layer of the sandwich has to be very thin (again, a few times 10 A), and such structures were first made by laying down an aluminum film by evaporation, oxidizing a surface layer of the film, then evaporating a second metal layer on top. The success of this technique, and then the striking and beautiful current-voltage functions which were found when one or both metals were superconductors, indicated that this variety of tunneling phenomenon was likely to be as important as that in *p-n* junctions. At this stage the timeliness of a research conference was recognized by Elias Burstein, who organized the symposium on which this article reports.

The Symposium on Electron Tunneling in Solids, sponsored by the Laboratory for Research on the Structure of Matter, University of Pennsylvania, and by ARPA, was held on January 30 and 31, 1961. The 134 conferees included five from Europe: P. Aigrain (Paris); V. A. Chuenkov and L. V. Keldysh (Moscow); W. Franz (Hamburg); and J. G. Park (London). Sessions and conferees were both housed in Philadelphia's new Sheraton Hotel, an arrangement which made an all-work conference—about 15 hours of sessions in two days—as painless as possible. The present report covers most, though not all, of the conference topics, but does not attempt to mention all contributions or contributors.

THE currently most successful research on junction tunneling—investigation of the "phonon assisted" processes—does not involve the details of the tunneling phenomenon, but only that it behaves like a quantum transition governed by the exclusion principle. The current from side "1" to side "2" is given by a sum, over the states on side "1", of $f_1(E_1)\,[1 - f_2(E_2)]$ where the f's are the Fermi functions of electron energy E, with $f_1(E) = f_2(E - eV)$ where the bias energy, eV, is reckoned positive for forward bias. In the ordinary (Zener) tunneling process, $E_1 = E_2$. It has been found, however,[3] that the tunneling transition between bands may also be accompanied by absorption or emission of a phonon whose wave vector connects the conduction and valence band edges in the Brillouin zone, and then E_1 and E_2 differ by the phonon energy. Consequently, at low temperatures the current due to this particular tunneling process flows only for applied potentials V such that eV exceeds the phonon energy. The phonon energies appear as inflections of the junction's current-voltage function, which thereby provides a spectrometer for the lattice frequencies at the appropriate wave vectors. The frequencies determined in this way, for germanium and silicon, have agreed well with those of Brockhouse from neutron spectroscopy and, as R. N. Hall and A. G. Chynoweth showed at the conference, the picture is now almost complete for these semiconductors. Many of the appropriate lattice frequencies for the III-V compounds have also been determined, and more recently some for the lead-sulphide series of semiconductors. In practice, one finds additional inflections due to tunneling transitions in which more than one phonon is emitted. Hall reviewed the results obtained, so far, by himself and collaborators with this new spectroscopy.

Photos courtesy University of Pennsylvania News

L. V. Keldysh (left) and W. Franz

R. N. Hall, lecturing on phonon spectroscopy

There is a similar sort of situation for the tunneling process depicted in Fig. 3, when one or both metals are superconductors, in that the current-voltage characteristic is indicative of the energy states of the two metals rather than of the tunneling process itself. The results obtained by Giaever and Shapiro and their collaborators for lead, tin, aluminum, and indium provide a dramatic confirmation of the BCS theory.[4] Sharp inflections in the current-voltage characteristic show directly the existence of the superconductor energy gap (the gap magnitudes evidently differ substantially from 3.5 kT_c, however). As the results presented at the symposium showed, for small biases (comparable with the energy gap) the net current J is given by

$$J = \text{const.} \int [f_1(E) - f_2(E)] \rho_1(E) \rho_2(E) dE$$

where the ρ's are the two density-of-states functions (also displaced, relative to each other, by the bias energy, eV). The $\rho(E)$ function obtained, by analyzing the characteristics on this basis, reproduces quite well the peculiar $\rho(E)$ function given by the BCS theory. The temperature and magnetic-field dependences of the gaps also are found to agree with the BCS theory. It will be interesting to see whether the "hard" superconductors do or do not show a similar conformity with BCS. In these experiments "normal quasielectrons" are injected at the interface. Burstein drew attention to the analogy with minority-carrier injection into a semiconductor, and the possibility of analogous properties such as photovoltaic effects.

An equation, such as the one above, for the current J, if derived from first principles, would contain a factor $|M_{12}|^2$ in the integral, with $M_{12}(E)$ the matrix element for transitions through the barrier. The experimental results reported at the conference showed that $|M_{12}|^2$ is independent of E (and, in fact, has the same value as for the normal state of the superconductor) throughout the neighborhood of the energy gap where $\rho(E)$ has its characteristic variation. This conclusion seems quite contrary to the results which have been arrived at for Esaki junctions, and it was accordingly the subject of some discussion. For Esaki junctions, the WKB solutions for the envelope wave functions give $|M_{12}|^2$ proportional to v_1v_2 where the v's are the normal components of electron velocity outside the space-charge region. Usually (both for an Esaki junction and for a barrier between metals) the current is contributed by a small solid angle of velocity directions nearly normal to the junction plane. The factor v_1v_2 in $|M_{12}|^2$ then just compensates the density-of-states factors in the integral giving the current.[5] The same cancellation for the barrier-and-superconductor(s) case would result in a $J(V)$ characteristic quite contrary to the experimental results. When the role of the neighborhood of the classical turning point (Fig. 2) is taken over by the inherently more complicated crystal interface, however, the analysis which gives the v_1v_2 factor no longer applies. The turning-point formula connecting allowed and forbidden Bloch waves becomes a more general formula expressing current conservation, but the latter is not a sufficient condition to determine the matching of states at the interface. The details of the interface region are involved also. It appears that, with a superconductor, some equivalently simple situation appears again. Bardeen expounded the view that this arises from the many-electron character of the superconductive state. The constancy of the effective matrix element recalls F. London's interpretation of the Meissner effect as a "rigidity of the wave function".

Esaki junctions are more complicated, in many ways, than the models on which theory has been based so far. A theme of current research, therefore, is to investigate the agreement between their properties and the existing tunneling theory. The central result of the latter, which follows from the WKB formula for the envelope wave functions in the junction region, is that the tunneling probability is proportional to exp $(-\lambda)$ where

$$\lambda = 2 \left| \int_A^B \boldsymbol{\alpha}(\mathbf{r}) \cdot d\mathbf{r} \right|$$

and $\boldsymbol{\alpha}$ is the imaginary part of the wavevector of the "forbidden" Bloch state at each point of the path AB. M. I. Nathan showed how this relation is partly, though not conclusively, verified experimentally by displaying the dependence of the current density, through the exponent λ, on junction field and on hydrostatic pressure. The results confirm that the exponent varies linearly with (relatively small) changes of field, or with pressure, at a rate proportional to its value at zero bias or pressure. Encouraging progress was also reported, at the conference, with another approach in which the effect of a static magnetic field is investigated. When the magnetic field is "longitudinal", so that the classical Larmor orbits are parallel to the junction plane, its effect should be not to change the character of the tunneling motion through the junction but, through the Landau-Peierls quantization, to change the effective energy gap and hence the exponent. With a transverse magnetic field, on the other hand, the resulting Lorentz force has a component parallel to the junction field. The character of the tunneling motion is thereby changed, and the theory of the effect becomes quite complicated. R. R. Haering showed that the magnetic-field dependence of the current for both orientations of magnetic field could be analyzed so as to determine λ. With Esaki's data for In-Sb junctions in a magnetic field, it was tentatively verified that the values so obtained for different reverse biases did satisfy the theoretical relation [5]

$$J/V = \text{const.} \exp\left[-\lambda(V)\right].$$

Thus, our main idea about the phenomenon seems to be confirmed.

As was mentioned above, the tunneling phenomenon may be understood in terms of the *envelope* wave function, in which the variation of the actual electron wave function on an atomic scale, with the lattice periodicity, is disregarded. The possibility of observing a manifestation of this lattice periodicity is indicated in some work which was presented by G. H. Wannier. In a crystal in a uniform electric field, the electron states of each band form a "Stark ladder" of states equally spaced in energy, the energy interval being (field) $\times$ (electron charge) $\times$ (lattice constant in the field direction). The states of the ladder are identical except that their wave functions are successively displaced, in the field direction, by the lattice constant. In the situation of Fig. 2 the energies of the states form a continuum, of course. Wannier argued, however, that, although the Stark ladder is in this sense lost, one should expect the states to retain a variation, having the periodicity of the lattice constant, with the position of the classical turning point. The states on each side of the junction would therefore vary correspondingly with energy, with a period proportional to the fields at A or B respectively. When the two fields at the Fermi energy were roughly equal, one should find a corresponding periodicity of tunneling current with bias. Wannier presented some data, obtained by Chynoweth and collaborators, for In-Sb diodes, which showed such a periodicity. If the proposed interpretation of this data is correct, the phenomenon provides a unique and con-

G. H. Wannier and "Stark ladder" characteristic of In-Sb diode

venient way of observing the lattice periodicity of Bloch states.

The electrons in a degenerately doped semiconductor are subject to strong coulomb forces in addition to the lattice potential, and so it is of dubious validity to describe their behavior in terms of Bloch states. Keldysh outlined a field-theoretic treatment of tunneling, designed to take account of such complications. Although we are now dealing with a true many-electron state on each side of the junction, the tunneling transition itself is induced by a one-electron perturbation (interaction with the lattice modes or with the junction field). Accordingly, the tunneling rate may be calculated in terms of this interaction and the two one-electron Green's functions for the conduction band on the n side of the junction and for the valence band on the p side. Keldysh's equations for these one-electron Green's functions involve the junction potential in a manner which generalizes the usual formulas of the independent-electron theory.

An important feature, for engineering purposes, of the current-voltage characteristic of Esaki diodes [2] is the decrease of current which follows its initial increase, as the bias voltage is increased from zero in the forward direction. This negative conductance may be attributed both to the "uncrossing of the bands" and to the increase of the exponent λ, as the bias is increased; but its true nature has not been established in any detail, so far. ("Uncrossing" refers to forward biases large enough for there to be no electron energies common to the conduction band on the n side and the valence band on the p side. It is believed, however, that in these heavily doped semiconductors the energy distribution of the states is modified, with a "tail" in place of a sharp band edge.) For considerably larger

E. O. Kane

Leo Esaki, L. V. Keldysh, and J. Bardeen

voltages the current again increases rapidly, evidently on account of the minority-carrier diffusion current. The "valley" between the decreasing tunnel current and this final rise is, normally, partly filled in by a third component, the so-called "excess current". The excess-current characteristics are not all alike: an exponential increase with voltage, and a broad "hump" with sides formed by increasing and decreasing currents, are common forms. A portion of the symposium was concerned with the nature of excess current.

If the bands are effectively uncrossed, then the electrons carrying the excess current must lose energy on their way through the junction. However, E. O. Kane pointed out that the conceivable tunneling processes in which this electron energy is given up by interaction with other degrees of freedom—with the lattice (emission of many phonons), with the radiation field (photon emission), and with other electrons (Auger process)—turn out to have theoretically predicted rates which are too small to account for the observed excess current. It is believed, rather, that excess current is due to electron transitions from the conduction band to, and then to the valence band from, localized states associated with impurities and lattice defects in the junction. It has indeed been found that additional excess current is induced by radiation damage and by doping with certain impurities. C. T. Sah reported that diodes made from silicon containing gold, in addition to the normal doping, have an excess current characteristic with several sharp features, which he identified with electron levels of various charge states of the gold atoms. This phenomenon is presumed to correspond to the naturally occurring humps. On the other hand, R. A. Logan and colleagues reported that fast-electron bombardment of diodes may induce either hump or

Elias Burstein (left), organizer of the symposium

exponential types of excess current. (Comparable work by R. S. Claassen was not represented at the conference.) This situation was clarified, however, by annealing studies.[6] The two types of radiation-induced excess current change into, or are superseded by, each other on annealing at different temperatures, and to some extent with identifiable activation energies. In this way the defects causing different components of the excess current may be disentangled and identified. The locations of the participating defects (inside the space-charge region or on its boundaries), and the natures of the transitions (both or only one inelastic;

both or only one a tunneling process) may, however, be more difficult to find out.

THE study of barrier tunneling between metals is at an early stage, and many fruitful future directions may not even have been delineated. Fabrication of single-crystal structures may allow a great advance in what can be done. As was reported at the conference, the fabrication technique is no longer confined to oxidized aluminum barriers. Not only have oxidized layers of other metal films been used, but also separate "sandwich fillings" of barium stearate and of polymer layers. With these, however, it may be that the electron transition is not describable in conventional tunneling terms. The latter do appear, tentatively, to apply to oxide barriers between normal metals, since the transition to the Fowler-Nordheim characteristic at large biases has been observed (and the appropriate "work functions" thereby determined).

Murray Lampert makes a point

Pierre Aigrain, J. Grossman, J. Pankove

For Esaki junctions, a number of interesting questions have emerged. It seems to have been established that the prefactor for the current (the "const." of the equation above for J/V) differs substantially from the value given by theory. Our conception of the tunneling transition may accordingly have to be revised, though much more experimental study is needed. The tunneling current characteristic for forward biases has not been accounted for. The development of the tunneling theory itself draws attention to gaps in our knowledge: The wave vector—energy relation for the forbidden energy ranges is of a different character than for the allowed ranges, and neither the formal theory nor the empirical data now available to us may be sufficient for an adequate description of the system of "forbidden" electron states. Finally, there is the problem of the nature of the tunneling process when coupling effects, beyond (or including?) the coulomb interaction of the electrons presumed to be included in the Bloch scheme, are important. (Such a coupling effect may underlie the dip in the conductance near zero bias, for Esaki junctions in polar semiconductors, found at low temperatures by Hall and his co-workers and ascribed by them to a failure of the lattice polarization energy to "adjust", in the tunneling process, from the value for the initial electron state in one band to that for the final state in the other.) For the case of coupling, by the phonon field, of the thermal band edge states to those of a higher minimum in the same band, Kane reported his conclusion that in tunneling with emission of the phonon the governing exponential factor is still that associated with the thermal gap. This result, which is contrary to what one anticipates naïvely from the form of the second-order perturbation theory, is suggestive of what might be found more generally by the Keldysh approach. For electron cooperative effects, in addition to superconductivity, one thinks naturally of magnetic materials. Here both barrier tunneling (with transition metals) and Esaki junctions (made by doping magnetic insulators) are obviously of great potential interest, and preliminary investigations of the former were mentioned at the conference.

The foregoing examples of potential developments suggest that, in Hall's words: "This is not to be the last conference on electron tunneling."

Bibliography

1. Zener, C., Proc. Roy. Soc. **A145**, 523 (1934).
2. Esaki, L., Phys. Rev. **109**, 603 (1958).
3. Holonyak et al., Phys. Rev. Letters **3**, 167 (1959); Esaki, L., and Miyahara, Y., Solid State Electronics **1**, 13 (1960).
4. See the article by R. W. Schmitt in Physics Today, December 1961.
5. Price, P. J., and Radcliffe, J. M., IBM Journal **3**, 364 (1959); Kane, E. O., J. Appl. Phys. **32**, 83 (1961).
6. Logan et al., J. Appl. Phys. **32**, 1201 (1961).