TUNNELING PHYSICS

Carver A. Mead
California Institute of Technology

I should like to discuss what I feel to be the two fundamental principles in our understanding of solid state physics to date. I think those of you who are experts on the subject will agree when I say that probably the outstanding features of solid state physics, as opposed to other branches of physics, are (1) that we deal with systems having a periodic structure of atoms or molecules of some sort - a crystal lattice, which is nearly perfect or nearly periodic over a large number of lattice spacings, and (2) that we deal with systems where the wave nature of the electron is of primary importance. I should like to put these two ideas together by considering what would happen, on a one-dimensional model, if we try to propagate an electron wave down through a periodic structure or lattice.

Let us first consider an electron nearly at rest. In this case, the wavelength is very long and an electron can propagate unimpeded through a lattice.

Let us then give the electron more and more momentum - push it harder, down through the lattice. As the momentum increases, the wavelength becomes shorter and we come to the point where the electron's wavelength becomes equal to the spacing between the crystal's lattice points. We may assume that there is some type of interaction between the lattice and the electron wave. This can be called "reflection," or "scattering," or whatever one likes.

When the electron's wavelength is equal to the spacing between the lattice points, this scattering, or reflection, from the first lattice point will be augmented by the scattering of the electrons from the second lattice point. A little bit more of it will be reflected from the next lattice point and a little bit more from the next. The reflections will all be in phase and the amplitude of the electron wave transmitted in the forward direction will approach zero. One can recognize immediately how this electron dies out going through the periodic structure of the lattice. If a certain fraction of the electron wave is scattered or reflected from each lattice point, this then is the amount of the decrement in the amplitude of the electron wave. When the decrement in the amplitude of the wave is proportional to the amplitude, the wave is damped out exponentially with distance. Our concept of the forbidden band in a crystal is basically that there are no propagating solutions for the electron wave with this certain momentum in this periodic structure. It does not mean that electrons with this energy cannot occur; it merely means that they will not propagate for any appreciable distance.

The tunneling physics that I would like to discuss is based on the behavior of electrons with this momentum. Figure 1 shows a generalized, solid state tunnel device. This device has two wires, one at each end with some material in between. This is a representation that could be a PN junction, where the material in between happens to be a crystal, half of which is P-type and half N-type. It could be only an insulator (which is what is shown) or it could be any one of a number of things. You can then ask: If there are electrons in the insulator, or a semiconductor which is in the middle, and if there are electrons on both ends, what sort of tunnel transitions can occur? The term "tunnel transition" as used here, means the transition in which the electron wave on passing (propagating) through a crystal (as from the left to the right in Figure 1), even though exponentially damped in the forbidden band of the crystal, is not completely damped because the distance through the forbidden band is very short. The electron wave is only completely damped if it must go an infinite distance
in the forbidden band. Therefore, if we make the "forbidden band" (the region through which it has to tunnel or through which it has to penetrate) short enough so that there is still some amplitude remaining at the other end, a certain fraction of the electrons manage to get through. There are many ways in which they can do this.

In a tunnel transition the electron wave can start in the metal, on the left of the diagram, and go through to the conduction band of the insulator or the semiconductor (3). They can start in the valence band of the semiconductor and move into the conduction band of the semiconductor (1) or they can start in the valence band and move into the other metal (4). All these mechanisms are possible wherever there is an electron which has only a finite distance to go until it reaches a position where it is allowed to have a propagating solution. These are so-called "tunnel transitions" - or electrons penetrating through the "forbidden region."

An Esaki tunnel diode is a device which looks very much like the energy diagram in Figure 1, except that there are some kinks in it. Basically, we are looking at transitions from the valence band of the semiconductor into the conduction band. One experiment, which I should like to tell you about has been done at numerous laboratories around the country and is quite illuminating to our understanding of tunnel physics.

Consider a device which has a forbidden band through which electrons are tunneling. In reality, we never have a one-dimensional type of periodic structure. We have a crystal which works in three dimensions; therefore, our placement of the conduction and valence bands, or the energies at which the electron has a wave length which just allows it to be totally reflected, vary in the different directions. Consequently, if we talk about valence and conduction bands, for example, it is not necessarily true that the valence electron with the highest energy is going in the same direction as a conduction electron with the lowest energy. If one wishes to make a transition in a semiconductor device which has an electric field in it where the electron has to penetrate the minimum distance, this might be a transition where the electron has to be deflected - where it is traveling in a different direction when it is through making the transition than when it started. If this is true, the electron must somehow give up momentum or be given momentum. Of course, at room temperature we have plenty of "strangers" in the lattice called "lattice vibrations" or "phonons," which are just quantized lattice vibrations, so that the electron has no trouble doing this. If we cool the diode down to say, 4° Kelvin, or somewhat less, we soon find that it is very difficult for electrons to do this. The number of electrons which are making tunneling transitions - from going in one direction in the valence band to going in another direction in the conduction band - is very small. If we apply a certain amount of voltage, there is a very small number of electrons which are allowed to do this - until we reach a certain critical voltage. This voltage corresponds to the energy required to create one of these little quantized lattice vibrations. At this point the electron has enough extra energy to enable it to excite one of these quantized lattice vibrations in order to change its direction of motion. Consequently, it is much easier for it to make this tunneling transition (and the current suddenly increases), as shown in Figure 2.

We can then go a little further and nothing much will happen until we reach the point where another one of these quantized lattice vibrations can be excited and the current again increases very rapidly. This study of tunneling and its relationship to quantized lattice vibrations has given us a better understanding of the tunneling phenomenon itself and also of the basic nature of the "strangers" within the semiconductors that we are talking about.
For the rest of this talk, I should like to concentrate on materials which are not semiconductors. I shall cite two experiments which I consider to be of some interest. The first of these experiments was done at the General Electric Company, and the second was performed here at the Institute.

Figure 3 shows the energy diagram for a structure which consists of a metal, a very thin insulating layer, and another metal. Can we obtain tunneling transitions from one metal to the other? Notice in this diagram that the number of electrons which can make tunneling transitions is only the number of electrons in the eV shown. The reason for this is that the electrons in the metal on the left, which are at energies below the Fermi level on the right, are all facing filled states and the exclusion principle does not allow them to make a transition. So it is only electrons above this energy which can tunnel into vacant states shown on the right. Since the number of electrons facing vacant states on the right is roughly proportional to the voltage applied, we would expect the volt ampere characteristic of this kind of device to be ohmic (obeys Ohm's law).

There is an exception to this. If we have a material which has no vacant states on the right-hand side - if by some mechanism we are not allowed to have electrons for a certain range of energies above the Fermi level - we will then have to apply higher voltage to get out of the region where there are no allowed states. This situation exists if, on the right-hand side, we have a superconductor. It is known nowadays, from some recent work done by Bardeen, Cooper and Shreiber, that a superconductor should have a very small energy gap or a very small increment of energy in which there are no allowed electron states just above the Fermi level. Figure 4 shows what happens in such a device. At room temperature or so, it has an ohmic characteristic as expected. However, if we make a superconductor out of it, you see that for a small number of millivolts - corresponding to the width of this forbidden gap on the right-hand side - we are not allowed to have electrons traveling over to the right because there are no vacant states into which these electrons can tunnel; therefore, there are no electrons making transitions and there is no current. After the applied voltage exceeds the forbidden gap, the current increases very rapidly because the electrons then have allowed states into which they can tunnel. In this way, it is possible to observe directly the energy gap in a superconductor. There have been some extensions of this experiment using two superconductors, and it turns out that a negative resistance can be observed in this way.

Figure 5 shows the conductance of this same device as a function of voltage. This conductance is simply the differential of the voltage current characteristic. It gives essentially the density of the vacant states that are allowed for electrons in the right-hand metal. There are none for some time, then the density of states goes up very rapidly as if, in a superconductor, the states that normally exist all the way down to the Fermi level are somehow crowded out just above the forbidden gap. The solid curve was computed from the Bardeen, Cooper-Shreiber theory. It exhibits almost too close agreement with the experimental points.

It will be noticed in Figure 4 that the distance through the forbidden gap through which the electrons had to tunnel was essentially constant. This being true, the number of the electrons left over after getting through the forbidden gap was essentially constant and the nonlinearities in the voltage current characteristic came from the difference in the density of states. Some work has been done here at the Institute in which we have looked at the other extreme. We have made the insulator relatively thick so that a sizeable electric field has been applied to get the path through which the electrons have to tunnel down to where a substantial number of electrons are tunneling. The electrons make the transition, labeled 3 in Figure 1, from the vicinity of the Fermi level in the left-hand metal to the conduction band in the semiconductor, or insulator.
Figure 6 shows this kind of experiment in a little more detail. We have a metal on the left, a metal on the right, an insulator in between, and an electron wave schematically representing the kind of transition taking place. An electron travels to the right with a large amplitude, is exponentially damped in the forbidden region, then gets into the conduction band of the insulator gathering energy as you can see when the wave lengths get short, and finally entering the second metal.

Now, there are two questions we should like to ask about this experiment. What will the voltage-current characteristic look like? What happens to the electron when it gets over into the second metal? The first question is quite easy. The amplitude of the electron wave, (the fraction of the electrons that get through) is exponentially dependent upon the inverse of the distance that it has to go. The distance is just the work function, (the height above the Fermi level where the semiconductor condition band begins ) divided by the electric field. So one would expect the current, which is proportional to how many electrons make this transition, to go like $1/\text{electric field}$.

Figure 7 shows some experimental points with a fit to a theoretical curve. It turns out that the people who do a lot of theoretical work have difficulty getting within factors of $10^4$ of the kind of current actually observed, so a constant, one way or the other, does not seem to bother anybody. I took the liberty of adjusting this curve up and down to make it fit the data. As you see, there is a reasonable fit between the data and the experimental points from this kind of device.

The temperature dependence is shown in Figure 8. If one takes the voltage required for a given tunneled current as a function of a temperature, it is found that it is constant at low temperatures and then falls off at $1/T$ at higher temperatures.

Another question is, of course, what sort of energy distribution do these electrons which make this tunneling transition have? That is shown in Figure 9.

Above the Fermi level in the left-hand metal we rapidly run out of electrons to make this transition, so the current goes to zero (this was actually done for absolute zero temperature); at higher temperatures, there is a little "tail" up there. At lower energies the distance through which the electrons have to tunnel is larger, so the current rapidly decreases. Hence, we have a source of electrons making a tunnel transition which are reasonably monoenergetic and whose current density is controlled by the voltage that applies to the diode.

I would like to come back to the question I asked before, and that is: If you have a metal-insulator-metal structure, what happens to the electron after it has gone through the insulator and it is over in the second metal? It has a lot of energy, corresponding to the voltage applied to the diode. Eventually one knows that the electron wanders around in the metal and ends up in equilibrium down by the Fermi level somewhere. But how long does this take? How far does the electron have to go before it has collisions which decrease its energy? I have asked a number of people and received the same number of answers. Some said that it went a very short distance - just a few atomic diameters. Others said it went a large distance; and the rest said that it was somewhere in between. It turns out that all answers are right, as you can see in Figure 10.

Figure 10 is a plot of some work done by Harry Thomas in Germany a few years back with potassium. We have good evidence that the same thing happens in other metals. Plotted is the mean free path (not the conductivity mean free path, but the distance the electron has to go into this metal before it actually has a collision and loses energy) as a function of electron energy. This mean free path is a very sensitive function of just how fast the electron is going. The electrons with low energies have very long mean free paths. It can be seen that they are on the order of 1000 angstroms;
then as the energy is increased toward the plasma resonance energy of the metal, the electron mean free path decreases very suddenly because the electron is exciting plasma oscillations in the metal. Then we ask ourselves the question: What would happen if we made that second metal thin? - compared to a thousand angstroms - which is not hard to do. Would the electron go right on through and come out the other side? The answer is, yes it would. The energy band structure of an experiment like this is shown in Figure 12.

Shown to the left in Figure 11 is the metal-insulator-metal structure. In the middle labeled "Base" is a metal layer that we have made very thin compared with the mean free path of the metal at that energy. Between the Base and the Collector we have put another insulator or a vacuum. On the right is something to collect the electrons which come out through the metal film (Base). It is hard to do these things, and Figure 12 shows how we go about building them. We start with the metal (Emitter) and cover up the edges so that there is just a little spot in the middle where we want to do the experiment. We anodize a very thin insulator on top of the metal and evaporate on it a thin metal layer (Base). Then we either add another insulator, or if we do not want another insulator, we put it in a vacuum system with a metal plate facing the structure and apply an electric field to that. It turns out that if we actually build these devices, they work. They do not work well yet, of course, because we have not built enough of them and we do not yet know very much about them. However, one of the particular characteristics we have observed (and we think we are learning a little of solid state physics out of this) is shown in Figure 13.

In figure 13 is plotted the fraction of the electrons which actually tunnel between the first and second metal layers - the fraction that comes out into the vacuum or into the second insulator layer and is collected at the other end as a function of the total emitted current - the current through the first metal layer. Universally, in any of these devices we have tested, the fraction gets bigger as the current gets bigger. We think this is because there are a lot of interface states and many ways for the electron to get trapped and not be able to get through. We cannot yet make these devices out of nice single crystals because it is a very difficult technique. If we force more and more current through the device, the electrons have a better chance of overcoming these trapped states and of getting through to the other side.

By the use of this technique we hope to learn something about the properties of solids, and in particular, about the behavior of electrons of a few electron-volts energy in metals and insulators. Also one would hope that practical application for these principles could be found. To date, there has been a great deal of speculation but it would not be surprising if cold cathodes of very high current density and high frequency triode amplifiers using this principle could be developed. However, the materials and technology problems involved are very difficult and a large effort will undoubtedly be necessary before practical devices become feasible.