

Essential Elements of Solid State Physics

Frederick J. Ernst

February 21, 2005

Abstract

This lecture was devised while I was a teacher of physics at the Illinois Institute of Technology in Chicago.

1 Metals

One of the most fascinating theorems of quantum mechanics concerns the passage of electron waves through perfectly regular crystal lattices. It may be shown that for certain energies such waves pass unimpeded through the crystal, while for other energies transmission is impossible. There are, therefore, certain allowed bands of energies and certain gaps between the allowed bands.

As the temperature approaches absolute zero all the electrons drop down to the lowest possible energy states consistent with the Pauli exclusion principle. In metals the valence electrons are freely able to move about the crystal lattice, which characteristically has the hexagonal close-packed or face-centered cubic structure, in which each lattice site has 12 nearest neighbors. Near the highest occupied energy level are unoccupied levels which can readily accept conduction electrons. Thus, the so-called *Fermi level* lies within an unfilled energy band.

In metals only electrons near the Fermi level participate in conduction. Thus, the number of participating electrons does not vary much with temperature. On the other hand, as the temperature rises thermal vibrations of the lattice cause the latter to become less and less ideal, so the conductivity

of metals decreases with increasing temperature. In fact, the conductivity of metals is frequently found to vary inversely as the absolute temperature.

2 Semiconductors

In intrinsic semiconductors such as germanium and silicon electrons are much more tightly bound to pairs of atoms. Typically the atoms of a semiconductor have four valence electrons which form covalent bonds with four other atoms. These semiconductors have the same lattice structure as diamond, with a tetrahedral bond arrangement. Because the electrons are relatively tightly bound, albeit not as strongly as in an insulator, we conclude that a complete energy band must be filled. The filled band in a semiconductor is called the valence band. The next unoccupied band above the valence band is called the conduction band. Whereas in diamond the gap between the two bands is 6 e.v., in silicon and germanium the gaps are 1.10 e.v. and 0.70 e.v., respectively.

If one reconsiders the Fermi-Dirac distribution law for a situation in which one has a completely filled energy band, it can be shown that the (unoccupied) Fermi level actually lies halfway between the valence and the conduction band.

Any mechanism which serves to raise electrons from the valence band to the conduction band enhances the conductivity of a semiconductor. As the temperature rises thermal collisions will accomplish this. It should be noted, in fact, that when an electron is raised to the conduction band, not only can that electron participate in conduction, but also the (positive) hole left behind can contribute to the current. In semiconductors one has in general two types of carriers, electrons and holes.

In an intrinsic semiconductor the number of carriers increases dramatically with temperature. As can be seen from the Fermi-Dirac distribution, the increase in number of carriers is exponential. Although at higher temperatures the crystal lattice becomes less ideal, the consequent decrease in conductivity by no means balances the increase in number of carriers. Hence in semiconductors the conductivity increases strikingly with temperature.

The number of carriers may be enhanced by the process called *doping*. If one desires a material in which the number of electrons far exceeds the number of holes (N-type) one adds minute amounts of a valence 5 impurity. The impurity atoms can masquerade as silicon or germanium atoms at occa-

sional lattice sites, but each such impurity atom willingly gives up its extra valence electron. The small amount of energy which is necessary is represented by the impurity (donor) level being placed a small distance below the conduction band.

If one desires a material in which the number of holes far exceeds the number of electrons (P-type) one adds minute amounts of a valence 3 impurity. When these impurity atoms masquerade as silicon or germanium atoms, they gladly accept an extra fourth electron, leaving a hole free to move about through the lattice. This situation is represented by the impurity (acceptor) level being placed a small distance above the valence band.

3 P-N Junctions

A single silicon or germanium crystal can be fabricated in such a way that it is P-type at one location and N-type at another location. Since it is nevertheless a single crystal the electrons and holes can both diffuse from one location to the other. This would not be possible if two distinct crystals were connected by a wire. Holes cannot leave the semiconducting crystal.

Holes tend to diffuse from the P-type material, where they are numerous, to the N-type material, where they are not so numerous. On the other hand, electrons tend to diffuse from the N-type material, where they are numerous, to the P-type material, where they are not so numerous. This diffusion is essential to the operation of any junction diode or transistor.

Of course, diffusion is a two-way process. Nothing prevents a hole from diffusing from N-type to P-type, for example. It is a matter of statistical balance. When equilibrium is established, as many holes diffuse from N-type to P-type and vice versa. In this equilibrium situation, however, there are extra holes in the N-type material adjacent to the P-type material, and there are extra electrons in the P-type material adjacent to the N-type material. The narrow junction layer is, in effect, a small capacitor. It is incidentally the capacity associated with the junction which limits the use of transistors and diodes at high frequencies.

Below we show the distribution of charge across a P-N junction.

FIGURE HERE

This charge distribution corresponds to a potential function of the form plotted below.

FIGURE HERE

One might get the impression that one could hook this crystal up to a wire with a bulb in it and use it to make the bulb light. After all the two ends of the crystal are at different potential! Such a possibility is, of course, ruled out by the laws of thermodynamics. The error lies in neglecting the contact potential between the crystal and the wire.

Nevertheless, we can ask what would be the effect of connecting a battery across our crystal. If one connects the positive side of the battery to the N-type material, and the negative side to the P-type material, then the barrier in the vicinity of the junction will be augmented. One would not expect any significant conductivity in this case.

On the other hand, if the positive side of the battery is connected to the P-type material, and the negative side is connected to the N-type material, then the barrier in the vicinity of the junction will be decreased. Holes, which formerly did not have sufficient energy to overcome the barrier and move from P to N, will now be able to move. The same will apply to certain electrons, which formerly could not make it from the N-type to the P-type material.

For any given temperature, the conductivity of a junction diode increases dramatically as the forward bias voltage is increased.

FIGURE

For silicon and germanium junction diodes the dramatic increase in conductivity typically occurs around 0.4-0.5 volt. On the other hand, with reverse bias only a small leakage current flows.

The magnitude of the leakage current depends very much upon the temperature. In terms of practical circuits this means that attention must be paid to the possible effects of this temperature dependence. In the case of high current diodes and power transistors, suitable heat sinks must be provided.

4 Transistors

A transistor is a single crystal of germanium or silicon in which there are three regions, in either a P-N-P or an N-P-N configuration. It is not usually symmetrical. In general there is say a large P-type region, within which is located a smaller N-type region, and within that is still another smaller P-type region. These three regions are commonly called the collector, base, and emitter, respectively.

In operation the collector-base junction is reversed biased, so no current flows. In order to “turn the transistor on” the emitter-base junction is forward biased. This causes a current to flow across the emitter-base junction. The base region is, however, rather small. About 95% of the carriers diffuse right across the base region and proceed into the collector region rather than leave by the wire connected to the base. This diffusion process provides a current in the collector in spite of the fact that the collector-base junction is reverse-biased. Once the extra carriers are in the collector region they proceed by normal conduction to the collector electrode.

A small emitter-base voltage controls a current which flows through a region across which is a substantially higher load voltage. Thus voltage amplification is obviously possible. Other connections of transistors permit one to build current or power amplifiers as well.

FIGURE