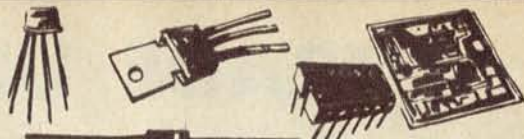


Fundamentals of SOLID STATE



Chapter 2

by Jamieson Rowe

Atoms in combination — energy interaction — crystalline solids and energy bands—the valence band—conductors and electrical conduction—insulators and semiconductors —the effect of excitation—electrons and holes—crystal conductivity and resistivity.

Having looked at the modern concept of atomic structure, and at what might be called the "internal" behaviour of individual atoms, let us now examine what happens when atoms link together to form molecules and "solid" matter. It should become apparent as we progress that knowledge of this "external" behaviour is essential for a clear understanding of electrical conduction.

We have seen that in an individual atom, the electrons surrounding the central nucleus can only occupy certain "allowed" orbits, each of which correspond to a particular value or level of total electron energy, and that in the unexcited or "ground" state the electrons of an atom are found occupying the orbits nearer the nucleus in numbers determined by the orbit capacities. We have also seen that in a practical situation involving light, heat and other forms of radiant energy, electrons are constantly transferred back and forth between allowed orbits as the atom absorbs and emits "lumps" or quanta of energy whose sizes correspond to the energy differences between the various levels.

Two individual and separate atoms of the same element will possess the same allowed orbit structure, or in other words the energy levels of their allowed orbits will be identical. Note that in saying this we make no reference to the electrons occupying the levels, but refer only to the allowed levels themselves. Hence it is not implied that at every instant of time each atom will have exactly the same excitation energy, with identical numbers of electrons at each level. In fact this would not be so even if their situations were equivalent, because the random nature of energy absorption and emission would produce instantaneous differences such that all we could say is that they had the same average excitation energy.

A most interesting thing happens if two such atoms are brought near to one another: the electric fields around the two nuclei interact in such a way that each of the allowed electron energy levels of both atoms progressively "splits" into a pair of levels (orbits), whose energy difference increases as the two atoms are brought closer together. At first sight, this may seem quite inexplicable; however a moment's

thought should show that it is no more so than many other similar effects with which the reader is likely to be familiar.

Recall, for example, that when two resonant circuits tuned separately to the same frequency are coupled together, they interact such that in the coupled state neither is resonant at the original frequency, but both are effectively resonant at two new adjacent frequencies whose separation depends upon the degree of coupling. It is this effect which produces the familiar "double humping" associated with large coupling factors.

Another example occurs in the case of loudspeakers fitted into tuned enclosures. Here a loudspeaker cone system and an enclosure, having the same resonant frequency when separated,

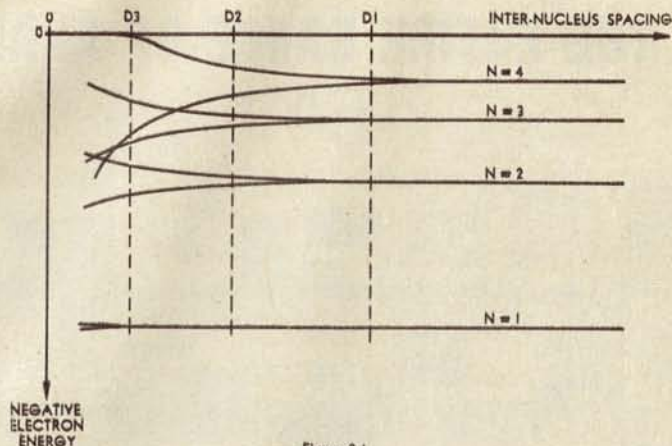


Figure 2.1

interact when together to produce the same sort of double resonance — which in this case is used to smooth the low-frequency response.

In fact, it is found that this sort of interaction effect is quite universal where oscillatory systems are concerned. Therefore it should not be surprising that it occurs between the allowed electron energy levels of "coupled" atoms, particularly as we have seen that each energy level corresponds to an orbit which represents a particular mode of "oscillation" associated with the wavelike aspect of electron behaviour.

As one might expect, it is the highest or least negative electron energy levels of two atoms which first split as they are brought nearer, because these correspond to the largest allowed orbits. For the same reason it will be the level pairs produced by these levels which will be found most widely separated for any given distance or spacing between the two atomic nuclei. This is illustrated in figure 2.1, which shows the splitting of the various energy levels as a function of the nucleus spacing.

Note particularly that this diagram applies equally to either atom, and that in the interests of clarity only the first four levels are shown. It may be seen that for large spacing, the levels are unaltered from their "individual atom" values, but as the spacing decreases they split progressively from the higher levels. At a spacing distance D1, for example, only the N=4 level has split, while at a smaller spacing D2 both the N=3 and N=2 levels have split also but by smaller amounts. At a still smaller spacing D3, the lower of the pair of levels corresponding to the N=4 level has moved below the higher of the N=3 pair. Such "overlapping"

occurs more and more as the spacing is reduced.

What does this mean? Simply that when two similar atoms are placed relatively near one another, their interaction effectively alters and increases the number of "allowed" orbits for the electrons surrounding each. Hence when the atoms, whose behaviour is represented by figure 2.1 are spaced at a distance D2 apart, each has two new allowed orbits in place of each of the orbits corresponding to its previous N=4, N=3 and N=2 energy levels. As splitting occurs progressively from the highest levels down, this will also mean

that all of the higher levels not shown will also have split into two, so that each atom will have very many more allowed orbits than it would have had alone. (In fact the number of allowed orbits will have almost doubled, as in this example only the $N=1$ level has remained unaltered at a spacing of $D2$.)

It so happens that, in the same way that the energy levels of two atoms split into pairs when they are brought together, the energy levels of larger numbers of relatively close atoms are found to split into a corresponding number of new levels. With three atoms, the levels each tend to split into triplets; with four atoms, into quadruplets, and so on.

Accordingly, if we have a lump of an element comprising a large number "M" of atoms regularly spaced at a particular distance, certain of the "individual" energy levels will be found to have split into the same large number of M new energy levels, forming bands. The number of levels which will have split into such bands, and the energy width of the bands, will depend upon the atomic spacing, with the higher levels splitting before the lower and to a greater extent.

An example may help in picturing this situation. A cube of metal measuring one centimetre on each side typically consists of something like 10^{23} atoms — one-hundred-thousand-million-million-million. This means that in place of certain of the higher energy levels of an individual isolated atom of the metal concerned, each of the atoms of the metal cube will have bands each containing no less than 10^{23} extremely closely spaced individual levels. A cube one-hundred-thousand times smaller in volume will similarly have 10^{23} levels in each of the atomic bands—still a very large number!

In both cases the number of bands present, and their "width" in terms of energy levels, will depend as before only upon the inter-atomic spacing. In fact the number of bands and their width is exactly the same as the number of "paired" levels and the separation widths produced for the simple case of only two atoms, illustrated in figure 2.1. Hence, although the size of a lump of material determines the number of discrete levels making up each of the energy bands, it does not affect either the number or the width of the bands.

The type of atomic interaction which we have been considering occurs almost only in the "solid" state of matter, as opposed to the "liquid" and "gaseous" states, because it is only in the solid state that the spacing between atoms is sufficiently small and relatively fixed. And as one might expect, the solid materials whose behaviour most closely conforms to this picture are those in which the atoms are arranged in very regular 3-dimensional "lattice" patterns—the crystalline solids.

The electron energy relationships inside a typical crystal structure are illustrated in figure 2.2, which is a two-dimensional energy/distance representation of the same type as that for a single atom given previously in figure 1.3.

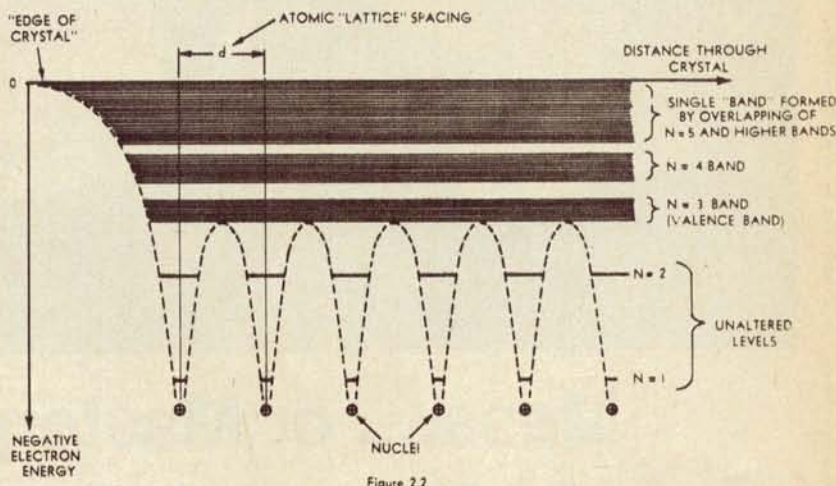
It may be noted that in this example the lattice spacing or distance between the atomic nuclei is such that the $N=1$ and $N=2$ energy levels have

remained unaltered, while the $N=3$ and higher levels have split into the expected bands each comprising M closely spaced new levels. In fact overlapping of the $N=5$ and higher bands has produced virtually a single "higher band," extending right up to the zero energy level. Such overlapping tends to occur with the higher levels in crystalline solids, both because the splitting is greater for these levels, and also because as we have seen previously the energy differences between the original atomic orbit levels decrease with increasing distance from the nucleus.

In this example the $N=3$ band is shown as the valence band, which corresponds to the valence electron energy

energy levels represented in figure 2.2 by the $N=1$ and $N=2$ levels.

On the other hand, electrons having higher or less negative energy can occupy any of the many levels comprising the valence and higher bands, in which they are no longer the "property" of individual atoms but belong only to the crystal as a whole. Those whose energy places them within the valence band are thus "shared" equally by all the atoms of the crystal, and it is in fact these electrons which effectively bind the crystal together. Any electrons in the higher bands are even less restrained than these, having at the same time less negative potential energy and more kinetic energy, and



level of the individual atoms concerned. Although shown here as an isolated band, not overlapped by higher bands, the valence band is not necessarily so isolated, and is in fact overlapped in certain crystals.

As a result of the interactions between the atoms of the crystal lattice, only the walls of the electron energy wells (dashed lines) surrounding the nuclei at the edge of the crystal rise fully to the zero energy level, as they do with an isolated atom. For the nuclei inside the crystal, the well "walls" fuse and cancel at a somewhat lower level, as shown. The level at which they fuse is in fact very close to the valence band, and this has considerable importance.

It may be noted that below the fusion level, the original electron energy levels are unaltered, and that they are shown as before separately for each nucleus. Conversely above the fusion level, all levels have become bands, and are shown extending continuously throughout the lattice. The significance of these distinctions is that electrons occupying energy levels below the fusion level are influenced almost solely by the individual atomic nuclei, whereas electrons occupying the energy bands above the fusion level are virtually uninfluenced by single individual nuclei, and are effectively "common property."

In other words, this means firstly that electrons having low or more negative energy can exist in the crystal lattice only in orbits closely surrounding the individual nuclei. These are the highly bound "core electrons," and they will be those occupying orbits corresponding to the unaltered

these can accordingly move with increasing freedom anywhere inside the crystal.

It is those electrons in the "common property" valence and higher energy bands of a crystalline solid which are responsible for its electrical behaviour, and the part played in this regard by such electrons is very largely determined both by the relative positions of these bands, and by the distribution of electrons in them. Hence in order to gain an insight into electrical conduction in a crystal, we must look closely at both the bands themselves and the ways in which electrons can occupy them.

There are a number of different ways in which atoms can link or "bind" together to form crystal structures. Depending upon the type of atomic bond involved, and the size of the atoms, a particular crystal lattice will have a definite inter-atomic spacing, and thus an appropriate number of the atomic electron energy levels will be split into bands of appropriate width. The disposition of electrons in the allowed levels and bands will depend, as before, upon both their disposition in the ground state of an individual atom, and on the excitation energy of the atom concerned.

From this it may be appreciated that each crystal structure formed by the various elements will tend to have a different and unique overall energy pattern, with different energy band widths and spacing, and each different with respect to the number of electrons occupying the various levels and bands at a given temperature.

Despite this, it happens that most crystalline solids fall into only two

broad categories when one considers the electron energy situation associated with the valence and higher energy bands. One of these situations applies in the case of metal crystals which are excellent electrical conductors; the other applies in the case of crystalline solids which are basically either semiconductors or insulators.

The first type of situation is basically that in which the valence electrons of the various atoms of the crystal do not completely fill the valence band in the ground state, as illustrated in figure 2.3.

This situation can arise if the electrons of an individual atom of the ele-

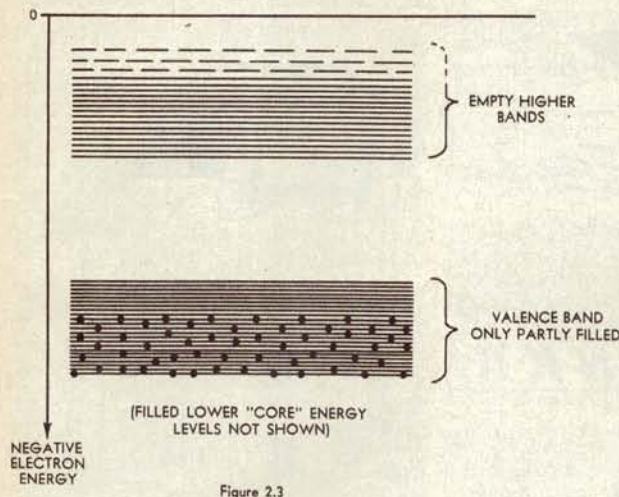


Figure 2.3

ment concerned do not fill the original valence level; it can equally be caused by a crystal lattice spacing which results in overlapping of the "true" valence band by a higher order band or bands, to produce a much wider effective valence band. For our purposes it does not matter which factor is responsible, the essential point being that the valence band is not completely filled.

In order to understand how this situation allows the crystal concerned to act as a good electrical conductor, consider for a moment what happens when an external source of EMF is connected across the crystal. Due to the applied EMF, an electric field is set up through the crystal; as a result one end of the lattice has an effective potential energy with respect to the other, so that the various electron energy levels and bands no longer remain horizontal but have a slope which corresponds to the electric field gradient. This is illustrated in figure 2.4, which shows the same valence and higher energy bands which were shown in equilibrium in figure 2.3.

Electrons are always in motion, and those in the valence band of a crystal are continually "sharing themselves around" among all the atoms of the lattice. The effect of the applied electric field, as one might expect, is to produce a tendency for the electrons to be accelerated in the "downhill" direction of the field, and slowed down or decelerated in the "uphill" direction.

Now deceleration of electrons by the field is in fact difficult, because this

implies loss in kinetic energy, and falling of the electrons concerned to lower levels; yet the lower levels are filled. However, the converse process of electron acceleration is quite easy, because this involves the transfer of electrons to higher energy levels, and such levels are in this case readily available in the form of the remaining empty upper levels of the partly filled valence band.

Acceleration of electrons thus occurs readily under the influence of the field, and there is the "nett flow of charge from one end of the crystal to the other" which we define as an electric current. In moving through the crystal the electrons exchange negative potential energy for kinetic energy, jumping

energy level which is completely filled with electrons, and in this case all the levels of the valence band are so filled.

The reason why a nett electron flow cannot occur in a completely filled energy level is that, for a nett flow to occur, there must be set up either an electron density or an electron velocity unbalance between one "end" of the level and the other. In a completely filled level a density unbalance is fairly obviously impossible; but so too is a velocity unbalance, because by definition all electrons in a given level have the same kinetic energy.

It may help in understanding this point if one imagines a filled level as something like a highway capable of carrying only a single lane of cars in each direction, and on which all the

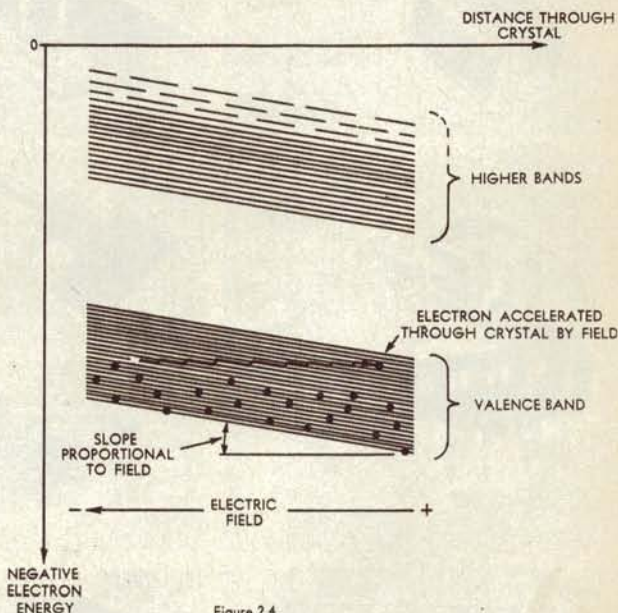


Figure 2.4

from level to level and effectively moving along the crystal energy diagram along paths such as that shown in figure 2.4.

A solid material can thus be defined as an electrical **conductor** if its energy band situation in the vicinity of the valence band corresponds to that shown in figure 2.3. In other words, it is one in which the valence band is only partly filled with electrons. This is the situation which applies in the case of metallic conductors such as copper, gold, silver and aluminium.

The second type of energy band situation which can occur in the vicinity of the valence band of crystals in the ground state is that illustrated in figure 2.5. It may be seen that the only essential difference between this situation and that for a good conductor shown in figure 2.3 is that the valence band is here **completely filled**. The only energy levels of the crystal unoccupied in the ground state are thus those in the higher bands, separated from those of the valence band by a relatively wide "forbidden energy gap."

It may seem surprising, but a crystalline solid in which this energy band situation occurs is completely unable to conduct electricity when in the ground state. This is because a nett electron flow from one region of the crystal to another is impossible in any

cars must travel at a fixed speed (corresponding to the particular energy level). If the highway is "filled" with both lanes carrying cars moving "bumper to bumper," there is no way in which more cars can travel in one direction than in the other; in other words, there can be no "nett car flow" in either direction.

The only ways in which a nett flow could occur would be either if the lanes of the highway were not filled, so that more cars could conceivably travel in one direction than in the other (a density unbalance), or if cars could travel at different speeds (a velocity unbalance), the latter implying the availability of additional "energy level" lanes.

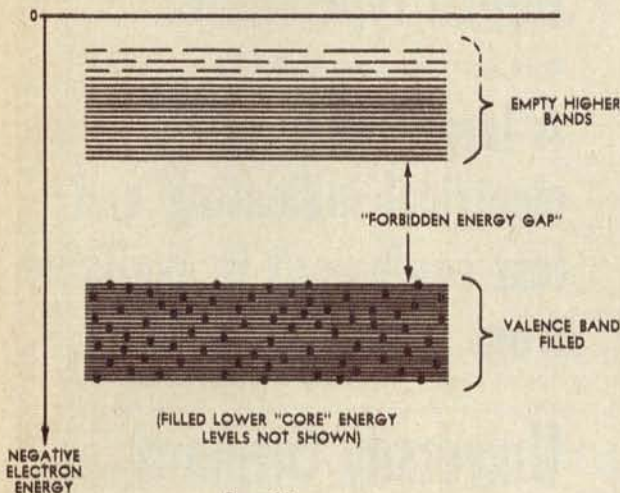
From the foregoing it may be seen that if the valence band of a crystalline solid is completely filled, the crystal concerned will be an electrical **insulator**. ALL crystals whose energy band situation in the vicinity of the valence band corresponds to that shown in figure 2.5 in the ground state are thus strictly insulators in that (hypothetical) state.

Into this category fall both those materials normally known as "insulators" and those which have relatively recently become known as "semiconductors," as noted earlier. In fact, there is **no essential difference** between these

two groups of materials, only a difference in the degree to which their behaviour changes with excitation level. To clarify this point, let us now look at the effect of excitation on the basic situation shown in figure 2.5.

We have seen previously that an individual atom would only be in its ground state if it could be maintained at a temperature of absolute zero (-273 deg. C), shielded against all forms of radiant energy such as heat and light; whereas in actual fact, an atom in a practical environment is taking part in a continual process of energy absorption and emission, involving the transfer of electrons between its various allowed energy levels. As one might expect, the same argument applies to a crystal lattice composed of a large number of such atoms.

A crystalline solid in a practical environment involving heat, light and other radiant energy is therefore



similarly involved in a continuous process of absorption and emission, with electrons now transferring both between levels within the crystalline energy bands, and also between the bands. The latter naturally involves absorption or emission of larger energy quanta than the former, as it involves transfer across the relatively large forbidden energy gaps between bands.

Under such conditions the "insulator" energy band situation shown in figure 2.5 will change. Absorption and emission of energy by the crystal lattice will reach a dynamic balance or equilibrium at an excitation level above the ground state, in which a small proportion of ever-changing electrons from the valence band have been transferred to higher energy bands. This is illustrated in figure 2.6.

The extent to which this will occur depends both upon the energy level of the environment in which the crystal finds itself, and also upon the width of the forbidden energy gap between the valence and next higher energy band. Naturally enough, the higher the temperature of the heat energy present in the crystal, the "bluer" the light incident on its surface, and so on, the greater will be the tendency of valence band electrons to acquire the energy necessary for them to be transferred to higher bands; but this granted, the proportion which do actually transfer

will depend upon the energy width of the forbidden gap.

The width of the forbidden energy gap varies widely among the crystalline solids whose valence band situation is represented by figures 2.5 and 2.6. Accordingly, such materials also vary widely in the degree to which electrons are transferred from the valence to higher bands under the influence of excitation. And as we shall see shortly, this behaviour determines directly their electrical characteristics.

In a crystal of diamond, the binding between the constituent carbon atoms is such that the forbidden energy gap is very wide. It amounts to some 6 electron-volts (eV), where an electron-volt is a convenient unit of energy used in atomic physics and other fields;

tion shown in figure 2.5, and both types of material behave as shown in figure 2.6 with excitation. The only difference is one of degree.

Hence by raising the temperature of an "insulator" crystal, for example, one could obtain a semiconductor, while conversely by cooling a "semiconductor" one produces an insulator.

From our earlier look at conduction in metallic crystals, the reader may by now have deduced that a semiconductor crystal in the excited state shown in figure 2.6 will become quite a good conductor, by virtue of the electrons which have transferred from the originally full valence band into the originally empty higher bands. And this is quite so, although it is only half the story.

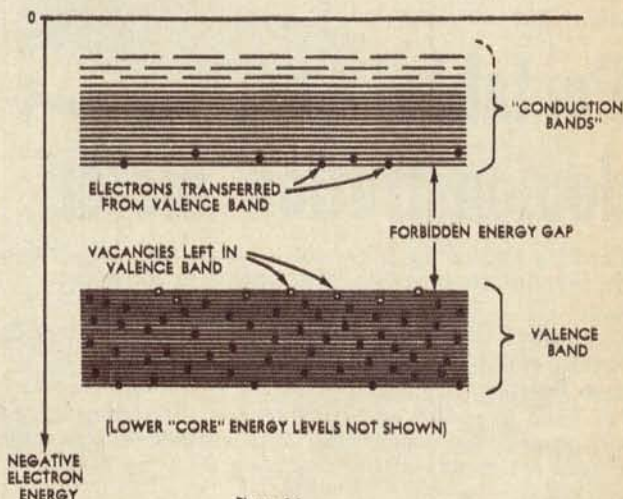


Figure 2.6

The electrons which have transferred into the higher bands, because these bands are largely empty, are certainly capable of forming a net carrier flow through the crystal under the influence of an applied electric field. In fact because of this, the higher bands are usually called the conduction bands, as shown in figure 2.6. However, as it happens, the "vacancies" which are left by transferred electrons in the valence band are also able to contribute to conduction.

In order to understand this, consider that when an electron is transferred from the valence band to a conduction band, this is actually equivalent to the weakening of a valence electron bond between two adjacent nuclei of the crystal lattice. Instead of the usual two-electron "covalent" bond which each nucleus shares with each of its four adjacent nuclei, there is left in the place concerned a weakened bond having only a single electron. This is illustrated in the two-dimensional picture of figure 2.7, where the weakened bond is shown consisting of the single remaining electron together with a hole or vacancy in place of the missing electron.

Because of the missing valence electron, the electrical charge balance of the crystal lattice is upset in the vicinity of the weakened bond. The positive charges of the relatively fixed atomic nuclei are no longer exactly balanced by the negative charges of the surrounding electron population, so that a localised net positive charge is produced.

one electron volt is the potential energy acquired by an electron when it is moved through an electric field for a distance corresponding to an increase of one volt.

In comparison, the forbidden energy gap of pure germanium crystal is only 0.72eV, while that of pure silicon is only a little larger at 1.11eV. It may be seen from this that in such materials the proportion of electrons which have transferred from the valence band to higher bands will be much greater, at a given degree of excitation, than for a material such as diamond.

Because under "normal" practical conditions crystals of materials like germanium and silicon have significant numbers of electrons which have transferred from the valence band to higher bands, whereas crystals of materials like diamond have not, and because this results in significant differences in the normal-conditions electrical behaviour of the two groups of materials, it has become convenient to distinguish between them. Crystals of germanium and silicon are thus known as semiconductors, while those of materials such as diamond are known as insulators.

It should perhaps be stressed again that there is no distinct division between the two groups of materials; as we have seen, in the ground state both have the "insulator" energy band situa-

In fact this positive charge is localised right in the "hole" originally occupied by the missing electron, and it has a value of charge equal and opposite to the negative charge of an electron. Neither of these facts are really surprising in view of the way in which the charge is produced.

The interesting thing is that such a hole is capable of moving through the crystal lattice, and as a moving positive charge it can thus effectively make a contribution to a current flow which is almost equal (but opposite) to that of an electron.

A hole tends to move through the crystal lattice because electrons in neighbouring valence bonds are attracted by its positive charge; when such an adjacent electron jumps across to "fill" the hole, it in turn leaves a hole in its own original bond to be filled by another electron, and so on. This "leapfrog" effect results in the effective movement of the hole through the lattice. Under the influence of an applied electric field, the hole movement will tend to take place in the direction opposite to that taken by a conduction band electron.

It may perhaps seem from this description that the concept of a hole is a redundant one, for the reason that "hole movement" in a particular direction through a crystal might seem to be "really nothing more" than a series of small jumps by electrons in the opposite direction; but this is not so. The fact is that the localised positive charge present in a crystal lattice at a weakened valence bond is no more and no less a reality than the "localised negative charge" which we are pleased to call an electron. It even has an effective mass, which is approximately equal to that of an electron.

To use an analogy, a hole in a crystal lattice valence band is rather like an air bubble in a test-tube almost filled with water. Both might be interpreted merely as "vacancies" whose effective movement takes place purely by means of movement in the opposite direction of something which superficially seems more "real" — like electrons, or water. Yet like the air bubble, a hole makes its existence apparent by means of its behaviour, in this case its electrical behaviour.

In a semiconductor crystal of the type whose valence band situation is shown in figures 2.5 and 2.6, then, for every electron which is transferred to the conduction bands and accordingly becomes available as a "negative current carrier," there is also produced a hole which remains in the valence band but is equally available as a "positive current carrier."

Because of this, it is usual to say that excitation of a semiconductor crystal lattice results in the production of electron-hole carrier pairs. Similarly the emission of energy by the lattice is visualised as a process whereby a wandering electron in the conduction band "accidentally" meets a hole wandering in the valence band, the two permanently cancelling or "annihilating" one another and producing a photon of appropriate energy. The latter process is usually termed recombination.

A pure or intrinsic semiconductor material such as we have been considering thus contains, in the excited state, equal numbers of conduction

band electrons and valence band holes available for electrical conduction. However, the two types of carrier do not contribute to current flow in an exactly equal manner, because holes are in the valence band and cannot move through the material at the same rate as conduction band electrons. In effect, whereas the electrons in the conduction band can move speedily through the lattice without having to conform to any orbit requirements, the holes in the valence band must "weave" their way through the crystal binding orbit system, and therefore travel at a lower rate.

This means that while the numbers of free electrons and holes present in an excited intrinsic semiconductor at any one time are equal, any nett cur-

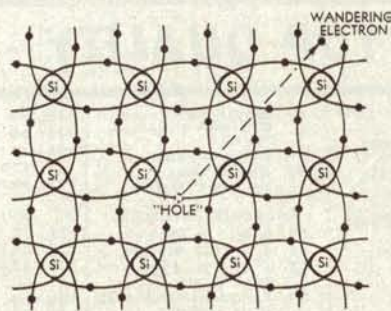


Figure 2.7

rent flowing through the material is carried more by the faster-moving conduction band electrons moving from negative to positive than by the slower-moving holes moving from positive to negative.

To use the analogy of a highway introduced earlier, but in a slightly different sense, the situation is now like a two-lane highway in which both lanes are packed with cars travelling in opposite directions, bumper to bumper but in this case at different speeds (corresponding to the two different energy bands). Although any given length of highway will contain equal numbers of cars in the two lanes, there will still be a greater car "flow" in the faster lane than in the slower lane.

Because the generation of electron-hole carrier pairs depends upon the excitation level of the crystal lattice, the number of such carriers available for conduction varies directly with the excitation level. Hence the conductivity of an intrinsic semiconductor crystal similarly varies directly with excitation. In the ground state, as we have seen, it will be zero; in more practical circumstances it will rise to a value which will depend directly upon both the temperature and the frequency/

intensity characteristics of any incident at its surface.

At this point it is perhaps worthwhile to pause briefly and note the contrast between the current picture of semiconductor-insulator conduction, which we have been examining, and earlier ones which held that these materials were merely those wherein the valence electrons were "harder for the electric field to pull free." It may be seen that the latter idea was quite wrong, because in fact such materials cannot conduct at all under the influence of an electric field alone; they become capable of conduction only when excited. Neither this fact nor the

TABLE 2.1

Material	Resistivity, Ohm-cM
Copper	1×10^{-6}
Bismuth	1×10^{-4}
Germanium	47
Silicon	200,000
Mica	1×10^{10}
Diamond*	3×10^{10}

*Theoretical resistivity. In fact unmeasurable.

existence of holes as additional current carriers in these materials could be explained by the earlier theories.

In talking about the electrical behaviour of a semiconductor at a particular excitation level, reference is often made to the resistivity, which is simply the reciprocal of the conductivity. Resistivity is usually defined as the resistance in ohms between opposite faces of a cube of material measuring one centimetre on each side: this gives units of ohms/cm² or ohm-cM.

As the conductivity of an intrinsic semiconductor rises from zero with excitation, this means that the resistivity effectively falls from a value of infinity. Table 2.1 gives the approximate resistivity figures for pure silicon and germanium under "normal" conditions, and also gives the equivalent figures for typical metallic conductors and insulators.

The fact that the resistivity of intrinsic semiconductors falls sharply with excitation is exploited by using them in thermistors, or temperature-dependent resistors which have a negative coefficient. This is in fact the main use of intrinsic semiconductors as such, their resistivity being rather too high and too temperature-dependent for direct use in most other semiconductor devices.

SUGGESTED FURTHER READING

- BURFORD, W. B., and VERNER, H. G., *Semiconductor Junctions and Devices*, 1965. McGraw-Hill Book Company, New York.
- MORANT, M. J., *Introduction to Semiconductor Devices*, 1964. George G. Harrap and Company, London.
- SCROGGIE, M. G., *Fundamentals of Semiconductors*, 1960. Gernsback Library Inc., New York.
- SHIVE, J. N., *Physics of Solid State Electronics*, 1966. Charles E. Merrill Books Inc., Columbus, Ohio.
- SMITH, R. A., *Semiconductors*, 1959. Cambridge University Press.