are included. The third lists important and useful physical constants and semiconductor material properties and the last gives silicon processing data necessary for working out problems related to diffusions or implantations.

Each chapter is followed by a list of questions to test the reader's grasp of the various ideas outlined in the chapter, and then by a list of problems to train the reader in applying the mathematical approaches and equations. Some of the problems actually supplement and broaden the information previously given.

To cover all the material a two-semester course is necessary in which the chapters composing parts 1 and 2 are covered in the first one and may be looked upon as a basic course, while parts 3 and 4 may be considered a more advanced course and covered in the second one. As already mentioned, parts 3 and 4 are independent of one another and the teacher may elect to cover only selected chapters from them.

A teacher's manual containing the solutions to the problems can be obtained from the publisher.

ACKNOWLEDGEMENTS

The author would like to thank his colleagues in the Technion with whom he worked when developing this course, and his many students, without whose questions many obscure points would have been left in the dark.

Special thanks are due to Dr D. Lubsens, Dr Shacham-Diamand and Mr Kornfield of the Microelectronic Laboratory at the Technion, IIT, who helped in obtaining some of the photographs.

I would also like to thank Motorola Inc. U.S.A., N.V. Philips Gloeilampenfabrieken, The Netherlands and National Semiconductors (Israel), who permitted the use of their published data sheets, characteristics and parameters for various devices used in this book to illustrate the connection between theory and practice.

A. B-L.

(1) Semiconductors and Electronic Devices, Adir Bar-lev, 2nd Edition, Prentice Hall, 1984

(2) Introduction to Solid State Physics, Charles Kittel, 6th Ed, John Wiley & Sons, 1986

* This package contains both sources.

SEMICONDUCTORS AND THEIR PREPARATION FOR ENGINEERING USE 1

1.1 Semiconductors

Semiconductor materials are distinguished by having their specific electrical conductivity somewhere between that of good conductors (10^6 $(\Omega\text{cm})^{-1}$) and that of good insulators ($10^{-5}(\Omega\text{cm})^{-1}$); hence the name. Among those materials, by far the most important in engineering use is silicon (Si). Of quite lesser importance is germanium (Ge), which, like silicon, is an element belonging to Group IV of the periodic table (Table 2.2). Becoming more important daily are the compound semiconductors, usually compounded of two elements (but sometimes more) of Groups III and V or II and VI of the periodic table. From those gallium arsenide (GaAs) is the most important. Also in use for specific purposes are indium antimonide (InSb), gallium phosphide (GaP), cadmium sulphide (CdS), lead–tin–telluride (PbSnTe) and others.

We shall mainly concentrate on Si, Ge and GaAs, from which the majority of present-day devices are being made, but the theoretical results apply to all of them.

Electronic devices necessitate use of almost absolutely pure semiconductor materials in which an exactly measured amount, usually extremely small, of a foreign dopant has been included to control its electrical properties. Also, the semiconductor must normally be in the form of a single crystal throughout the device, since, as we shall see, its desired electrical properties depend on the ordered, periodic nature of the crystal structure and any faults would be detrimental.

Let us give a short review of present-day engineering solutions to the purity and single crystal requirements.

1.2 Purification

Silicon, one of the most abundant elements on earth, is always found in a compound form in nature, usually combined with oxygen (sand is mainly SiO_2). It is first purified as far as possible by chemical methods. Reduction with carbon, according to $SiO_2+2C \rightarrow Si+2CO$, yields metallurgical grade silicon of up to 99% purity. By combining it with HC1, it is converted to liquid $SiHC1_3$. This is further purified using fractionation processes similar to those employed in the petroleum industry, and then reduced in hydrogen according to $SiHC1_3 + H_2 \rightarrow Si+3HC1$ and vapor deposited on thin silicon rods used as hot substrates. The deposition forms thick, semiconductor grade, polysilicon rods in which the concentration of troublesome impurities is about 1 in 10^9 silicon atoms. In single crystal form such silicon would have a resistivity of about 200 Ωcm which is sufficient for most applications. For special uses very high purity (and consequently resistivity) Si is needed. This is obtained by a method called zone refining, also used for germanium. It is based on the tendency of most impurities to remain in the liquid part when the melted semiconductor gradually solidifies. The ratio of the impurity concentration on the solid side of the liquid–solid interface C_s to that on the liquid side C_l is called the *segregation* (or distribution) coefficient K of that specific impurity and is usually much smaller than one.

Figure 1.1(a) describes such a purification system for Si using a floating zone; a solid Si bar is held vertically inside a fused silica (also called quartz) tube without touching it. (This is important, since melted Si, at 1420°C, is extremely active chemically and combines with or sticks to everything it touches.) Surrounding the tube there is a short copper coil in which a high-frequency current (about 0.5 MHz) passes, generated by an induction heating generator. Strong eddy currents are induced in the section of the Si bar inside the coil and this section melts. If this section is short, the strong surface tension of molten Si, combined with its low density, is sufficient to support the molten zone in its place. The quartz tube, being an insulator, is not affected.

The coil is slowly moved vertically relative to the Si, with the region immediately in front of it melting and that behind it solidifying. Because of the small segregation coefficient, most impurities stay in the melt and are therefore 'swept' along the bar towards one end. This may be repeated several times, the relatively dirty edge sawn off, the remainder recast into a new bar and the whole process repeated. A very high degree of purity results, with the remaining undesired impurity concentrations ten orders of magnitude or more below that of the Si. The zone refining is usually done in a hydrogen atmosphere to reduce the oxygen content. Germanium can be zone refined in horizontal graphite boats as it is much less active at its melting temperature of 937°C.

It can easily be shown (Problem 1.4) that if the original impurity concentration C_0 in the semiconductor is uniform, then, after a single molten zone

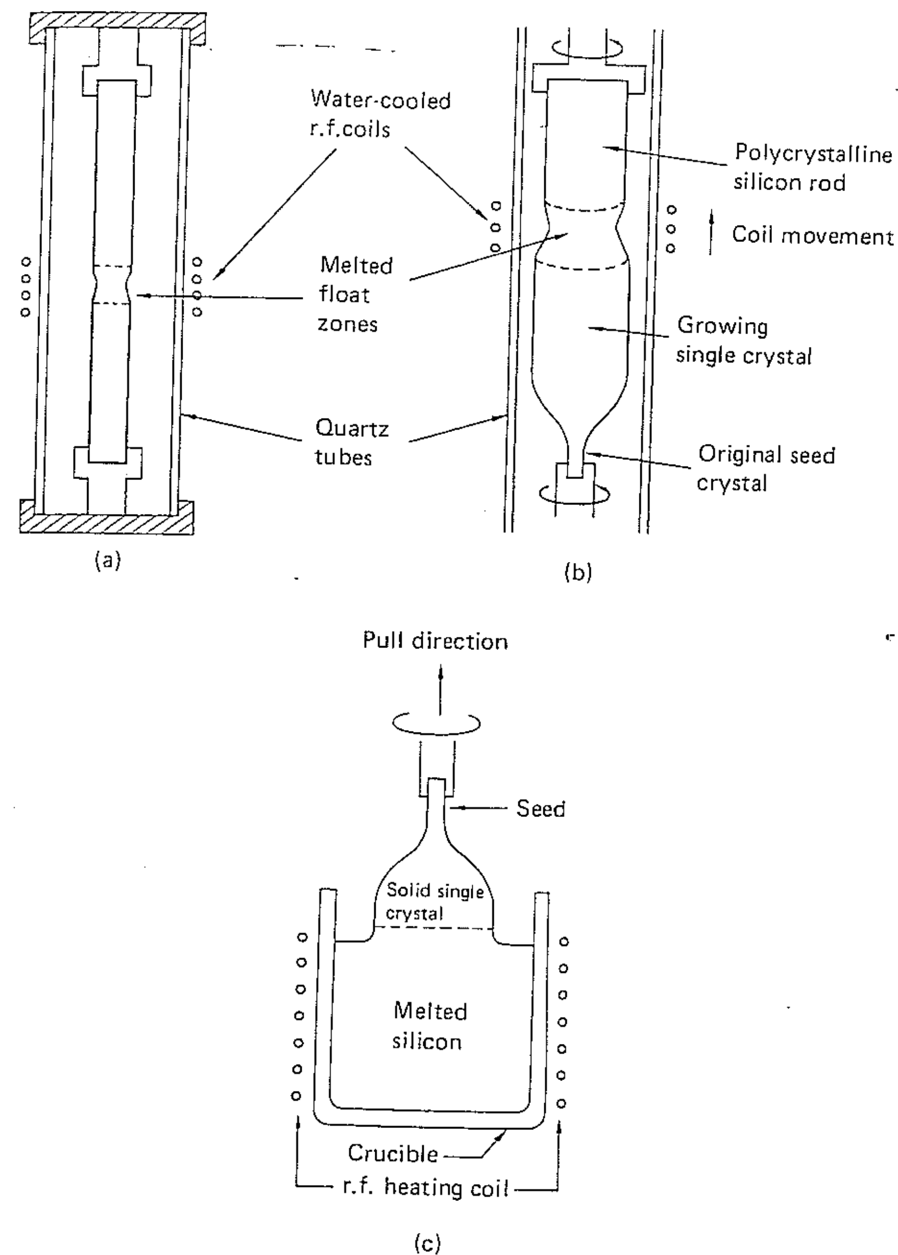


Figure 1.1. (a) Float zone (FZ) purification of Si; (b) single crystal growth by FZ technique; (c) the Czochralski single crystal pulling technique.

pass starting at $x=0$, one gets a new impurity concentration profile of

$$C_s(x)=C_0[1-(1-K)e^{-Kx/L}], \tag{1.1}$$

where L is the length of the molten zone.

This equation is the initial impurity distribution for the second pass. For copper ($K=4\times10^{-4}$) or iron ($K=8\times10^{-6}$), e.g., only a few passes suffice to appreciably reduce the impurity content.

1.3 Single-crystal Formation

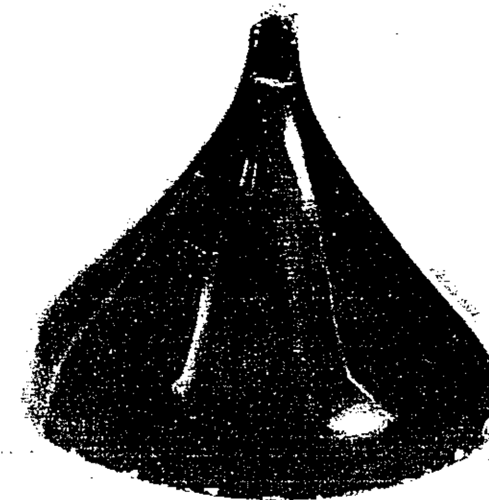
Conversion of the polysilicon rods to single crystal may be done by a similar float zone technique shown in Fig. 1.1(b). The polyrod is mounted vertically over a piece of single crystal, Si, called *the seed*, that is pre-cut in the desired crystallographic orientation. RF heating is used to melt the top of the seed and the bottom of the polyrod and form a molten float zone. If the RF coil is now moved very slowly upwards, the bottom of the molten zone would start to solidify on top of the seed, continuing its single crystal structure and orientation. As the molten zone traverses the polyrod, it transforms the rod's polycrystalline structure into a single crystal.

Both rod and seed are slowly rotated during growth to preserve uniformity of temperature and composition. Crystals of 10 cm diameter and 50 cm long are grown routinely today.

The most common method for silicon crystal growing in use today is the Czochralski pulling technique. The purified Si is remelted in a quartz-lined graphite crucible, shown schematically in Fig. 1.1(c). The *seed*, attached to a holder, is dipped into the molten Si and then very slowly pulled up again, turning at the same time to preserve uniformity. Molten Si sticking to the seed will start to solidify, if the temperature is properly controlled, and its crystal structure will follow that of the seed. The growth is performed with the growing crystal containing an exactly known amount of some specific impurity previously added to the melt, which, as we shall see, determines the electrical properties. Long crystals of 10 and 12.5 cm diameter can be grown.

Since Si in molten form is chemically very reactive, Czochralski (CZ) pulled crystals are not as pure as FZ grown ones and contain some undesired impurities, like carbon and oxygen, absorbed from the crucible and its lining, and sometimes even minute traces of heavy metals. As we shall learn, this affects the electrical properties of devices built in such a crystal.

Compound semiconductors are much more difficult to grow in single crystal bulk form. The difficulties stem from the usually very different vapor pressures of the compound constituents causing the more volatile one to evaporate away from the melt.



Top section of a pulled silicon single crystal still attached to the seed crystal.

A technique known as Liquid Encapsulated Czochralski (LEC) is often used today to grow such crystals: the melt of the compound constituents is covered on top by a liquid encapsulant, which floats on it, is not missible in it and has a low vapor pressure. B_2O_3 is often used for GaP growth. The encapsulant, coupled with inert gas overpressure in the crucible, prevent the volatile constituent from evaporating away during growth. Another method uses an hermetically sealed tube that can carry the pressure at the growth temperature, with different parts of it maintained at different temperatures. A second problem is that variations from stoichiometry (proper ratio of the constituents) and crystal defects may dominate the electrical properties. Since bulk semiconductor purification and crystal growth is done by special material suppliers and is not an in-house technology in the electronic device industry, we shall not deal with it further, however, integrated circuit and GaAs device technology often require an in-house growth of a thin single crystal layer on a matching single crystal substrate with different electrical properties. This is done by a method called *epitaxy* which will be described in Chapter 16.

Additional information on semiconductor material technology can be found in Reference 1.

QUESTIONS

1.1 Will there be an effect on the induction heating process of Fig. 1.1 if either hydrogen or argon gas flows in the silica tube instead of its being evacuated?

1.2 A specific impurity has a segregation constant in Ge which is smaller than 1 but larger than 0. Can Ge containing a lot of that impurity be purified just by repeating the zone refining process many times on the same Ge bar?

1.3 A sample of single-crystal Si is grown from the melt by the Czochralski pulling method. In order to obtain the desired electrical conductivity a certain amount of the impurity antimony (Sb) with $K \simeq 0.02$, is also melted in the crucible of the liquid Si before the growth is started. Will the content of Sb in the grown crystal be uniform along its length?

PROBLEM

1.4 Assuming an originally uniform impurity distribution C_0 along a Si bar, show that eq. (1.1) results after one zone pass. How will you proceed to obtain the impurity distribution after the second pass? After the next pass?

CRYSTAL STRUCTURE AND VALENCE MODEL OF A PURE AND DOPED SEMICONDUCTOR 2

2.1 A Simplified Picture of a Semiconductor Crystal

We shall now give a very simplified first picture of the situation inside a semiconductor, with and without impurities. Qualitative explanations will be used here and any quantitative results must therefore be considered as rough approximations only. In the next chapters we shall return to the various concepts appearing here for the first time, using a more quantitative (and, unfortunately, more difficult) approach.

Both Si and Ge atoms have 4 valence electrons, i.e. electrons belonging to the outer shell of the atom. Those are the electrons which participate in chemical bonding when the atoms form compounds. Upon solidifying from the liquid phase both Si and Ge crystallize in the diamond structure (the crystalline form of carbon), because for those elements this is the structure which minimizes the free energy. The basic unit cell of the diamond structure is shown in Fig. 2.1.

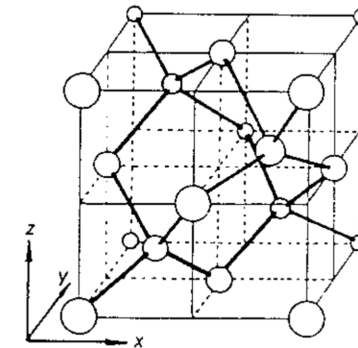


Figure 2.1. The diamond crystal structure. (Atoms are not drawn to scale; they should be envisaged as touching each other.)

Each atom is seen to have 4 nearest neighbors, set in a tetrahedral structure (GaAs and many other compound semiconductors crystallize in the zincblende structure, which is similar to that of diamond, with Ga and As atoms occupying alternating positions).

The basic unit cell length a and the distance between two neighboring atoms (which is equal to the atomic diameter D) are given in Table 2.1, with several additional properties of Si, Ge and GaAs.

Table 2.1. Some of the properties of the most important semiconductors (m_0 = Free electron rest mass)

Property	Si	Ge	GaAs
Atomic No.	14	32	—
Atomic weight	28.08	72.60	144.6
Density (kg m^{-3})	2.33×10^3	5.33×10^3	5.32×10^3
Melting point (°C)	1420	937	1237
Atoms per unit volume (m^{-3})	5×10^{28}	4.42×10^{28}	2.21×10^{28}
$\varepsilon/\varepsilon_0$	11.8	16	12.5
m_e^*/m_0	0.26	0.12	0.07
m_h^*/m_0	0.49	0.28	0.5
Unit cell length a (nm)	0.543	0.566	0.565
Atomic diameter (nm)	0.235	0.246	—

For additional information see Appendix 3.

Each of the 4 valence electrons of an atom in the diamond structure can be looked upon as shared between it and one of its four nearest neighbors, belonging to both of them. Each atom has then, in effect, 8 shared electrons in its outer shell in four pairs with opposite spins. This is a very stable low-energy structure. The bonding created by equal numbers of shared electrons is called *covalent* bonding. It does not involve electric charge transfer between different locations in the lattice. In a compound semiconductor like GaAs, however, the internal bonding in the crystal is also partly ionic, since the As atom has higher electronegativity than the Ga and the electron cloud shifts towards it. There is more negative charge then near the As while the vicinity of the Ga is positive. This charge transfer results in the semiconductor being polarized under electric fields. This is important when dealing with the processes that hinder charge carrier movements (scattering mechanisms) in such semiconductors.

Let us now simplify things and describe the semiconductor crystal, say Si, in a two-dimensional form as in Fig. 2.2(a). The little dots represent pairs of electrons with opposing spins, shared by the atom and its nearest neighbor. This is the so-called valence model, which is helpful in obtaining a simple, qualitative explanation of many properties of the semiconductor.

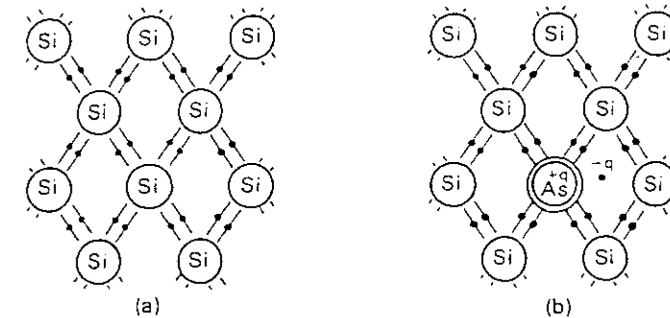


Figure 2.2. A two-dimensional picture of a Si crystal: (a) pure; (b) including some arsenic atoms (a donor impurity, see Section 2.3).

2.2 Free Charge Carriers

First let us consider a completely pure semiconductor and look for availability of charge carriers in it. At very low temperatures one can expect such material to behave like an insulator, since a shared valence electron is bound to its locality and there are no sources from which it can obtain the extra energy necessary to free itself from its bonds and make it available for current-carrying purposes. Electrons in inner shells nearer to the nucleus are even more tightly bound. The extra energy necessary to free a valence electron can be obtained from absorption of light photons arriving from the outside, if such light is available, or from the thermal vibrations of the crystal lattice atoms around their proper positions. These vibrations, called phonons, are a measure of the temperature and are very weak at low temperatures.

An idea of the available thermal energy can be obtained by comparing the vibrating crystal atoms to the molecular motion in an ideal gas inside a hypothetical container, as in Fig. 2.3.

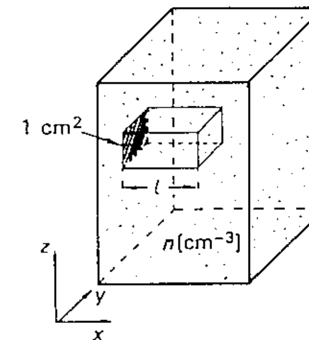


Figure 2.3. An ideal gas container for obtaining the thermal energy associated with molecular motion.

Let us designate:

n—number of molecules per unit volume (concentration)
v_{th}—average thermal velocity
l—distance travelled by a molecule in a unit time.

Thermal motion is equally probable in every direction. Therefore at any moment $\frac{1}{6}n$ molecules in a unit volume are moving in any of the six possible directions ($\pm x$; $\pm y$; $\pm z$). In a unit time, the number of molecules hitting a unit area of the container wall is equal to one-sixth of the number contained in a parallelepiped whose base is a unit area of the wall and whose height is $l = v_{th} \cdot 1 = v_{th}$ (m), as marked in Fig. 2.3. This number is therefore

$$\frac{1}{6} nl \cdot 1 = \frac{nv_{th}}{6} \text{ (molecules).} \tag{2.1}$$

But from the kinetic theory of gases we know that the pressure on the wall p (force per unit area) is given by

$$p = nkT, \tag{2.2}$$

where

k = Boltzman constant (1.38×10^{-23} J °K^{-1}), and
T = absolute temperature in °K.

Our $\frac{1}{6}nv_{th}$ molecules impinging on a unit area of the wall in each second exert on it a force equal to the pressure. This force is the total change in molecular momentum per unit time. Assuming a rigid wall and elastic collision, this change, for one molecule, is $2mv_{th}$ (the velocity changes direction). Therefore

$$p = \tfrac{1}{6}nv_{th} \cdot 2mv_{th} \text{ (N m}^{-2}\text{)}.$$

Combining with (2.2) and rearranging, we obtain

$$\frac{mv_{th}^2}{2} = \frac{3}{2} kT. \tag{2.3}$$

Hence kT is a measure of the thermal energy. At a normal room temperature of 300°K (27°C) it is equal to 0.026 (eV) (1 eV, or one electron-volt, is the energy given to an electron by accelerating it through 1 V potential difference; it is equal to 1.6×10^{-19} J). Substituting numerical values (from Appendix 3) into eq. (2.3), we get $v_{th} \simeq 10^5$ (m s^{-1}) at room temperature, which is approximately the measured value.

How does the thermal energy compare with the internal ionization energy, i.e., the energy necessary to free a valence electron from its bond to a specific locality, so it can become a free charge carrier and contribute to current if external fields are applied? As we shall see in subsequent chapters this energy

is about 1.1 eV for Si and 0.67 eV for Ge. At room temperature the average thermal energy, kT, is much lower, and therefore only very few electrons in Ge will manage to break free and even fewer in Si. The reason that few such electrons do exist is that the numbers mentioned are average values. There is a very small probability that an electron may accumulate a much higher thermal energy than the average. Although the probability is small, the number of valence electrons is high (about 10^{23} (cm^{-3})) and therefore the concentration of free carriers won't be completely negligible even at room temperature, and is expected to grow as the temperature increases.

When a valence electron is ionized, two charge carriers are actually created. The second one, called a *hole*, is the charge located in the vicinity vacated by the electron. That vicinity is left with a net positive charge of $+q$. Any one of the other valence electrons moving nearby can step into the vacated state with very little additional energy (less than the thermal energy) thereby shifting the net positive charge, i.e. the hole, to a new location. Both the free electron and the hole can therefore move around in the semiconductor crystal.

When one applies an electric field to the semiconductor, the random movement of the free electrons and holes is immediately affected. Before the field is applied this movement is the result of successive collisions with various scattering centers in the crystal lattice—for instance defects, ionized impurity atoms and thermal vibrations—which cause frequent changes in direction and velocity of the moving charge carriers. This random movement results in zero average current. With an external field the movement is very similar but will have an average component in the direction of the external field. This average velocity depends on the field and is called *drift velocity* and we shall return to it in the next chapter.

The moving carriers, though called 'free', are confined inside the crystal, subject to the potential fields of the atomic nuclei arranged periodically at the lattice sites. The electronic motion depends on both the externally applied field and these internal periodic fields. In Chapter 6 we shall see that the overall effect of the internal fields on the electronic movement can be taken into account by assigning the electron an *effective mass*, $m_e^\star$, different from its mass in vacuum m_0. The effective mass may be dependent on the direction of movement in respect to the crystallographic axes in some crystals (though in Si or Ge this averages out).

The hole too, has an effective mass $m_h^\star$, positive and different from that of the electron. Superficially, this may look surprising: it is easily understood that hole movement is accompanied by positive charge movement as explained. But where does the positive mass come from? In fact, this mass is a mathematical concept resulting from application of quantum-mechanical principles to the crystal structure, as will be shown in Chapter 6. However, some understanding of it may be obtained by remembering that hole movement

consists actually of valence electrons moving from one atom to the next. A force (an external field) is necessary to accelerate the hole, and if there is a force and resulting acceleration, directed in the correct way for a positive charge, their ratio can be looked upon as a positive mass. The values of the effective masses for electrons and holes in various semiconductors is given in Table 2.1.

2.3 Impurities in Semiconductors; Carrier Concentrations

It is clear that the number of free electrons, n, in a pure semiconductor is exactly equal to the number of holes, p, in it. Such pure semiconductors are called *intrinsic*, and an index 'i' is appended to n and p in this case:

$$n_i = p_i \tag{2.4}$$

The intrinsic carrier concentrations are very small and depend strongly on temperature. In order to fabricate devices such as diodes or transistors, it is necessary to increase the free electron or hole population by a large amount (henceforth the adjective 'free' applied to the electrons and holes will be omitted, although always implied, since only the free ones are of importance in device behavior). This is done by intentionally *doping* the semiconductor, i.e., adding specific impurities in controlled amounts. Such a doped semiconductor is called *extrinsic*.

The choice of the proper impurity and its effect can be understood by examination of Table 2.2, which is a section of the periodic table of elements, including some of those that belong to Groups III, IV and V (the Roman numeral being equal to the number of valence electrons). The atomic number, representing the total number of electrons for each element, is also included.

Table 2.2. A section of the Periodic Table of Elements

Group II	Group III	Group IV	Group V	Group VI
	5 B (boron)	6 C (carbon—diamond)	7 N (nitrogen)	
	13 Al (aluminum)	14 Si (silicon)	15 P (phosphorus)	16 S (sulfur)
30 Zn (zinc)	31 Ga (gallium)	32 Ge (germanium)	33 As (arsenic)	34 Se (selenium)
48 Cd (cadmium)	49 In (indium)	50 Sn (tin)	51 Sb (antimony)	52 Te (tellurium)

Thus, silicon and germanium have 4 valence electrons in their outer shell while boron and gallium have 3 and phosphorus, arsenic or antimony have 5.

If a Si (or Ge) atom in the crystal lattice is replaced by an impurity atom from Group V of about the same size, say arsenic, four of the valence electrons of the arsenic will take part in the covalent bonding with the neighboring Si atoms while the fifth one will be only weakly attached to the arsenic atom location. Because its orbit is relatively large, encompassing many Si atoms, this electron can be looked upon as moving in a silicon-filled space, with the relative dielectric constant of Si (about 12). This reduces the Coulomb attraction between it and its nucleus twelve-fold. Such an impurity atom is shown schematically in Fig. 2.2(b).

In order to get a quantitative estimate of the additional energy necessary to free this fifth electron altogether, to make it available as a free charge carrier, let us regard the arsenic nucleus, screened by all its electrons but that distant fifth one, as equivalent to a large hydrogen nucleus with a net positive charge of one proton $+q$ and with this single fifth electron moving around it in a space of a relative dielectric constant of 12. This is shown schematically in Fig. 2.4.

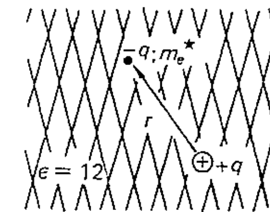


Figure 2.4. A hydrogen-like atom, embedded in a dielectric material with $\varepsilon = 12$ for calculating the ionization energy of the fifth electron (such as an As atom in Si).

Bohr's model of the atom can be used to find the ionization energy of our hydrogen-like atom of Fig. 2.4. By comparing the balanced centripetal and electric attraction forces, we obtain

$$\frac{q^2}{4\pi\varepsilon r^2} = \frac{m_e^* v^2}{r}. \tag{2.5}$$

According to De Broglie, the electron wavelength λ and its momentum p are related to its orbit radius r and Planck's constant h:

$$2\pi r = \lambda = \frac{h}{p} = \frac{h}{m_e^* v}. \tag{2.6}$$

On separating ν out, this gives

$$\nu=\frac{h}{2\pi m_e^\star r}. \tag{2.7}$$

Substituting ν into eq. (2.5) and separating out r gives

$$r=\frac{h^2\varepsilon}{q^2 m_e^\star \pi}=\frac{h^2\varepsilon_0}{q^2 m_0 \pi}\left(\frac{\varepsilon}{\varepsilon_0}\right)\left(\frac{m_0}{m_e^\star}\right)=r_H\left(\frac{\varepsilon}{\varepsilon_0}\right)\left(\frac{m_0}{m_e^\star}\right), \tag{2.8}$$

where r_H is the radius of the electron orbit in the hydrogen atom and ε_0 the permittivity of empty space. Using the numbers of Table 2.1 one finds that the orbital radius of the fifth electron in our hydrogen-like impurity atom is about 45 and 130 times that of hydrogen in the cases of Si and Ge respectively. Since the size of the impurity atom is only about 2.5 times that of hydrogen, the path is indeed large enough to include many of the host material atoms.

To find the ionization energy of the fifth electron, we remember that this energy E is the algebraic sum of the kinetic and the potential energies, with the latter negative since it is taken as zero for the free electron:

$$E=\frac{m_e^\star \nu^2}{2}-\frac{q^2}{4\pi\varepsilon r}. \tag{2.9}$$

Substituting $m_e^\star \nu^2$ and r_H from eqs. (2.5) and (2.8) respectively yields:

$$E=-\frac{q^2}{8\pi\varepsilon_0 r_H}\left(\frac{\varepsilon_0}{\varepsilon}\right)^2\left(\frac{m_e^\star}{m_0}\right)=E_H\left(\frac{\varepsilon_0}{\varepsilon}\right)^2\left(\frac{m_e^\star}{m_0}\right), \tag{2.10}$$

where E_H is the ionization energy of hydrogen and is equal to -13.6 eV. Using this and the values in Table 2.1 one gets $E\simeq 0.025$ eV and 0.01 eV for impurities in Si and Ge respectively.

We have mentioned that the average thermal energy at room temperature is $kT=0.026$ eV, i.e., about equal to the values we found. Moreover, as we shall later find by more exact analysis, it is enough for the thermal energy to be about one-tenth of the ionization energy to free the extra impurity electron for conduction. Therefore, above 100°K most of the impurity atoms are already ionized and the electrons they contribute are available as current carriers. Silicon doped with As, Sb or P will, therefore, be rich in electrons compared to holes (note: ionization of the fifth, weakly bound electron, does not create a hole, i.e. an empty state into which a valence electron can step without an appreciable increase of its energy, since the valence electrons are much more tightly bound and consequently at a much lower energy state). Such a semiconductor is called *N-type* on account of its numerous negative charge carriers. The corresponding doping impurity is called a *donor*, having 'donated' an electron for conduction. Donor impurities for Si and Ge belong to Group V in Table 2.2.

If Si or Ge are doped with an atom of Group III of Table 2.2, having only 3 valence electrons, then at the location of that impurity one of the covalent bonding electrons is missing. The location is electrically neutral, but with a little additional energy (comparable to the ionization energy of the donor atom) one of the other valence electrons can cross over and complete the missing bond. When this happens—and it does happen, thanks to the available thermal energy—a hole is created at the position vacated by that valence electron. The impurity atom, having accepted an additional electron, is called an *acceptor* and is now ionized with a net negative charge of $-q$. A semiconductor doped with acceptors is rich in holes, i.e., positive charge carriers, and therefore called *P-type*.

Energy levels of valence and conduction electrons, donors and acceptors, are shown in Fig. 2.5. The donor ionization energy is E_c-E_d. The acceptor ionization energy is E_a-E_v. E_g is called the gap energy. We shall return to this energy picture in Chapter 6.

Donor and acceptor impurities which are useful for Si and Ge doping and their measured ionization energies are listed in Table 2.3.

Table 2.3. Ionization energies of common impurities: (a) in Si and Ge; (b) in GaAs

Impurity (Type)	Ionization energy (eV) Si	Ge
B (P)	0.045	0.0104
Al (P)	0.057	0.0102
Ga(P)	0.065	0.0108
In (P)	0.16	0.0112
P (N)	0.044	0.0120
As(N)	0.049	0.0127
Sb (N)	0.039	0.0096

(a)

Impurity (Type)	Ionization energy (eV)
Zn(P)	0.0307
Cd(P)	0.0347
Si (P)	0.0345
Si (P)	0.00581
S (N)	0.00610
Se (N)	0.00589
Te (N)	0.0058

(b)

A compound semiconductor, like GaAs, is doped N-type by tellurium (Group VI), P-type by cadmium (GroupII) and either P or N by Si (Group IV). (See Question 2.3). One vital point must be emphasized: the addition of impurities does not violate the electrical neutrality of the semiconductor. N-type material would have many mobile electrons but also the same number of ionized, positively charged donor impurity atoms, which are, however, fixed in their lattice positions. The same holds for P-type material with reversed charge signs. Any tendency of the mobile charge cloud to vacate a certain region would immediately lead to net charges forming and the creation of internal fields (by Poisson's equation) to counteract that tendency till neutrality is restored.

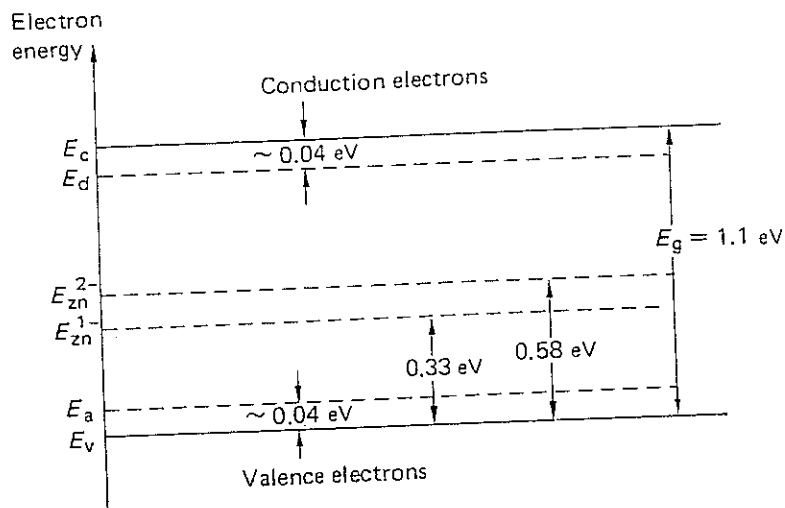


Figure 2.5. Electron energy levels in doped silicon:

E_v is the highest level of valance electrons bonding the crystal lattice atoms together;

E_a is the typical level of electrons occupying the empty bonds on acceptor atoms sites;

E_{Zn}^{1-}, E_{Zn}^{2-} — are deep impurity levels created by Zn doping;

E_d is the typical level of the fifth donor electron when still bound to the donor atom;

E_c is the lowest levèl of conduction electron, free to move inside the semiconductor;

$E_g = E_c - E_v$ is the extra energy a valence electron must obtain to become free for conduction. It is much larger than the available thermal energy $kT = 0.026$ eV at room temperature.

Electrons in N-type material and holes in P-type are called *majority carriers*, while holes in N-type and electrons in P-type are called *minority carriers*, for reasons that will now be discussed.

An important relation will now be shown to hold: the product of the concentrations of majority and minority carriers in thermodynamic equilibrium is independent of the doping impurity concentration and is a function only of the temperature and the semiconductor material.

We shall henceforth denote concentrations at thermodynamic equilibrium by a bar above the letter ($\bar{n}$, $\bar{p}$). These concentrations are the average result of two opposing dynamic processes:

(a) Generation rate G of electron–hole pairs per unit volume and time. In equilibrium G is the result of thermal energy only manifested by vibrational waves in the crystal lattice (phonons), which may sometimes be large enough to ionize a semiconductor atom, thereby creating a hole–electron pair. In equilibrium, G depends only on temperature and the intrinsic material properties and not on the impurities, since these are already ionized in the range of useful temperatures. Hence G in doped material is equal to that of an intrinsic material.

$$G = G_i(T) \text{ (pairs m}^{-3}\text{ s}^{-1}) \quad (2.11)$$

(b) The recombination rate R of electrons with holes (pairs per unit volume and time). When carriers recombine they disappear, giving off their kinetic energy and momentum by emission of phonons (i.e. heating the crystal) or photons (emission of radiation). Recombination rate must be equally dependent on the concentrations of electrons and holes and of course, must depend on T as well:

$$R = R(\bar{n}, \bar{p}, T)$$

under equilibrium conditions the recombination and generation rates must be equal:

$$R(\bar{n}, \bar{p}, T) = G_i(T). \quad (2.12)$$

By (2.12) the value of R is independent of the impurity content.

Obviously, no recombination is possible if either $\bar{n}$ or $\bar{p}$ are zero. Therefore R must contain some product of $\bar{n}$ and $\bar{p}$. The simplest such function is

$$\bar{R}(\bar{n}, \bar{p}, T) = \bar{n}\bar{p}r(T), \quad (2.13)$$

where $r(T)$ is a function of T only. (2.13) also means that electrons and holes have equal effects on the recombination rate which is reasonable. Assuming (2.13) is correct then by substituting it in eq. (2.12) one gets

$$\bar{n}\bar{p} = \frac{G_i(T)}{r(T)}. \quad (2.14)$$

Since this product is a function of T only and independent of the impurity content we can apply it to intrinsic material, and by virtue of eq. (2.4) get

$$\bar{n}\bar{p} = n_i^2. \quad (2.15)$$

This important relation will be proved in a more rigorous way in Chapter 7 to find its functional dependence on the material parameters and temperature. The numerical values obtained at 300°K are

$$n_i\,(\text{Si}) = 1.4 \times 10^{10} (\text{cm}^{-3})$$
$$n_i\,(\text{Ge}) = 2.4 \times 10^{13} (\text{cm}^{-3})$$
$$n_i\,(\text{GaAs}) \simeq 1.7 \times 10^{6} (\text{cm}^{-3}).$$

Relative to the number of atoms in the crystal (about 5×10^{22} atoms per cm^3), these numbers are indeed very small. Adding impurities with concentrations typical of semiconductor devices will usually make n_i negligible in comparison to the majority carrier concentration, especially in GaAs and Si.

Thus, if Si is made P-type by adding $N_A=10^{16}$ cm^{-3} boron (acceptor) atoms, the available number of majority holes around room temperature will be essentially equal to N_A, since the intrinsic hole concentration is six orders of magnitude lower and negligible. The electron (minority) concentration in this material will be given by eq. (2.15).

$$\bar{n}=\frac{n_i^2}{\bar{p}}\simeq\frac{n_i^2}{N_A}=1.9\times10^4\ \ cm^{-3}.$$

We see that the minority concentration is drastically reduced from the intrinsic state by the addition of impurities. This can be intuitively understood considering that the probability that an electron will meet and combine with a hole is very much increased by the addition of acceptor impurity, since the number of holes has increased tremendously and the few intrinsic electrons find themselves in a sea of holes with which to recombine. This makes no measurable change in the number of holes but has a very noticeable effect on the minority concentration, which is the concentration of electrons in this case.

2.4 Compensation and Deep Impurities

A semiconductor doped by equal concentrations of donor and acceptor impurities is said to be fully *compensated*. The free electrons donated by the donors are 'grabbed' by the acceptors since a free electron comes down in energy when it occupies an acceptor state, as can be seen in Fig. 2.5, and each system tends in equilibrium towards its lowest possible energy.

In a fully compensated semiconductor, therefore, all the donated electrons have been caught by acceptor states and none is available for conduction. All the valence electrons stay in the valence levels since no acceptor state is left unoccupied, so there are no holes either. The number of available charge carriers will be very low, like in an intrinsic, undoped, semiconductor. Contrary to intrinsic material, however, a compensated semiconductor has a lot of positively and negatively charged ions (the donors and acceptors) embedded in it and though macroscopically it is electrically neutral, these charges would affect its conductivity as we shall see in Chapter 3.

If an atom from Group II, like Zn, is used to dope a Group IV semiconductor like Si, two bonds in the lattice will be missing in the vicinity of the Zn atom. This atom can therefore accept either a single electron from the valence band and become singly ionized, or accept two and become doubly ionized. The energy levels that these valence electrons must attain to be 'accepted', however, are much higher than for a Group III acceptor like boron (E_a in Fig. 2.5) and are shown in the figure too. Such levels are called *deep*. Thermal energy is not sufficient to excite valence electrons into them but they can catch (trap) free electrons from the conduction level. As we shall learn, this enhances the recombination of such trapped electrons with holes that may come by. Thermal energy is not sufficient to liberate an electron trapped in a deep impurity. One therefore finds that deep impurities drastically reduce the number of electrons free for conduction and an N-type material so doped behaves like an insulator. Deep levels may be created in semiconductors by crystalline defects and by many heavy metal atoms. Especially useful are gold (Au), which in minute quantities is used in Si to enhance recombination and increase the operating speed of switching devices (Chapter 11), and chromium (Cr), used to dope p-type GaAs, create deep donor levels that cause compensation and turn it into a practical insulator called *semi-insulating GaAs*. Such GaAs has resistivities of up to $10^9\Omega$cm and is used as a single crystal substrate on which digital circuits or microwave devices of GaAs are made. As we shall see, such circuits benefit greatly from the insulating property of the substrate. Another method by which semi-insulating GaAs may be obtained is by growing it undoped which gives it a very high specific resistivity to begin with. This can be further increased in selected regions by bombarding them with high energy protons in vacuum. The protons (hydrogen nuclei) wreck the crystalline structure, create many deep levels, and make the bombarded region amorphous and highly resistive.

Other regions of such a substrate can be doped to a lower resistivity by ion implantation (to be described in Chapter 16) so that transistors can be built there. Undoped GaAs substrates pose fewer processing problems than chromium-doped ones and their use is increasing.

2.5 High Doping Densities and Degenerate Semiconductors

In all our treatment up to now and in the theories that are developed in the next chapters there is one implicit assumption and that is that the impurity atoms are few and far between in the semiconductor crystal. If their concentration is made large enough for the orbits of the fifth distant electrons of neighboring donor atoms to start to overlap, then they begin to be influenced by each other and then the semiconductor properties, such as the behavior of its conductivity with temperature, will change. Such a material is called *degenerate*, i.e., one can no longer consider the allowed orbits of the electrons (and the energies associated with them) independently. They unite into a single system in which the Pauli exclusion principle, forbidding electrons to have the same allowed state, holds. We shall return in Section 7.3 to this principle and to its consequences on the allowed energy states of those extra impurity

electrons. We can, however, obtain an estimate of the impurity concentration, N, above which energy levels of impurity atoms become affected by the nearness of other impurity atoms in their neighborhood, as follows:

The average distance between neighboring impurity atoms is $N^{-1/3}$ if N is their concentration. If this distance becomes comparable to the diameter of the fifth electron orbit as given by eq. (2.8), degeneracy sets in. This leads to a value of approximately 10^{19}cm^{-3} as the limiting concentration (about three orders of magnitude less than the semiconductor atom concentration). The technological limit to impurity inclusion is usually higher and is called the *solid solubility* limit. This is a property of the semiconductor, the impurity and the temperature at which the impurity is introduced. Attempts to increase the impurity concentration further will fail because the excess impurity will segregate, form inclusions in the crystal and will not be electrically active, i.e., will not contribute carriers. High doping densities also introduce mechanical stresses in the crystal because of accumulated differences in atomic sizes and increase the number of crystal faults.

QUESTIONS

2.1 When electrons in a semiconductor accumulate high energies, e.g., when the semiconductor is heated, some of them may be energetic enough to be emitted out of the material. Can this also happen to holes?

2.2 What will happen to a semiconductor doped by equal amounts of donors and acceptors?

2.3 Why are donors appropriate to Si and Ge chosen from the elements in Group V of the periodic table? What role do you think a Si atom has in a GaAs crystal if it replaces an As atom? A Ga atom? (Such dopants are called *amphoteric*).

2.4 Will a good or a bad match between the sizes of an impurity and the host semiconductor atom have any effect? Will it affect the solid solubility?

2.5 How do you think the number of majority and minority carriers will change in Si doped by Sb when the temperature is changed from 0°K to near melting temperature? Do you expect exactly the same behavior if the Sb is replaced with As? (*Hint*: Make use of data in Table 2.3.)

2.6 Can the electron's effective mass be measured by applying magnetic or electric fields to an electron while it moves in vacuum?

PROBLEMS

2.7 Using Fig. 2.1, calculate the following properties of Si:

(a) The atomic diameter (unit cell size for Si crystal is 0.543 nm).

(b) The number of atoms in 1 cm^3.

(c) The number of atoms per unit area in the crystal planes (111), (110), (100). These are Miller indices, which are used to indicate specific planes in the crystal.) These three planes are the ones most frequently encountered in semiconductor technology and are shown in Fig. 2.6.

(d) The specific weight (atomic weight of Si is 28). Compare your results with those found in Table 2.1.

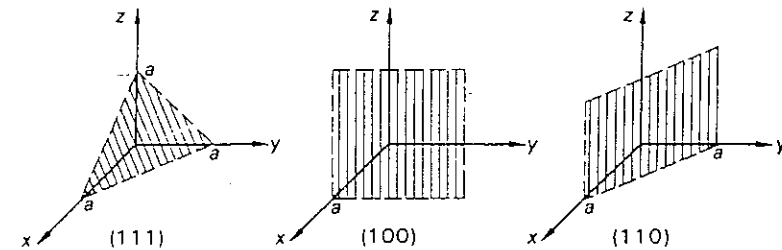


Figure 2.6. The most commonly encountered crystal planes.

2.8 Find the electric field strength E necessary to accelerate an electron from rest to the thermal limiting velocity of 10^5 m s^{-1} within a distance equal to the unit cell dimension in Si.

2.9 How many electron-volts are there (a) in 1 kg m, and (b) in 1 J?

2.10 What are the equilibrium concentrations of holes and electrons at 300°K in

(a) Si doped with $N_D = 3 \times 10^{14}\text{cm}^{-3}$ donors;

(b) Ge doped with the same density of acceptors as in (a)?

(c) State your conclusions regarding (i) acceptable engineering assumptions in each case, and (ii) the relative importance of the thermally generated carriers compared to those contributed by the impurities.

MOBILITY AND ELECTRICAL CONDUCTIVITY 3

3.1 Scattering Mechanisms

If a constant voltage source is connected to the two sides of a semiconductor chip, an electrical field $\mathbf{E}$ (V cm^{-1}) is created in it. This field acts upon the free charge carriers and causes them to drift in the direction of the force it applies, thereby creating a current called *drift current*.

When a charge carrier, say an electron, is acted upon by a constant electrical field in a vacuum, its ensuing acceleration, $\mathbf{a}$, is (Newton's law):

$$\mathbf{a} = \frac{q\mathbf{E}}{m_e}, \tag{3.1}$$

and its velocity $\mathbf{v}$ at time t, if it started from rest:

$$\mathbf{v} = \int_0^t \mathbf{a}\, dt = \frac{q\mathbf{E}t}{m_e}, \tag{3.2}$$

i.e., velocity increases linearly with time.

Inside a semiconductor, on the other hand, the movement of the charge carrier is not smooth but is perturbed by various obstacles, causing what is known as *scattering*. There are two main types of scattering mechanism (there are more but we shall neglect the less important):

(a) *Lattice scattering* is caused by collisions of the moving carrier with disturbances in the periodic internal potential inside the semiconductor

crystal. These disturbances are due to the vibrations of the crystal lattice atoms around their 'proper' place in the lattice because of their thermal energy. The effect of the internal periodic potential itself, which exists in any crystal, can be taken into consideration by assigning an effective mass m^* to the moving electron or hole, as we shall see in Chapter 6. These masses are different from the mass m_0 of the electron outside the crystal. Therefore it is only the disruptions in the periodic potential, caused by the thermal vibrations of the atoms, that scatters the drifting free carriers: at a certain moment an electron can bump into a region where the crystal atoms are more densely packed than usual, yet a moment later it may find itself in a sparsely packed region. The dense and sparse regions form pressure waves existing inside the crystal.

These thermally-caused, lattice pressure waves have also corpuscular or particulate properties, just as photons do. Their energy distribution also follows the same statistical law obeyed by photons, as we shall learn in Chapter 6.

These vibrational wave–particle entities are called *phonons*. They can exist only inside the lattice and have wavelength and velocity like any other wave, or energy and momentum like any other particle. The energy and momentum, however, are not independent. They are related by the so-called dispersion equation, which is obtained when one considers the movement of atoms in lattice under the elastic forces existing between each atom and its near neighbors, like a number of point masses tied together with springs.

The wave or particle nature that phonons exhibit depends on the type of experiment performed. Their interaction with current carriers inside the semiconductor results from the local disturbances introduced by the existence of phonons in the otherwise periodic lattice potential. The interaction can also be looked upon as collisions between the current carriers and phonons. In these collisions, the total energy and momentum are conserved, but become redistributed, changing magnitude and direction of carrier velocity, i.e., they cause the carrier to be scattered. The carrier thus loses the kinetic energy it has accumulated through being accelerated by the external field since the previous scattering event.

Following the collision the carrier is again accelerated in the direction of the field and, inevitably, scattered again, and this repeats all the time. It is obvious that lattice scattering will grow more severe with increasing temperature.

(b) *Impurity scattering* is caused by the presence of ionized impurity atoms in various positions in the crystal lattice. Due to their net charge, they exert a force on the free carrier passing nearby, causing it to change its direction (like a comet entering the gravity field of a star). This type of scattering is less severe if the free carrier is moving faster (i.e. at higher temperatures), and spends less time in the vicinity of the ionized impurity atom.

3.2 Average Drift Velocity; Mobility

Our purpose now is to find an expression for the average velocity of a charge carrier under the effect of a field E, when there is a scattering mechanism such that if we start with n_0 carriers at time $t=0$, measured for each carrier from the moment of its last collision, then a time t later there are still $n(t)$ that have not suffered a second scattering collision and that are still accelerating in the direction of the field. Between t and $t+dt$ an additional dn carriers will suffer a second collision and lose their momentum in the field direction. The number of still accelerating carriers will therefore be reduced by dn which, to a first approximation, is proportional to the number of still uncollided carriers n and to the time increment dt but not to the moving carriers' energy. Let us call the proportionality constant $1/\tau$; then

$$-dn = \frac{1}{\tau} n \, dt. \tag{3.3}$$

The solution of this equation, by separation of variables, is

$$n = n_0 \exp\left(-\frac{t}{\tau}\right), \tag{3.4}$$

where τ has the dimensions of time. Actually it is the average free time between collisions. To see this, let us find the probability for a collision during the period dt, which is dn/n_0. From (3.3), (3.4) we get

$$-\frac{dn}{n_0} = \frac{1}{\tau} \exp\left(-\frac{t}{\tau}\right) dt. \tag{3.5}$$

The right-hand side of (3.5) gives the distribution of the time t that a carrier moves till it collides (a Poissonian distribution). We see, for instance, that a relatively large fraction of the starting n_0 carriers will collide again near $t=0$, at the beginning of their movement, because at $t=0$ the right-side of (3.5) is maximum. But there will be some, on the other hand, whose time t to the next collision will be very long because $\exp(-t/\tau)$ never quite reaches zero in a finite time. Therefore some electrons or holes will be accelerated to very high velocities, while most will reach only a low velocity before colliding and being scattered again. To find an average velocity, or an average time between collisions, we must sum the times to the second collision for the various carriers, assigning a proper weight to each time t depending on the relative number of carriers colliding at that time. This weighting function is $|dn/n_0|$, given by eq. (3.5). Let us also assume that the carrier's velocity $\mathbf{v}_0$ (magnitude and direction) immediately after a collision is completely random (that is a good assumption).

Let $\mathbf{r}$ be the radius vector. Then, at time t, accelerated by the influence of an electric field E, the carrier will be at:

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{v}_0 t + \frac{q\mathbf{E}}{m^*}\frac{t^2}{2}$$

where $\mathbf{r}_0$ is its original position and $q\mathbf{E}/m^*$ is its acceleration.

The average distance travelled by a carrier during t is therefore:

$$\langle \mathbf{r}-\mathbf{r}_0 \rangle = \langle \mathbf{v}_0 t \rangle + \frac{q\mathbf{E}}{2m^*}\langle t^2 \rangle.$$

Since the starting random velocity $\mathbf{v}_0$ and the time t to the next collision are unrelated, the average of their product is equal to the product of their averages, and since the average of the random quantity $\mathbf{v}_0$ is zero, the first term on the right-hand side of the last equation is zero. The average velocity, i.e., the carrier drift velocity $\mathbf{v}_d$, will be given by

$$\mathbf{v}_d = \frac{\langle \mathbf{r}-\mathbf{r}_0 \rangle}{\langle t \rangle} = \frac{q\mathbf{E}}{2m^*}\frac{\langle t^2 \rangle}{\langle t \rangle} \tag{3.6}$$

The averages of t^2 and t can be calculated using their proper weighting function $|dn/n_0|$, as already explained:

$$\langle t \rangle = \int_0^{n_0} t \left|\frac{dn}{n_0}\right| = \int_0^{\infty} t \frac{e^{-t/\tau}}{\tau} dt = \tau$$

$$\langle t^2 \rangle = \int_0^{n_0} t^2 \left|\frac{dn}{n_0}\right| = \int_0^{\infty} t^2 \frac{e^{-t/\tau}}{\tau} dt = 2\tau^2.$$

(The integrals are easily solved using integration by parts.) We see that indeed τ is the average or mean free time between collisions.

Substitution of those values into eq. (3.6) of $\mathbf{v}_d$ will give

$$\mathbf{v}_d = \frac{q\tau}{m^*}\mathbf{E}. \tag{3.7}$$

We obtained a velocity that is proportional to the field and is constant in a constant field. The proportionality factor is called *mobility* and is usually assigned the letter μ:

$$\mu_{e,h} = \frac{q\tau}{m^*_{e,h}}; \qquad \mathbf{v}_d = \mu\mathbf{E}. \tag{3.8}$$

The index e or h indicates whether one is referring to an electron or hole, respectively. In most semiconductors the electron's mobility is higher.

Sometimes, as in GaAs or InSb, it is higher by one or more orders of magnitude. Typical values are given in Table 3.1.

Actually the proportionality between v_d and E (the magnitudes of v and E) is gradually lost at higher and higher fields. When a charge carrier accumulates a very high kinetic energy (as it will do in a high field), it may fall prey to additional scattering processes, previously ignored, which depend on its energy. As a result v_d will gradually saturate to a value equal to the average thermal velocity of about 10^7 cm s^{-1} at fields of some several tens of kilovolts per centimeter. The concept of mobility is therefore meaningful, and it can be considered a constant only at relatively low fields where v_d is still proportional to E. As already mentioned, the mobility is a result of both lattice and impurity scattering, and both depend on temperature but in opposing manners. At temperature T the carriers rush around in the semiconductor with an average thermal velocity v_{th} given by (2.3) and proportional to $T^{1/2}$. The lattice vibrations, due to temperature, increase their amplitude with T, which makes the mean free path of a carrier between lattice scattering events l_L become proportional to T^{-1}. Substituting $\tau = l_L/v_{th}$ in (3.8) makes μ_L, the mobility due to lattice scattering alone:

$$\mu_l \propto T^{-3/2}. \tag{3.9}$$

Impurity scattering becomes severe if the thermal energy kT, of a carrier moving a distance r from a charged impurity, becomes comparable to the potential the charge creates there which is proportional to r^{-1}. The distance r for effective impurity scattering, therefore, is proportional to T^{-1}. Each impurity atom thus presents a scattering area (called a scattering cross section) of πr^2 to the moving carrier. The mean free path between impurity scattering events l_I, is inversely proportional to that area and therefore proportional to T^2. Substituting $\tau = l_I/v_{th}$ in (3.8) give μ_I, the mobility due to impurity scattering alone:

$$\mu_I \propto T^{3/2}. \tag{3.10}$$

When both scattering mechanisms are present, as in any doped semiconductor operating at a temperature above abolute zero, the group of carriers dn colliding between t and $t+dt$ will be composed of two subgroups, completely independent of each other and therefore summable. Equation (3.3) should be written in this case in the form:

$$-dn = -(dn_1 + dn_2) = \frac{n}{\tau_1}\,dt + \frac{n}{\tau_2}\,dt = \left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right) n\, dt,$$

where τ_1 is the average time between collisions of the first kind (lattice), and τ_2 between collisions of the second kind (impurity). Comparing to (3.3), we see that we can repeat the whole calculation with

$$\frac{1}{\tau'} = \frac{1}{\tau_1} + \frac{1}{\tau_2}$$

substituted for $1/\tau$. The resulting mobility will then be

$$\frac{1}{\mu} = \frac{1}{\mu_L} + \frac{1}{\mu_I}. \tag{3.11}$$

The dependence of μ on the temperature will therefore be determined by μ_I at low temperatures (when μ_I is the smaller of the two), and by μ_L at high temperatures (when μ_L is smaller). The higher the doping level the lower μ_I becomes, and consequently the lower μ becomes. In fact it may become so small that even at room temperature and above μ_I would still determine the total mobility.

For real semiconductors these are only rough approximations since additional scattering mechanisms might exist. Thus in a partially ionic material like GaAs the electrons of the gallium atom are drawn nearer to the arsenic nucleus which renders the gallium locations in the lattice slightly positive while the arsenic locations become slightly negative with an electric field in between. We call such material *polarized*. A free electron moving in the crystal will interact with the polar field and may lose energy and momentum to it (i.e., be scattered). In practice, therefore, the exponent of T usually differs from 3/2.

In compensated semiconductors, there are a lot of charged ions in the crystal lattice. This increases impurity scattering and reduces the total mobility compared to uncompensated material with the same density of free carriers.

Table 3.1 gives numerical data for μ in various, almost pure semiconductor materials, around room temperature.

Table 3.1. Mobilities of electrons and holes in various intrinsic semiconductors at 300°K

Material	μ_e(cm² V⁻¹ s⁻¹)	μ_h(cm² V⁻¹ s⁻¹)
Ge	3900	1900
Si	1350	480
GaAs	8500	480
GaP	450	20
InSb	80,000	200
InAs	23,000	100

3.3 Conductivity

Having found the charge carriers' drift velocities in a semiconductor containing electrons and holes under the influence of a field E, we can calculate the ensuing drift current densities:

$$J_e\ (\text{drift}) = -qn(v_d)_e = +qn\mu_e E,$$

$$J_h\ (\text{drift}) = +qp(v_d)_h = +qp\mu_h E,$$

and the total drift current density will be their sum, since even though electrons and holes have charges of opposite sign, their velocity in the field E has opposite direction also and therefore

$$J\ (\text{drift}) = J_e + J_h = q(n\mu_e + p\mu_h)E. \tag{3.12}$$

(We have been repeating the word 'drift' because, as will be seen in the next chapter, the current may have an additional component caused by diffusion of the carriers.)

The *specific conductance* σ of the semiconductor is defined as the ratio of the magnitudes of J and E.

$$\sigma \triangleq \frac{J}{E} = q(n\mu_e + p\mu_h)\ [\Omega\text{cm}]^{-1}. \tag{3.13}$$

In the case of intentionally doped extrinsic material, one of the terms in the brackets of (3.13) is usually negligible compared to the other. In the case of an intrinsic material, where the number of electrons equals the number of holes,

$$\sigma_i = qn_i(\mu_e + \mu_h). \tag{3.14}$$

The specific conductivity varies with temperature for two main reasons. One is the dependence of the free charge carriers concentration on temperature which will be felt either at very low temperatures (when not all the impurity atoms are ionized) or at very high temperatures (when the rate of generation of thermally-created free carriers becomes very high). The second reason is the dependence of mobility on temperature already mentioned, which has less effect and determines the conductivity in the intermediate range. Figure 3.1 gives the general shape of this dependence on temperature for an

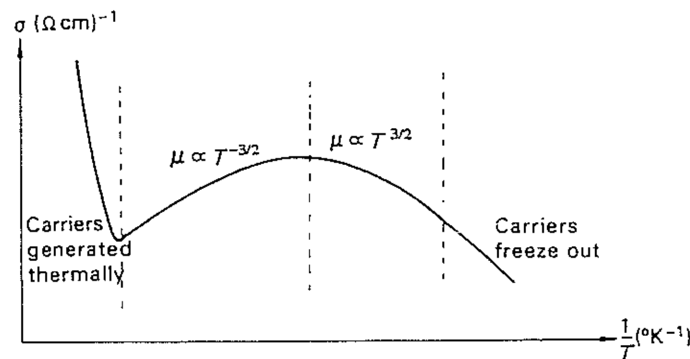


Figure 3.1. The general dependence of specific conductivity on temperature for a semiconductor with an average (10^{15}–10^{16} cm^{-3}) doping.

extrinsic semiconductor with medium doping. It should be mentioned that if the doping is high and the material degenerate, the conductivity becomes more or less constant with temperature in the low and intermediate ranges. At low temperatures this results from the ability of the extra electron or hole associated with the dopant atom to hop from one impurity atom to its very near neighbor without the necessity for thermal energy to ionize it.

The mobility obtained from conductivity measurements is called *conductivity mobility* and refers to majority carrier movement, i.e. electrons in N-type material and holes in P-type. The mobility of the minority species is usually termed *drift mobility*, and is approximately equal to the conductivity mobility for the same doping. As we shall see in Chapter 5, mobility can also be measured by a special effect, called the Hall effect, and one gets then slightly different values. Mobility so measured is termed *Hall mobility*.

QUESTIONS

3.1 A potential difference is applied between two parallel metal plates held in vacuum. An electron is put into the electric field region and starts to move. Can its mobility be calculated?

3.2 The bonding between atoms in a GaAs crystal is partially ionic. Does this introduce an additional scattering factor? If so, why?

3.3 Imagine a semiconductor in which the electron's effective mass is reduced when the applied external field is increased (and this really happens in GaAs or InP). How will this affect the resistivity of such a semiconductor?

3.4 A *compensated semiconductor* is one doped with both donors and acceptor atoms in the same concentration. In such a material $\bar{n} = \bar{p} = n_i$, as in an intrinsic material. Does its resistivity also equal that of an intrinsic undoped material? If not, why not?

PROBLEMS

3.5 Using Table 3.1, calculate the resistance at room temperature (300°K) between the contacts of the silicon chip shown in Fig. 3.2, for the following cases:

(a) the silicon is intrinsic;

(b) it is doped with donors, with $N_D = 10^{16}$ cm^{-3};

(c) it is doped with acceptors, with $N_A = 10^{16}$ cm^{-3}.

6.4 Conductors, Semiconductors and Insulators

In a real three-dimensional crystal, the E–k relation is much more complicated and depends on the orientation of the momentum vector k with respect to the lattice axes, since interatomic distances and the internal potential field shape depend on direction inside the lattice. The basic effect of energy bands formation, however, still remains.

The electrons in a real crystal occupy all the allowed states, starting from the lowest energy, till all the electrons are accommodated. Above the highest occupied level there are more allowed states which are usually empty.

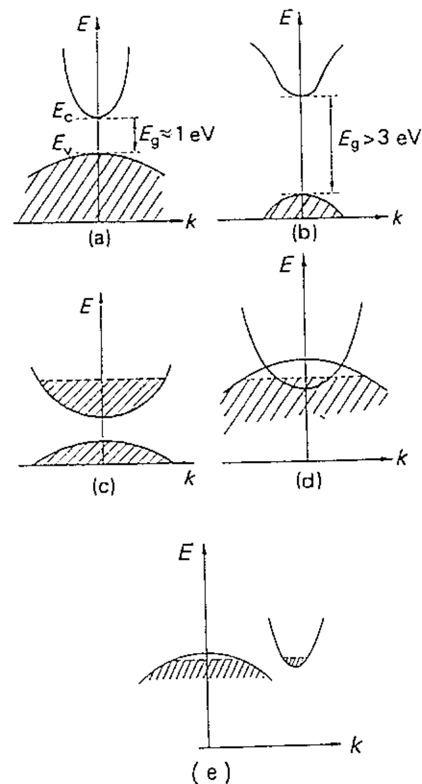


Figure 6.7. The energy bands and their occupation in different materials (hatched regions: levels filled with electrons):
(a) a semiconductor. E_v—the top of the valence band, E_c—the bottom of the conduction band, $E_g = E_c - E_v$ is the forbidden band gap;
(b) an insulator, E_g is large;
(c) a metal conductor; the top band is only partially filled;
(d) and (e) conductors with band overlap.

An intrinsic semiconductor crystal, i.e. a crystal containing no impurities, has the very special property that the upper occupied band, called the *valence band*, is completely full. Above this band comes a relatively narrow (about 1 eV) forbidden energy band called the *band gap*, and above this comes another allowed but normally empty band called the *conduction band*. This band picture is shown in Fig. 6.7(a). The band-gap width is usually denoted by E_g.

Applying an electric field to such a semiconductor, when held at low temperature, does not result in any current, since the field must accelerate the electrons, i.e., increase their kinetic energy, which means that they have to transfer to slightly higher, empty energy levels. No such levels are available. E_g is much too wide a gap for an electron to jump across by virtue of field acceleration. Therefore, no current flows. If the temperature is raised a few of the valence electrons (electrons populating the valence band) will gain sufficient thermal energy to overcome E_g and appear in the conduction band. The electron in the conduction band and the hole left in the valence band can now be accelerated and so carry current. The conductivity, however, is small, and hence the name 'semiconductor'. In an insulating material, such as diamond or silicon dioxide, the band gap is several times wider and even near melting temperatures the thermal energy is not sufficient to transfer an appreciable number of electrons to the conduction band. Its resistivity, therefore, is extremely high. The energy bands in an insulator are as shown in Fig. 6.7(b).

In a metallic conductor, on the other hand, the top occupied band is only partially filled, as in Fig. 6.7(c). Electrons can increase their energy continuously even at low temperatures, and this results in very high conductivity.

Other possible conduction mechanisms are shown in Fig. 6.7(d) and (e). Here the two top bands overlap to some extent, resulting in somewhat different conduction properties. Band overlap occurs in divalent metals, transition metals and semimetals.

6.5 The Effective Mass and the Concept of a Hole

Let us now examine the relation between the electric field strength F and the resulting acceleration of an electron occupying an energy level E near the bottom of an almost empty conduction band. From (6.10) the velocity of the electron is

$$v_g = \frac{1}{\hbar}\frac{\partial E}{\partial k},$$

i.e., it is determined by the shape of the E–k curve. The power supplied by the field, which is the electric field force qF times the velocity v_g, must equal the

increase in the energy of the electron per unit time:

$$\frac{dE}{dt} = qFv_g = \frac{qF}{\hbar}\frac{\partial E}{\partial k}.$$

But

$$\frac{dE}{dt} = \frac{\partial E}{\partial k}\frac{dk}{dt};$$

therefore

$$\frac{d(\hbar k)}{dt} = qF. \tag{6.26}$$

We find that the time derivative of $\hbar k$ gives the applied external force, as if $\hbar k$ were a momentum. This is an interesting result, since qF is not the whole force, only its external part, while the effects of the internal potential fields are already included in the E–k curve. The quantity $\hbar k$ is therefore called *crystal momentum*, since it plays the same role as momentum does in its relation to *external* forces. It is not, however, the total momentum, and the effects of internal forces are included in the dispersion curve and through it in the $\partial E/\partial k$ term. Henceforth, whenever the momentum of an electron or hole is mentioned, this will mean the crystal momentum.

The acceleration caused by the field is

$$a = \frac{dv_g}{dt} = \frac{\partial v_g}{\partial k}\frac{dk}{dt} = \frac{qF}{\hbar}\frac{\partial v_g}{\partial k} = \frac{qF}{\hbar^2}\frac{\partial^2 E}{\partial k^2}.$$

Comparing this to Newton's law relating acceleration and force, we can define an *effective mass* $m_e^\star$ for the electron

$$m_e^\star = \hbar^2 \left(\frac{\partial^2 E}{\partial k^2}\right)^{-1}. \tag{6.27}$$

The effective mass is therefore determined by the radius of curvature of the E–k curve at the given energy level and already includes all the internal potential field effects. In general, $m_e^\star$ varies with k. In fact, (6.27) predicts that whenever the E–k curve is concave, as at the bottom of the top band of Fig. 6.8 (which is the conduction band), $m_e^\star$ is positive, but when it is convex, as near the top of the lower band in that figure (the valence band) $m_e^\star$ is negative. This means that a particle in that state will be accelerated by the field F in the reverse direction expected for a negatively charged electron, i.e., it will behave as if it had a positive charge and mass. This is the concept of *the hole*.

To clarify this concept let us consider the current carried by an almost full valence band.

As can be seen from (6.24), cos kl is an even function, which makes E an even function of k and therefore symmetric in respect to $k=0$. The derivative $\partial E/\partial k$ will therefore be an odd, antisymmetric function and from (6.10)

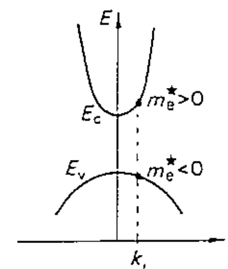


Figure 6.8. The effective mass in the conduction and the valence bands.

this means that

$$v_g(-k_0) = -v_g(+k_0) = -\frac{1}{\hbar}\frac{\partial E}{\partial k}\bigg|_{k=k_0}. \tag{6.28}$$

The total current density carried by n electrons in the valence band is

$$J = \frac{1}{V}\sum_{i=1}^{n}(-q)v_{gi} \tag{6.29}$$

where V is the volume and v_{gi} is the velocity of the ith electron. If the band is completely full, i.e., if $n=2N$, then for each electron in state $+k_0$ there corresponds another in $-k_0$, whose momentum and velocity (by 6.28)) have the opposite direction. The contributions of these two electrons to the current cancel each other. We have arrived at what we already knew: in a full band the total current must be zero.

$$J = -\frac{q}{V}\sum_{i=1}^{2N} v_{gi} = 0. \tag{6.30}$$

If the band is not completely full but contains only $n=2N-p$ electrons, there are p states near the top of the band that are empty. For a positive field direction all the empty states have positive k values (the field pushes the electrons in the negative k-direction and so they fill all the available states with negative k by acquiring momentum in that direction).

The total current density can then be written

$$J = -\frac{q}{V}\sum_{i=1}^{2N-p} v_{gi} = -\frac{q}{V}\left[\sum_{i=1}^{2N} v_{gi} - \sum_{j=2N-p}^{2N} v_{gj}\right] \tag{6.31}$$

$$= -\frac{q}{V}\left[0 - \sum_{j=2N-p}^{2N} v_{gj}\right] = +\frac{q}{V}\sum_{j=2N-p}^{2N} v_{gj}.$$

Since all the p empty states have positive velocities associated with them, J is a positive quantity, as if it were carried by p positive charge carriers of charge $+q$, i.e. by p holes.

Had there been an electron in one of those empty states, it would have been accelerated by the field with

$$a = -\frac{qF}{m_e^*(k_j)}.$$

But, we have seen that $m_e^*(k_j) < 0$ for k_j near the convex top of the valence band in Fig. 6.8. So, if we assign our hole a positive charge $+q$ and a positive effective mass given by

$$m_h^*(k_j) = -m_e^*(k_j), \tag{6.32}$$

it would have exactly the proper acceleration.

The concept of a hole is very convenient to describe the electrical behavior of an almost full band. Since the valence band in a semiconductor is just that, we always prefer to consider the behavior of those few empty states, the holes, rather than add the contributions of all the electrons in that almost full band.

As already mentioned, the effective masses, as defined by (6.27), are not necessarily constant. They are constant only if the relationship between E and k is parabolic, i.e. $E = ck^2$. Since our interest is in semiconductors, with either an almost empty conduction band (N-type) or an almost full valence band (P-type), only the shapes of the bottom of the conduction band or the top of the valence band need be considered and these can be approximated by parabolas such as

$$E \simeq E_c + \frac{\hbar^2 k^2}{2m_e^*}, \tag{6.33a}$$

$$E \simeq E_v - \frac{\hbar^2 k^2}{2m_h^*} \tag{6.33b}$$

which, by applying (6.27), give the proper constant effective masses. These approximations are shown in Fig. 6.13(a).

6.6 Band Shapes of Real Semiconductors

Real semiconductors are usually doped to N- or P-types. The addition of donors actually adds allowed energy states in a level E_d in the forbidden gap, slightly below the bottom of the conduction band, E_c, as shown in Fig. 6.9(a). The abscissa in that figure has no physical meaning and is used

for convenience only, while the ordinate is the electron energy. This is a common, simplified way of showing the important levels without sketching the whole $E(k)$ curve. At very low temperatures the extra donor electrons are still attached to their atoms and occupy the E_d levels. Since the impurity atoms are relatively far apart, they do not affect each other and the E_d states are localized and therefore marked on the figure by a broken line. Around 100°K the thermal energy already enables the extra impurity electron to shift

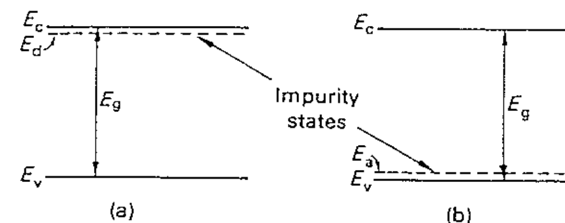


Figure 6.9. A simplified energy band picture with allowed states created by impurity atoms: (a) in N-type semiconductor; (b) in P-type semiconductor.

into one of the many empty states of the nearby conduction band, where it has an effective mass m_e^* and mobility μ_e and can carry current. Addition of acceptor impurities means the addition of allowed but empty (at very low temperatures) states slightly above the valence band at E_a in Fig. 6.9(b). Again at increased temperatures around 100°K valence electrons, from the originally full valence band, will transfer into those states, leaving holes behind.

In real semiconductors the E-k relation is much more complicated than our simplified model. Figure 6.10 shows the valence and conduction bands of three important semiconductors. The curves are not symmetrical since different k-directions are used for each side, so as to show more than one crystal orientation and also the various minima in the conduction bands.

It is only in GaAs that the lowest conduction band minimum occurs at $k=0$, but in this material too there is another local minimum, 0.36 eV higher and situated near the edge of the Brillouin zone, in the <100> direction. In both Si and Ge the lowest conduction-band minimum occurs near the zone edges. The valence band maximum is always at $k=0$. The band gap E_g is the energy difference between the minimum of E_c and the maximum of E_v and only in GaAs they occur at the same k-value. GaAs is therefore called a *direct band gap* semiconductor, as are also InSb, InP, CdS and others. Here an electron transferring from the top of the valence band to the bottom of the conduction band or vice versa changes only its energy and not its momentum $\hbar k$. In Si, Ge and also GaP, AlSb and others both energy and momentum

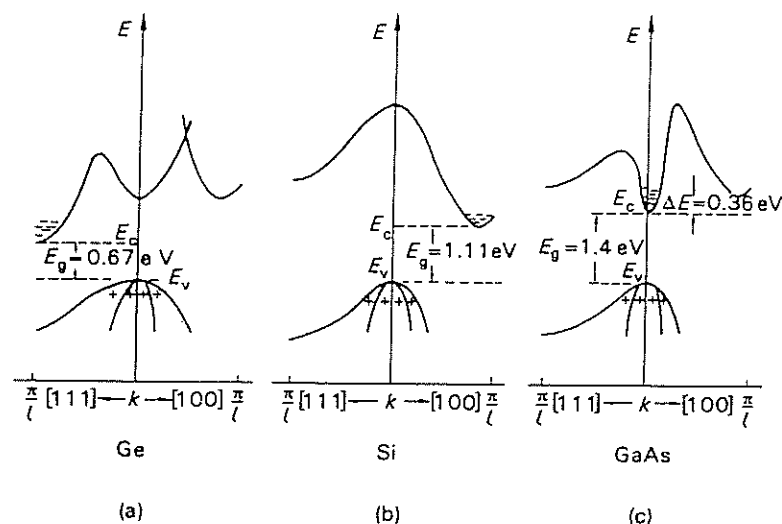


Figure 6.10. The energy bands in momentum space for the important semiconductors, at 300°K: (a) Ge; (b) Si; (c) GaAs.

must be simultaneously changed so they are called *indirect band gap* semiconductors. The type of band gap is very important from the utilization point of view, since both energy and momentum conservation laws must hold during carrier generation or recombination.

When, for example, the generation is by a photon of energy $E_{ph}=hf$, we must have $E_{ph}>E_g$ (the extra energy appears as the kinetic energies of the generated hole and electron) but the momentum must also be conserved. Now, the momentum of the photon is only

$$p_{ph}=mc=\frac{mc^2}{c}=\frac{E_{ph}}{c}=\frac{hf}{c}=\frac{h}{\lambda}=\hbar k_{ph}.$$

Therefore

$$k_{ph}=\frac{E_{ph}}{\hbar c}\,, \tag{6.34}$$

where c is the velocity of light.

In an indirect semiconductor such as Ge, the change in k involved in a transfer across E_g is approximately from $k=0$ to $k=\pi/l$. For $l\simeq 0.2$ nm this means $\Delta k \simeq 1.5\times 10^{10}$ m^{-1}. When compared with (6.34) for $E_{ph}\simeq 1$ eV we get a value about three orders of magnitude too low. The photon cannot account for the necessary momentum change in either generation or recombination. This means that other quantum particles, such as phonons, must participate and either remove or contribute the necessary excess momentum. This makes radiative recombination (i.e. direct electron–hole recombination with

photon emission) an unlikely process in indirect semiconductors. It is, however, a very likely process in direct band gap semiconductors such as GaAs, and it is therefore from such materials that lasing and other light emitting devices are being made, as we shall see in Chapter 10.

The phonons, i.e., the vibrational modes of the lattice mentioned in Section 3.1. propagate in the crystal with the acoustic velocity v_{ac}, which is about 10^5cm s^{-1}. The energy of the phonon is of the order of kT, i.e. about 0.026 eV at room temperature. Using (6.34) to compare the momenta of the photon and the phonon, we find for Ge (E_g=0.67 eV).

$$\frac{k_{phon}}{k_{phot}}=\frac{E_{phon}/\hbar v_{ac}}{E_{phot}/\hbar c}\simeq\frac{kT/v_{ac}}{E_g/c}\simeq 12{,}000.$$

A phonon has therefore a relatively high momentum but relatively low energy. In an indirect semiconductor the most likely recombination process involves transfer of excess energy and momentum to several phonons, i.e., heating the lattice. The carrier is usually first trapped in a trapping center, such as a crystal defect, i.e., it becomes localized, gradually losing its energy and momentum to phonons until it finally recombines with an opposite carrier that passes by. Excess carrier lifetime in such materials is therefore long, and they are used for transistor production where this property is important. Sometimes recombination involves both phonon and photon, as in GaP, which is indirect. There the emitted phonon takes most of the momentum and the emitted photon takes most of the energy and is in the visible red-light range. GaP is therefore extensively used for devices emitting visible light.

Careful examination of Fig. 6.10 reveals another interesting point: the valence band consists actually of two overlapping bands with a common maximum at $k=0$ but with different curvatures (and therefore different effective masses). Consequently, there are two types of holes, called *light and heavy*. In Si the mass of a light hole is $0.16m_0$ and that of a heavy hole is $0.5m_0$. There are many more available states in the heavy band, so most holes are heavy.

From Fig. 6.10(b) we learn that conduction electrons in Si are actually distributed among six equivalent minimums (valleys) of the conduction band (since there are six <100> directions in the crystal). Collisions will scatter the electrons from one valley to another. Due to the large number of available states an external field will not be able to increase the electrons energy by much, they will just scatter into approximately equal vacant energy states; losing their excess momentum and energy to phonons.

This is not the case in the N type GaAs of Fig. 6.10(c) and an even more interesting effect can be foreseen. There are two local minima in the conduction band. The lowest is at $k=0$ and has low effective mass. The other, 0.36 eV higher, is again a six-fold minimum and has a high effective mass and low mobility (eq. (3.8)). In equilibrium, at room temperature, practically all the

MEASURING THE ELECTRICAL PARAMETERS OF A SEMICONDUCTOR 5

The basic electrical semiconductor parameters of main interest are as follows:

(a) the conductivity σ (or resistivity $\rho = 1/\sigma$)
(b) the majority carrier concentration
(c) the type (N or P)
(d) the mobility
(e) the diffusion constant
(f) the lifetime.

In this chapter several laboratory techniques for measuring those parameters will be described.

5.1 The Resistivity

This measurement is the simplest to understand, though not necessarily to perform. It can be done by making four metal contacts to the semiconductor, as shown in principle in Fig. 5.1(a).

The two outside contacts are used to pass a known current I (supplied by a current source) while the two inner ones are used to measure the ensuing voltage V, using a voltmeter with a high input impedance compared to the measured value. In this way the metal–semiconductor contact resistance, which may be high, nonlinear and dependent on current direction, does not affect the results: The contact resistances at points A, B of Fig. 5.1(a) are

immaterial, since they do not control the constant current forced through them, while the voltage developed across them is not measured anyway. The contact resistance at C and D is also unimportant because the high input impedance voltmeter draws negligible current. From the figure,

$$V=\rho \frac{L}{Wd} I;$$

therefore

$$\rho=\frac{Wd}{L}\frac{V}{I}. \tag{5.1}$$

The voltage-measuring contacts should be as narrow as possible in the current direction since they short out the semiconductor region over which they lie. The current contacts on the two sides should be uniform to insure uniform current density.

A more practical, simpler way of getting the same results is by using a commercially obtainable probe, which has a head with four springy metal points, arranged in a line, usually less than 1 mm apart, as in Fig. 5.1(b). When the probe is lowered onto the semiconductor surface, the two outside points are used for passing the constant current while the two inner ones are used for the voltage measurement. The current density in this case is not uniform but if the proper current-flow field problem is solved, one finds that for a semiconductor wafer with thickness d much bigger than the distance s between points, and for a measurement performed far enough from the edge of the wafer, the resistivity is given by

$$\rho=2\pi s \frac{V}{I} \quad (\Omega\text{cm}), \qquad (s \ll d).$$

If the thickness d of the wafer or of the measured semiconductor layer is much less than s, then:

$$\rho=\frac{\pi}{\ln 2}\frac{V}{I} d=4.53 \frac{V}{I} d \quad (\Omega\text{cm}), \qquad (s \gg d).$$

In between these two extremes a correction factor is needed.

This method is especially useful for very thin semiconductor layers, such as are obtained when dopant atoms of one type of impurity are allowed to diffuse from the surrounding ambient into a semiconductor already doped with the opposite type of impurity (a basic process in modern transistor technology). After about an hour of diffusion at high temperature, a thin layer on the semiconductor surface changes type because the density of the new impurity there exceeds that of the original. The diffused impurity density, however, is not uniform in depth. Instead of average resistivity we then prefer

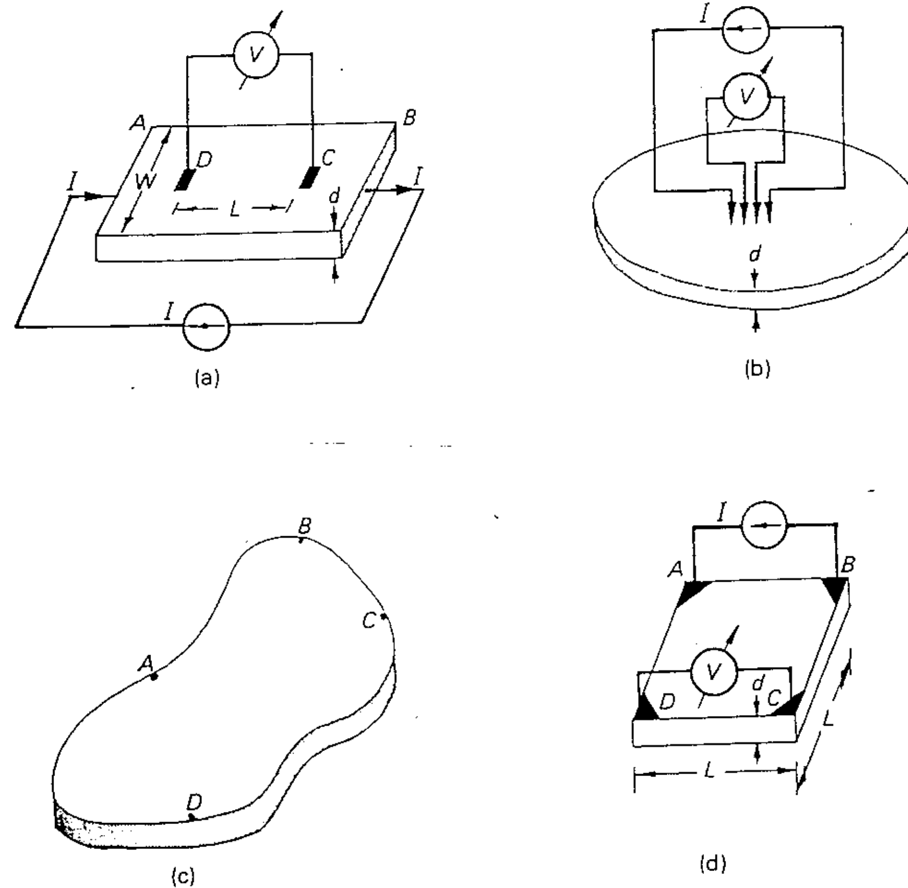


Figure 5.1. Methods for measuring resistivity: (a) the four-contact method; (b) the four-points probe; (c) the Van der Pauw method; (d) a symmetrical sample for Van der Pauw's method.

to speak of the layer *sheet resistivity* R_s, measured in 'ohms per square', defined as the resistance of an arbitrary size square of that layer:

$$R_s = \frac{\rho}{d} = 4.53 \frac{V}{I} \quad (\Omega/\square). \tag{5.2}$$

Such diffused layers are used to form resistors in integrated circuits. If a resistor of R (Ω) is required, one gives the diffused layer the surface geometry of a rectangle, of length L and width W, so chosen that

$$R = R_s \frac{L}{W} \quad (\Omega).$$

A second method, known as Van der Pauw's method, enables one to measure ρ of a semiconductor wafer of thickness d, and an arbitrary shape, like in Fig. 5.1(c). Four small contacts, A, B, C, D arranged along the periphery, are used to pass current and measure voltage as before.

If by $R_{AB,CD}$ one denotes the ratio of the voltage V_{CD} measured between contacts C and D, to the current I_{AB} flowing between A and B, then Van der Pauw has shown that

$$\exp\left(-\frac{\pi d}{\rho} R_{AB,CD}\right) + \exp\left(-\frac{\pi d}{\rho} R_{BC,DA}\right) = 1. \tag{5.3}$$

If the contacted wafer is completely symmetrical, like a circle or the square in Fig. 5.1(d), this equation yields ρ explicitly:

$$\rho = \frac{\pi d}{\ln 2} \frac{V_{CD}}{I_{AB}}. \tag{5.4}$$

5.2 Majority Carrier Concentration, Type and Mobility by the Hall Effect

The Hall effect, discovered in 1874 by E. H. Hall, is an often used experimental tool to measure the mobile carrier density, together with the sign of their charge. If the conductivity is known, one can also calculate the mobility.

The Hall effect results from the force **F**, with which a magnetic field **B** ($\mathrm{W\,m^{-2}}$) acts on a current density **J**, according to the vector product:

$$\mathbf{F} = \mathbf{J} \times \mathbf{B} \tag{5.5}$$

If **J** and **B** are perpendicular, as in Fig. 5.2, then **F**, **J**, **B** form a right-handed Cartesian coordinate system and the direction of the force is as marked in the figure.

This force acts on the charge carriers, which happen to be electrons here, moving at a drift velocity **v**, opposite to the conventional current direction. Equation (5.5) shows, however, that the force depends on the current **J** and *not* on the type of carriers, so it will remain in the same direction, pushing the carriers towards the back, even if the N-type semiconductor is replaced by a

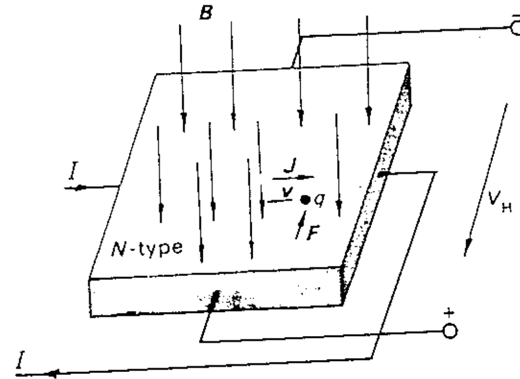


Figure 5.2. The Hall effect. An N-type semiconductor with current density **J** in a normal magnetic field **B**. A force **F** acts on an electron q carrying the current.

P-type, with hole carriers moving in the same direction as the current. When the mobile carriers are pushed towards the back, the front becomes depleted and the semiconductor loses its local neutrality. There is now an excess of the mobile type of charge at the back and an excess of the opposite ionized impurity charge at the front. In the case of Fig. 5.2, the electrons, pushed to the back, will make the back contact negative with respect to the front contact, which will be made positive. This gives rise to a measurable voltage V_H, called the *Hall voltage*. For a P-type semiconductor the polarity of V_H is in the reverse direction. This method also provides an experimental demonstration of the existence of holes.

The value of the Hall voltage will be determined by the balance that must be reached between the forces that the transverse electric field E_H, related to V_H, and the magnetic field **B** exert on the carriers. They exactly oppose each other and their vector sum must be zero.

For **J** normal to **B** one can dispense with the vector product sign and consider magnitudes only:

$$F\,(\text{magnetic}) + F\,(\text{electric}) = 0,$$

or

$$JB + nqE_H = 0$$

(5.6)
$$E_H = -\frac{1}{nq} JB.$$

The Hall coefficient, R_H, is defined as

(5.7)
$$R_H \triangleq \frac{E_H}{JB}.$$

Therefore for electrons

(5.8)
$$R_H = -\frac{1}{nq}.$$

While for holes, where the electrical force F(electric) is in the opposite direction,

(5.9)
$$R_H = +\frac{1}{pq}.$$

Actually, the drift velocity of the carriers, responsible for the magnetic force, is not a constant but a statistical average as seen in Chapter 3. Also there are sometimes nonlinear effects dependent on the value of $|\mathbf{B}|$. Consequently, a correction factor r_H (usually between 1 and 1.5, sometimes as small as 0.5), called the *Hall coefficient factor* should be added to the magnetic force equation making

(5.10)
$$R_H = -\frac{r_H}{nq} \quad \text{or} \quad R_H = +\frac{r_H}{pq}.$$

For our purposes r_H can be assumed to be about 1.

Measurement of R_H, the Hall coefficient, gives us the carrier concentration directly. It can be measured at any desired temperature and, if that is low enough for some of the impurities not to be ionized, it tells us the fraction that is (by comparison with its value at higher temperatures when all impurities are ionized). Also, in defective crystals full of dislocations, a sizable fraction of the impurities may not be ionized because they are located at grain boundaries or dislocations and not at proper lattice sites. Then R_H yields the active fraction only. Hall measurements can be made highly sensitive, enough to sense concentrations as low as $10^{12}(\text{cm}^{-3})$.

For semiconductor materials that contain both electrons and holes at about the same concentrations, i.e. approximately intrinsic, the Hall voltage caused by one is cancelled by the other; the net result is a function of both concentrations and mobilities and may even be zero.

From the definition of the conductivity of an extrinsic semiconductor,

(5.11)
$$\mu_h = \frac{\sigma}{q\bar{p}} \quad \text{or} \quad \mu_e = \frac{\sigma}{q\bar{n}}.$$

Comparing with (5.8) and (5.9) we see that mobilities can be found from the Hall coefficient and conductivity:

(5.12)
$$\mu_{e,h} = \sigma \,|\, R_H \,|_{e,h}.$$

Mobility measured in this way is called *Hall mobility*, while that found directly from σ, $\bar{n}$ or $\bar{p}$ or from the Haynes–Shockley experiment, to be described in Section 5.4, is called *conductivity mobility*. The ratio of the two is the Hall

coefficient factor r_H of eq. (5.10) which, as mentioned, may be slightly different from 1.

To make a Hall measurement, the sample is prepared in a bridge form as in Fig. 5.3(a). The d.c. current is fed via contacts A and B and the Hall voltage is measured between C and D, or E and F (or both with an average taken). The same sample may be used for conductivity measurement too.

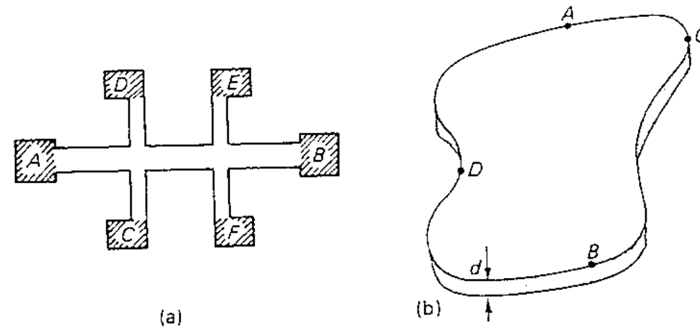


Figure 5.3. Semiconductor samples for Hall and conductivity measurements: (a) using a bridge form; (b) using an arbitrary form by Van der Pauw's method.

A sample of an arbitrary form may also be used. Van der Pauw has shown that application of a perpendicular magnetic field would result in a change in $R_{AB,CD}$, defined as in (5.3), compared to its zero magnetic field value and

$$\mu_H = \frac{\sigma d}{B} \Delta R_{AB,CD}. \tag{5.13}$$

Note (Fig. 5.3(b)) that here current is applied and voltage measured at diagonally opposite pairs of contacts.

The Hall effect has many uses today, besides measurement of semiconductor parameters. The vector product property, for example, may be utilized for angular position sensing; the ability to sense *static* magnetic fields by a tiny device is exploited in mapping magnetic fields in cramped spaces.

Hall devices for such uses are usually made from N-type InSb which, as Table 3.1 tells us, has a very high electronic mobility, and therefore high R_H and high sensitivity.

5.3 Carrier Type by the Hot Probe Method

This is a simple, frequently used method, unencumbered by the necessity for the preparation of a special sample.

One simply touches the unknown semiconductor surface by two identical metal probes, between which a galvanometer is connected as in Fig. 5.4(a). One of the probes is heated while the other is at room temperature. The hot probe heats the semiconductor immediately under it, with a consequent rise in the kinetic energy of the free carriers there. These then move with higher thermal velocities than their cooler neighbors. The carriers therefore diffuse out of the hot region faster than their slower neighors can diffuse back into it from the vicinity. This results in the hot region becoming slightly depleted of majority carriers and acquiring the potential of the ionized impurities there, while the vicinity of the cold probe remains neutral. Current will therefore flow in the galvanometer, the direction of which depends on the sign of the charge of the ionized impurity. Thus, on an N-type semiconductor, the hot probe is the more positive one, while on a P-type it is the more negative. The cold probe polarity therefore indicates the type.

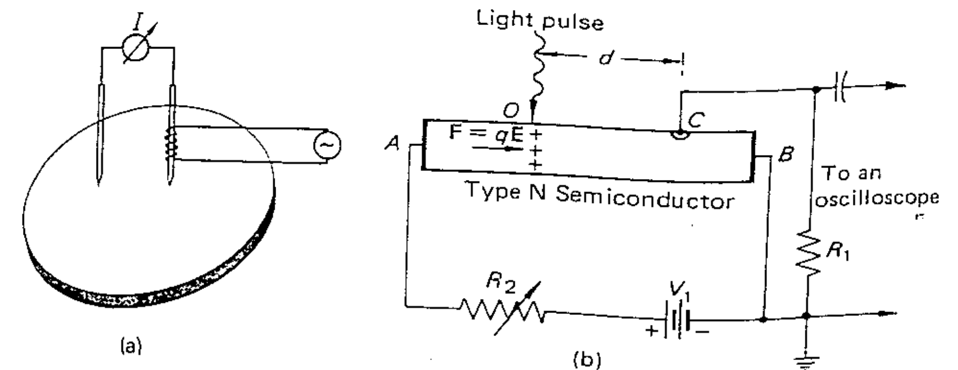


Figure 5.4. (a) The hot probe method; (b) the Haynes–Shockley experiment.

5.4 The Haynes–Shockley Experiment: Mobility, Diffusion Constant and Lifetime of Minority Carriers

The mobility of the minority carriers, the diffusion constant and the lifetime were measured together in a classical experiment, performed in 1949 by Haynes and Shockley. The experiment is very illuminating. The set-up is as in Fig. 5.4(b): A semiconductor bar of a constant cross section (Haynes and Shockley used N-type Ge) with three contacts, is mounted on a holder. Contacts A and B are ohmic, i.e. they have low contact resistance that is independent of current polarity and magnitude and the equilibrium near them is not disturbed. Those contacts are used for passing current and creating a

An important situation is the following: let a static magnetic field $\mathbf{B}$ lie along the z axis. Then the equations of motion are

(CGS)
$$m\left(\frac{d}{dt}+\frac{1}{\tau}\right)v_x=-e\left(E_x+\frac{B}{c}v_y\right) ;$$
$$m\left(\frac{d}{dt}+\frac{1}{\tau}\right)v_y=-e\left(E_y-\frac{B}{c}v_x\right) ; \qquad (51)$$
$$m\left(\frac{d}{dt}+\frac{1}{\tau}\right)v_z=-eE_z .$$

The results in SI are obtained by replacing c by 1.

In the steady state in a static electric field the time derivatives are zero, so that the drift velocity is

$$v_x=-\frac{e\tau}{m}E_x-\omega_c\tau v_y ; \qquad v_y=-\frac{e\tau}{m}E_y+\omega_c\tau v_x ; \qquad v_z=-\frac{e\tau}{m}E_z , \qquad (52)$$

where $\omega_c \equiv eB/mc$ is the **cyclotron frequency**, as discussed in Chapter 8 for cyclotron resonance in semiconductors.

Hall Effect

The Hall field is the electric field developed across two faces of a conductor, in the direction $\mathbf{j}\times\mathbf{B}$, when a current $\mathbf{j}$ flows across a magnetic field $\mathbf{B}$. Consider a rod-shaped specimen in a longitudinal electric field E_x and a transverse magnetic field, as in Fig. 14. If current cannot flow out of the rod in the y direction we must have $\delta v_y = 0$. From (52) this is possible only if there is a transverse electric field

(CGS)
$$E_y=-\omega_c\tau E_x=-\frac{eB\tau}{mc}E_x ; \qquad (53)$$

(SI)
$$E_y=-\omega_c\tau E_x=-\frac{eB\tau}{m}E_x$$

The quantity defined by

$$\boxed{R_H=\frac{E_y}{j_xB}} \qquad (54)$$

is called the **Hall coefficient**. To evaluate it on our simple model we use $j_x = ne^2\tau E_x/m$ and obtain

(CGS)
$$R_H=-\frac{eB\tau E_x/mc}{ne^2\tau E_xB/m}=-\frac{1}{nec} ; \qquad (55)$$

(SI)
$$R_H=-\frac{1}{ne}$$

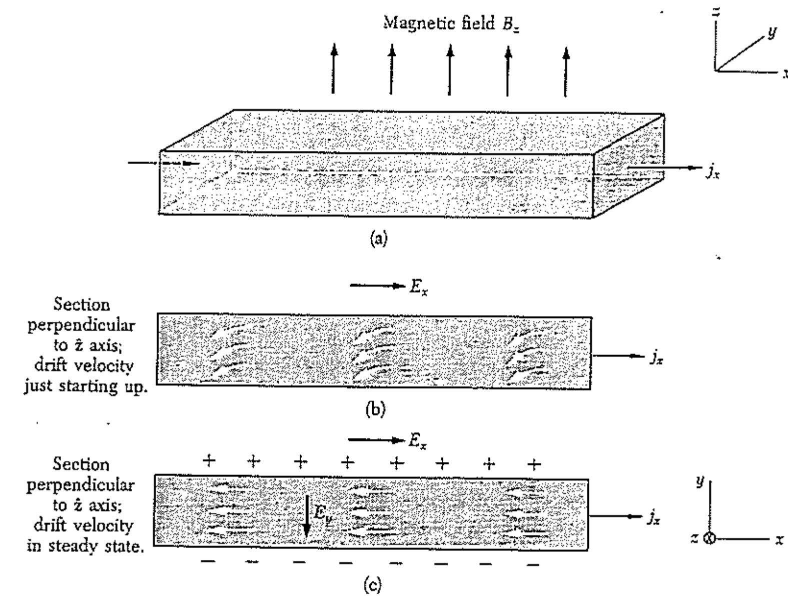


Figure 14 The standard geometry for the Hall effect: a rod-shaped specimen of rectangular cross-section is placed in a magnetic field B_z, as in (a). An electric field E_x applied across the end electrodes causes an electric current density j_x to flow down the rod. The drift velocity of the negatively-charged electrons immediately after the electric field is applied as shown in (b). The deflection in the y direction is caused by the magnetic field. Electrons accumulate on one face of the rod and a positive ion excess is established on the opposite face until, as in (c), the transverse electric field (Hall field) just cancels the Lorentz force due to the magnetic field.

This is negative for free electrons, for e is positive by definition.

The lower the carrier concentration, the greater the magnitude of the Hall coefficient. Measuring R_H is an important way of measuring the carrier concentration.

The symbol R_H denotes the Hall coefficient (54), but it is sometimes used with a different meaning, that of Hall resistance in two-dimensional problems. When we treat such problems in Chapter 19, we shall instead let

$$\rho_H = BR_H = E_y/j_x \qquad (55a)$$

denote the **Hall resistance**, where j_x is the surface current density.

The simple result (55) follows from the assumption that all relaxation times are equal, independent of the velocity of the electron. A numerical factor of order unity enters if the relaxation time is a function of the velocity. The expression becomes somewhat more complicated if both electrons and holes contribute to the conductivity. The theory of the Hall effect again becomes simple in high magnetic fields such that $\omega_c\tau \gg 1$, where ω_c is the cyclotron frequency and τ the relaxation time. (See *OTS*, pp. 241–244.)

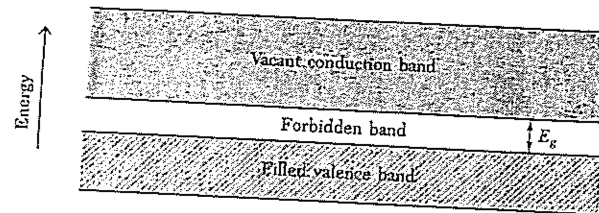


Figure 2 Band scheme for intrinsic conductivity in a semiconductor. At 0 K the conductivity is zero because all states in the valence band are filled and all states in the conduction band are vacant. As the temperature is increased, electrons are thermally excited from the valence band to the conduction band, where they become mobile.

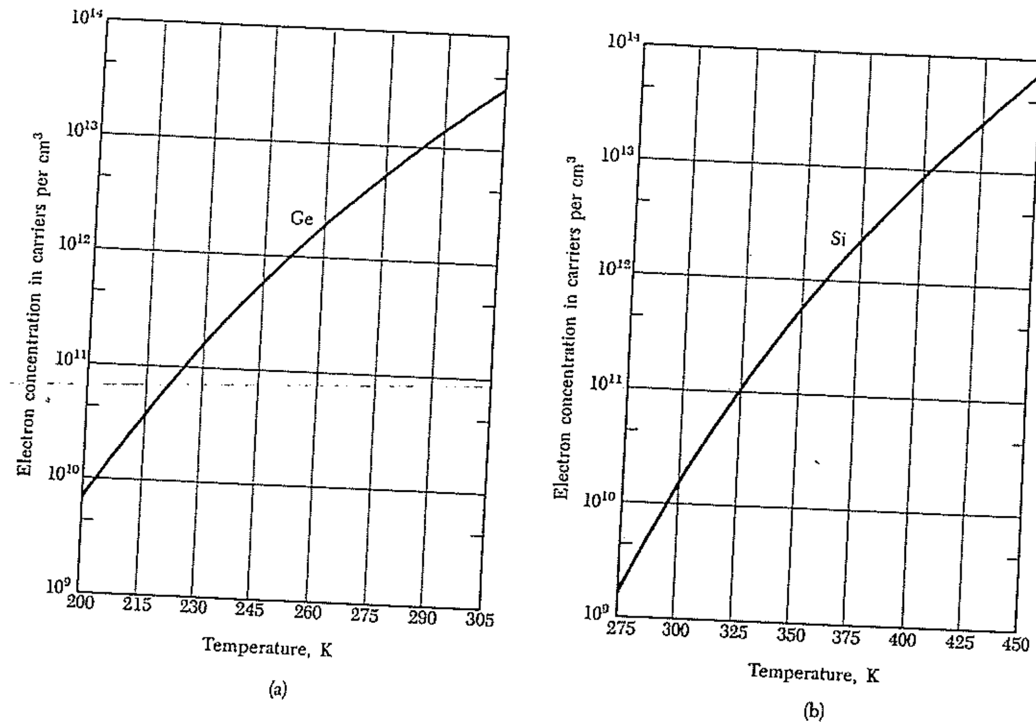


Figure 3 Intrinsic electron concentration as a function of temperature for (a) germanium and (b) silicon. Under intrinsic conditions the hole concentration is equal to the electron concentration. The intrinsic concentration at a given temperature is higher in Ge than in Si because the energy gap is narrower in Ge (0.66 eV) than in Si (1.11 eV). (After W. C. Dunlap.)

Table 1 Energy gap between the valence and conduction bands

(i = indirect gap; d = direct gap)

Crystal	Gap	E_g, eV 0 K	E_g, eV 300 K	Crystal	Gap	E_g, eV 0 K	E_g, eV 300 K
Diamond	i	5.4		HgTe[a]	d	−0.30	
Si	i	1.17	1.11	PbS	d	0.286	0.34–0.37
Ge	i	0.744	0.66	PbSe	i	0.165	0.27
αSn	d	0.00	0.00	PbTe	i	0.190	0.29
InSb	d	0.23	0.17	CdS	d	2.582	2.42
InAs	d	0.43	0.36	CdSe	d	1.840	1.74
InP	d	1.42	1.27	CdTe	d	1.607	1.44
GaP	i	2.32	2.25	ZnO		3.436	3.2
GaAs	d	1.52	1.43	ZnS		3.91	3.6
GaSb	d	0.81	0.68	SnTe	d	0.3	0.18
AlSb	i	1.65	1.6	AgCl		—	3.2
SiC(hex)	i	3.0	—	AgI		—	2.8
Te	d	0.33	—	Cu_2O	d	2.172	—
ZnSb		0.56	0.56	TiO_2		3.03	—

[a]HgTe is a semimetal; the bands overlap.

The threshold of continuous optical absorption at frequency ω_g determines the band gap $E_g = \hbar\omega_g$ in Figs. 4a and 5a. In the **direct absorption process** a photon is absorbed by the crystal with the creation of an electron and a hole.

In the **indirect absorption process** in Figs. 4b and 5b the minimum energy gap of the band structure involves electrons and holes separated by a substantial wavevector $\mathbf{k}_c$. Here a direct photon transition at the energy of the minimum gap cannot satisfy the requirement of conservation of wavevector, because photon wavevectors are negligible at the energy range of interest. But if a phonon of wavevector $\mathbf{K}$ and frequency Ω is created in the process, then we can have

$$\mathbf{k}(\text{photon}) = \mathbf{k}_c + \mathbf{K} \cong 0 \ ; \qquad \hbar\omega = E_g + \hbar\Omega \ ,$$

as required by the conservation laws. The phonon energy $\hbar\Omega$ will generally be much less than E_g: a phonon even of high wavevector is an easily accessible source of crystal momentum because the phonon energies are characteristically small (~0.01 to 0.03 eV) in comparison with the energy gap. If the temperature is high enough that the necessary phonon is already thermally excited in the crystal, it is possible also to have a photon absorption process in which the phonon is absorbed.

he semiconductor
the range can be

CHAPTER 8: SEMICONDUCTOR CRYSTALS

Carrier concentrations representative of metals, semimetals, and semiconductors are shown in Fig. 1. Semiconductors are generally classified by their electrical resistivity at room temperature, with values in the range of 10^{-2} to 10^{9} ohm-cm, and strongly dependent on temperature. At absolute zero a pure, perfect crystal of most semiconductors will be an insulator, if we arbitrarily define an insulator as having a resistivity above 10^{14} ohm-cm.

Devices based on semiconductors include transistors, switches, diodes, photovoltaic cells, detectors, and thermistors. These may be used as single circuit elements or as components of integrated circuits. We discuss in this chapter the central physical features of the classical semiconductor crystals, particularly silicon, germanium, and gallium arsenide.

Some useful nomenclature: the semiconductor compounds of chemical formula AB, where A is a trivalent element and B is a pentavalent element, are called III-V (three-five) compounds. Examples are indium antimonide and gallium arsenide. Where A is divalent and B is hexavalent, the compound is called a II-VI compound; examples are zinc sulfide and cadmium sulfide. Silicon and germanium are sometimes called diamond-type semiconductors, because they have the crystal structure of diamond. Diamond itself is more an insulator rather than a semiconductor. Silicon carbide SiC is a IV-IV compound.

A highly purified semiconductor exhibits intrinsic conductivity, as distinguished from the impurity conductivity of less pure specimens. In the **intrinsic temperature range** the electrical properties of a semiconductor are not essentially modified by impurities in the crystal. An electronic band scheme leading to intrinsic conductivity is indicated in Fig. 2. The conduction band is vacant at absolute zero and is separated by an energy gap E_g from the filled valence band.

The **band gap** is the difference in energy between the lowest point of the conduction band and the highest point of the valence band. The lowest point in the conduction band is called the **conduction band edge**; the highest point in the valence band is called the **valence band edge**.

As the temperature is increased, electrons are thermally excited from the valence band to the conduction band (Fig. 3). Both the electrons in the conduction band and the vacant orbitals or holes left behind in the valence band contribute to the electrical conductivity.

BAND GAP

The intrinsic conductivity and intrinsic carrier concentrations are largely controlled by E_g/k_BT, the ratio of the band gap to the temperature. When this ratio is large, the concentration of intrinsic carriers will be low and the conductivity will be low. Band gaps of representative semiconductors are given in Table 1. The best values of the band gap are obtained by optical absorption.