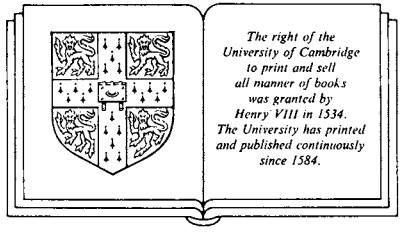
Principles of modern technology

Adrian C. Melissinos
Department of Physics
University of Rochester



CAMBRIDGE UNIVERSITY PRESS
Cambridge
New York Port Chester Melbourne Sydney

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1990

British Library cataloguing in publication data
Melissinos, Adrian C.
Principles of modern technology.
1. Technology
I. Title
600

Library of Congress cataloguing in publication data available

ISBN 0 521 35249 5 hard covers
ISBN 0 521 38965 8 paperback

Transferred to digital printing 2003

MP

Contents

1

THE TRANSISTOR

1.1 Intrinsic semiconductors

It is well known that certain materials conduct electricity with little resistance whereas others are good insulators. There also exist materials whose resistivity is between that of good conductors and insulators, and is strongly dependent on temperature; these materials are called *semiconductors*. Silicon (Si), germanium (Ge) and compounds such as gallium arsenide (GaAs) are semiconductors, silicon being by far the most widely used material. Solids, in general, are crystalline and their electrical properties are determined by the atomic structure of the overall crystal. This can be understood by analogy to the energy levels of a free atom.

A free atom, for instance the hydrogen atom, exhibits discrete energy levels which can be exactly calculated. A schematic representation of such an energy diagram is shown in Fig. 1.1(*a*). If two hydrogen atoms are coupled, as in the hydrogen molecule, the number of energy levels doubles as shown in part (*b*) of the figure. If the number of atoms that are coupled to each other is very large – as is the case for a crystal – the energy levels coalesce into *energy bands* as in Fig. 1.1(*c*). The electrons in the crystal can only have energies lying in these bands.

When an atom is not excited the electrons occupy the lowest possible energy levels. In accordance with the Pauli principle only two electrons (one with spin projection up and the other down) can be found at any one particular energy level. Thus the levels – or states – become progressively filled from the bottom. The same holds true in the crystal. The electrons progressively fill the energy levels within a given band, and only when the band is completely filled do they begin to populate the next band. The energies of the electrons are typically few *electron-Volts* (eV).

In an insulator the occupied energy bands are completely filled. As a result the electrons cannot move through the crystal. This is because motion implies slightly increased energy for the electrons but the next available energy level is in the conduction band which is far removed from the valence band. Thus the electron must acquire enough energy to overcome the *energy gap* between the *valence band* and the *conduction band* as shown in Fig. 1.2(*a*). In a conductor the valence and conduction bands overlap and the outermost electron of the atom is free to move through the lattice (Fig. 1.2(*c*)). In a semiconductor the energy gap is much smaller than for insulators and due to thermal motion electrons have a finite probability of finding themselves in the conduction band.

Fig. 1.1. Energy levels of an atomic system: (*a*) single atom, (*b*) two coupled atoms, (*c*) in a many-atom system the energy levels coalesce into energy 'bands'.

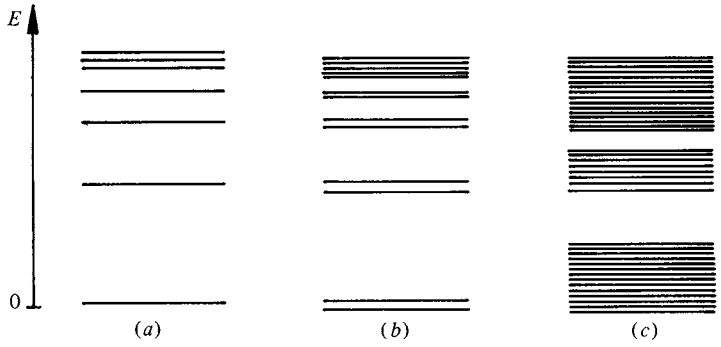
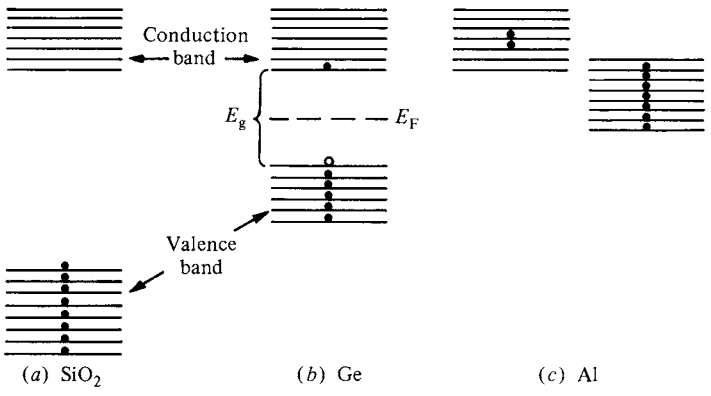


Fig. 1.2. Energy band structure for: (*a*) an insulator such as SiO_2, (*b*) a semiconductor, such as Ge, (*c*) a good conductor such as Al.



Furthermore when an electron makes a transition from the valence to the conduction band it leaves a vacancy in the valence band. This vacancy can move through the lattice (just as a bubble 'moves' through a liquid) and contribute to the flow of current; we speak of transport of electric charge by the motion of *holes* (Fig. 1.2(*b*)).

To obtain a feeling for the occupancy of the energy levels in a solid we can consider the following simple model for a conductor. We assume that one electron in each atom is so loosely bound that it is practically free inside the crystal. This is the case for copper which has $Z = 29$, and thus every atom has 29 electrons. Of these, 28 electrons completely fill the $n = 1$ (2 electrons), $n = 2$ (8 electrons) and $n = 3$ (18 electrons) shells, leaving one electron outside the closed shells. Such an electron is loosely bound to the atom and in fact it occupies a level in the conduction band; thus it can move freely through the crystal. It is simple to calculate the density of free electrons in copper. We have $Z = 29$, $A \simeq 63$, $\rho = 8.9\ \mathrm{g/cm^3}$ and assume one free electron per atom; then

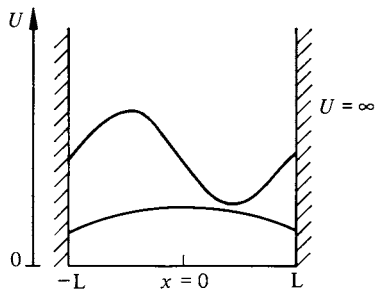
$$n_e = \frac{N_0}{A}\left(\frac{\text{atoms}}{\text{g}}\right) \times \rho\left(\frac{\text{g}}{\text{cm}^3}\right) = \frac{6 \times 10^{23}}{63} 8.9 = 8.5 \times 10^{22} \frac{\text{electrons}}{\text{cm}^3} \tag{1.1}$$

where $N_0 = 6 \times 10^{23}$ is Avogadro's number.

The free electrons in a metal can be described approximately as particles confined within a cubic box but with no other forces acting on them. This situation is depicted for one dimension in Fig. 1.3 and we speak of a 'potential well' of length $2L$. In this case the solution of Schrödinger's equation leads to wave functions of the form

$$\psi_n(x) = (1/\sqrt{L}) \cos(k_n x) \qquad \text{or} \qquad (1/\sqrt{L}) \sin(k_n x)$$

Fig. 1.3. The wave function for the lowest and next to lowest energy states of a particle confined to the region $-L < x < L$ by an infinitely high potential.



where the wave number k_n can take only the discrete values

$$k_n = n\frac{\pi}{2L} \qquad n = 1, 2, 3, \ldots \tag{1.2}$$

so as to satisfy the boundary conditions $\psi(-L) = \psi(L) = 0$. Thus the allowed energies of the particles in the potential well are quantized and given by

$$E_n = \frac{p^2}{2m} = \frac{\hbar^2 k_n^2}{2m} = \frac{\hbar^2 \pi^2}{8mL^2} n^2 \tag{1.3}$$

If we generalize to three dimensions, we must use three quantum numbers, n_x, n_y and n_z and the energy is given by

$$E_n = \frac{\hbar^2 \pi^2}{8mL^2}(n_x^2 + n_y^2 + n_z^2) \qquad n_x, n_y, n_z = 1, 2, 3, \ldots \tag{1.3'}$$

Every particular combination of n_x, n_y, n_z represents a different energy level and only two electrons can occupy it. Note that several energy levels (different combinations of n_x, n_y, n_z) can have the same energy; we say that these levels are degenerate.

We can use Eq. (1.3') to calculate the energy of the highest filled level given the density of free electrons n_e in the crystal. This level is called the *Fermi level* and its energy is the Fermi energy for the system. It is given by

$$E_F = \frac{\hbar^2}{2m}(3\pi^2 n_e)^{2/3} \tag{1.4}$$

where m is the mass of the electron. For Cu we use the result of Eq. (1.1) and

$$\hbar c = 2 \times 10^{-5} \text{ eV-cm}$$

$$mc^2 = 0.5 \times 10^6 \text{ eV}$$

to find $E_F = 7.1$ eV, in good agreement with observation. To see how Eq. (1.4) is derived we must count the number of (n_x, n_y, n_z) combinations available when the maximal value of $(n_x^2 + n_y^2 + n_z^2)$ is specified. In Fig. 1.4 every combination of (n_x, n_y, n_z) is indicated by a dot in 3-dimensional space. When n_x, n_y, n_z are large, a given value of $(n_x^2 + n_y^2 + n_z^2)^{1/2} = \text{constant}$ defines the surface of a sphere in this space; all levels on the surface of the sphere have the same energy. The number of levels inside the sphere equals its volume, because the dots are spaced one unit apart from one another. Since n_x, n_y, n_z must be positive the number of combinations N_c is given by the volume of one octant

$$N_c = \tfrac{1}{8}(\tfrac{4}{3}\pi R^3) = \frac{\pi}{6}[(n_x^2 + n_y^2 + n_z^2)_{\max}]^{3/2}$$

Because of the Pauli principle the number of electrons occupying the N_c

levels is $N_e = 2N_c$. Thus the energy of the highest occupied level is

$$E_{max} = \frac{\hbar^2\pi^2}{8mL^2}(n_x^2 + n_y^2 + n_z^2)_{max}$$

$$= \frac{\hbar^2\pi^2}{8mL^2}\left(\frac{3}{\pi}N_e\right)^{2/3} = \frac{\hbar^2}{2m}\left[3\pi^2\frac{N_e}{8L^3}\right]^{2/3} \tag{1.4'}$$

Note that the N_e electrons are confined in a volume of size $V = (2L)^3$ and therefore in Eq. (1.4') $(N_e/8L^3) = n_e$ is the free electron density establishing the result of Eq. (1.4).

Let us now return to the free electron model. In the absence of excitations, that is at very low temperature, only the levels below the Fermi energy, E_F will be occupied. Let $f(E)$ indicate the probability that a level at energy E is occupied; clearly $f(E)$ is bounded between 0 and 1. If we plot $f(E)$ as a function of E for $T = 0$ it must have the square form indicated by curve A in Fig. 1.5. As the temperature increases some of the levels above E_F will become occasionally occupied, and correspondingly some levels below E_F will be empty. The probability of occupancy, $f(E)$ for a finite temperature $T_1 \neq 0$ is indicated by curve B in Fig. 1.5. The

Fig. 1.4. Counting the number of states labeled by the integers n_x, n_y and n_z such that $(n_x^2 + n_y^2 + n_z^2) \leqslant R^2$. Each state is represented by a dot.

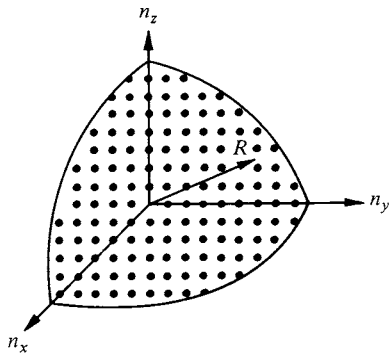
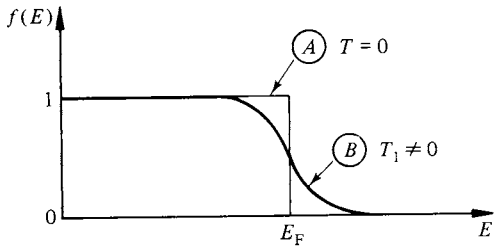


Fig. 1.5. The Fermi distribution function for zero temperature (A) and for finite temperature (B); E_F is the Fermi energy.



function $f(E)$ is known as the *Fermi function* and is given by

$$f(E) = \frac{1}{e^{(E-E_F)/kT} + 1} \tag{1.5}$$

In the limit $T \to 0$ Eq. (1.5) reduces to $f(E) = 1$ if $E < E_F$ or to $f(E) = 0$ if $E > E_F$ in agreement with curve A of Fig. 1.5.

For finite T, consider an energy level E_k lying above E_F; we define $\varepsilon = (E_k - E_F)$. As long as $\varepsilon \gtrsim 3kT$ Eq. (1.5) can be approximated by

$$f(\varepsilon) \simeq e^{-\varepsilon/kT} \tag{1.6}$$

For a level E_j lying below E_F we define $\varepsilon' = (E_F - E_j)$. We are now interested in the probability that the level E_j is empty, namely in $f'(\varepsilon') = 1 - f(\varepsilon')$. As long as $\varepsilon' \gtrsim 3kT$ a valid approximation to Eq. (1.5) is

$$1 - f(\varepsilon') = 1 - \frac{1}{e^{-\varepsilon'/kT} + 1} = e^{-\varepsilon'/kT} \tag{1.6'}$$

Eqs. (1.6) show that at finite temperature there are as many occupied states above the Fermi level as there are empty states below it. This result can serve as a rigorous definition of the Fermi level. Finally we note that the expansions of Eqs. (1.6, 6') coincide with the classical Boltzmann distribution.

To get a better feeling for the implications of the Fermi function on the distribution of carriers in a semiconductor we first calculate kT at room temperature. Boltzmann's constant

$$k = 1.38 \times 10^{-23} \text{ J/K}$$

and if we take

$$T = 300 \text{ K}$$

$$kT = 4.1 \times 10^{-21} \text{ J} = 0.026 \text{ eV}$$

The energy gap for an insulator is of order $\Delta E \sim 5$ eV whereas for semiconductors it is $E_g \sim 1$ eV. Thus for semiconductors at room temperature a small fraction of the electrons in the valence band can be thermally excited into the conduction band.

For a pure semiconductor we designate the number density of (intrinsic) electrons in the conduction band by n_i. For an intrinsic semiconductor the density of holes will also equal n_i and therefore the Fermi level will lie in the middle of the energy gap as shown in Fig. 1.2(*b*). The *intrinsic carrier density* is then given by the probability of occupancy $f(\varepsilon)$ multiplied by N_s the number of available states per unit volume. Using Eq. (1.6) we find

$$n_i = N_s e^{-E_g/2kT} \tag{1.7}$$

(N_s is an effective density of states near the band edge and for silicon it

is of order $\sim 10^{19}\,\mathrm{cm}^{-3}$). In general n_i is much smaller than the free electron density in a good conductor. For instance, for silicon where $E_g = 1.1$ eV, at room temperature $n_i \simeq 10^{10}\,\mathrm{cm}^{-3}$, whereas for germanium ($E_g = 0.7$ eV), $n_i \simeq 10^{13}\,\mathrm{cm}^{-3}$. This should be compared to the free electron density in copper which we calculated to be $n_f \sim 10^{23}\,\mathrm{cm}^{-3}$. Of course the crystal as a whole remains electrically neutral, but if an electric field is applied the carriers will be set in motion and this will lead to the transport of charge. It is evident from Eq. (1.7) that the conductivity of a pure semiconductor will be highly temperature dependent.

1.2 Doped semiconductors

We saw in the previous section that the intrinsic carrier densities are quite small. Thus, unless a semiconductor is free of impurities to a high degree, the phenomena associated with the motion of the intrinsic carriers will not be manifest. On the other hand, by introducing a particular impurity into the semiconductor one can greatly enhance the number of carriers of one or of the other kind (i.e. of electrons or of holes). The great technical advances in selectively and accurately controlling the concentration of impurities in silicon have made possible the development of microelectronics. We speak of doped semiconductors.

To understand the effect of doping we note that the electronic structure of Si or Ge is such as to have four electrons outside closed shells; they are elements of chemicals valence 4.

			Filled shells	Valence
Si	$Z = 14$	$A \sim 28$	$(n=1, n=2)_{10}$	$(3s)_2\ (3p)_2$
Ge	$Z = 32$	$A \sim 72$	$(n=1, n=2, n=3)_{28}$	$(4s)_2\ (4p)_2$

If one examines the periodic table in the vicinity of Si and Ge, one finds the valence 3 elements boron (B, $Z = 5$), aluminum (Al, $Z = 13$) or indium (In, $Z = 49$). On the other side are valence 5 elements such as phosphorus (P, $Z = 15$), arsenic (As, $Z = 33$) or antimony (Sb, $Z = 51$). What will happen if impurities from these elements are introduced into pure silicon?

If valence 5 elements are introduced into the silicon lattice the extra electron will be loosely bound and can be easily excited into the conduction band. We say that these elements are *donor* impurities. If valence 3 elements are introduced they will have an affinity for attracting an electron from the lattice, creating a vacancy or hole in the valence band. We say that valence 3 elements are *acceptor* impurities.

Because of their different electronic structure as compared to that of the crystal lattice, the donor levels are situated just below the conduction

band, as shown in Fig. 1.6(*a*). The acceptor levels are instead located slightly above the valence band (Fig. 1.6(*b*)). This energy difference is so small that at room temperature the impurity levels are almost completely ionized. Thus in the case of donor impurities the charge carriers are electrons and we speak of an *n-type* semiconductor whereas for acceptor impurities the carriers are the holes and we speak of a *p-type* semiconductor. Recall that the crystal is always electrically neutral and that the charge of the carriers is compensated by the (opposite) charge of the ionized impurity atoms, the ions however, remain at fixed positions in the lattice.

In the presence of impurities the position of the Fermi level is determined by the concentration of the impurities and moves toward the conduction band if the dominant free carriers are electrons, toward the valence band if the dominant free carriers are holes. This is sketched in Fig. 1.6. The position of the donor level is indicated by the plus signs in Fig. 1.6(*a*), of the acceptor level by the minus signs in (*b*) of the figure.

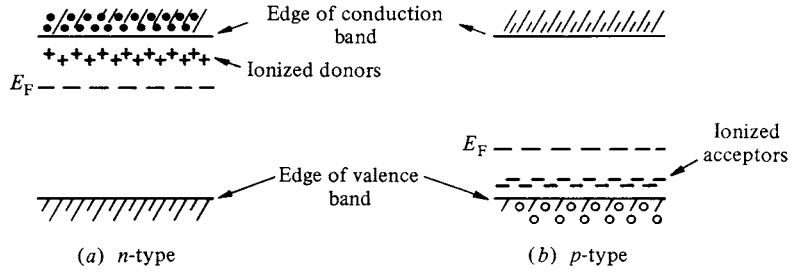
As an example we consider an *n*-type semiconductor, and as usual, designate the (extrinsic) conduction electron density by n, and the hole density by p. Then according to Eqs. (1.6, 6′)

$$\left.\begin{aligned} n &= N_c e^{-(E_c - E_F)/kT} \\ p &= N_v e^{-(E_F - E_v)/kT} \end{aligned}\right\} \tag{1.8}$$

Here we introduced a new concept, the effective density of states N. This is the number of available energy states per unit volume, the subscripts c and v referring to the condition and valence band correspondingly. In general N_c and N_v need not be equal to one another.

Similar relations hold for the intrinsic carriers except that we designate

Fig. 1.6. Energy band diagram for doped semiconductors. Dots represent electrons and open circles holes: (*a*) for an *n*-type semiconductor (note the position of the donor level), (*b*) for a *p*-type semiconductor (note the position of the acceptor level).



the corresponding Fermi level by E_i, and n_i must equal p_i. Thus

$$n_i = N_c e^{-(E_c - E_i)/kT} = p_i = N_v e^{-(E_i - E_v)/kT} \tag{1.9}$$

This relationship can be solved to yield the exact value of E_i

$$E_i = \tfrac{1}{2}(E_c + E_v) + \tfrac{1}{2}kT \ln(N_v/N_c) \tag{1.9'}$$

as well as a convenient expression for n_i

$$n_i = (N_v N_c)^{1/2} e^{-(E_c - E_v)/2kT} \tag{1.9''}$$

Finally we can multiply the two Eqs. (1.8) with one another

$$np = N_c N_v e^{-(E_c - E_v)/kT}$$

and by comparing with Eq. (1.9″) obtain the very important relation

$$np = n_i^2 \tag{1.10}$$

The product of the electron and hole densities is independent of the doping and depends on the intrinsic properties of the semiconductor and the temperature. This is true under equilibrium conditions and provided the intrinsic carriers are not highly excited.

In a doped semiconductor we have majority and minority carriers. For instance in an *n*-type semiconductor the electrons are the *majority* carriers and the holes the *minority* carriers; the opposite is of course true for *p*-type semiconductors. The density of ionized donors and acceptors is designated by N_D and N_A respectively. Then if the electrons are the majority carriers it holds

$$N_D \gg N_A \qquad \text{and} \qquad n \sim N_D$$

We can obtain more accurate relations by taking into account the electrical neutrality of the crystal. The charge density ρ must equal zero and therefore

$$\rho = q(p - n + N_D - N_A) = 0 \tag{1.11}$$

Solving Eq. (1.11) for p and inserting the result in Eq. (1.10) we obtain a quadratic equation for n, whose solution is

$$n = \tfrac{1}{2}(N_D - N_A) \pm \tfrac{1}{2}[(N_D - N_A)^2 + 4n_i^2]^{1/2}$$

For an *n*-type semiconductor, where $(N_D - N_A) > 0$ we must keep the solution with the positive radical. And if $(N_D - N_A) \gg n_i$ we have

$$\begin{aligned} n_n &= (N_D - N_A) \sim N_D \\ p_n &= \frac{n_i^2}{(N_D - N_A)} \sim \frac{n_i^2}{N_D} \end{aligned} \tag{1.12}$$

where the subscript indicates the type of semiconductor. For instance, n_n or p_p are majority carrier concentrations. Similar relations are valid for *p*-type semiconductors.

Thus we see that by the controlled introduction of impurities we can create materials with a particular type of majority carriers. It is the junction of two or more such materials that makes possible the control and amplification of electric current by solid state devices.

1.3 Charge transport in solids

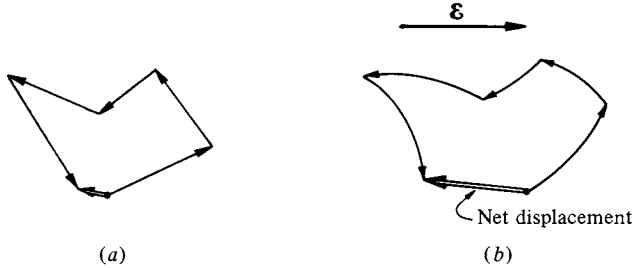
One is familiar with the notion of an electric current 'flowing' through a wire. What we are referring to is the transport of electric charges through the wire, and this in turn is a consequence of the motion of the carriers in the wire. In a good conductor the carriers are electrons, while in a gas discharge or in a liquid the carriers are both electrons and positive ions. In a semiconductor the carriers are electrons, or holes, or both, depending on the material. The current at a point x along the conductor is defined as the amount of charge crossing that point in unit time $I = \Delta Q/\Delta t$. It is more convenient to use the current density $\mathbf{J}$ which is the amount of charge crossing unit area (normal to the direction of $\mathbf{J}$) per unit time. By definition then

$$\mathbf{J} = qn\mathbf{v}_d \tag{1.13}$$

Here q is the charge of the electron (carrier), n is the carrier density, and $\mathbf{v}_d$ is the *drift velocity* of the carriers.

The carriers in a solid are in continuous motion because of their thermal energy. This motion is completely random as the carriers scatter from the lattice and it does not contribute to *net* transport of charge. Thus to transport charge a drift velocity must be superimposed on the random motion. This can be achieved by applying an external electric field. The motion is then modified as shown schematically in Figs. 1.7(*a*), (*b*). (We have assumed that the carriers are electrons so their

Fig. 1.7. Idealized motion of free electrons in a metal: (*a*) in the absence of an external electric field, (*b*) in the presence of an external electric field a net drift current is established.



motion is opposite to the direction of $\mathscr{E}$.) Another cause for net carrier motion is the presence of density gradients. The carriers will then move so as to equalize the density and we speak of diffusion. Finally, carriers can be lost by recombination with impurities, or conversely, they can be created by photo-ionization or thermal excitation.

We first examine the motion of carriers under the influence of an electric field $\mathscr{E}$. The acceleration of a charged particle will be $\mathbf{a} = \mathbf{F}/m = q\mathscr{E}/m^*$ where we have replaced the mass, m, of the particle by an *effective mass* m^* because the carriers do not move in free space but in the lattice. If the time between collisions is t_{coll} then the average or drift velocity in the direction of the electric field will be

$$\mathbf{v}_d = \tfrac{1}{2}\mathbf{a}t_{\text{coll}} = q\frac{t_{\text{coll}}}{2m^*}\mathscr{E} \tag{1.14}$$

Namely, the drift velocity is proportional to the electric field $\mathscr{E}$. This is a very general result and the proportionality coefficient is called the *mobility* μ. Thus

$$\mathbf{v}_d = \mu\mathscr{E} \tag{1.14'}$$

From Eqs. (1.14, 14') we can express the current density as

$$\mathbf{J} = qn\mu\mathscr{E} \tag{1.15}$$

This result is equivalent to Ohm's law which states that the current density is proportional to the electric field and is related to it by the *conductivity* σ

$$\mathbf{J} = \sigma\mathscr{E} \tag{1.15'}$$

Thus

$$\sigma = qn\mu \tag{1.16}$$

Conductivity has dimensions of $(\text{ohm}^{-1}\,\text{m}^{-1})$ and has been recently defined as the 'siemens'. When both types of carriers contribute to the transport of charge, Eq. (1.16) must be modified to read

$$\sigma = q(n\mu_- + p\mu_+) \tag{1.16'}$$

The inverse of the conductivity is the *resistivity*, ρ, and the resistance of a conductor of cross sectional area A and length L is given by

$$R = \rho\frac{L}{A} = \frac{1}{\sigma}\frac{L}{A} \tag{1.16''}$$

We can evaluate the mobility if we knew the time between collisions t_{coll}. Instead, it is convenient to introduce the mean free path (m.f.p.), l, between collisions. Then, $t_{\text{coll}} = l/v_{\text{rms}}$ where v_{rms} is the velocity due to thermal motion. We can write

$$\tfrac{1}{2}m^*(v_{\text{rms}})^2 = \tfrac{3}{2}kT$$

and therefore $v_{rms} = (3kT/m^*)^{1/2}$, so that from Eq. (1.14)

$$\mu = \frac{ql}{2(3kTm^*)^{1/2}}$$

Thus the mobility is a property of the crystal and depends on the temperature. As the electric field is increased the drift velocity increases and reaches a saturation value v_s. Typical values for negative carriers in silicon are

$$\mu \sim 10^2 \text{ cm}^2/\text{V-s} \qquad v_s \sim 10^7 \text{ cm/s}$$

In general, the mobility of the positive carriers is much smaller than that of the negative carriers.

When density gradients are present in the solid, the carriers will *diffuse* from regions of high concentration to those of lower concentration. The flux of carriers is proportional to the density gradient. In one dimension we have

$$\mathscr{F}_x = -D\frac{dn}{dx} \tag{1.17}$$

where D is the diffusion coefficient. We can expect that the diffusion coefficient is related to the mobility of the carriers, and this relationship was first established by Einstein. One finds that

$$D = \frac{kT}{q}\mu \tag{1.17'}$$

Therefore the current due to diffusion is given by

$$J_x = -\mu kT\left(\frac{dn}{dx}\right) \tag{1.17''}$$

a result that should be compared to Eq. (1.15).

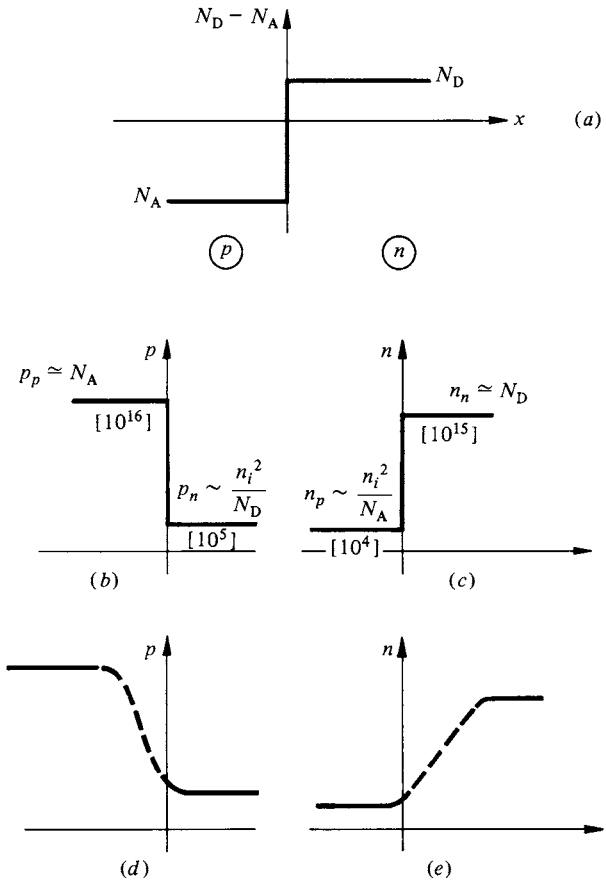
In addition to the drift and diffusion currents, carriers may be being lost due to recombination. Recombination often takes place at traps, which are locations in the crystal where a hole is trapped near the conduction band. In many semiconductors, under the influence of light or other radiation, an electron can become excited from the valence to the conduction band, increasing the density of carriers; thus a photocurrent can flow through the circuit if it is suitably biased.

1.4 The *p–n* junction

So far we have considered current flow in semiconductors which were uniformly doped to make *n*-type or *p*-type material. If two such

semiconductor materials of different type are joined the current flow through the junction depends on the polarity of the external bias. The technology for making *p–n* junctions is an important development which we discuss later. For the analysis of the junction it is sufficient to use a one-dimensional approximation as shown in Fig. 1.8. We assume that for the *p*-type material (to the left of the junction) the ionized acceptor density is N_A, while for the *n*-type material the ionized donor density is N_D and that $N_A > N_D$; this is shown in Fig. 1.8(*a*). In the idealized case the distribution of positive carriers follows the impurity distribution and would be as in Fig. 1.8(*b*), where the numbers in brackets are typical

Fig. 1.8. Carrier density distribution in the immediate vicinity of a *p–n* semiconductor junction: (*a*) donor and acceptor densities define a step junction, (*b*) positive and (*c*) negative carrier densities for the idealized case, (*d*) and (*e*) represent the realistic equilibrium distribution of carrier densities. Values in brackets are typical concentrations per cm^3.

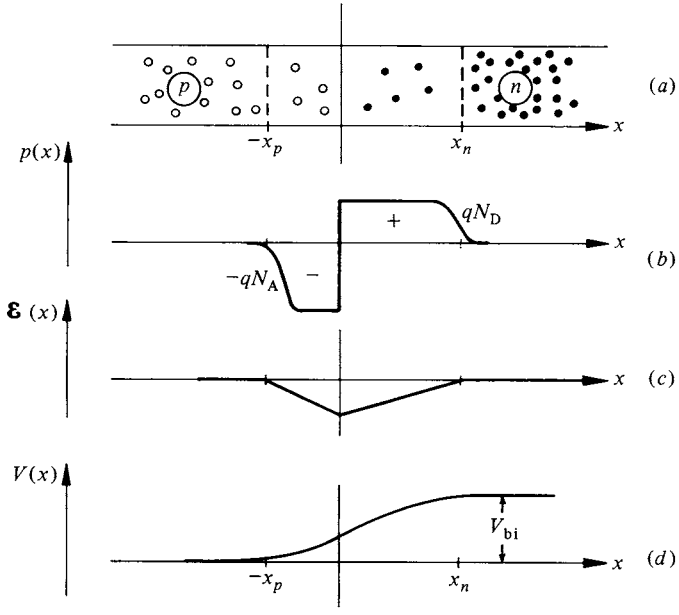


values. To the left of the junction the holes are the majority carriers while they are minority carriers to the right of the junction; the converse is true for the negative carriers as indicated in Fig. 1.8(*c*). For silicon at room temperature $n_i = 10^{10}$, so we assume that $np = n_i^2 = 10^{20}$.

The idealized distributions shown in Figs. 1.8(*b*), (*c*) are modified in practice because the majority carriers diffuse across the junction. As the holes move into the *n*-type material they very quickly recombine with the free electrons and this results in a *reduction* of the majority carriers to the right of the junction; similarly, as the electrons diffuse into the *p*-type material recombination takes place reducing the majority carriers to the left of the junction. Thus the carrier distribution has the form shown in (*d*), (*e*) of Fig. 1.8; a finite *depletion* zone is created in the vicinity of the junction.

Following the above discussion we sketch our model junction as in Fig. 1.9(*a*) where we indicate the holes by open circles and the electrons by dots; the junction is at $x = 0$ and the boundaries of the depletion region are labeled by $-x_p$ and x_n. For $x < 0$ there exists an excess of ionized acceptors, that is an excess of negative charge. For $x > 0$ there exists an excess of ionized donors, that is an excess of positive charge. Thus the

Fig. 1.9. The electrostatic parameters in the vicinity of a junction: (*a*) definition of the depletion region, (*b*) electric charge density, (*c*) electric field, (*d*) electrical potential.



charge density $\rho(x)$ is distributed as shown in Fig. 1.9(*b*). A non-zero charge distribution gives rise to an electric field, which in its simplest form, is directed from the positive to the negative charge. Thus the field is negative and as shown in Fig. 1.9(*c*). Finally by integrating the electric field we can find the potential in the vicinity of the junction; this is indicated in (*d*) of the figure.

Clearly, the electric field 'pushes' the electrons towards positive x, and the holes toward negative x; that is, *against* the direction in which the carriers tend to diffuse. The electric field can be calculated by integrating Gauss' law

$$\mathscr{E} = \frac{1}{K_s \varepsilon_0} \int_{-\infty}^{x} \rho(x)\, dx \tag{1.18}$$

Here K_s is the dielectric constant of silicon; $K_s \simeq 11.8$. In our example peak field is reached at $x = 0$ and $\mathscr{E}(x = 0)$ is negative. Similarly, the potential $V(x)$ is given by integrating the electric field

$$V(x) = -\int_{-\infty}^{x} \mathscr{E}(x)\, dx \tag{1.18'}$$

The difference in potential across the junction is designated by V_{bi} (where bi stands for 'built-in') as shown in Fig. 1.9(*d*).

We can evaluate V_{bi} by noting that under equilibrium conditions both the electron current and the hole current across the junction must be zero. The total current is the sum of the drift current J_{dr} and the diffusion current J_D. Looking at the current we have, using Eqs. (1.15, 1.17″)

$$J_{dr/n} + J_{D/n} = qn\mu\mathscr{E} - \mu kT \frac{dn}{dx} = 0$$

or

$$V_{bi} = \int_{-\infty}^{+\infty} \mathscr{E}\, dx = \frac{kT}{q} \int_{-\infty}^{+\infty} \frac{dn}{n} = \frac{kT}{q} \ln\left(\frac{n_\infty}{n_{-\infty}}\right)$$

The electron densities at large positive x and large negative x can be taken as $n_\infty = N_D$ and $n_{-\infty} = n_i^2/N_A$ so that

$$V_{bi} = \frac{kT}{q} \ln\left[\frac{N_D N_A}{n_i^2}\right] \tag{1.18″}$$

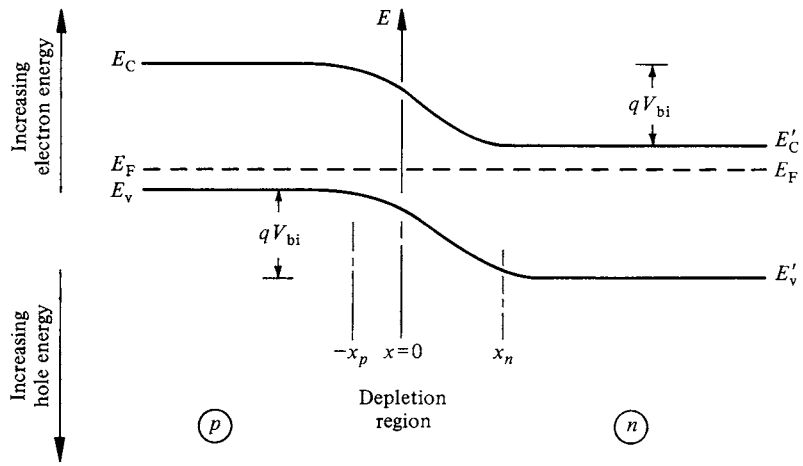
For the densities used in Fig. 1.8 we find $V_{bi} = 0.65$ eV which is typical of most commercial junctions. The typical thickness of the depletion region is of the order of 1 micron (10^{-6} m) or less.

A most convenient way for looking at the potentials and the carrier motion at a junction is to consider the energy band diagram. This is shown in Fig. 1.10. As we recall the position of the Fermi level with respect

to the edge of the valence or the conduction band is different for *p*-type and *n*-type materials (see Fig. 1.6). However when the two materials are joined, the Fermi levels *must be* at the *same* energy when the system is in *equilibrium*; otherwise there would be flow of charge until the Fermi levels equalized.* Thus the band diagram takes the form shown in Fig. 1.10, the relative displacement of the bands being given by qV_{bi}; (here negative potential is toward the top of the page in contrast to Fig. 1.9). The importance of this diagram is that the electrons must gain energy to move upwards, thus their motion from the *n*-region to the *p*-region is impeded. Similarly the holes must gain energy to move downwards (uphill for the holes is down), thus their motion from the *p*-region to the *n*-region is impeded. The depletion region is characterized by the sloping part of the band edges and in fact the electric field is proportional to that slope.

When an external voltage is applied between the ends of the *p*-type and *n*-type material, the potential difference will appear across the junction and modify the energy diagram by *displacing* the relative position of the Fermi levels. We say that the junction is *biased*. There are two possibilities: if a positive voltage is applied to the *n*-type material the potential difference across the junction will increase and there can be no current flow across the junction as shown in Fig. 1.11(*a*). The junction is *reverse* biased. If

Fig. 1.10. Energy band diagram for a *p*–*n* junction; at thermal equilibrium the Fermi level must be at the same energy in both parts of the junction.



* As an analogy one can think of two containers of fluid which are filled to different heights; when the containers are put in communication the water will flow until the levels equalize.

the negative voltage is applied to the *n*-type material the potential across the junction is reduced and there is flow of electrons toward the left end of holes toward the right. When the carriers cross the junction they recombine but the flow is sustained because the potential drives the majority carriers on both sides toward the junction. In this configuration the diode is *forward biased*, as shown in Fig. 1.11(*b*). In the case of a forward biased junction there are enough carriers lying sufficiently high in the conduction band to have energy $E' > E_C$; these carriers drift across the junction under the influence of the external potential. As can be deduced from Eq. (1.18″) for a typical *n*–*p* junction a bias of 0.5–1.0 volt is sufficient to reach saturation. We remind the reader that the current is carried by the majority carriers in each part of the bulk material, i.e. by holes in the *p*-region and by electrons in the *n*-region.

The simple *p*–*n* junction such as described here forms a very useful device widely used in electrical circuits. It is referred to as a *diode* and represented by the symbol shown in Fig. 1.12(*a*). For positive voltage the junction is forward biased and the current flow grows exponentially until it reaches saturation. The current v. voltage (*I*–*V*) characteristic of a typical diode is shown in Fig. 1.12(*b*); the *non-linear* nature of the diode is clearly exhibited. The *I*–*V* curve can be described analytically by an

Fig. 1.11. Energy band diagram for a biased *p*–*n* junction: (*a*) reverse bias, (*b*) forward bias. The physical connection to the voltage source is indicated in the sketches.

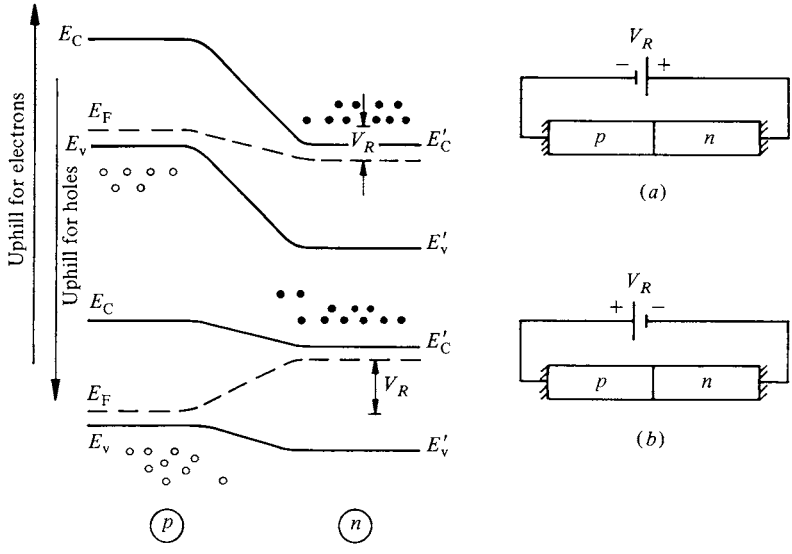
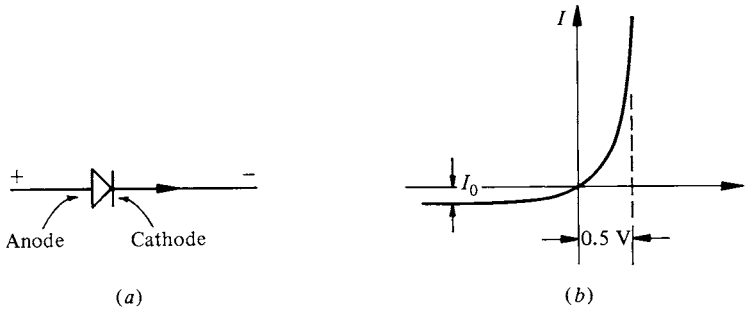


Fig. 1.12. A *p–n* junction forms a diode: (*a*) circuit symbol, (*b*) *I–V* characteristic.



equation of the form

$$I = I_0(e^{qV/kT} - 1)$$

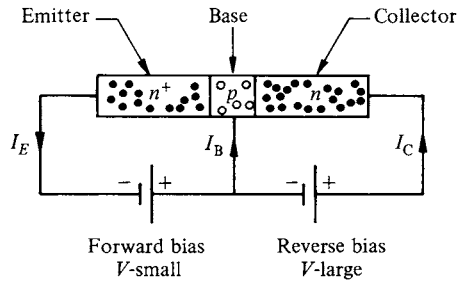
where V is the biasing voltage.

1,5 The junction transistor

The junction transistor consists of two *p–n* junctions connected back to back with the common region between the two junctions made very thin. A model of the *n–p–n* transistor is shown in Fig. 1.13. Note that one junction is forward biased at a relatively low voltage, whereas the other junction is reverse biased at a considerable voltage. The three distinct regions of differently doped material are labeled *emitter*, *base* and *collector* respectively.

According to the biasing shown in the figure, electrons will flow from the emitter into the base; one would expect a large positive current I_B

Fig. 1.13. A *n–p–n* transistor consists of two back to back diode junctions. The biasing scheme and current flow are indicated.

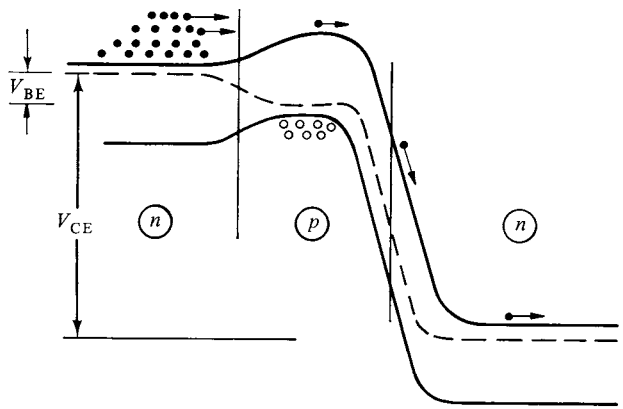


from the base to the emitter. If however the base is thin enough, the electrons injected into the base will reach the base–collector junction before recombining or diffusing in the base. Once electrons cross the base–collector junction they can move freely in the collector since they are majority carriers. Furthermore, the base–collector voltage difference is large so that the electrons gain much more energy than they lost in overcoming the voltage difference between the collector and base. Such a system can provide power amplification. Thus I_B is a small current, while I_E and I_C are much larger when the transistor is in the conducting state.

The energy band diagram for the *n–p–n* transistor is shown in Fig. 1.14. The majority carriers are electrons and therefore once they reach the collector they fall through the potential hill. The *small* voltage between base and emitter can be used to control the flow of current across the much *larger* base–collector voltage. For the device to operate in this fashion the electrons injected from the emitter must traverse the junction without attenuation. In a good transistor ~ 0.95 to 0.99 of the injected carriers traverse the base. Typical widths for the base are of order $W_B \sim 1$–$5\ \mu$m. The emitter is heavily doped with donors so as to be able to provide the necessary current even with small base emitter bias. Junction transistors are referred to also as 'bipolar' transistors to distinguish them from field effect devices.

The symbols for a transistor are shown in Fig. 1.15. The arrows indicate the direction of *positive* current flow so that on the left of the figure we recognize an *n–p–n* transistor (as in Fig. 1.13) and a *p–n–p* transistor on

Fig. 1.14. Energy band diagram for a biased *n–p–n* transistor. For electrons, positive energy is toward the top of the page (uphill); thus electrons flow in the direction indicated.



the right. Typical values are $V_{BE} \sim 0.2$ V whereas $V_{CE} \sim 5$–10 V. The performance of a transistor can be characterized by the *current transfer ratio* α, which is defined as

$$\alpha = \frac{\Delta I_C}{\Delta I_E} \tag{1.19}$$

Here we use Δ to indicate changes in current rather than steady state currents that may result from any particular biasing arrangement. Clearly $\alpha < 1$, but for a good transistor α must be close to one.

To calculate the *current gain* of a transistor, we recall that the emitter current is the sum of the collector and base currents.

$$\Delta I_E = \Delta I_C + \Delta I_B \tag{1.19'}$$

The current gain β is defined by

$$\beta = \Delta I_C / \Delta I_B \tag{1.20}$$

Using this definition and that of Eq. (1.19) in Eq. (1.19′) we obtain

$$\frac{1}{\alpha} \Delta I_C = \beta \Delta I_B + \Delta I_B = (1 + \beta) \Delta I_B$$

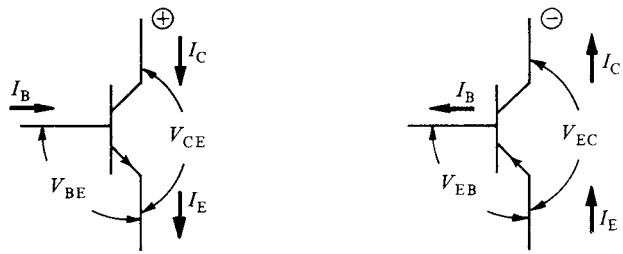
or

$$\beta = \alpha(1 + \beta), \qquad \beta = \frac{\alpha}{1 - \alpha} \tag{1.20'}$$

For a typical value of $\alpha \simeq 0.98$ one finds $\beta \simeq 49$. Namely, we can use a small current into the base of the transistor to control the flow of a much larger current from the collector to the emitter. Thus a transistor is a device that controls current flow.

The transistor is a three-terminal device and thus there are more biasing possibilities than for a diode. There are three basic biasing configurations which can be classified as: common or *grounded base*; common emitter;

Fig. 1.15. Circuit symbol for a junction transistor; the current flow and voltage definitions are indicated: (*a*) *n–p–n*, (*b*) *p–n–p*.



(*a*) *n-p-n* (*b*) *p-n-p*