Assuming a drop of 0.4 V across the diode when it conducts, a base–emitter voltage of 0.7 V and $h_{fe} = 100$ (see Sec. 8.3), calculate the base and collector currents and also the current in each resistor.

Assuming an operating temperature of 300 K, determine the saturation current of the diode.

*Ans.* 0.028 mA, 2.75 mA, 0.43 mA, 2.35 mA, 78 pA

**26.** A hypothetical device consists of a p-type semiconductor, resistivity 0.1 Ω m, diameter 100 $\mu$m, thickness 2 $\mu$m and workfunction 1 eV, sandwiched between two different metal contacts, $M_1$ with workfunction 1.4 eV and $M_2$ with workfunction 0.6 eV. Explain in general terms how the structure would behave electrically.

When a voltage of 0.5 V is connected between $M_1$ and $M_2$ with $M_2$ positive, 10 nA flows. Estimate the current flowing with the polarity reversed, assuming operation at 290 K, explaining the reasoning behind the calculation. Neglect all surface-charge effects.

*Ans.* 6.5 mA

# 8 The bipolar junction transistor

## 8.1 Introduction

The physical processes that determine the electrical behaviour of bipolar junction transistors (BJT) will be discussed in this chapter. It will become evident that the operation of transistors in this class depends on the interaction of both majority and minority carriers, and because two carrier types are essential such devices are classified as *bipolar*. This term serves to differentiate the devices discussed here from another main class, unipolar transistors, in which the current is transported by majority carriers only; these will be discussed later.

## 8.2 Phenomenological description of current transport in the abrupt-junction bipolar transistor

The bipolar junction transistor consists essentially of a single-crystal semiconductor, most often of silicon or germanium, which contains a narrow central region of opposite conductivity type to that of the rest of the material. For example, an npn transistor contains a narrow p-type layer sandwiched between two n-type layers with ohmic contacts made to each region, as shown diagrammatically in Fig. 8.1(a). Now, the physical construction of a modern

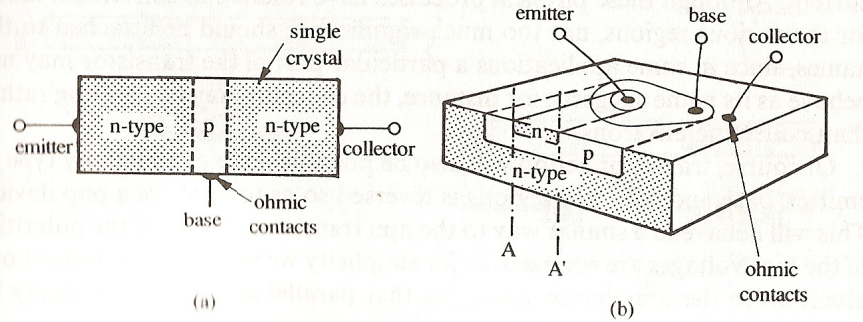



**Fig. 8.1 (a) Model of an npn transistor; (b) a planar transistor structure.**

transistor, which will be discussed in more detail later, may well be very different in appearance from the simple model shown; for example, a planar version of a discrete transistor structure may be as shown in Fig. 8.1(b). However, it will be seen that sections at A, A′ encompass an npn sandwich not unlike the model and for the moment at least the simple model will be sufficient to help understand the basic transistor action.

The transistor thus consists of two back-to-back pn junctions closely coupled electrically by a narrow region of material common to both. In normal operation, one of the junctions is forward-biased and the other reverse-biased, as shown in Fig. 8.2. Briefly, what happens is that minority

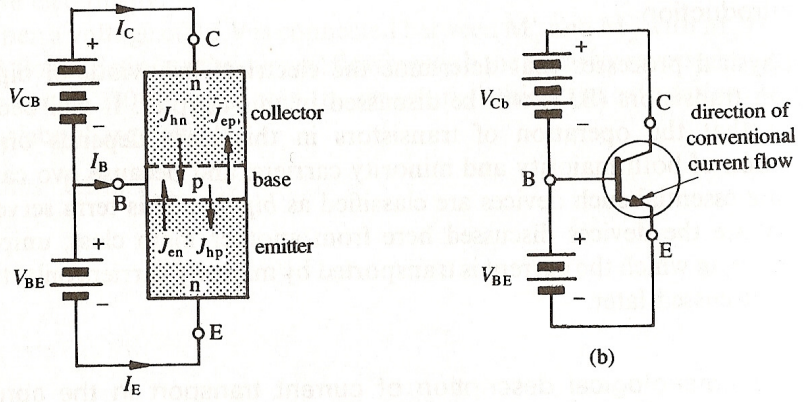



**Fig. 8.2 (a) Normal biasing arrangement for an npn transistor; (b) circuit element representation.**

carriers are injected from the *emitter* into the *base* region as a consequence of the forward bias appearing across the emitter–base junction and, because the base is deliberately made very thin, nearly all the injected carriers reach the reverse-biased base–collector junction, eventually determining the *collector* current. Although these physical processes have resulted in convenient labels for the various regions, not too much significance should be attached to the names, since in some applications a particular part of the transistor may not behave as its name suggests; for instance, the collector may be emitting rather than collecting electrons.

Of course, transistor action will also be possible if the conductivity type of emitter, base and collector regions is reversed so as to produce a pnp device. This will behave in a similar way to the npn transistor, provided the polarities of the bias voltages are reversed, so for simplicity we shall initially restrict our attention to the npn device, assuming that parallel arguments can easily be developed for the complementary device.

The npn transistor is usually preferred in Si since the mobility and diffusion



coefficient for minority electrons in the base are higher than for the holes in the base of a pnp device, which, as we shall see, leads to a higher operating speed. Typically, pnp transistors have a turn-over frequency at which the gain falls to unity of around 1.5–2 GHz. The corresponding maximum operating frequency for an npn transistor approaches 10 GHz. There are also technological reasons for the preferential choice of the npn configuration. For example, the highly doped buried layer beneath collectors in planar epitaxial transistors used in ICs, which is necessary, as will become apparent, for low-resistance access to the collector, must be produced using a dopant that does not diffuse further during subsequent processing stages. Arsenic is an ideal material for such layers, creating an $n^+$ buried layer, which dictates that collectors must be n-type, in an overall npn device.

In most modern transistors, transport of carriers across the base is by diffusion and drift in the presence of density gradients of carriers and of an electric field. It will be convenient to consider these two processes independently so that initially the base region will be considered field-free, thus ensuring that transport of minority carriers across it is exclusively by diffusion. It will also be simpler at first to assume that the junctions are abrupt by virtue of a doping profile that changes rapidly with distance, as discussed in Chapter 7.

The band structure of an unbiased, symmetrically doped npn structure in equilibrium will be as shown in Fig. 8.3(a). The relative widths of base and depletion regions are much exaggerated for clarity. As a result of the necessary

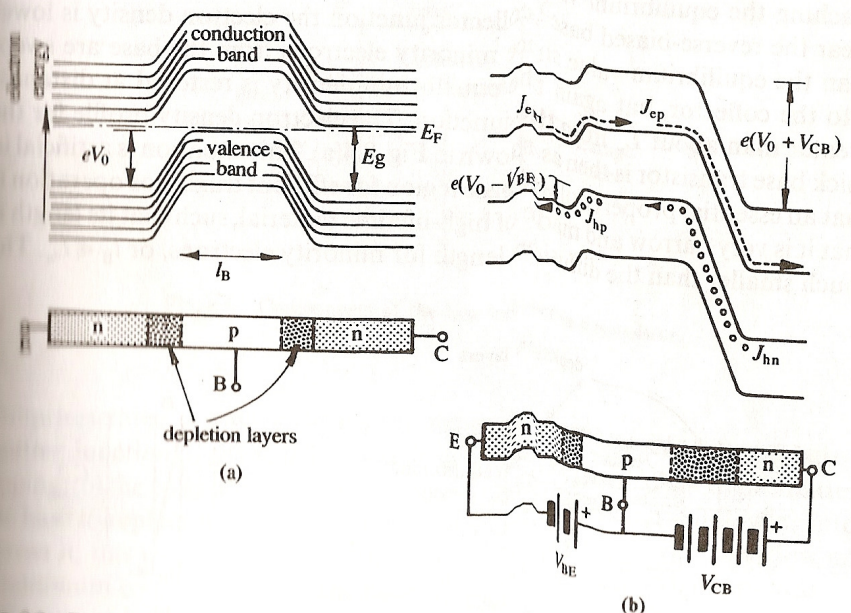
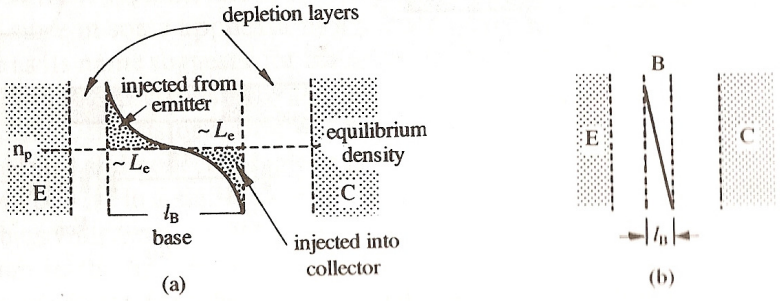



**Fig. 8.3 Band structure of an npn transistor: (a) with no bias applied and (b) with external bias voltages.**

continuity in the Fermi level through the transistor in equilibrium, contact potentials and potential barriers are established at the emitter–base and base–collector junctions which prevent diffusion of majority carriers across the junctions, as described earlier. When the correct d.c. bias voltages are applied the band structure is modified to that shown in Fig. 8.3(b). The potential barrier between emitter and base is reduced by virtue of the forward bias to $e(V_0 - V_{BE})$ and an electron current is injected from emitter in to base region, $J_{en}$, and a hole current in the reverse direction, $J_{hp}$. We shall see later that efficient transistor operation is achieved if most of the current across this junction is carried by carriers originating in the emitter, i.e. $J_{en} \gg J_{hp}$. This is achieved by doping the emitter to a higher degree than the base. The reverse bias at the collector–base junction causes an increase in barrier height to $e(V_0 + V_{CB})$; as a consequence, there is no majority-carrier diffusion and the only current flow across the junction is due to the motion of minority carriers. Thus a very small saturation current, $I_{co}$, called the collector leakage current, flows across the collector–base junction, which has contributions $J_{ep}$ and $J_{hn}$ from minority carriers in the base and collector regions.

Turning our attention now to the base region, let us first assume that its width $l_B$ is longer than the diffusion length for minority carriers $L_e$ and $L_h$. The density of minority electrons in the p-type base will be increased above the equilibrium value near the emitter–base junction because of the presence there of electrons injected from the base. This excess of electrons decays exponentially away from the junction because of recombination with majority holes, reaching the equilibrium value at a distance of order $L_e$ from the junction. Near the reverse-biased base–collector junction the electron density is lower than the equilibrium value since minority electrons from the base are swept into the collector, but again the equilibrium density is restored at distances greater than about $L_e$ from the junction. The electron density profile for the thick base transistor is then as shown in Fig. 8.4(a). This situation is artificial in that an essential property of the base region for efficient transistor operation is that it is very narrow and made of high-lifetime material, such that its length is much smaller than the diffusion length for minority electrons, or $l_B \ll L_e$. The






minority-carrier concentration in the base region is then as shown in Fig. 8.4(b). Under these circumstances there is little recombination in the base region and nearly all the electrons injected from the emitter diffuse across the base because of the steep electron density gradient existing in it and are eventually swept across the base–collector junction down the potential hill and into the collector. If, as is usually the case, the emitter is more heavily doped than the base, so that the majority of the current across the emitter–base junction is transported by electrons, then the collector current, $I_C$, is only slightly less than the emitter current, $I_E$.

It will be noticed that the electron current flowing into the collector is largely independent of $V_{CB}$, provided this is large enough to prevent majority-carrier diffusion across the junction; as a consequence, the collector circuit has a high impedance. This is in contrast to the emitter circuit, which has a very low impedance because small voltage changes at the emitter–base junction cause large changes in the current flowing across it. Since the collector current is nearly equal to the emitter current, as explained, the large difference in impedance level between collector and emitter circuits can result in potentially high power amplification. Incidentally, this description of the action of the device in terms of the transfer of current from a low- to a high-impedance circuit accounts for its original name, transfer resistor, which was subsequently contracted to transistor.

If we now consider the base current, $I_B$, it can be seen from Fig. 8.5 that it

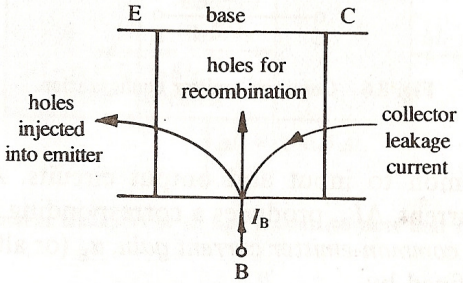



**Fig. 8.5 Components of the base current in a transistor.**

comprises three components: (a) the hole current flowing across the base–emitter junction, $J_{hp}$, which is made as small as possible by high emitter doping; (b) the collector leakage current, $I_{co}$; and (c) a hole current flowing into the base to replace holes that are lost by recombination with electrons flowing across it; this current can be made very small by reducing recombination to a minimum by ensuring that the base is thin and that the lifetime of minority electrons in it is high. Thus the base current is normally much smaller than either collector or emitter currents.

## 8.3 Gain parameters of the bipolar transistor

Let us assume initially that the external circuit of the transistor is as shown in Fig. 8.2; the arrangement is the *common-base* connection, since the base contact is common to both collector and emitter circuits. If the emitter current is changed by increment $\Delta I_E$, there will be a corresponding incremental change in collector current, $\Delta I_C$, and the relative change is described in terms of a gain parameter, which is defined by

$$\frac{\text{change in collector current}}{\text{change in emitter current}} = \left(\frac{-\Delta I_C}{\Delta I_E}\right)\Bigg|_{V_{CB}\text{const.}} = \alpha_B \qquad (8.1)$$

where $\alpha_B$ is the *common-base current gain.* Since by previous arguments a change in emitter current results in a change in collector current that is only marginally smaller, $\Delta I_C \simeq \Delta I_E$ and $\alpha_B$ is only slightly less than unity, typical values being in the range 0.900–0.999.

Although it has been convenient to discuss the common-base operation of a transistor first, it is far more usual for the device to be operated in the *common-emitter* circuit configuration shown in Fig. 8.6, in which the emitter

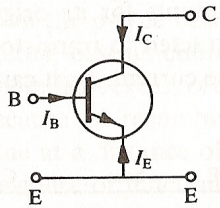



**Fig. 8.6 Common-emitter configuration.**

connection is common to input and output circuits. Arguing as before, a change in base current, $\Delta I_B$, produces a corresponding change in collector current, $\Delta I_C$, and a *common-emitter current gain,* $\alpha_E$ (or alternatively $h_{fe}$ – see Sec. 8.5), is then defined by

$$\alpha_E = \left(\frac{\Delta I_C}{\Delta I_B}\right)\Bigg|_{V_{CE}\text{const.}} \qquad (8.2)$$

Since by Kirchhoff's law

$$I_B + I_C + I_E = 0$$

then

$$\Delta I_B = -(\Delta I_C + \Delta I_E)$$

Using Eq. (8.1) to eliminate $\Delta I_E$ gives

$$\Delta I_B = \Delta I_C[1 - (1/\alpha_B)]$$



and finally

$$\Delta I_C/\Delta I_B = \alpha_E = \alpha_B/(1-\alpha_B) \qquad (8.3)$$

Thus, we arrive at a relationship between the common-base and common-emitter current gains. Since $\alpha_B$ is usually very nearly unity, it is evident from Eq. (8.3) that $\alpha_E$ is much greater than unity, typically being in the range 10–1000.

It is possible to estimate the value of the gain parameter $\alpha_B$, and hence $\alpha_E$, through Eq. (8.3), in terms of the physical processes occurring in the transistor, its structure and its composition. It is evident that $\alpha_B$ will be dependent on (a) the number of electrons injected from the emitter into the base and (b) the proportion of these which diffuse across the base, without recombination, to the collector. It is therefore convenient to subdivide the gain parameter, $\alpha_B$, into two components:

$$\alpha_B = \eta_E \beta \qquad (8.4)$$

where $\eta_E$, the *injection efficiency,* is the ratio of the electron current injected into the base from the emitter to the total emitter–base junction current, and $\beta$, the *base transport factor,* is the ratio of the electron current at the collector junction to that at the emitter junction. Current transport in an npn transistor in terms of these components of $\alpha_B$ is shown schematically in Fig. 8.7. Since the

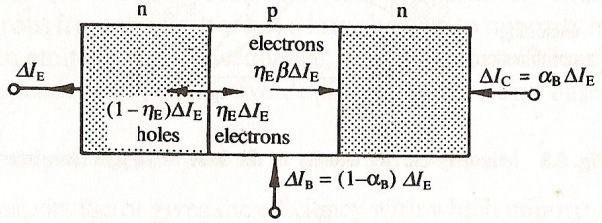



**Fig. 8.7 Current transport in an npn transistor (note that $I_{co}$ is omitted).**

current flowing into the base to replace holes lost by recombination is $\eta_E \Delta I_E(1-\beta)$, and the base current flowing to provide the hole current across the emitter junction is $\Delta I_E(1-\eta_E)$, then, ignoring the collector leakage current, the total base current is

$$\Delta I_B = \eta_E \Delta I_E(1-\beta) + \Delta I_E(1-\eta_E) = (1-\alpha_B)\Delta I_E$$

which is the same as that required to satisfy Kirchhoff's law.

We now consider the components of $\alpha_B$ in more quantitative detail.

### 8.3.1 Emitter injection efficiency, $\eta_E$

From the definition given above and by referring to Fig. 8.2(a) it follows that

the emitter injection efficiency can be written as

$$\eta_E = J_{en}/(J_{en} + J_{hp}) \tag{8.5}$$

It is tempting to use the arguments outlined is Sec. 7.6 and assume that the ratio of electron to hole currents at the junction is approximately equal to the ratio of emitter to base conductivities, but this would apply only if the base and emitter widths were substantially longer than the minority-carrier diffusion lengths, which we know is not usually the case in a transistor. A rigorous derivation of $\eta_E$ would follow the analysis for the pn junction with finite dimensions, which is outlined in Sec. 7.7, but the following approximate approach will be sufficient for our purposes.

Consider first the density variation of minority carriers to the base as illustrated in Fig. 8.8. Electrons injected from the emitter raise the local

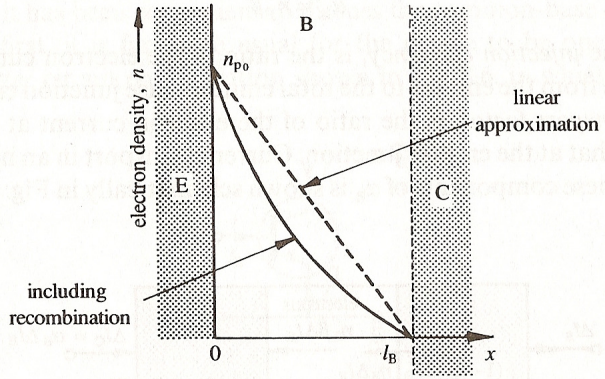



**Fig. 8.8 Minority-carrier density in the base of an npn transistor.**

electron density just inside the base from $n_p$, the equilibrium value, to $n_{p0}$, given by Eq. (7.11) as

$$n_{p0} = n_p \exp(eV_{BE}/kT) \tag{8.6}$$

At the collector end of the base region the electron density is depressed below the equilibrium value by virtue of the reverse bias there, to a value $n_p \exp(-eV_{BC}/kT)$, which can be assumed zero since usually $V_{BC} \gg kT$. Since the length of the base, $l_B \ll L_e$, the diffusion length for minority carriers, there will be little recombination in the base and the minority-carrier density falls off almost linearly with distance. Assuming such a constant density gradient, the electron diffusion current is

$$J_{en} = eD_e n_{p0}/l_B \tag{8.7}$$

On comparing this equation with (7.20) for the electron current flow in a wider n-type region of a pn junction, it will be noticed that the difference between the



two expressions is that $l_B$ in Eq. (8.7) replaces $L_e$ in the earlier equation. It follows that Eq. (7.22) is applicable to the thin-base, thin-emitter junction provided that $L_e$ and $L_h$ are replaced by $l_B$ and $l_E$, respectively. Using this modified equation and applying similar arguments to those used in Sec. 7.6 gives

$$\frac{J_{en}}{J_{hp}} \simeq \frac{D_e n_p}{l_B} \div \frac{D_h p_n}{l_E} \simeq \frac{\sigma_E}{\sigma_B}\frac{l_E}{l_B} \tag{8.8}$$

Substituting this expression in Eq. (8.5) gives the emitter injection efficiency as

$$\eta_E \simeq \left(1 + \frac{\sigma_B}{\sigma_E}\frac{l_B}{l_E}\right)^{-1} \simeq 1 - \frac{\sigma_B l_B}{\sigma_E l_E} \tag{8.9}$$

Thus, the emitter efficiency is largely controlled by the relative doping of base-emitter regions but it is also influenced by the ratio of the lengths of such regions. For example, if the conductivity of the emitter is made, say, 100 times that of the base, then provided the base length is less than the emitter length, which is usually the case, $\eta_E$ is in excess of 99 per cent, and such values can be achieved in practice.

An alternative approach is used in the heterostructure bipolar transistor (HBT) described in Sec. 12.6. Semiconductors having different energy gaps are chosen to form the emitter and base such that the resulting heterojunction, as will be seen, contains band offsets that create a relatively small barrier to injected electrons from the emitter but a large barrier to minority holes flowing from base into emitter. As a consequence, a high emitter efficiency is ensured, which is independent of the relative doping of emitter and base.

### 8.3.2 Base transport factor, $\beta$

The base transport factor gives the efficiency with which minority carriers are transported across the base region, and for the particular case of an npn transistor

$$\beta = \frac{J_e|_{\text{collector junction}}}{J_e|_{\text{emitter junction}}} \tag{8.10}$$

Since in this expression the numerator is only slightly less than the denominator because of the usually slight recombination that takes place in the base, a more accurate method for determining $J_e$ is required than that described in the previous section; the linear approximation to the electron density profile in the base region is no longer good enough to calculate the electron current, and the lower curve of Fig. 8.8 is now applicable. The shape of this curve can be found by applying the continuity equation for excess electrons in the base region, $\delta n$, which gives

$$d^2(\delta n)/dx^2 = \delta n/L_e^2$$

which, as we have seen, has the general solution

$$\delta n = C_1 \exp(-x/L_e) + C_2 \exp(+x/L_e) \tag{8.11}$$

The constants $C_1$ and $C_2$ can be determined from the boundary conditions, which are

$$\delta n = n_{p0} - n_p \qquad \text{at } x = 0$$

and

$$\delta n = -n_p \simeq 0 \qquad \text{at } x = l_B, \text{ since } n_p \ll n_{p0}$$

Applying these boundary conditions to Eq. (8.11) gives

$$\delta n = \frac{n_{p0} - n_p}{1 - \exp(2l_B/L_e)} [\exp(x/L_e) - \exp(2l_B/L_e)\exp(-x/L_e)] \tag{8.12}$$

Now, since the electron diffusion current is proportional to the gradient of the electron density given by Eq. (8.12), performing the differentiation and substituting in Eq. (8.10) gives

$$\beta = \frac{J_e|_{x=l_B}}{J_e|_{x=0}} = \frac{[\mathrm{d}(\delta n)/\mathrm{d}x]|_{x=l_B}}{[\mathrm{d}(\delta n)/\mathrm{d}x]|_{x=0}} = \frac{2\exp(l_B/L_e)}{1+\exp(2l_B/L_e)}$$

Since $l_B \ll L_e$ usually, this expression simplifies to

$$\beta = [1 + \tfrac{1}{2}(l_B/L_e)^2]^{-1} \simeq 1 - \tfrac{1}{2}(l_B/L_e)^2 \tag{8.13}$$

Thus, for high base transport factors giving current gains as near to unity as possible, the lifetime of minority electrons in the base region must be high so as to make $L_e$ large. An advantage of Si or Ge in this respect is that relatively high lifetimes in the range 1–100 $\mu$s are realizable; for base widths of 1–5 $\mu$m, which are readily fabricated using modern technology, transport factors in the range 0.95 to in excess of 0.99 can result.

## 8.4 Non-ideal transistor structures

Although the model used so far is satisfactory for explaining basic transistor mechanisms, it needs considerable modification before it is applicable to physically realizable transistors. Some of the additional effects that occur in real transistors and the manner in which these limit their electrical performance will now be discussed.

### 8.4.1 Avalanche breakdown and multiplication

An upper limit is set on the collector voltage $V_{CB}$ by avalanche breakdown in the reverse-biased collector–base junction, as discussed previously in Sec. 7.12. Fields of order $10^6\,\mathrm{V\,m^{-1}}$ are required in the depletion layer for breakdown to occur, which usually limits $V_{CB}$ to a maximum of several tens of volts.



For collector voltages lower than that necessary for the onset of avalanche breakdown, $V_{BD}$, there is a voltage range in which minority electrons in the base–collector depletion layer are accelerated sufficiently to cause electron-hole pair production by ionizing collisions with the lattice, but the holes produced by the process are not sufficiently energetic when accelerated in the field to produce the secondary ionization that is essential to maintain a self-sustained breakdown. In this voltage range, although no complete breakdown occurs, there is some electron multiplication and the number of electrons collected is greater than the number arriving at the base–collector layer edge. Under these conditions the current gain, $\alpha_B$, is increased by a factor, $\eta_c$, called the *collector efficiency*, which has been found to be given empirically by

$$\eta_c = (1 - V_{CB}/V_{BD})^{-n} \tag{8.14}$$

where $n$ is usually in the range 2–4. It is thus possible for the effective current gain, which includes this additional factor due to avalanche multiplication, to exceed unity; the emitter circuit may then display negative-resistance effects, which can lead to undesirable instabilities.

### 8.4.2 Base-width modulation and punch-through

It has been tacitly assumed that the base width, $l_B$, is always constant, but in a real device the effective width of the base, $l_{B,eff}$, is dependent on $V_{CB}$ and to a lesser extent on $V_{BE}$. This is because the boundaries that delineate the extent of the base are the edges of the depletion layers at the junction, which vary in position with changing bias voltages. For example, the assumed almost-zero minority-carrier concentration occurs at the edge of the collector–base depletion layer, which in turn is dependent on the collector voltage $V_{CB}$. If the base is relatively lightly doped, the collector–base depletion layer extends mostly into the base and the effective base width, using Eq. (7.49), is given approximately by

$$l_{B,eff} \simeq l_B - \left(\frac{2\varepsilon V_{BC}}{eN_a}\right)^{1/2} \tag{8.15}$$

where $l_B$ is now the distance between the metallurgical junctions and $N_a$ is the acceptor concentration in the base. Therefore, as $V_{CB}$ is increased, the effective base width is reduced; there is a corresponding increase in the slope of the minority-carrier density profile as shown in Fig. 8.9, which leads to a higher collector current. It follows that the emitter efficiency, the base transport factor and current gains $\alpha_B$ and $\alpha_E$ are dependent in a non-linear way on the voltage $V_{CB}$.

The variation of collector current with changing $V_{CB}$, an effect known as *base-width modulation*, also affects the output characteristics of a transistor since the $I_C$–$V_{BC}$ curves are no longer horizontal but take on a positive slope

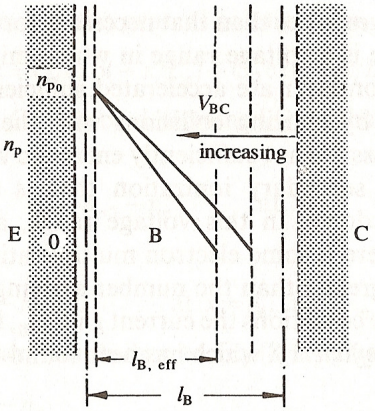



**Fig. 8.9 Variation of effective base width with collector voltage $V_{BC}$.**

indicating that the device has a finite output impedance that is voltage-dependent. The input characteristics are also affected; since the input circuit behaves as a forward-biased diode and the input current is given by

$$I \simeq I_0 \exp(eV_{BE}/kT)$$

where $I_0$ is dependent on the effective base width, the input impedance is then to some extent influenced by $V_{CB}$.

At high collector voltages or for low doping concentrations in the base, it is possible for the collector–base depletion layer to extend completely across the base region, thus effectively short-circuiting the collector to the emitter. The minimum voltage for this *punch-through* effect to occur can be obtained by letting $l_{B,eff}$ go to zero in Eq. (8.15), which gives

$$V_{BC}|_{max} = eN_a l_B^2/2\epsilon \qquad (8.16)$$

Punch-through may thus set an upper limit to the permissible collector voltage but in many transistors the maximum value of $V_{BC}$ is set by the onset of avalanche breakdown, which often occurs at lower voltages. It is of course possible to raise the punch-through voltage by increasing the doping concentration in the base, but since this automatically reduces the emitter efficiency, $\eta_E$, the particular choice of $N_a$ is an engineering compromise.

### 8.4.3 Base resistance

It has been assumed in the simple transistor model that all externally applied voltages appear across the relatively high-resistance depletion layers and that voltages dropped across the bulk semiconductor regions are negligible. While it is usually permissible to ignore the resistance of collector and emitter regions in this way, the base region is relatively lightly doped in modern transistors and effects due to its finite resistance must normally be taken into account.



Consider, for example, the planar transistor shown in Fig. 8.1(b) and in section in Fig. 8.10. Base current, $I_B$, flowing to the active base region, passes through a region that can have a significantly high resistance, which is represented by a lumped resistance, $r_B$, in the diagram. As a consequence, the effective

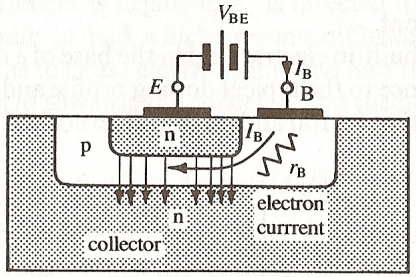



**Fig. 8.10 Base resistance and emission crowding in a real transistor.**

emitter–base voltage is given by

$$V_{BE,eff} = V_{BE} - I_B r_B \qquad (8.17)$$

Unfortunately, the situation is more complicated than is suggested by Eq. (8.17), since $I_B$ itself is dependent on $V_{BE,eff}$.

It should be noted that the arguments so far are also applicable to a.c. signals; indeed, whereas it is often possible to omit $r_B$ when calculating biasing conditions, it is usually necessary to include an effective base resistance in the small-signal equivalent circuit.

A more serious consequence of a finite base resistance is that it causes *emission crowding* in the base. This arises because the base current, which is moving laterally in the active base region, i.e. perpendicular to the minority electron flow, causes a voltage drop across the face of the emitter that is in such a direction as to reduce the effective forward bias voltage in the centre of the emitter relative to that at the edges. Electron current from the emitter thus tends to concentrate towards the periphery of the emitter, as shown diagrammatically in Fig. 8.10. Under high-current conditions, irreversible damage can be caused by excessive current densities at emitter edges. At more modest current levels, emission crowding is not so serious and can be accounted for by inclusion of an additional resistance in the effective lumped base resistance.

### 8.4.4 Graded-base or drift transistors

The transistors discussed so far have been assumed to possess a uniformly doped base in which negligible electric fields exist; minority electron transport in this case is predominantly determined by diffusion effects. However, many modern transistors, particularly those made by the diffusion process, have an

inherently non-uniform base doping profile. For many applications this is advantageous and in some devices, for example, high-frequency and switching transistors, such a doping profile is introduced deliberately. Devices with a non-uniform impurity concentration in the base are known as minority-carrier *graded-base* or *drift* transistors; the latter name arises since an electric field always exists in the base which causes the minority-carrier current in it to have a drift component.

The origins of the built-in electric field in the base of a drift transistor can be explained with reference to the typical doping profile and band structure of an npn device shown in Figs 8.11(a) and (b). It will be noticed that the net acceptor

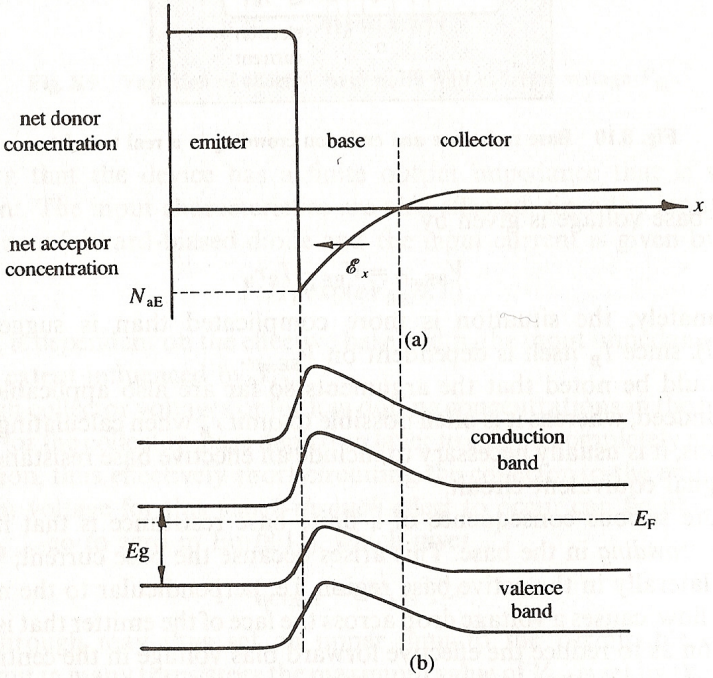



**Fig. 8.11 (a) Idealized doping profile and (b) equilibrium band structure of a drift transistor.**

level in the base has its largest value, $N_{aE}$, next to the emitter and falls to zero at the base–collector junction. The corresponding negative gradient in majority hole density causes holes to move towards the collector, thus exposing fixed ionized acceptors and creating negative space charge. The resulting electric field is in such a direction as to oppose further migration of holes and so maintains the net hole current due to drift and diffusion at zero. The base current due to majority holes will, in the absence of recombination, be given by Eq. (6.52):

$$J_h = e\mu_h p\mathscr{E}_x - eD_h \, dp/dx \tag{8.18}$$

For low-level injection, the hole density at any point in the base, $p$, is



approximately equal to the acceptor concentration there, so, for zero hole current, Eq. (8.18) yields

$$\mathscr{E}_x = \frac{D_h}{\mu_h N_a}\frac{dN_a(x)}{dx} = \frac{kT}{e}\frac{1}{N_a}\frac{dN_a}{dx} \tag{8.19}$$

Since the impurity gradient is negative, $\mathscr{E}_x$ is directed towards the emitter as expected. Now, this built-in field, which prevents diffusion of majority carriers, is in such a direction as to cause electrons injected into the base to drift under the influence of the field towards the collector, and the net electron current in the base is given by

$$J_e = ne\mu_e\mathscr{E}_x + eD_e dn/dx \tag{8.20}$$

Further, since carriers usually drift with faster velocities than they diffuse, the transit time of electrons across the base is very much shorter in a drift transistor than in a corresponding diffusion transistor in which $\mathscr{E}_x$ is assumed to be zero. This property has enabled drift transistors to be operated at frequencies in the gigahertz range.

It is desirable for some applications that the doping gradient be arranged such that the built-in field in the base is everywhere uniform. If the transistor is constructed so that the doping profile is exponential and of the form

$$N_a(x) = N_{aE}\exp(-Cx)$$

then Eq. (8.19) gives

$$\mathscr{E}_x = \frac{kT}{e}\frac{1}{N_a}(-C)N_a = -\frac{kTC}{e}$$

which satisfies the condition that $\mathscr{E}_x$ is constant and independent of position.

### 8.4.5 Geometrical effects

The flow of minority carriers through the base of a transistor has been assumed to be one-dimensional. Of course, in real transistors this is not so and the minority-carrier current flow from the emitter spreads out laterally to some extent in the base, before it arrives at the collector. The proportion of the current from the emitter that is collected thus depends in some part on the geometry of the transistor. For example, if the base region is thin and the area of the collector is made much greater than the area of the emitter, then the base transport factor approximates to the ideal value found for the one-dimensional model.

### 8.4.6 Transistors in the switching mode

So far, the bipolar transistor has been considered in its near-linear amplifying mode, but it is often used as an ON/OFF device in switching circuits, this being the dominant regime for the majority of integrated circuits.

To study switching behaviour, consider the drain characteristics of a bipolar transistor in the common-emitter configuration, shown in Fig. 8.12. The intersection of the load line corresponding to the load resistor, $R_L$, with a particular drain characteristic determines the collector voltage, $V_{CE}$, for a given base current, $i_B$. The amplifying region for a common-emitter amplifier

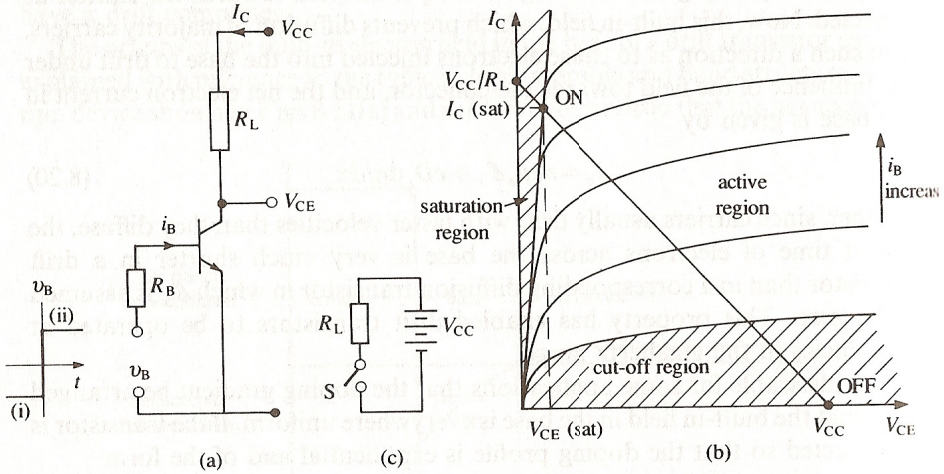



**Fig. 8.12 (a) An npn transistor in common-emitter mode with a step voltage applied to the base; (b) collector characteristics; and (c) the ideal equivalent circuit for switching.**

in which the output current responds almost linearly to changes in input current is shown as the active region (unhatched). However, if the base current is made zero or negative, by making the base voltage, $v_B$, zero or negative, region (i) in Fig. 8.12(a), thereby reverse biasing both transistor junctions, the operating point for the transistor enters the *cut-off* region (shown hatched). In such an OFF state, at the bottom end of the load line, the collector current becomes almost zero and the collector voltage almost equals $V_{CC}$, the collector supply voltage. The transistor is virtually open-circuit between collector and emitter, corresponding to switch, S, in Fig. 8.12(c) being open.

If the base current is subsequently driven large and positive, for example by a positive pulse change in $v_B$ as shown in (ii) Fig. 8.1(a), the transistor switches into the *saturation region*, again shown hatched, via the active region, which is traversed at a rate that is dependent on factors such as gain and frequency response. In this ON condition, large collector currents flow and the collector voltage falls to a very low value, called $V_{CE}$(sat), typically around 0.2 V for a silicon transistor, most of the supply voltage now being dropped across $R_L$. The transistor is virtually a short-circuit in this state, which almost duplicates the closed condition of the ideal switch, S, in Fig. 8.12(c).

In the ON state, the collector current, $I_C$(sat), is dependent on the load resistance and is given by

$$I_C(\text{sat}) = [V_{CC} - V_{CE}(\text{sat})]/R_L \qquad (8.21)$$



This corresponds to a base current required to drive the transistor into saturation of

$$i_B \geq I_C(\text{sat})/h_{fe} \qquad (8.22)$$

As $i_B$ is increased above this minimum value, to drive the transistor hard into saturation, the bias across the collector–base junction changes from zero at the outset to become *forward* biased when $i_B$ is bigger. Under these conditions, both junctions are injecting electrons into the base, which accumulates an excess of stored electronic charge. The distribution of minority electrons in the base for the three operating regimes is shown in Fig. 8.13, the excess stored

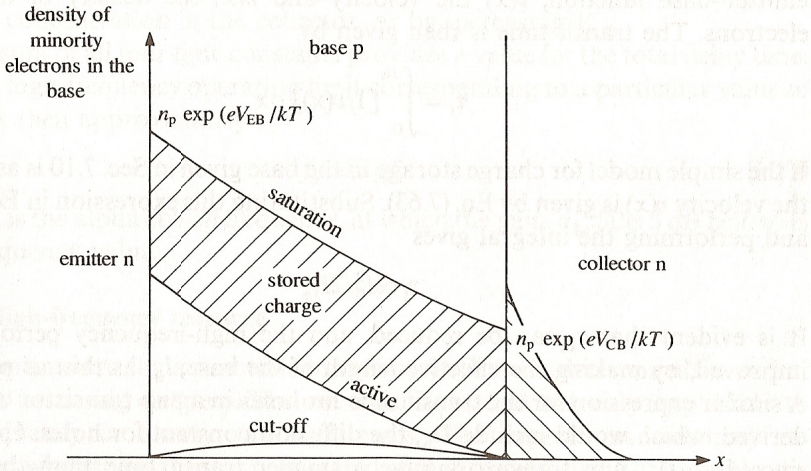



**Fig. 8.13 Minority electron distribution in the base for an npn transistor for cut-off, active and saturated operating conditions.**

electron charge in saturation being denoted by the hatched region. The electron density at the emitter–base junction, from Eq. (8.6), rises from around zero at cut-off to $n_p \exp(eV_{EB}/kT)$ for saturation, as shown, but the corresponding densities at the collector–base junction are $n_p\exp(-eV_{CB}/kT)$, which is again near zero, rising to $n_p\exp(+eV_{CB}/kT)$, corresponding to electron injection from the collector and the change in bias polarity. There will also be stored minority-carrier charge in the collector region of an ON transistor as shown, but this is much smaller and can often be ignored.

As for the pn diode discussed earlier, switching between ON and OFF states has to be accompanied by corresponding changes in the electronic charge stored in the base. For example, if the transistor is hard ON and is then switched into the OFF mode, all the stored charge shown in Fig. 8.13 has to be discharged before the current can adjust to the very small value corresponding to the cut-off condition. Such a process cannot occur instantaneously, so removal or insertion of stored charge out of or into the base leads to finite switching times. High-speed switching circuits are often designed to avoid this

difficulty by arranging that transistors are not allowed to saturate, thus reducing switching times.

### 8.4.7 High-frequency performance of transistors

The minority carriers take a finite time to traverse the base region of a transistor and it is this *transit time*, $\tau_t$, that is usually the major factor limiting high-frequency performance. As the operating frequency approaches $\tau_t^{-1}$, the transistor becomes inoperative.

The transit time for electrons to cross the base region of length $l_B$ in an npn transistor can be estimated as follows. Let $x$ be the distance from the emitter–base junction, $v(x)$ the velocity and $n(x)$ the density of minority electrons. The transit time is then given by

$$\tau_t = \int_0^{l_B} [1/v(x)]\,dx \tag{8.23}$$

If the simple model for charge storage in the base given in Sec. 7.10 is assumed, the velocity $v(x)$ is given by Eq. (7.63). Substituting this expression in Eq. (8.21) and performing the integral gives

$$\tau_t \simeq l_B^2/2D_e \tag{8.24}$$

It is evident that $\tau_t$ can be reduced, and the high-frequency performance improved, by making the effective length of the base, $l_B$, as thin as possible. A similar expression for the transit time for holes in a pnp transistor could be derived, which would include $D_h$, the diffusion constant for holes. However, since $D_e > D_h$, npn transistors have a smaller transit time than their pnp equivalents and are to be preferred for high-frequency or high-speed operation. The base transit time can be reduced further if electrons are accelerated across the base by an electric field, rather than moving solely under the influence of diffusion. This property is exploited in drift transistors, which have a built-in drift field in the base region arising from a deliberately introduced doping profile there (see Section 8.4.4).

There is an additional transit time, $\tau_C$, associated with carriers moving through the collector–base depletion layer. For collector voltages, $V_{CB}$, greater than a few volts, electrons in this region reach a saturated velocity, $v_{sat}$, of around $10^5$ m s$^{-1}$. If the depletion-layer thickness, which itself is dependent on the collector voltage, resistivity and so on, is $d_{CB}$, then

$$\tau_c \simeq d_{cB}/v_{sat} \tag{8.25}$$

Finally there are two more time constants, associated with the capacitance of the pn junctions in a transistor. First, the time constant due to the emitter–base junction, $\tau_{EB}$, is given by

$$\tau_{EB} \simeq r_B C_{EB} \tag{8.26}$$

where $C_{EB}$ is the capacitance of the forward-biased junction and $r_B$ its slope



resistance. Substituting approximate values for these gives

$$\tau_{EB} \simeq A\left(\frac{kT}{eI_e}\right)\left(\frac{e\epsilon_r\epsilon_0 N_{0B}}{2V_j}\right)^{1/2} \tag{8.27}$$

where $N_{0B}$ is the acceptor concentration in the base near to the emitter and $V_j$ the net junction voltage. Hence in order to reduce $\tau_{EB}$, the junction area, $A$, and $N_{0B}$ must be kept as small as possible.

The time constant related to the base–collector junction, $\tau_{BC}$, is given by

$$\tau_{BC} \simeq R_{cc} C_{CB} \tag{8.28}$$

which can be reduced by decreasing the collector access resistance, $R_{cc}$, or the doping concentration in the collector, or by increasing $V_{CB}$.

The sum of all four time constants provides a value for the total delay time, $\tau_T$. The high-frequency operating limit corresponding to a particular value of $\tau_T$, $f_\alpha$, is then approximately

$$f_\alpha \simeq (\tau_T)^{-1} \tag{8.29}$$

Here, $f_\alpha$ is the alpha cut-off frequency, at which the gain, $\alpha_B$, falls 3 dB below its low-frequency value.

### 8.4.8 High-frequency response

Gain parameters for a junction transistor have been determined in earlier sections assuming near-equilibrium conditions and are only applicable to low-frequency or slowly varying signals. The expressions obtained are modified when the transistor is operated at high frequencies. For example, the current gain in the grounded-base connection, $\alpha_B$, is not constant but falls off with increasing frequency in the manner shown in Fig. 8.14. The fall-off in gain

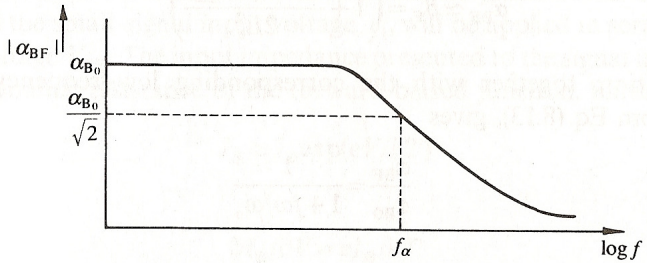



**Fig. 8.14 High-frequency common-base gain of a junction transistor.**

arises principally because of the finite time taken for minority carriers to diffuse across the base; when the transit time becomes comparable to the periodic time of an applied signal the minority carriers can no longer respond fast enough and the gain falls to zero. The *cut-off frequency*, $f_\alpha$, is defined as that frequency at which the magnitude of the current gain, $\alpha_{BF}$, falls by 3 dB to $1/\sqrt{2}$

of its low-frequency value, $\alpha_{B0}$. The value of $f_\alpha$ can be estimated by considering the continuity equation for minority electrons in the base, assuming no electric field exists there, which is, from Eq. (6.63)

$$\frac{\partial(\delta n)}{\partial t} = -\frac{\delta n}{\tau_{Le}} + D_e \frac{\partial^2(\delta n)}{\partial x^2} \tag{8.30}$$

The term on the left-hand side of the equation, which was neglected in the time-independent solution, must be retained when discussing high-frequency effects. If small sinusoidal alternating signals of angular frequency $\omega$ are superimposed on the d.c. currents flowing in the base under equilibrium conditions, the solution to the continuity equation is expected to be of the form

$$\delta n(x,t) = \delta n(x) \exp[j(\omega t + \phi)] \tag{8.31}$$

Substituting this trial solution into Eq. (8.30) gives

$$0 = -\delta n \frac{(1+j\omega\tau_{Le})}{\tau_{Le}} + D_e \frac{\partial^2(\delta n)}{\partial x^2}$$

or

$$\frac{\partial^2(\delta n)}{\partial x^2} = \frac{\delta n(1+j\omega\tau_{Le})}{L_e^2} \tag{8.32}$$

Comparing this equation with that given in Sec. 8.3.2 it will be noticed that $L_e^2$ in the d.c. continuity equation has to be replaced by $L_e^2/(1+j\omega\tau_{Le})$ to give the a.c. version of the equation. Equation (8.13) will therefore still be valid for the base transport factor in the a.c. case, provided $L_e^2$ is replaced by $L_e^2/(1+j\omega\tau_{Le})$. Thus, the gain at high frequencies, $\alpha_{BF}$, neglecting the frequency dependence of the emitter efficiency and assuming it to be near unity, becomes

$$\alpha_{BF} \simeq \beta_F = \left(1 + \frac{l_B^2(1+j\omega\tau_{Le})}{2L_e^2}\right)^{-1} \tag{8.33}$$

This equation, together with the corresponding low-frequency equation derived from Eq. (8.13), gives

$$\frac{\alpha_{BF}}{\alpha_{B0}} = \frac{1}{1+j\omega/\omega_c} \tag{8.34}$$

where

$$\omega_c = \frac{2L_e^2 + l_B^2}{\tau_{Le} l_B^2} \simeq \frac{2L_e^2}{\tau_{Le} l_B^2} = \frac{2D_e}{l_B^2} \tag{8.35}$$

It is evident from Eq. (8.34) that $|\alpha_{BF}|$ falls to $\alpha_{B0}/\sqrt{2}$ when $\omega = \omega_c$ and, hence, from Eq. (8.35),

$$f_\alpha = \omega_c/2\pi \simeq D_e/\pi l_B^2 \tag{8.36}$$



Although a transistor can be operated at higher frequencies, Eq. (8.36) gives an indication of the frequency at which the gain is falling off rapidly and at which phaseshift distortion becomes apparent. It will be noticed that, according to the approximate analysis presented, the $\alpha$ cut-off frequency is dependent only on the diffusion coefficient in and the width of the base. A thin base is again advantageous for good high-frequency performance and there is also some advantage in using npn transistor structures because the diffusion rate for minority electrons is higher than for holes. However, there are physical limitations to the reduction in base width and a more useful method of reducing the transit time, and hence increasing the high-frequency operating capabilities of a transistor, is to introduce a drift field in the base region by means of some degree of impurity grading, as discussed in Sec. 8.4.4.

## 8.5 Small-signal equivalent circuit

Although the bipolar transistor is inherently non-linear for large-amplitude signal variations, it may be considered to bahave in a linear manner over a limited range of its operating characteristics and a small-signal equivalent circuit can be derived to represent its electrical performance. Strictly, such a circuit will only therefore be applicable to small-amplitude a.c. or incremental d.c. signals. Many different equivalent circuits have been proposed; an equivalent circuit based on our discussions of the physical processes that take place in a transistor in the grounded-emitter configuration will be derived as an example. The parameters of the circuit will then be compared with those of a more usually encountered, more generally applicable circuit, that is based on a four-terminal network, 'black-box' approach.

Consider the base-emitter current of an npn bipolar transistor in the common-emitter connection (see, for example, Fig. 8.5). In the normal operating mode of the transistor, the base–emitter junction will be forward-biased and the small-signal input voltage, $v_i$, will be applied in series with the d.c. bias voltage $V_{BE}$. The input impedance presented to the signal includes the effective dynamic resistance of the forward-biased junction. Since

$$I_B \simeq I_0 \exp(eV/kT)$$

and

$$\partial I_B/\partial V = eI_B/kT$$

it follows that the dynamic resistance is given by

$$r = kT/eI_B \simeq 1/(40I_B) \qquad \text{at room temperature} \tag{8.37}$$

which is a low resistance at room temperature and normal bias voltage. For example, for typical standing currents, $I_B$, of order milliamps, $r$ is a few tens of ohms. The input circuit also includes the bulk resistance of the base region and

this is usually lumped together with the dynamic resistance to give an effective input resistance, $r_{BE}$.

Turning now to the output circuit, a large dynamic impedance will exist at the collector–base junction by virtue of the reverse d.c. bias across it and this results in a high value for the output resistance, $r_{CE}$, which is the resistance looking back into the collector–emitter terminals.

The signal current in the base, $i_B$, will be amplified and appear in the collector circuit as a current $\alpha_E i_B$, where $\alpha_E$ is the common-emitter current gain. Therfore, a simple, low-frequency, small-signal equivalent circuit based on the physical processes discussed so far might be of the form shown in Fig. 8.15(a), where a current generator $\alpha_E i_B$ is included in the collector circuit.

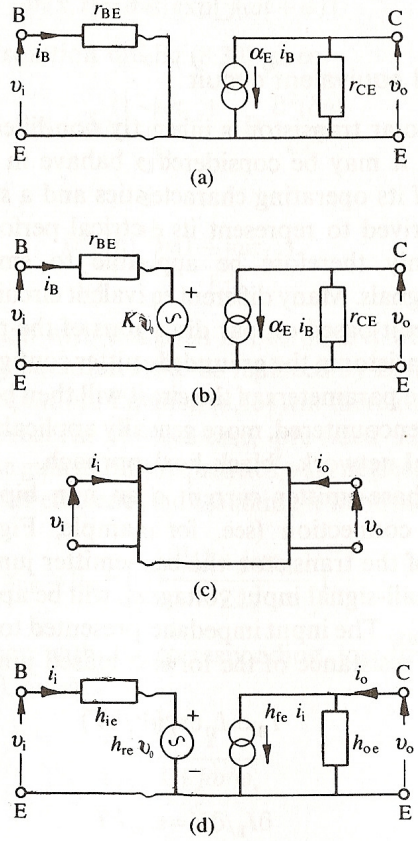



**Fig. 8.15 Possible low-frequency, small-signal equivalent circuits for a bipolar transistor.**

The simple equivalent circuit shown in Fig. 8.15(a) is highly idealized and can be refined somewhat so as to include base-width modulation effects by the addition of an extra component in the circuit to account for the built-in feedback between output and input circuits. The feedback arises because a change in output voltage changes the effective base width, resulting in



changes in collector, emitter and hence base currents. The sign of the change is such that an increase in $v_0$ leads to a decrease in $i_B$. The effect is accounted for in the more comprehensive equivalent circuit shown in Fig. 8.15(b) by including a voltage generator proportional to the output voltage in the base circuit, which is of such a polarity as to cause a reduction in the base current when the output voltage is increasing.

It is sometimes more convenient to treat the transistor as a two-port active network, as shown in Fig. 8.15(c), and develop an equivalent circuit from measurements that can be made at the two ports, rather than devising a circuit based on the physical processes occurring in a particular device. Again, there are several different forms for the defining equations of the equivalent circuit, depending on which set of terminal characteristics is considered. As an example, one possible set of measured parameters might be

(a) the *input* impedance with output short-circuited,

$$h_{ie} = (v_i/i_i)|_{v_o = 0}$$

(b) the *reverse* open-circuit voltage amplification

$$h_{re} = (v_i/v_o)|_{i_i = 0}$$

(c) the *forward* current gain with output short-circuited

$$h_{fe} = (i_o/i_i)|_{v_o = 0}$$

and (d) the *output* conductance with input open-circuited

$$h_{oe} = (i_o/v_o)|_{i_i = 0}$$

These are the *hybrid parameters* of a transistor, so-called because they do not all have the same dimensions. The additional subscript 'e' is added to designate the circuit configuration, in this example common *emitter*. It follows that the small-signal voltages and currents at the input and output terminals of the equivalent circuit are then related by the equations

$$\begin{aligned} v_i &= h_{ie} i_i + h_{re} v_o \\ i_o &= h_{fe} i_i + h_{oe} v_o \end{aligned} \qquad (8.38)$$

The equivalent circuit shown in Fig. 8.15(c) used in conjunction with Eqs (8.38) can be used to define the small-signal performance of a transistor completely when it is included in a particular circuit.

The equivalence of the two circuit representations discussed can be seen by noting that Eqs (8.38) are also valid for the circuit shown in Fig. 8.15(d). This has obvious similarity to the equivalent circuit based on the internal physical processes occurring in a transistor, Fig. 8.15(b), and it follows by direct comparison of the two circuits that

$$h_{ie} \equiv r_{BE} \qquad h_{re} \equiv K \qquad h_{fe} \equiv \alpha_E \qquad h_{oe} \equiv 1/r_{CE}$$

## 8.6 Fabrication of junction transistors

Discrete transistors are usually made using one of the IC processes discussed more fully in Chapter 10, using various combinations of diffusion, epitaxy and ion-implantation techniques. Many transistors are usually fabricated per slice, the number being dictated by area-dependent considerations such as power rating and device complexity, transistor chips being separated subsequently and individually packaged.

Cross sections of a selection of possible transistor structures are shown in Fig. 8.16. The diffusion process for fabricating transistors of the type shown in

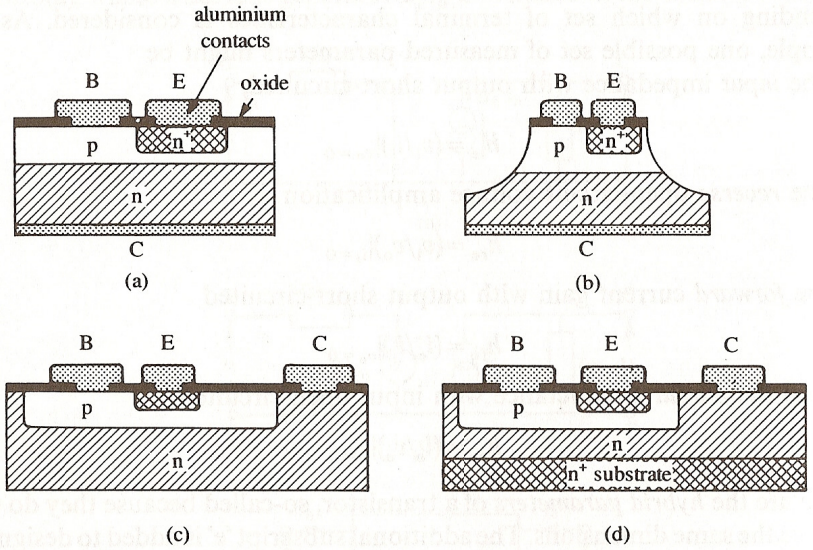



**Fig. 8.16 Construction of some discrete transistor types: (a) diffused; (b) mesa; (c) planar; (d) epitaxial planar.**

Fig. 8.16(a), which is described in Chapter 10, has many advantages. Until recently such technology was dominant but now newer techniques for junction formation, such as ion implantation and polysilicon emitters, are becoming progressively more important. In the *mesa* version, Fig. 8.16(b), the area of the collector junction is formed by an etching process, leaving the active portions isolated on a tapered plateau or mesa, which provides improved high-voltage performance. The planar discrete transistor, Fig. 8.16(c), is a direct derivative of the IC counterpart, all contacts being accessible from the top plane. Similarly for the epitaxial variant, Fig. 8.16(d), where the lightly doped epitaxial collector layer is grown on a highly doped supporting substrate, which provides a low-resistance path from collector terminal to the active collector region. Epitaxial diffused transistors have the additional advantage of reduced collector capacitance and higher breakdown voltages.



## 8.7 Silicon controlled rectifier

The silicon controlled rectifier, SCR, sometimes called a *thyristor*, is a four-layer, three-pn-junction, single-crystal silicon device, as shown in Fig. 8.17, which shows the principal features of a practical device, a possible model for the active silicon slice and the relative doping intensities in it. In an $n^+pnp^+$

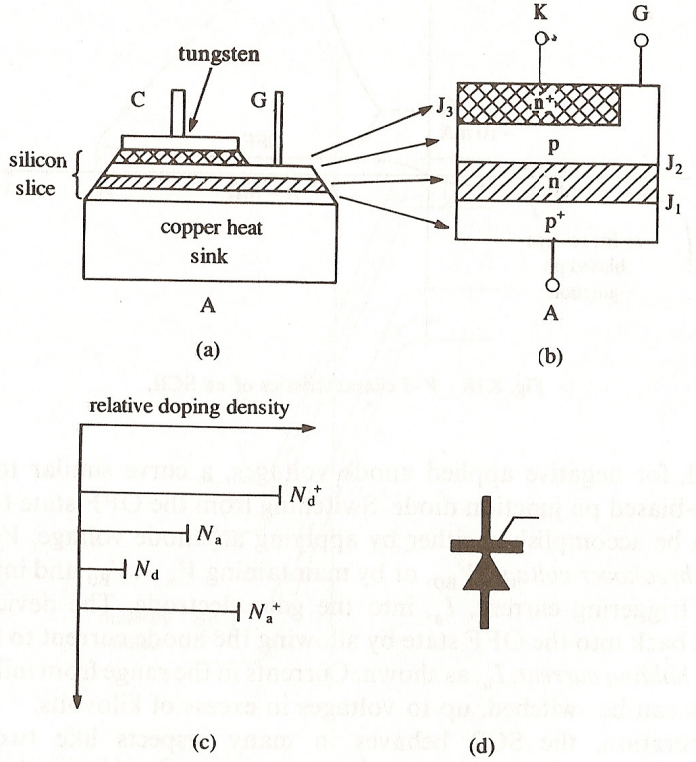



**Fig. 8.17 The silicon controlled rectifier: (a) a cross section; (b) a model of the silicon slice; (c) relative doping in the layers; and (d) its symbolic representation.**

device an *anode* connection, A, is made to the $p^+$-layer and contacts are also formed with the $n^+$ *cathode*, K, and the p-type *gate* region, G, as shown. Briefly, the SCR usually operates in a switching mode, when small injected gate currents control the point in the anode voltage cycle at which the device switches and large currents are allowed to flow.

SCRs are characterized by a very high-resistance OFF state in which the current flow between anode and cathode is very small, of order 1 mA, which on switching changes rapidly to a low-resistance ON state, via a negative-resistance region, in which large currents, typically kiloamps can flow. The general $I$–$V$ characteristics for such operations are therefore as indicated in Fig. 8.18. Three distinct operating regions are apparent, an OFF state, an ON

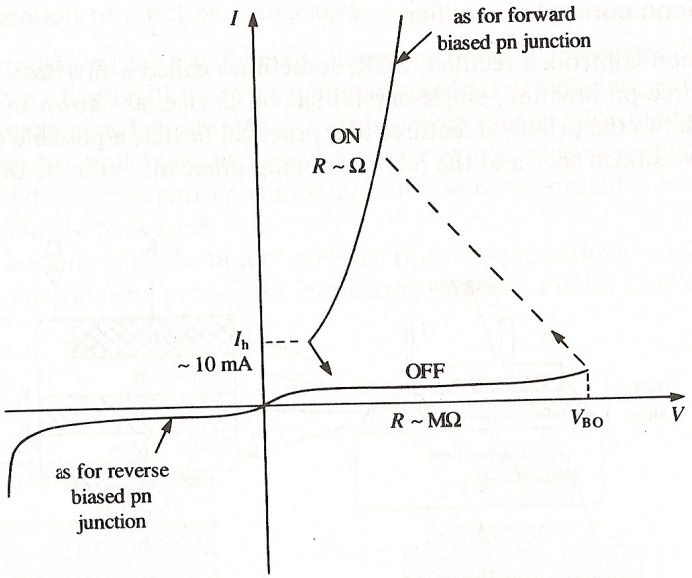



**Fig. 8.18** *V–I* **characteristics of an SCR.**

state and, for negative applied anode voltages, a curve similar to that for a reverse-biased pn junction diode. Switching from the OFF state to the ON state can be accomplished either by applying an anode voltage, $V_A$, greater than the *breakover voltage*, $V_{BO}$, or by maintaining $V_A < V_{BO}$ and injecting an external triggering current, $I_g$, into the gate electrode. The device can be switched back into the OFF state by allowing the anode current to fall below a critical *holding current*, $I_h$, as shown. Currents in the range from milliamps to kiloamps can be switched, up to voltages in excess of kilovolts.

In operation, the SCR behaves in many respects like two bipolar transistors, $n^+pn$ and $p^+np$, the properties of which are intimately linked by common regions, as shown in Fig. 8.19. Consider first the situation in which the gate current is zero, $I_g = 0$, for simplicity. In the forward bias condition

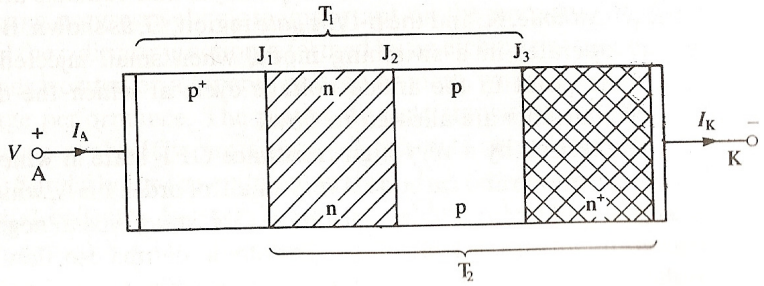



**Fig. 8.19 The four-layer diode, $I_g = 0$.**



assumed in the diagram when the anode voltage, $V_A$, is positive with respect to the cathode, junctions $J_1$ and $J_3$ are forward-biased but the current at low voltages is controlled by the reverse-biased junction, $J_2$. Under these conditions, only a very small saturation current flows across $J_2$, which consists of thermally generated minority electrons at either side of $J_2$ that are swept across the junction by the energetically favourable energy barrier that prevents majority carrier diffusion; see for example Fig. 8.20. However, since

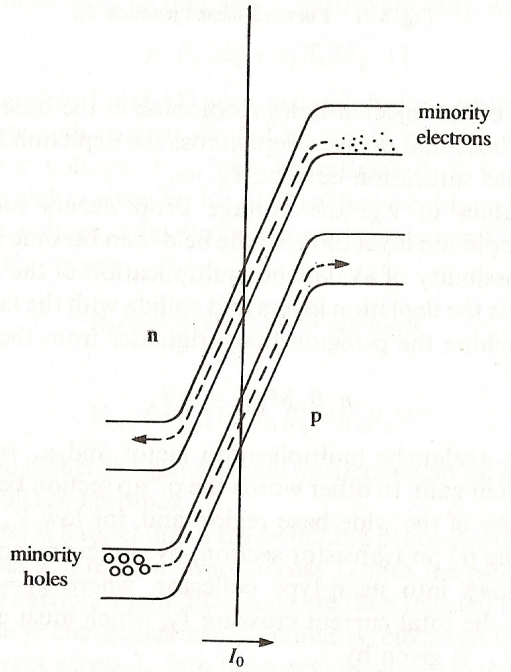



**Fig. 8.20 Carrier transport in the reverse-biased junction $J_2$.**

$J_2$ also behaves as the base–collector junction of both overlapping transistors, there is an additional component of current at $J_2$ originating from the effective emitter sections. Consider, for example, the $p^+n$ junction shown in Fig. 8.21. Since the acceptor concentration in the $p^+$-region, $N_a^+ \gg N_d$, the donor concentration in the n layer, most of the current crossing the forward-biased $J_1$ is transported by majority holes. The proportion of the junction current injected by holes is therefore $\eta_1 I_A$, where $\eta_1$ is the emitter injection efficiency of $T_1$, which will be almost unity. The injected holes become minority carriers in the n-region and the number of holes that survive recombination and reach the depletion layer of $J_2$ is $\beta_1(\eta_1 I_A)$, where $\beta_1$ is a 'base' transport factor. Contrary to the situation in a bipolar junction transistor (BJT), the length of the n-type 'base' region in an SCR, $l_n \gg L_h$, the diffusion length for minority holes, so

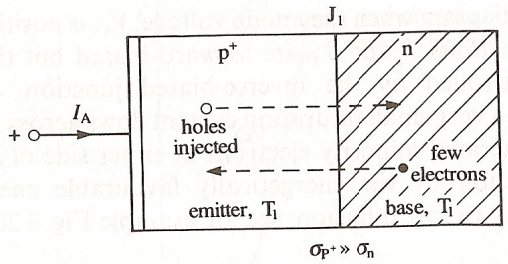



**Fig. 8.21 Forward-biased junction $J_1$.**

$\beta_1 \ll 1$ and most of the injected holes recombine in the base region and only a few reach $J_2$. Those that do are swept across the depletion layer at $J_2$ to add to the reverse bias saturation current, $I_0$.

At higher values of $V_A$, the voltage drop occurs mostly across the reverse-biased depletion layer of $J_2$, so the fields can become so high there that there exists a possibility of avalanche multiplication of the holes as they are accelerated across the depletion layers and collide with the lattice. So the total hole current reaching the p-region that originates from the $p^+$ 'emitter' is

$$\eta_1 \beta_1 M_1 I_A = \alpha_1 I_A \tag{8.39}$$

where $M_1$ is an avalanche multiplication factor and $\alpha_1$ is a common-base large-signal current gain. In other words the $p^+np$ section behaves like a poor transistor because of the wide base region and, for low $V_A$, $\alpha_1 \ll 1$.

Turning to the $n^+pn$ transistor section, by similar reasoning, an electron current $\alpha_2 I_K$ flows into its n-type collector, where $\alpha_2 = \eta_2 \beta_2 M_2$. Hence, provided $I_g = 0$, the total current crossing $J_2$, which must equal the external circuit current $I_A$, is given by

$$I_A = \alpha_1 I_A + \alpha_2 I_K + I_0 \tag{8.40}$$

as illustrated in Fig. 8.22. It follows that, since $I_A = I_K$ if $I_g = 0$ as has been

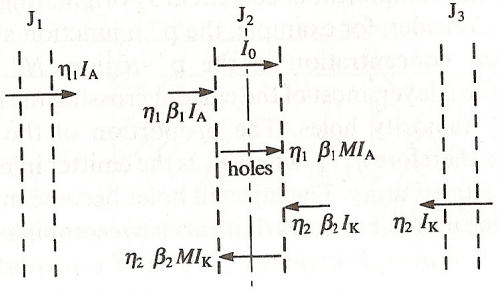



**Fig. 8.22 Currents crossing junction $J_2$.**



assumed, then the anode current becomes

$$I_A = \frac{I_0}{1-(\alpha_1+\alpha_2)} = \frac{I_0}{1-(\eta_1\beta_1 M_1 + \eta_2 \beta_2 M_2)} \tag{8.41}$$

At voltages below the breakover voltage, $V_A < V_{BO}$, the $M$ and $\eta$ factors are almost unity but the transport factors, $\beta$, are small so $I_A \simeq I_O$ and a very small reverse bias saturation current flows in this OFF condition.

At higher anode voltages, avalanche multiplication becomes more probable in the depletion layer of $J_2$, so that $M > 1$, until eventually, when $V_A = V_{BO}$ and

$$\eta_1 \beta_1 M_1 + \eta_2 \beta_2 M_2 \to 1 \tag{8.42}$$

then $I_A$ as indicated in Eq. (8.41) approaches infinity and the device switches into the ON state with current limited by circuit conditions.

The breakover voltage, $V_{BO}$, can be estimated, if it is assumed that $M_1 = M_2 = M$, which is given by Eq. (7.70) where $n \simeq 3$ for silicon and $V$ is the applied junction voltage. Substituting this expression in Eq. (8.42) at breakover, when $V = V_{BO}$, gives

$$\eta_1 \beta_1 + \eta_2 \beta_2 = 1 - (V_{BO}/V_{BD})^n$$

or

$$V_{BO} = V_{BD}(1 - \eta_1 \beta_1 - \eta_2 \beta_2)^{1/n} \tag{8.43}$$

It follows that the higher the gains, which are dependent on $\eta$ and $\beta$, the lower the breakover voltage becomes.

Physically what is happening in the switching process is that the holes that are injected across $J_1$ are then swept across $J_2$ into the p-region where they become majority carriers again. Electrons are then injected across $J_3$ into the p-region to preserve charge neutrality. Similarly, electrons from the $n^+$-layer are eventually swept across $J_2$ into the n-region where neutrality is preserved by extra injected holes. Clearly, this is a regenerative process and it is possible for the total current to rise rapidly.

The holes swept into the p-region compensate some of the negative exposed acceptors fixed in the reverse-biased depletion layer of $J_2$, as shown in Fig. 8.23, thus reducing the effective depletion-layer charge. As the voltage is increased so that $V_A$ approaches $V_{BO}$, the number of holes reaching the depletion layer increases, for the reasons explained, until sufficient ionized acceptors are compensated so that the *effective* depletion layer becomes similar to that for a forward-biased junction. In other words when the SCR is switched on, the potential barrier of $J_2$ is reduced until the junction is effectively forward-biased, as shown in Fig. 8.24. In this ON condition, large currents can flow but the anode voltage remains small.

The mode of operation discussed so far is known as *voltage triggering*. When gate currents are injected, so that $I_g > 0$, it is possible for an SCR to be *current triggered*, with $V_A < V_{BO}$. With gate current the situation is as depicted

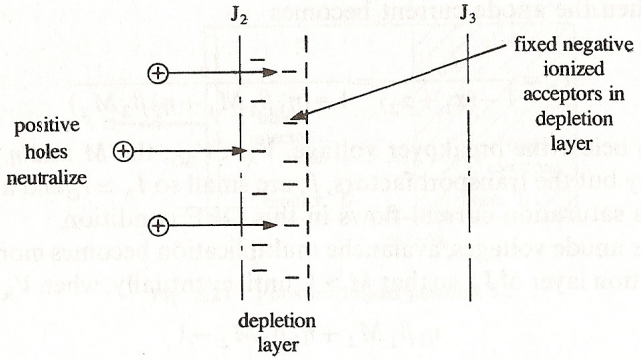



**Fig. 8.23 Reduction in effective depletion-layer charge by injected carriers.**

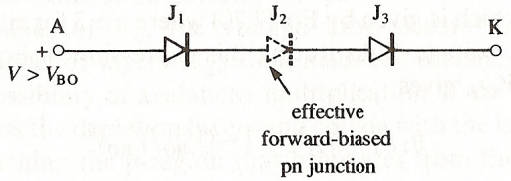



**Fig. 8.24 Equivalent circuit of SCR for $V > V_{BO}$.**

in Fig. 8.25. The current crossing $J_2$ is still given by Eq. (8.40) but this time Kirchhoff's law indicates that

$$I_K = I_A + I_g$$

so this can be substituted in the equation to give

$$I_A = \frac{\alpha_2 I_g + I_0}{1-(\alpha_1+\alpha_2)} \tag{8.44}$$

In the situation when $V_A < V_{BO}$ and $I_g$ is increasing from zero, injection of $I_g$ increases $I_K$ and hence $I_A$. Now for gains $\ll 1$, the gain of a transistor is very

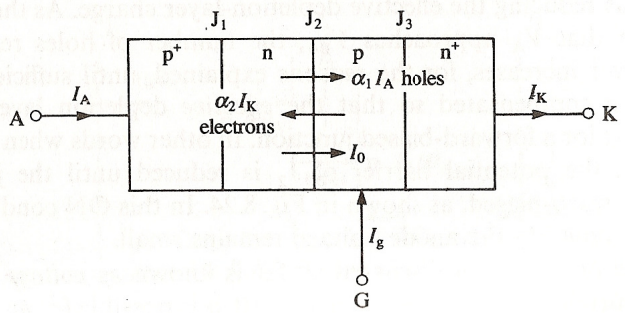



**Fig. 8.25 SCR with gate triggering.**



dependent on the emitter current, in this case $I_A$ and $I_K$, so as $I_g$ increases, so do $I_A$ and $I_K$ and hence $\alpha_1$ and $\alpha_2$. This continues until $(\alpha_1+\alpha_2)=1$ in Eq. (8.44), $I_A \to \infty$ and is only limited by the circuit resistance as the device switches ON. Since switching has occurred at $V_A < V_{BO}$, it follows that the gate current can be used to control the point at which switching occurs, and as $I_g$ is increased, $V_A$ for switching is reduced, as shown in Fig. 8.26. This is the most

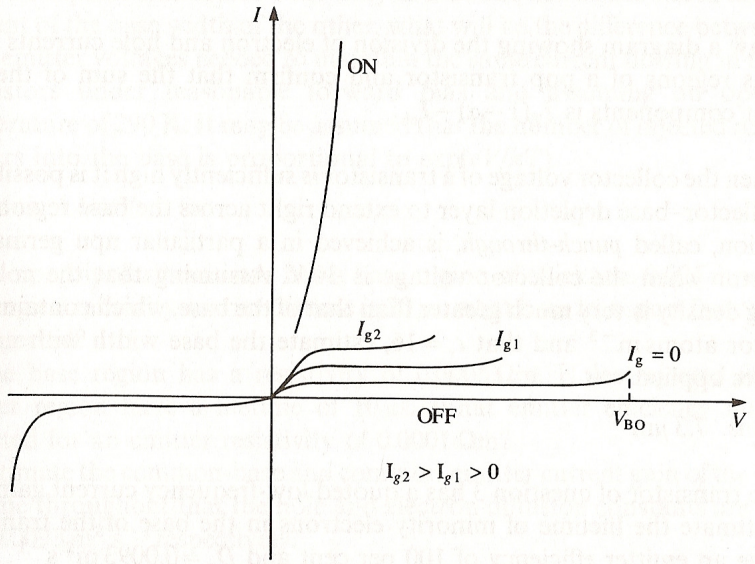



**Fig. 8.26 $I-V$ characteristics of SCR with gate triggering.**

common triggering mode since it only requires low levels of power to control the switching of large currents. The ratio $I_A/I_g$ at switching is of order $10^5$, so only a very small gate current, typically of order 10–100 mA, is required to initiate switching.

Since the dissipation in the forward-biased gate junction is limited typically to a fraction of a watt, which it is possible to exceed with continuous gate currents of say 100 mA and corresponding gate voltages of a few volts, the trigger current is often pulsed, to prevent possible device failure. Provided the pulse of $I_g$ is supplied for a sufficiently long time for the current multiplication process to occur and switching to take place, which is typically a few microseconds, then the duty cycle of the pulses can be made such that the average gate dissipation is never exceeded.

Dynamically the SCR is relatively slow. Its switching speed is typically of order 0.1 $\mu$s to turn ON and around 10 $\mu$s to turn OFF, which is much slower than a BJT. Its principal applications, therefore, are in the low-frequency, high-current field, for example for motor control in trains and drills or in light dimmers, and so on.

## Problems

**1.** A certain pnp transistor has an effective base width of 20 $\mu$m under certain biasing conditions. The thickness of the emitter region is 5 $\mu$m and its resistivity is $50 \times 10^{-6}\,\Omega$m. The effective base lifetime is 20 $\mu$s and $D_h = 0.0047\,m^2\,s^{-1}$. Estimate the common emitter current gain of the device.

*Ans.* 81

**2.** Draw a diagram showing the division of electron and hole currents in the various regions of a pnp transistor and confirm that the sum of the base current components is $I_E(1-\alpha)-I_{co}$.

**3.** When the collector voltage of a transistor is sufficiently high it is possible for the collector–base depletion layer to extend right across the base region. This condition, called *punch-through*, is achieved in a particular npn germanium transistor when the collector voltage is 30 V. Assuming that the collector doping density is very much greater than that of the base, which contains $10^{21}$ acceptor atoms $m^{-3}$ and that $\epsilon_r = 16$, estimate the base width with no bias voltages applied.

*Ans.* 7.3 $\mu$m

**4.** The transistor of question 3 has a quoted low-frequency current gain $\alpha_E$ of 30. Estimate the lifetime of minority electrons in the base of the transistor. Assume an emitter efficiency of 100 per cent and $D_e = 0.0093\,m^2\,s^{-1}$.

*Ans.* 0.09 $\mu$s

**5.** Show that the $h_{fe}$ of a high-gain bipolar transistor is approximately of the form

$$h_{fe} \cong [C_1(\sigma_B/\sigma_E) + C_2/L_e^2]^{-1}$$

where constants $C_1$ and $C_2$ depend on the geometry.

Under certain conditions of bias, the $h_{fe}$ of a certain bipolar transistor is 450. Assuming the ratio of emitter to base conductivities is 100:1, a diffusion coefficient for carriers in the base of $3 \times 10^{-3}\,m^2\,s^{-1}$ and emitter and base lengths of 10 and 20 $\mu$m, estimate the lifetime of carriers in the base.

*Ans.* 3 $\mu$s

**6.** A certain npn transistor has a cross-sectional area of $10^{-8}\,m^{-2}$ and a doping density in the emitter of $10^{23}\,m^{-3}$. It is operated at 20°C with a base–emitter junction voltage of 0.65 V and a collector voltage of 10 V, when the collector current is 1.35 mA.

Assuming electron and hole mobilities of 0.13 and 0.05 $m^2\,V^{-1}\,s^{-1}$ estimate: (a) the base–emitter junction current, assuming a saturation current density of 1 $\mu A\,m^{-2}$; (b) the density of majority carriers in the base, assuming that recombination in the base and collector leakage currents can be neglected; and (c) the ratio of electron to hole currents flowing across the base–emitter junction.



*Ans.* (a) 1.5 mA, (b) $2.5 \times 10^{22}\,m^{-3}$, (c) 10.4:1

**7.** If two bipolar transistors differ only in that one has a base width that is 90 per cent of the base width of the other, what will be the difference between the base–emitter voltages needed to maintain the same current flowing in the two transistors under reasonable forward bias and assuming an operating temperature of 290 K. It may be assumed that the number of injected majority carriers into the base is proportional to $\exp(eV/kT)$.

*Ans.* 1.1 mV

**8.** An npn bipolar transistor has a base region of thickness 0.025 mm. If the base region minority-carrier lifetime is 20 $\mu$s, what base transport factor might be expected?

The base region has a resistivity of 0.0005 $\Omega$m. Holes injected into the emitter region have a lifetime of 10 $\mu$s. What emitter efficiency might be expected for an emitter resistivity of 0.0001 $\Omega$m?

Estimate the common-base and common-emitter current gain of the device. Assume throughout that the hole and electron diffusion constants are 0.0044 and $0.0093\,m^2\,s^{-1}$ respectively.

*Ans.* 0.998, 0.975, 0.975, 39

**9.** A certain npn transistor has effective emitter and base lengths of 100 $\mu$m and 20 $\mu$m and the ratio of emitter to base conductivities is 50:1. Find the emitter efficiency of the transistor.

If the forward current gain of the transistor is 50 and the diffusion constant for carriers in the base is $3 \times 10^{-3}\,m^2\,s^{-1}$, estimate the lifetime of carriers in the base.

*Ans.* 0.996, 4 $\mu$s

**10.** A certain npn transistor has an effective base width of 20 $\mu$m under given bias conditions. Estimate its forward current gain in the common-emitter configuration, $h_{fe}$, assuming a lifetime of minority carriers of 1 $\mu$s and a diffusion constant of minority carriers of $0.01\,m^2\,s^{-1}$ in the base. Assume, for simplicity, an emitter efficiency of unity.

*Ans.* 49

**11.** In a particular bipolar transistor, the effective length of the base, $l_B$, under certain bias conditions, is 1 $\mu$m. When the bias is changed, so as to change $l_B$,