8.2 Thyristors : Semiconductor Power Devices

High power device design implies high current and high voltage capabilities alongwith the ability to operate at relatively high temperatures and conduct power dissipated in the form of heat to outside the device with a minimum of thermal (and hence electrical) resistance.

Most of the semiconductor power devices in use are thyristors. Also there are bipolar and power MOS (VMOS structure) and static induction transistors (SIT) but they are in use duly for moderately high power applications nowhere in comparison to thyristors. Power devices for microwave frequencies are usually made of GaAs.

Thyristors are switching devices used mainly in the control systems. The main current can be switched ON when desired and in some types also OFF but not controlled in between. In other devices the output current can be continuously controlled by base current or gate voltage and they can therefore be used for either switching or linear power amplification.

A power switch must have the following properties :

(i) Its ability to carry large currents with a uniform current density over a large device area.

(ii) While the device is in ON state, the device resistance must be as low as possible to reduce joule heating.

(iii) While the device is in OFF state, the device resistance must be very high as well as capable of holding high voltage ($\sim 10^3$ volts) across it.

(iv) The switching speed should be high. This means low capacitances and minimum minority storage effect.

(v) For power amplification, the forward β (or gm) should be constant and independent of the signal amplitude. This is necessary to avoid excessive distortions in large signal applications.

However, we shall be concerned only with thyristors in this book. The reader may consult any book on power electronics for further details.

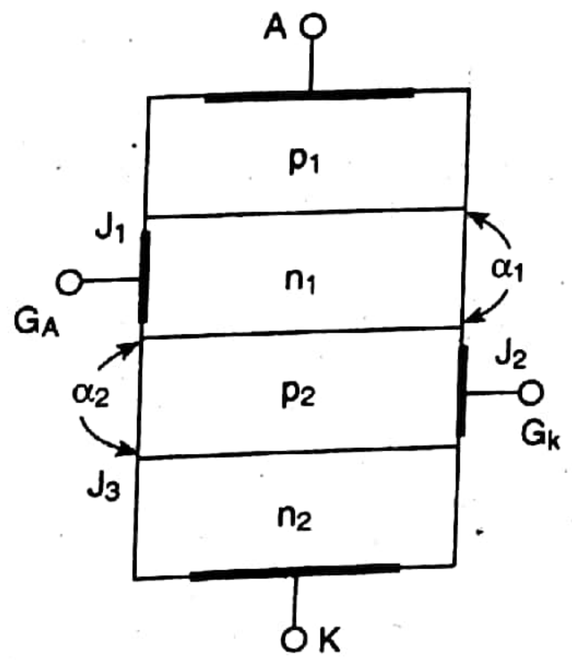
The thyristor family is one of the most important semiconductor devices today in the industrial or power electronics field. Thyristor is a generic name for a family of *pnpn* semiconductor switches though it is very often used to refer to a particular member—the silicon controlled rectifier. Different devices may have two, three, or four accessible electrodes and may conduct uni- or bidirectionally.

The name thyristor applies to a general family of semiconductor devices that exhibit bistable characteristics and can be switched between a high-impedance, low curent OFF state and a low-impedance, high-current ON state. The principle of operation of thyristors are intimately connected to BJT in which both electrons and

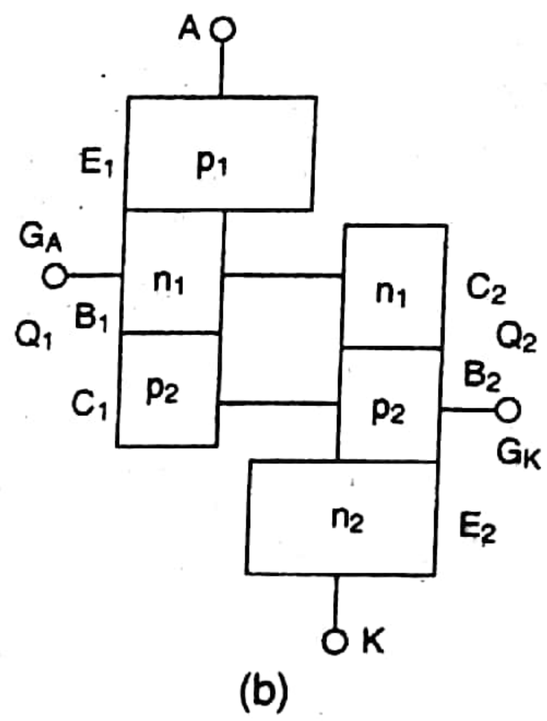
holes are involved in the transport processes. The term 'thyristor' is derived from 'gas thyratrons', because of the similarity in the output characteristics of these two. Because of the bistable nature, thyristors have found unique usefulness in applications ranging from speed control in home appliances to switching and power inversion in high-voltage transmission lines. Thyristors are now available with current ratings from a few mA to over 5000 amphere and voltage ratings extending above 10,000 V.

8.2.1 Shockley Diode

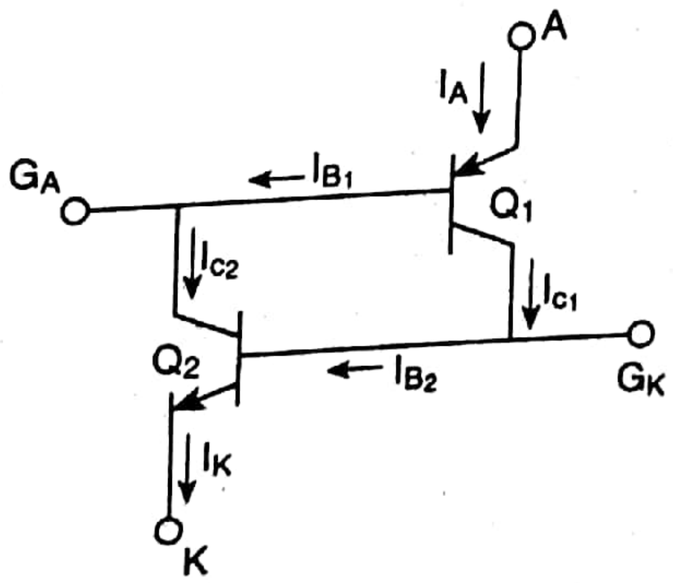
We shall first consider the basic operation of *pnpn* devices, and then the characteristics of the main types of switch. The four-layer $p_1n_1p_2n_2$ configuration with the *anode* terminal A at the outside p_1-region and a *cathode* terminal K at the outside n_2-region is shown in Fig. 8.14(a). We shall refer to the junction nearest the anode as J_1, the center junction as J_2, and the junction nearest to the cathode as J_3. Fig. 8.14(a) may be viewed as a combination of a *pnp* and the *npn* transistor (Fig. 8.14(b) and (c)). The circuit symbol is shown in Fig. 8.14(d). The significance of two-transistor analogy is that it shows the current gains of the sections (α_1 for $p_1n_1p_2$ transistor and α_2 for $n_1p_2n_2$ transistor), and the existence of feedback. These two-factors are necessary to obtain the desired switching and latching operation.



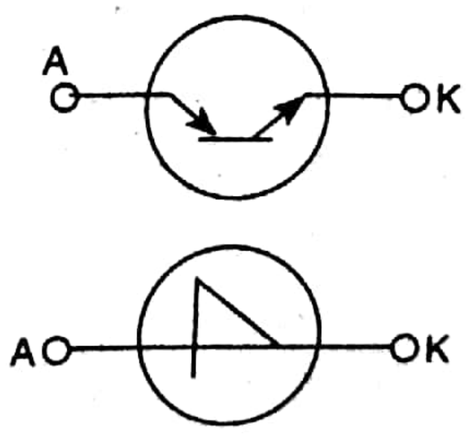
(a)



(b)



(c)



(d)

Fig. 8.14 Thyristor (a) A four-terminal pnpn device, (b) and (c) Two transistor analogy of the p–n–p–n device, (d) Symbol of Shockley diode

The operation of a thyristor depends on the polarity and voltage applied between the various electrodes. The simplest case leaves the two gate electrodes G_A and G_K open and the resultant device is called a *Shockley diode*. The applied voltage biasing the anode A is positive relative to the cathode K. Positive feedback is provided by the interconnection of the collectors and bases, so the gain around the loop $L = \beta_1\beta_2$. From section 7.1 it is expected that if $L \geq 1$, then regenerative positive feedback will cause a switching action. If the collector-base leakage current is I_{CO}, then we can write the equations (see eqn. (3.1.11)), from Fig. 8.14(c)

$$I_{B1} = I_{C_2}, \quad I_{C_1} = \beta_1 I_{C_2} + (\beta_1 + 1) I_{CO_1} \tag{8.2.1}$$

$$I_{B2} = I_{C_1}, \quad I_{C_2} = \beta_2 I_{C_1} + (\beta_2 + 1) I_{CO_2} \tag{8.2.2}$$

$$I_A = I_{C_1} + I_{C_2} = I_K$$

$$= \frac{(\beta_1 + 1)(\beta_2 + 1)(I_{CO_1} + I_{CO_2})}{1 - \beta_1\beta_2} \tag{8.2.3}$$

after some manipulation. Eqn. (8.2.3) is more commonly written interms of transistors' $\alpha = \beta/(\beta + 1)$ or $\beta = \alpha/(1 - \alpha)$:

$$I_A = \frac{I_{CO_1} + I_{CO_2}}{1 - (\alpha_1 + \alpha_2)} \tag{8.2.4}$$

With only small leakage currents flowing, the values of β (and also of α) will be very small, usually less than 1. Hence, $\beta_1\beta_2 \ll 1$, and the current through the Shockley diode

$$I_A \approx I_{CO_1} + I_{CO_2}$$

is also small. If the loop gain, β_1, β_2, can be increased to ≈ 1 then I_A increases rapidly, becoming *nominally* infinite for $\beta_1\beta_2 = 1$ (or $\alpha_1 + \alpha_2 = 1$). The positive feedback drives both transistors into saturation with a fast switching time; the saturation reduces the gain to just maintain $L = 1$ and, hence, the stability of the ON state. The current flowing will be primarily limited by the external circuit, unless the external inpedance is particularly small.

Such a development is not unexpected because of the unusual connection in Fig. 8.14(c). A change in current at any point in the loop is amplified and returned to the starting point with the same phase. For instance, if the Q_1 base current increases, the Q_2 collector current increases. This enhances base current of Q_2. This, in turn, produces a larger Q_2 collector current, which drives the Q_1 base harder. This build-up in current will continue until both transistors are driven into saturation. In this case, the diode acts as a closed switch.

On the other hand, if something causes the Q_1 base current to decrease, the Q_1 collector current will decrease. This reduces the Q_2 base current ($i_{C1} = i_{B2}$). In turn, there is less Q_2 collector current. which reduces the Q_1 base current even more.This regeneration continues until both transistors are driven into cut-off. At this time, the diode acts as an open switch.

The diode, thus, can be in either of two states : closed or open. It will remain in one state indefinitely. If closed, it stays closed until something causes the currents to decrease. If open, it stays open until something else forces the currents to increase.

In fact all thyristor devices can be explained in terms of an ideal *complementary latch*, the upper transistor being *pnp* while the lower one is *npn*. They differ in the way of triggering for opening or closing the latch.

The reason why the Shockley diode can exist in either of two states is explained with the help of eqn. (8.2.3) or (8.2.4). Now, we shall discuss its *I-V* characteristics. When the anode is forward biased i.e., A positive with respect to K, the junctions J_1 and J_3 are under forward bias, and the center junction under reserve bias. Hence a very small forward current (leakage current I_{C_0} of J_2) start to flow at a very small value and then increases with increasing reverse bias across J_2 because of avalanche multiplication (but not avalanche break-down). This increase in current increases β_1 and β_2. When the product of the small-signal *avalanche-enhanced* betas equal unity, $\beta_1\beta_2 = 1$, break-over occurs. Huge current flows through the diode and the diode is said to be switched to ON state (as shown in Fig. 8.15). This happens at a voltage V_{B0}, called the *break-over voltage.* Before attaining V_{B0} the diode is said to be in the *forward blocking* state. When the current reaches the critical value I_B for $L = 1$, break-down occurs, the characteristics having a negative resistance region which is traversed rapidly ($\sim \mu s$) to reach the current I_H, called the *holding current*, on the *diode* characteristics. It is so called because if the current does not reach I_H, the device will, then, switch back to the non-conducting state when the voltage is reduced. The situation is somewhat more complex than this and, depending on capacity, oscillations may occur.

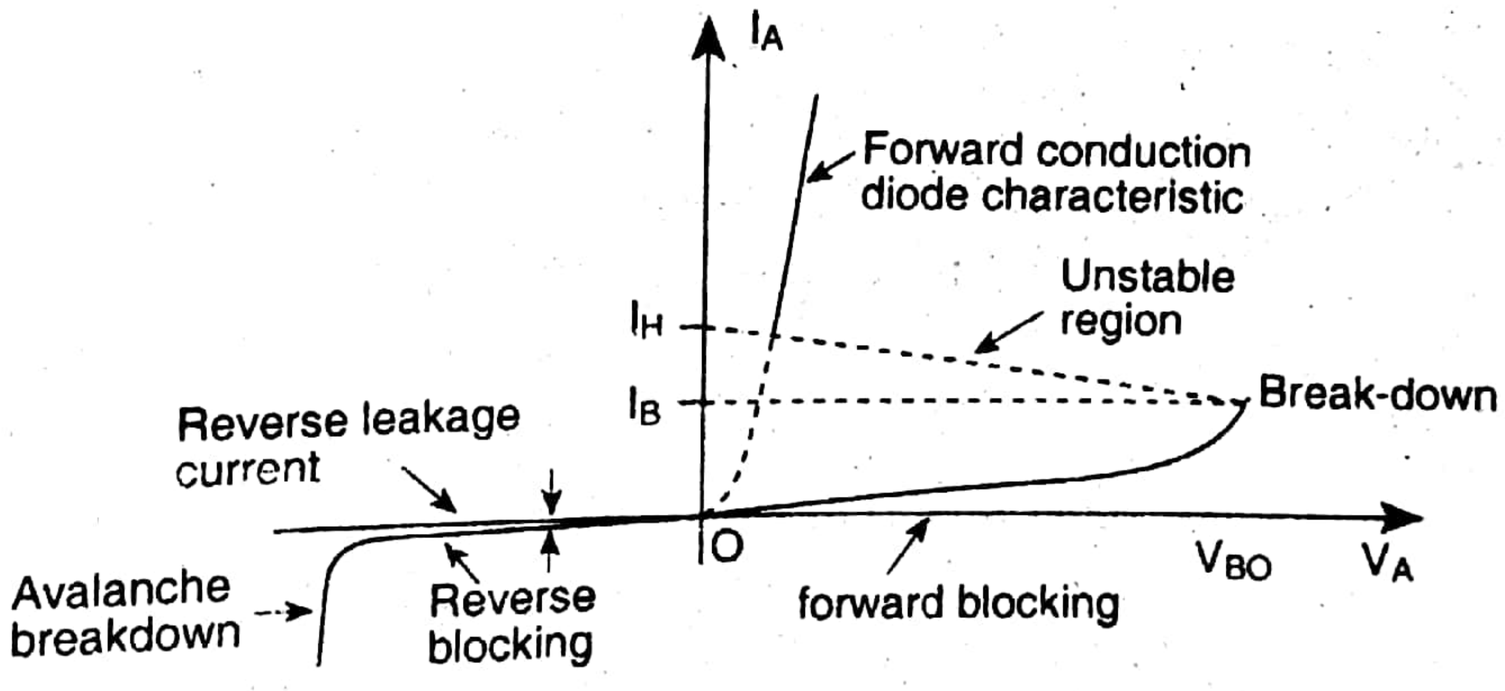


Fig. 8.15 I–V characteristic of Shockley diode

In the ON state all junctions are forward-biased, and so the total voltage across the device is equal very nearly the algebraic sum of these three saturation junction voltages. It should be kept in mind that the voltage drop across the central junction J_2 is in a direction opposite to the voltages across the junctions J_1 and J_2. Thus the total voltage drop across the diode in the ON state is $2V_{BE,sat} - V_{CB,sat} \approx 1.0$ V.

The operation of the Shockley diode depends on the fact that at low currents, the current gain α may be less than 0.5, a condition which is necessary if $(\alpha_1 + \alpha_2)$ is to

be less than unity. This characteristic of α is not encountered in Ge but is distinctive of Si. Hence Ge structures incline to settle immediately in the ON state and have no stable OFF state. Accordingly, Ge, *pnpn* switches are not available.

The value of β (or α) can be changed by several mechanisms :

(1) Overall voltage, which eventually causes avalanche multiplication (and no breakdown) of the leakage currents and hence an increase of β. This is the mechanism used in DIAC (section 8.2.3).

(2) Direct injection of current or more properly charge, via a gate electrode (G_A or G_K) to increase the conduction current by transistor action beyond I_H. This is the commonest technique and is applicable in all three-terminal devices.

(3) Photo generation of carriers in a base region to give an effect similar to (2). This is used in photo SCR, or light activated SCR, abbreviated as LASCR.

(4) Temperature : Leakage current in Si devices doubles for every 10°C rise in temperature, causing increase of β.

(5) Rate of change of overall voltage : If the rate is high, enough current will flow through the junction capacitances to raise β and cause switching. This known as *dv/dt effect*. It introduces an important limitation on the use of thyristors.

8.2.2 Silicon Controlled Rectifier (SCR)

The most commonly encountered thyristor in power circuit applications is the *silicon controlled recitifier* (SCR), which has the structure discussed above with the gate electrode G_K. Fig. 8.16(a) shows different current components in SCR which has two emitters e_1 and e_2 and two bases b_1 and b_2 with one floating. Let a current I flows through the device due to the application of a voltage V with the polarities shown. I_G is the gate current applied to G_K. Let also a hole current I_{pE_1} is injected from e_1 to b_1, a hole current I_{pC_1} is injected from b_2 to b_1, an electron current I_{nE_2} is injected from e_2 to b_2, and an electron current. I_{nC_2} is injected from b_1 into b_2. We further consider that the corresponding current amplification factors are $\alpha_{f_1}, \alpha_{r_1}, \alpha_{f_2}$ and α_{r_2}, respectively. Consequently, we can write for the three junctions :

$$I = I_{pE1} - \alpha_{r_1} I_{pC1} \tag{8.2.5}$$

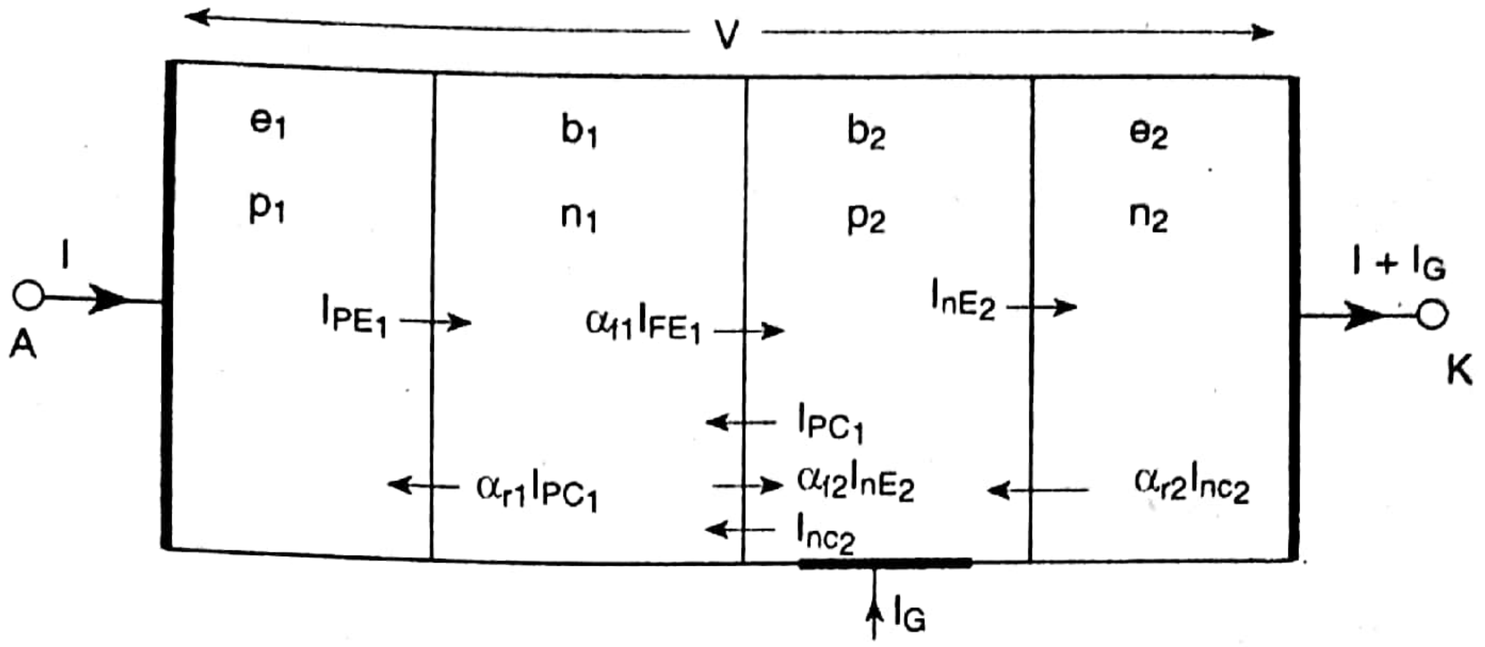
$$I = \alpha_{f_1} I_{pE1} - I_{pC1} + \alpha_{f_2} I_{nE2} - I_{nC2} \tag{8.2.6}$$

$$I + I_G = I_{nE2} - \alpha_{r_2} I_{nC2} \tag{8.2.7}$$

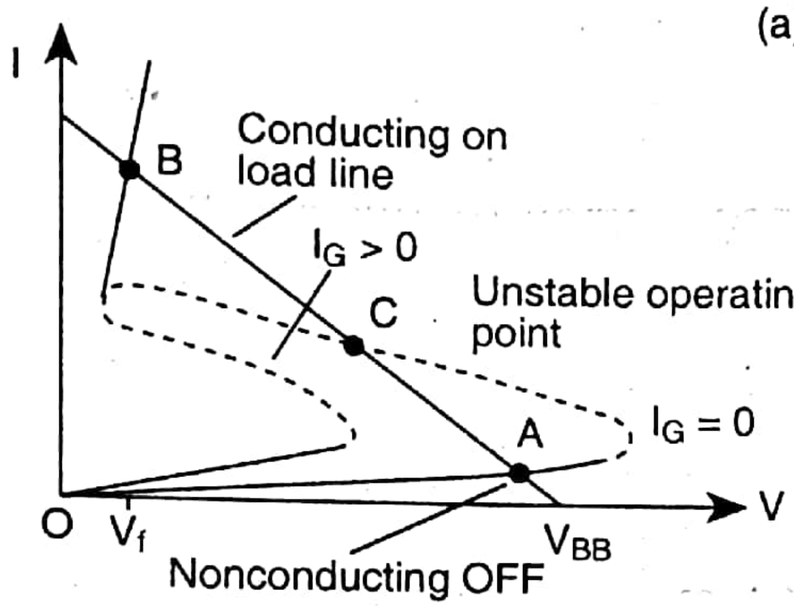
Substituting the value of I_{pE1} from eqn. (8.2.5) and I_{nE2} from (8.2.7) into the eqn. (8.2.6) we get

$$I = \alpha_{f_1}(I + \alpha_{r_1} I_{pC1}) - I_{pC1} + \alpha_{f_2}(I + I_G + \alpha_{r_2} I_{nC2}) - I_{nC2}$$

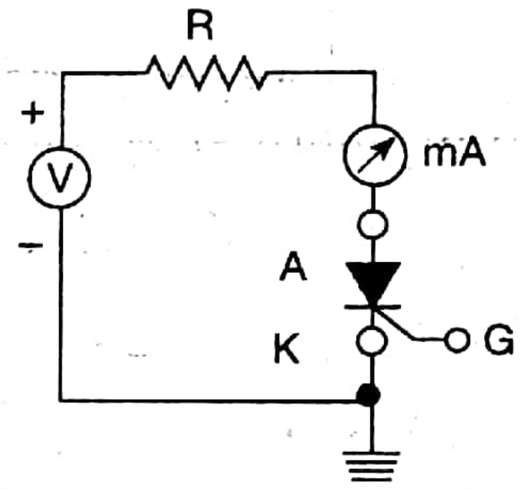
$$\text{or,} \quad -I(1 - \alpha_{f_1} - \alpha_{f_2}) + \alpha_{f_2} I_G = (1 - \alpha_{f_1}\alpha_{r_1}) I_{pC1} + (1 - \alpha_{f_2}\alpha_{r_2}) I_{nC2} \tag{8.2.8}$$



(a)



(b)



(c)

Fig. 8.16 SCR (a) Various junction currents (b) (U–V) characteristics (c) Symbol with circuit arrangement

Since $\alpha_{f_1}\alpha_{r_1}$ and $\alpha_{f_2}\alpha_{r_2}$ are each less than 1, the junction b_1b_2 will be reversed biased if

$$-(1-\alpha_{f_1}-\alpha_{f_2})I+\alpha_{f_2}I_G<0$$

and forward biased if

$$-(1-\alpha_{f_1}-\alpha_{f_2})I+\alpha_{f_2}I_G>0$$

Hence, switching will occur if

$$-(1-\alpha_{f_1}-\alpha_{f_2})I+\alpha_{f_2}I_G=0$$

$$\text{or,}\quad \alpha_{f_1}+\alpha_{f_2}+\alpha_{f_1}\frac{I_G}{I}=1 \tag{8.2.9}$$

Therefore the switching is easier if $I_G > 0$. Hence, when current is injected into the base region it may happen that there are *three* points of intersection (A, B and C) between the load line (assuming a resistance R in series with the battery, not shown in Fig. 8.16(b)), and the characteristic for $I_G = 0$; whereas there is only *one* point of intersection (B) if $I_G > 0$. It is easily seen that point C is an unstable operating point. Thus, if the operating point is at A (OFF state) for $I_G = 0$, then the operating point will switch to B (ON state) if I_G is switched on and is sufficiently large to satisfy the condition (8.2.9). If I_G is then turned OFF, the operating point will stay at the point B. The SCR can be turned ON by a small current I_G ($\sim mA$) because just before

turning ON the current I in eqn. (8.2.9) was reverse leakage current which is in μA range.

It is not necessary to maintain the gate current once the SCR switches to the conducting state; in fact, the gate essentially loses control of the device after regenerative transistor action is initiated. For most devices a gate current pulse lasting a few μs is sufficient to ensure switching. Ratings of minimum gate pulse height and duration are generally provided by the manufacturers for particular SCR devices.

Physical Explanation

The physical explanation for SCR operation is as follows :

Application of a positive voltage between anode and cathode will not help in conduction because the junction J_2 (Fig. 8.14(a)) will be under reverse bias and hence blocking, so long as $V_A < V_{B0}$ (Fig. 8.15). But looking at Fig. 8.14(c), we find that J_2 is the collector-base junction in both equivalent transistors Q_1 and Q_2 and reverse voltage across it is normal to transistor operation in the active region. Hence, current can flow in Q_1 if a base current is supplied i.e., if a current pulse is fed into the gate G_K (which is I_G in Fig. 8.16(a)) which serves as the base of Q_1. The resulting collector current of Q_1 serves as the base drive for Q_2; which also starts conduction then. The collector current of Q_2 serves as additional base derive for Q_1 and takes over when the *original gate trigger pulse ends*. This constitutes a positive feedback loop, and the two transistors very quickly become heavily saturated and the device turns ON. In saturation, the junction J_2 becomes forward biased (like any other collector junction of a saturated BJT) so that all three junctions J_1, J_2 and J_3 are now forward biased. The total forward voltage drop V_f of the SCR consists of the sum of these voltages, as in the case of Shockely diode, in addition to the ohmic voltage drop in the bulk of the SCR. However, J_1 and J_3 contribute a voltage of oppositive sign to that contributed by J_2, resulting in a total $V_f \sim 1$ volt depending on the triggering current I_G.

While the device conducts in the forward direction its current divides between the collectors of Q_1 and Q_2. If this current falls below the holding current I_H, then the two transistors come out of saturation and regain their amplifying properties. Any further reduction in current is immediately amplified by the positive feedback loop (reduction of one collector current reduces the base drive of the second—which, in turn, reduces collector current of the second—which again reduces the base drive of the first and so on). Both transistors are quickly ($\sim \mu s$) cut off and the SCR is turned OFF.

If a sinusoidal voltage is applied to the SCR anode, the device will be turned OFF once each alternate half cycle (i.e., when the voltage falls below the holding voltage) provided it is triggered ON regularly. The average rectified current can be varied over wide limits by controlling the point in each half cycle at which the SCR is turned ON. This is called *phase control* and is widely used in controlling ac power, for example, for heating, lighting, motor drive, electric welding, and a veriely of other industrial control applications (see section 8.2.5)

As an example of power control let us consider the circuit of a battery-charger regulator as shown in Fig. 8.16(d). $D1$ and $D2$ provide fullwave rectified dc voltage somewhat greater than the voltage of the battery to be charged. The medium-current SCR ($\approx 25A$) is in series with the battery being charged. The zener diode, V_z, provides

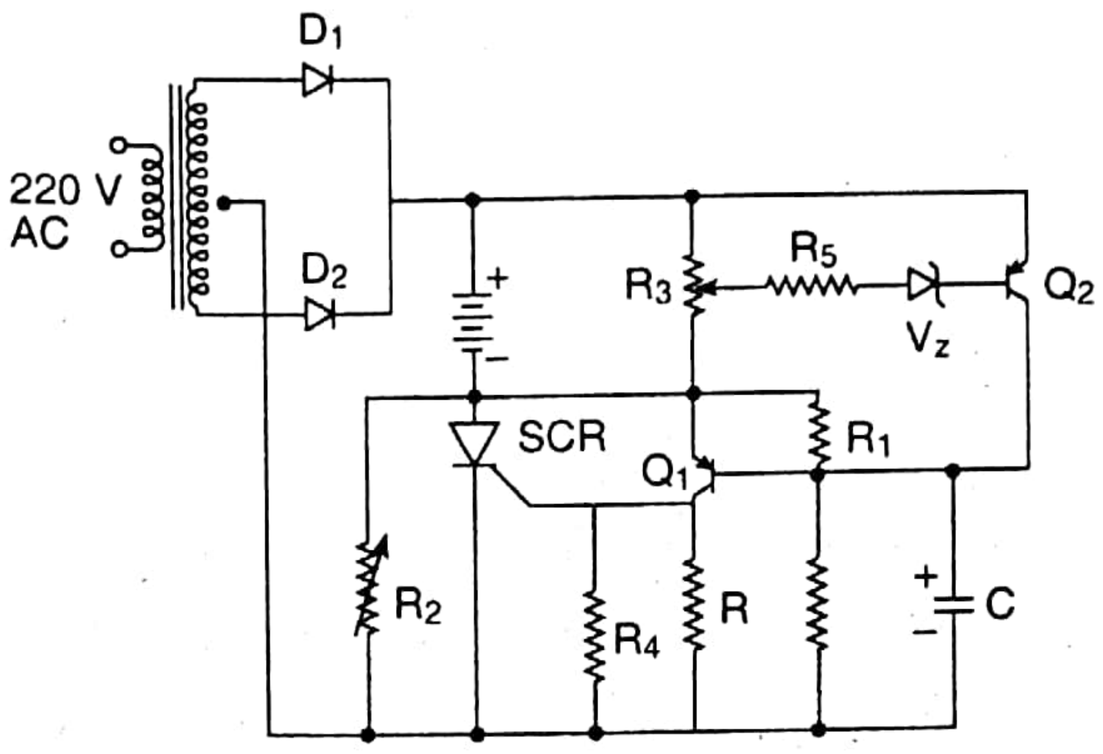


Fig. 8.16(d) Battery-charger circuit with SCR

a reference voltage for the transistor Q_2. If the battery voltage is lower than V_z, Q_2 will be cut off. At the same time, the emitter of Q_2 is sufficiently positive to conduct into the gate of the SCR, firing it and charging the battery. When the battery is sufficiently charged, Q_2 begins to conduct, reversing the voltage drop across R_1 and cutting off Q_1 and removing the gate signal to the SCR, resulting in the cut off of the SCR. R_2 allows a trickle charge to be maintained across the battery; R_5 and R_4 are used for current limiting. R and C provide Q_1 bias and the RC time constant during SCR firing. This circuit has the advantage of charge-to-trickle-charge changeover with tittle power loss.

8.2.3 DIAC

The diac (means *di*ode *ac* switch) and triac (*tri*ode *ac* switch) are bidirectional thyristors. They have ON and OFF states for positive or negative anode voltages and are therefore useful in ac applications.

The diac has mainly two structures :

(i) The ac trigger diode or $p-n-p$ diode switch.

(ii) The bilateral $p-n-p-n$ diode switch.

The former is simply a three layer device similar to a BJT without the base lead (Fig. 8.17(a)) and except that doping concentrations at the two junctions are approximately the same to have a symmetrical, bidirectional characteristics (Fig. 8.17(b)). No contact is necessary for the base region. When a voltage of any polarity is applied to a diac, one junction will be forward biased and the other reverse biased. The current will obviously be limited by the leakage current of the reverse-biased junction. At sufficiently large reverse bias avalanche multiplication occurs

followed by an electrical break-down at $V_{BR}(1-\alpha)^m$ (see Fig. 2.7. section 2.3.4), where V_{BR} is the avalanche break-down voltage of the $p-n$ junction, α is the common-base current gain, and m is a constant. As the current increases after break-down, α increases, causing a reduction in the terminal voltage. This reduction gives rise to a negative differential resistance.

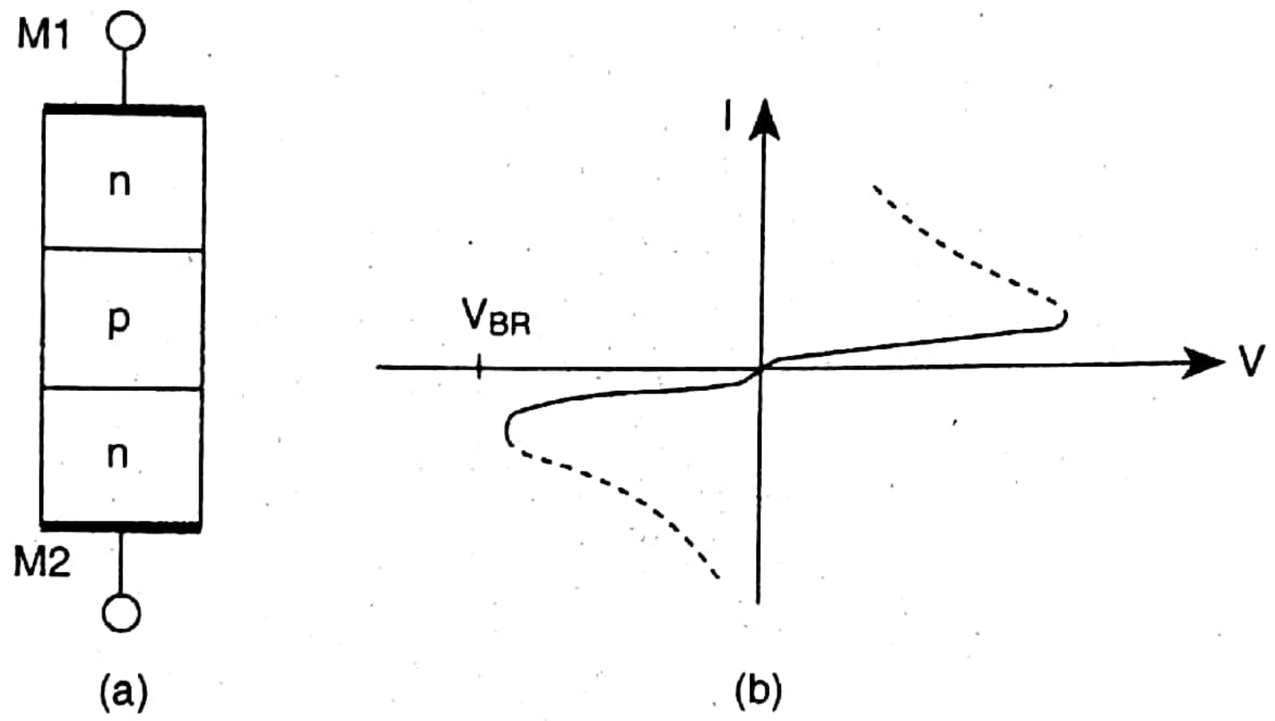


Fig. 8.17 AC trigger diode (a) Structure, (b) Typical characteristics of a diac

However, in practice, a diac is fabricated from $p-n-p-n$ thyristor for its greater efficiency and higher break-over voltages. The bidirection at $p-n-p-n$ diode switch behaves like two conventional Shockley diodes connected in antiparallel to permit the accomodation to voltage signals of two polarities, as in Fig. 8.18(a). Diac has two terminals, called the first main terminal $MT1$ and the second main terminal $MT2$.

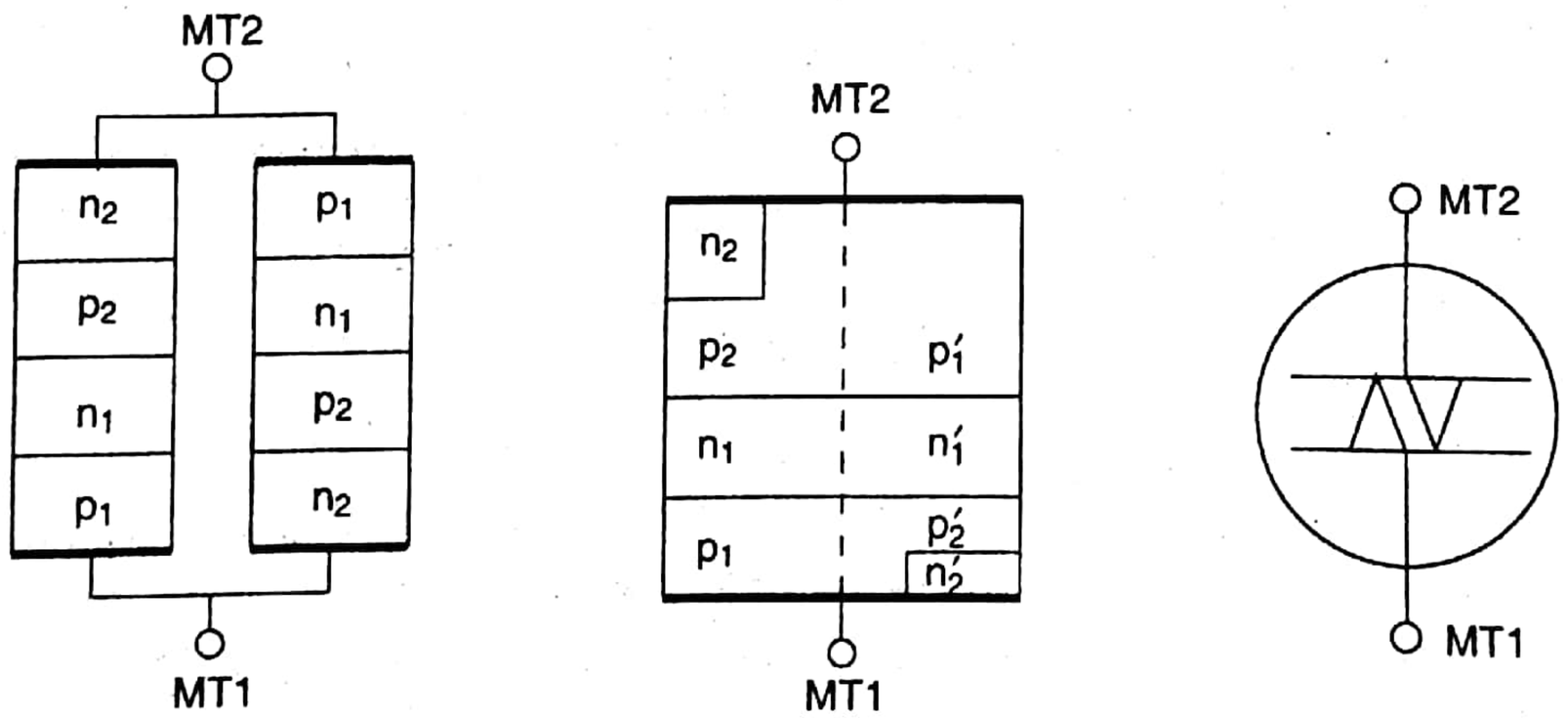


Fig. 8.18 Diac (a) Two Shockley diodes connected antiparallel, (b) integration of the diodes into a single two-port device, (c) Device symbol

This device has a short-circuited emitter structure, where n-type emitters are shorted to their neighbouring p regions by the overlying metallizations as shown in Fig. 8.18(b). The voltage across those junctions is therefore zero. The symmetry of this structure results in identical performance for either polarity of applied voltage. The circuit symbol is shown in Fig. 3.18(c) and the $I-V$ characteristics in Fig. 8.17(b). The diac

can be triggered into conduction by exceeding the break-over voltage. Because of its regenerative action by positive feedback, the bidirectional diode thyristor has a larger negative resistance and smaller forward voltage drop than that of an ac trigger diode (Fig. 3.17(a)).

8.2.4 TRIAC

AC power can be controlled by two SCRs, but since it is a very common requirement, a device consisting of two SCRs connected in inverse parallel has been developed. This is exactly a *triac*. The triac has three terminals, called the first main terminal $MT1$, the second main terminal MT2, and the gate G (Fig. 8.19(a)). Unlike SCR, the triac can conduct in both directions. It is very useful in motor speed control, temperature control, light dimmers, and other applications.

The triac structure is considerably more complicated than an SCR due to its five junctions J_1, J_2, J_3, J_4, J_5. In addition to the basic $p_1n_1p_2n_2$ structure, there are one junction gate n_3 and one n_4 region in contact with $MT2$. It may be noted that p_1 is shorted to n_4, p_2 to n_2 and n_3. Thus, like the diac, this device has also a short-circuited emitter structure, where n-type emitters are shorted to their neighbouring p regions by the overlying metallization. The voltage across those junctions is therefore zero. The gate metallization extends over two regions—n_3 and p_2.

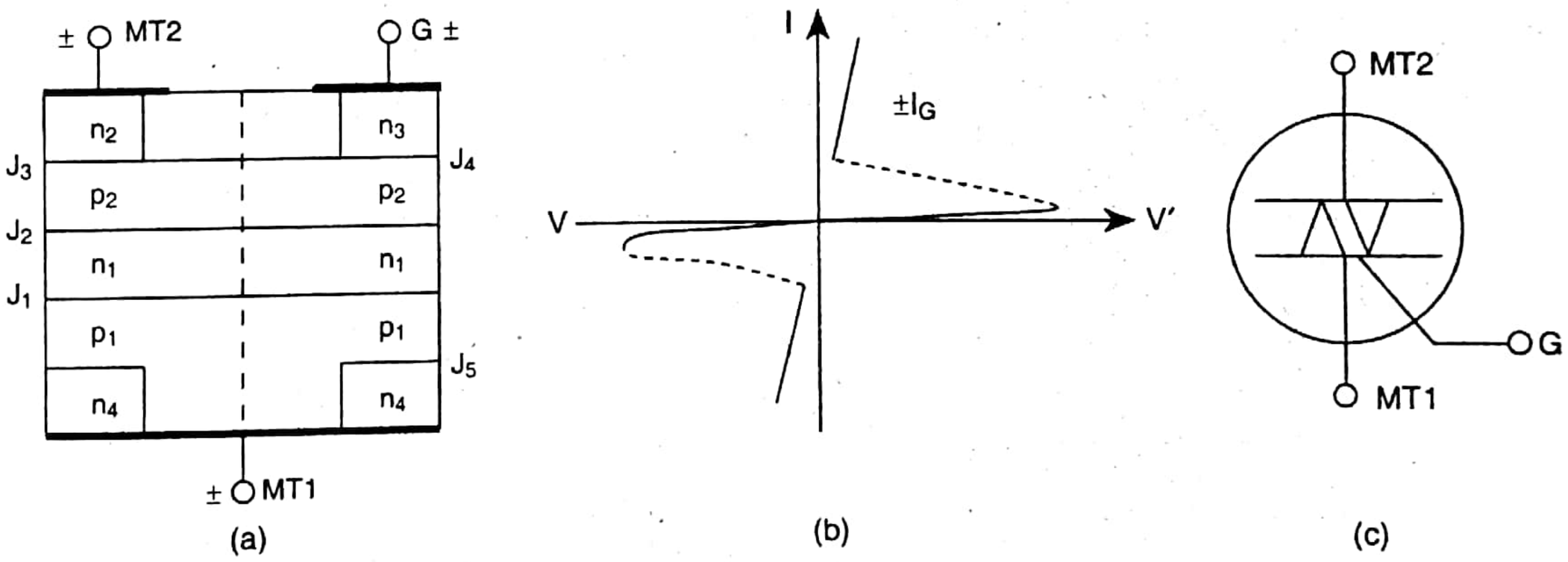


Fig. 8.19 Triac (a) Cross-section. (b) i–V characteristics, (c) Symbol

It is easily seen that there are four possible ways to trigger a triac : two polarities of overall voltage (i.e., $MT1$ is either + or −) and two possible gate currents (+ or −). When the main terminal $MT1$ is + with respect to $MT2$ and positive voltage (or current) is applied to the gate (also with respect to $MT2$), the device behaviour is identical to that of an SCR. The junction J_4 is reverse biased and is non-conducting : the gate current is supplied through the gate close to n_3 region; since junction J_5 is also reverse-biased and hence inactive, the main current flows through the left side of the $p_1 - n_1 - p_2 - n_2$ section.

Now, let a negative voltage is applied to the gate in the above case. The junction J_4 will now be forward-biased and electrons will be injected from n_3 to p_2. The auxilliary SCR $p_1-n_1-p_2-n_3$ will be turned ON by a flow of lateral base current in p_2 toward the n_3 region because of the increase in gain in the transistor $n_1-p_2-n_3$. Full conduction of this auxilliary SCR results in the current flowing out of this device and toward the n_2 region. This current provides the necessary gate current and trigger the left-side $p_1-n_1-p_2-n_2$ SCR into conduction.

We now consider that $MT1$ is negatively biased with respect to $MT2$, and the gate is positively biased. The junction J_3 will be forward biased between $MT2$ and the gate shorted to P_2. Electrons are injected from n_2 to p_2 and diffuse to n_1, resulting in an increase of forward bias of J_2 by the regenerative action, eventually full current is carried through the short at $MT2$. The gate junction J_4 is reverse-biased and is inactive. The full device current is carried through the right-hand side thyristor $p_2-n_1-p_1-n_4$.

Finally, $MT1$ is made negative with respect to $MT2$, and the gate is also made negative. Under this circumstance, the junction J_4 is forward-biased, and triggering is initiated by injection of electrons from n_3 to n_1 region. This action lowers the potential at n_1, causing holes to be injected from p_2 to n_1 region. This hole current provides the base drive for $p_2-n_1-p_1$ transistor, and the right-side thyristor $p_2-n_1-p_1-n_4$) is eventually turned ON. Since J_3 is reverse biased, the main current is carried from the short at $MT2$ through the n_4 region.

The $I-V$ characteristics and the circuit symbol of a triac are shown in Fig. 8.19(b) and (c) respectively. The triac can pass an ac current if triggered twice in each period. To turn if OFF the $MT1$–$MT2$ voltage should stay around zero for some minimum time, meaning that there is a maximum value of dv/dt (see section 8.2.1) that can be tolerated around $V=0$ (or a maximum sinusoidal wave frequency) so that the triac will not re-trigger itself.

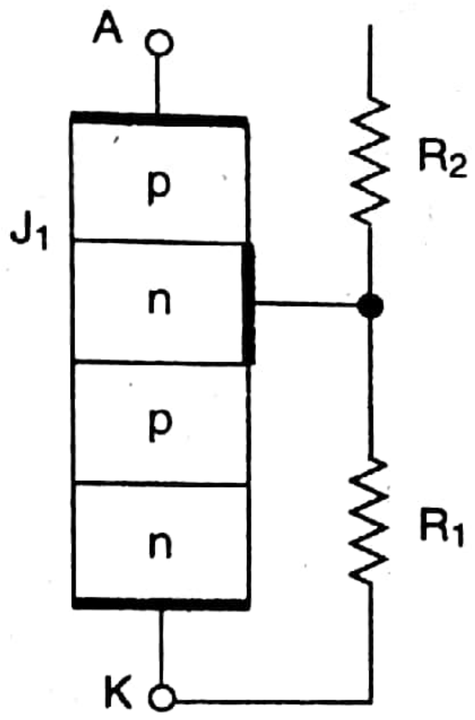
As in the SCR, the currents and temperatures have upper limits. At the maximum temperature (usually 125°C) the leakage currents are high (recall that for every 8°C rise in temperature the reverse saturation current is doubled i.e., it will be 2^{12} or 4096 times the value at 25°C) and the danger of unwanted triggering is to its highest. The thyristors are therefore mounted on heat sinks and have low junction-to-case thermal resistance.

8.2.5 Programmable Unijunction Transistor (PUT)

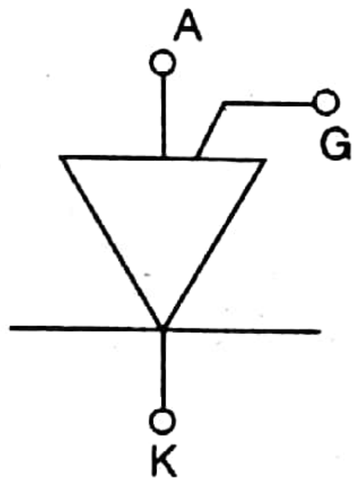
The conventional unijunction transistor has been discussed in section 7.4 as a negative resistance oscillator. But a *programmable unijunction transistor* (abbreviated as PUT) is of rather different internal construction. It is a four-layer $p-n-p-n$ structure having the same characteristics as the standard type. The difference is that the parameters η, R_{BB}, I_P, and I_V can be set by means of two external resistors, as shown in Fig. 8.20.

The structure of PUT is similar to that of an SCR except that the gate is brought out from the n-region adjacent to the anode, as shown in Fig. 8.20(a). It's symbol is shown in Fig. 8.20(b). This pn junction controls ON and OFF states of the device. The gate is always biased positive with respect to the cathode. When the anode voltage V_A exceeds the gate voltage V_G by approximately 0.7 V, the pn junction J_1 is forward biased and the PUT turns on. When the anode voltage falls below this level, the PUT turns OFF.

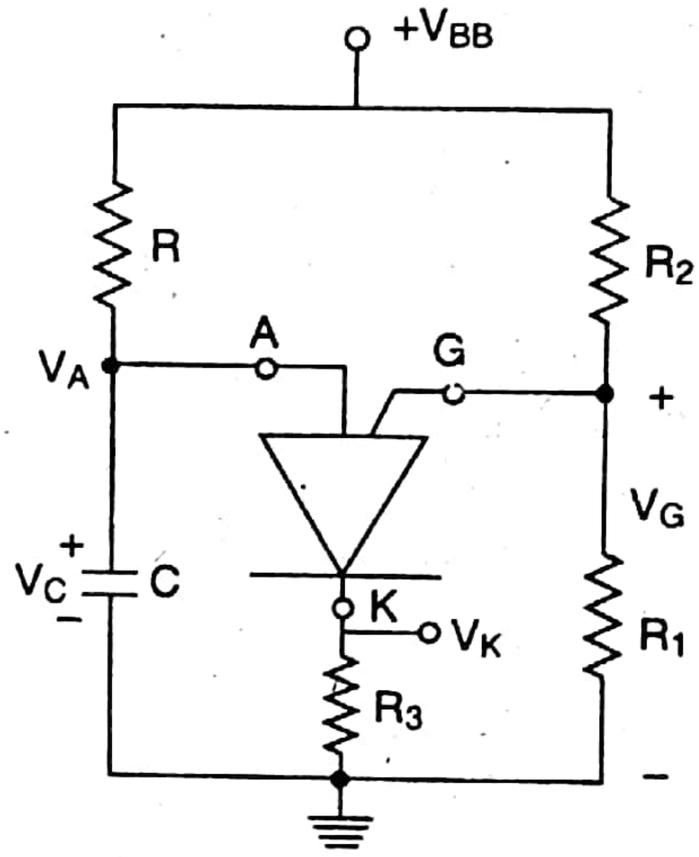
The gate can be biased to a desired voltage with an external voltage divider, as shown in Fig. 8.20(c), so that when the anode voltage exceeds this *'programmed'* level, the PUT turns ON. The anode current versus anode voltage characteristics are similar to UJT but can be more conveniently controlled. One such application is the relaxation oscillator (more commonly known as *saw-tooth generator*), as shown in Fig. 8.20(c).



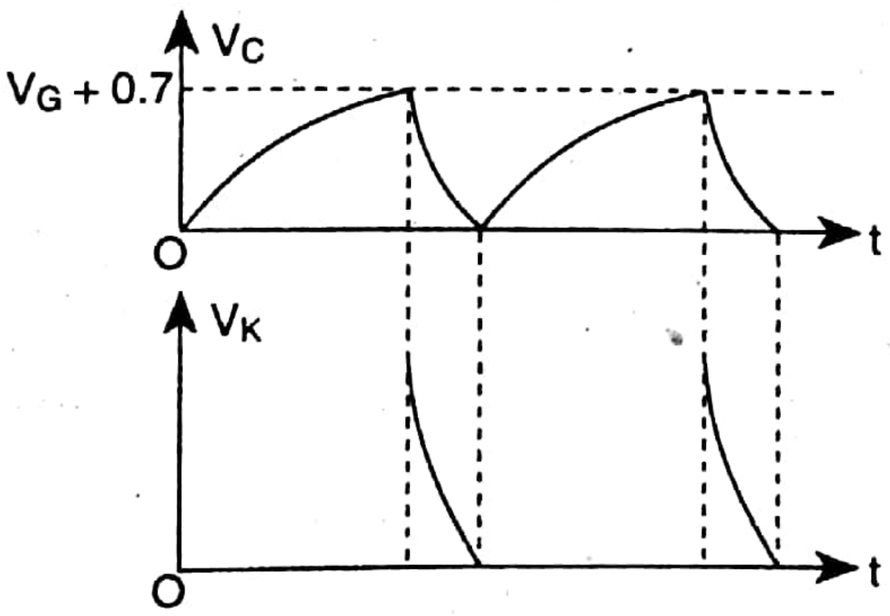
(a)



(b)



(c)



(d)

Fig. 8.20 Programmable unijunction transistor (PUT), (a) Basic structure, (b) Circuit symbol, (c) Relaxation oscillator circuit, (d) Output waveforms

The gate is biased at

$$V_G = \frac{R_1}{R_1 + R_2} V_{BB}$$

by the voltage divider network containing resistances R_1 and R_2. When dc power is applied, the PUT is OFF and the capacitor C starts to charge up to $+V_{BB}$ through the

resistance R. When the capacitor voltage V_C reaches $V_G + 0.7$ V, the PUT turns ON and the capacitor then discharges rapidly through the low ON resistance of the PUT and a small resistance R_3. A voltage spike is developed across R_3 during the discharge. As soon as the capacitor discharges completely the PUT turns OFF and the charging cycle starts on, as shown in the waveform shown in Fig. 8.20(d).

The intrinsic stand-off ratio (section 7.4) is

$$\eta = \frac{V_G}{V_{BB}} = \frac{R_1}{R_1 + R_2}$$

The period of oscillation T is given approximately by

$$T = \frac{1}{f} \approx RC \ln \frac{V_{BB}}{V_{BB} - V_G} = RC \ln \frac{1}{1-\eta} = RC \ln \left(1 + \frac{R_1}{R_2}\right)$$

The gate current I_G is given by

$$I_G = \frac{(1-\eta)}{R_G} \cdot V_{BB}$$

where $R_G = (R_1 \parallel R_2) = \frac{R_1 R_2}{(R_1 + R_2)}$

The restrictions on the choice of R_1 and R_2 can be found from the definition of η and R_G

$$R_2 = \frac{R_G}{\eta}$$

$$R_1 = \frac{R_G}{(1-\eta)}$$

η and R_G are normally supplied by the manufacturer.

8.3 CCD—A Novel MOS Device

Silicon MOSFET is probably the most important solid-state device for modern electronics. In the family tree of FET (section 4.1) we have just emphasized the importance of MOSFET. It is unfortunate that we have no scope to discuss them in detail. However, a device, which has not even been mentioned there, will be discussed here. It is *charge-coupled device* or simply CCD. These belong to a wider class of what are known as *charge transfer devices* (CTDs). The first CCD was fabricated in 1969, though the idea of a memory made up of a chain of capacitors was put forward as early as 1934.

To visualize the principle of operation of CCDs, let us imagine a MOS transistor with an extremely long channel and with many gates closely spaced between source and drain. Such an arrangement is in effect a long chain of MOS capacitors each of which is

formed by one of the gates and the substrate, and the input signal is transferred along this chain. Hence, these are *dynamic devices* in the sense that they move charge along a predetermined path under the control of clock pulse. In recent years, CCDs have come to be used in microelectronics as memory elements, delay lines, filters, signal-processing circuits, logic gates, and solid-state imaging devices (SSIDs) used instead of conventional TV tubes. Most recent use of them are in digital cameras. A special structure of a CCD that once charged can retain its gate charge *for years* is also in use and is known as *Floating* gate *Avalanche* injection MOS or *FAMOS* (see family tree in section 4.1). The gate in the FAMOS is made of doped polysilicon completely embedded in SiO_2 and electrically unconnected, i.e., its potential is floating.

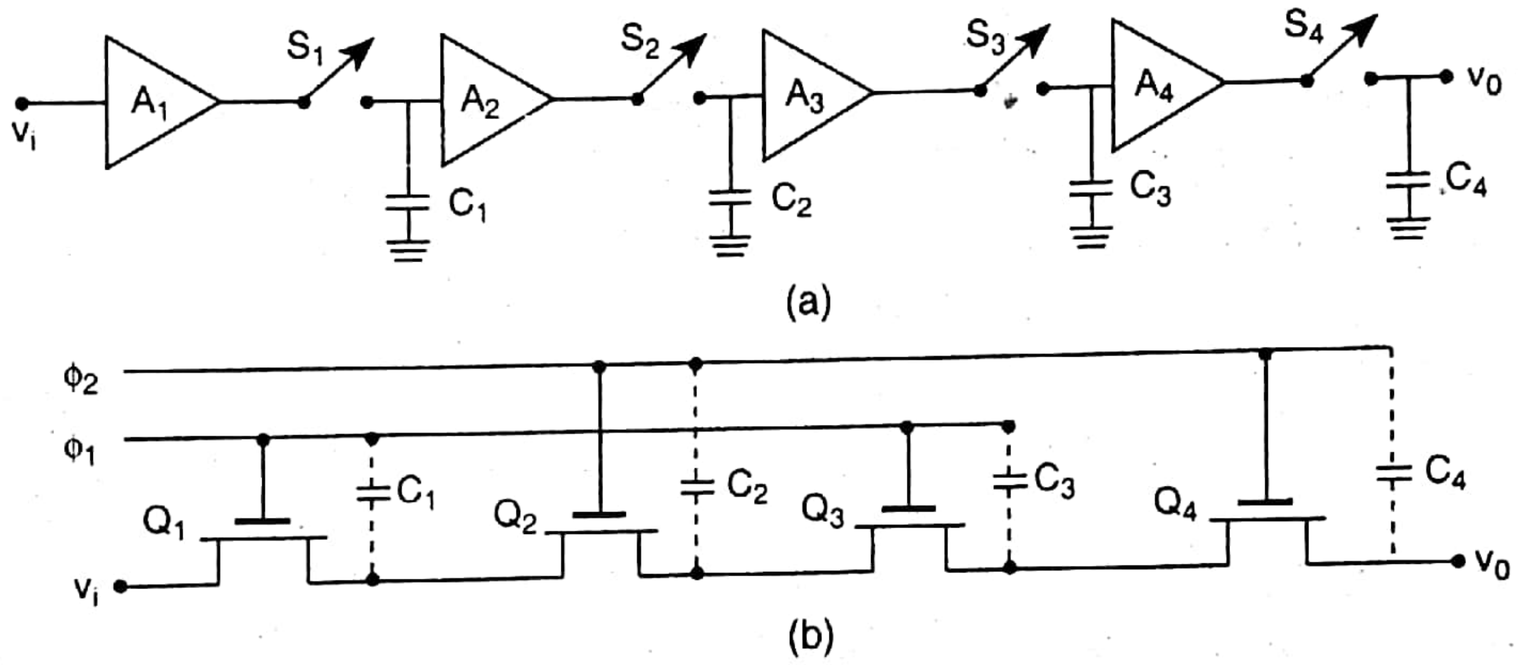


Fig. 8.21 Charge transfer systems using (a) OP AMPs, (b) MOST

The concept of charge transfer can be explained by using a chain of amplifiers with unity gain and infinite input impedance connected as shown in Fig. 8.21(a). On closing the switch S_1, the input signal is stored in the form of a charge pack $Q = \int CdV$ in the capacitor C_1. Now, let the switch S_1 be opened and then S_2 is closed; obviously the stored charge will be transferred to the capacitor C_2. Following the same procedure, the charge will eventually reach the output terminal and come out as a voltage v_0. In Fig. 8.21(b), each amplifier-and-switch pair is replaced by a MOS transistor. The transistors can be turned ON and OFF sequentially by applying clock pulses at the respective gate electrodes, and the charge is stored and transferred just as in Fig. 8.21(a). In practice, gates 1 and 3 are connected and pulsed: in a similar way gates 2 and 4 are joined. The circuit shown in Fig. 8.21(b) is known as the *bucket-brigade device*. It is a *two-phase* system because two clock pulses ϕ_1 and ϕ_2 are used separately.

A cross-section of a typical three-phase CCD is shown in Fig. 8.22(a). Here, the minority carriers are stored in potential wells created at the surface of a semiconductor. There carriers are transported along the interface of insulator (SiO_2) and semiconductor by filling and emptying a series of potential wells sequentially. Hence, they are called *surface-channel* CCD (or simply SCCD). There is another class of CCDs called *buried-channel* CCD (or BCCD) in which the doping of the semiconductor substrate is modified so that the storage and transfer of the charge packet takes place in the bulk of the

semiconductor just beneath the semiconductor surface. However, we shall consider only SCCD.

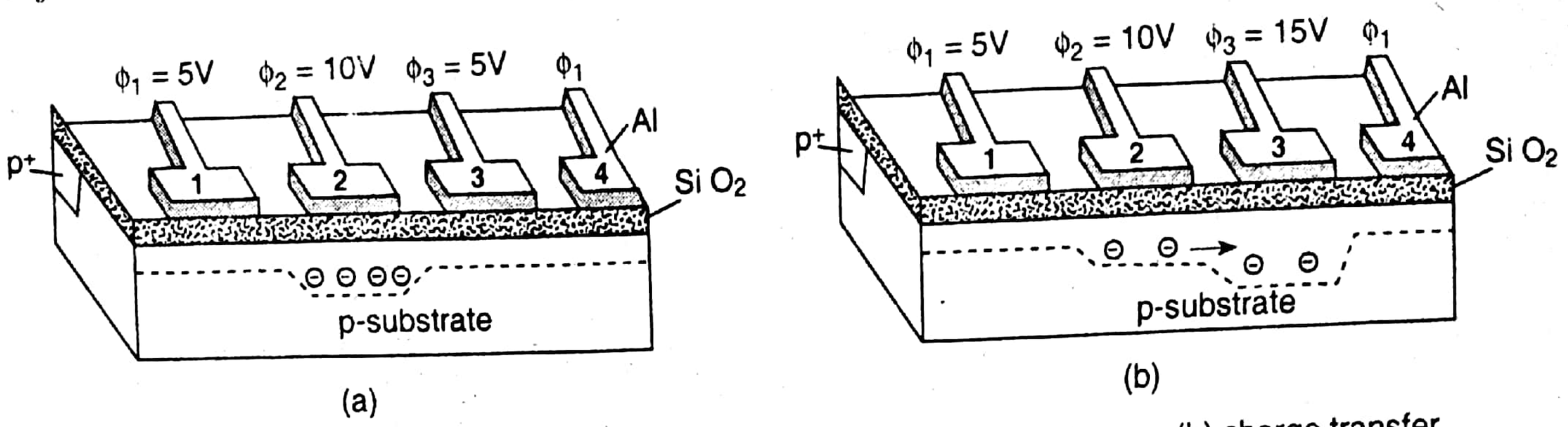


Fig. 8.22 Basic operation of a three-phase CCD, (a) charge storage, (b) charge transfer

In its simplest form, the CCD is a string of closely spaced array of MOS capacitors (or diodes) on a continuous insulator (SiO_2) layer that covers the semiconductor substrate, as in Fig. 8.22. If electrode 2 is biased at, say 10 V, more positively than its two adjacent electrodes (at 5 V), a *potential well* is set up, as depicted by the dotted line, and charge is stored under this electrode, as in Fig. 8.22(a). Now if the electrode 3 is biased with 15 V, a deeper potential well is established under electrode 3 (Fig. 8.22(b)). The stored charge seeks the lowest potential and therefore travels along the surface when the potential wells are moved electronically. It may be noticed that three electrodes are needed in this structure to facilitate charge storage and transfer in one direction only. The three electrodes will be referred to as *one stage* or *cell* of the device.

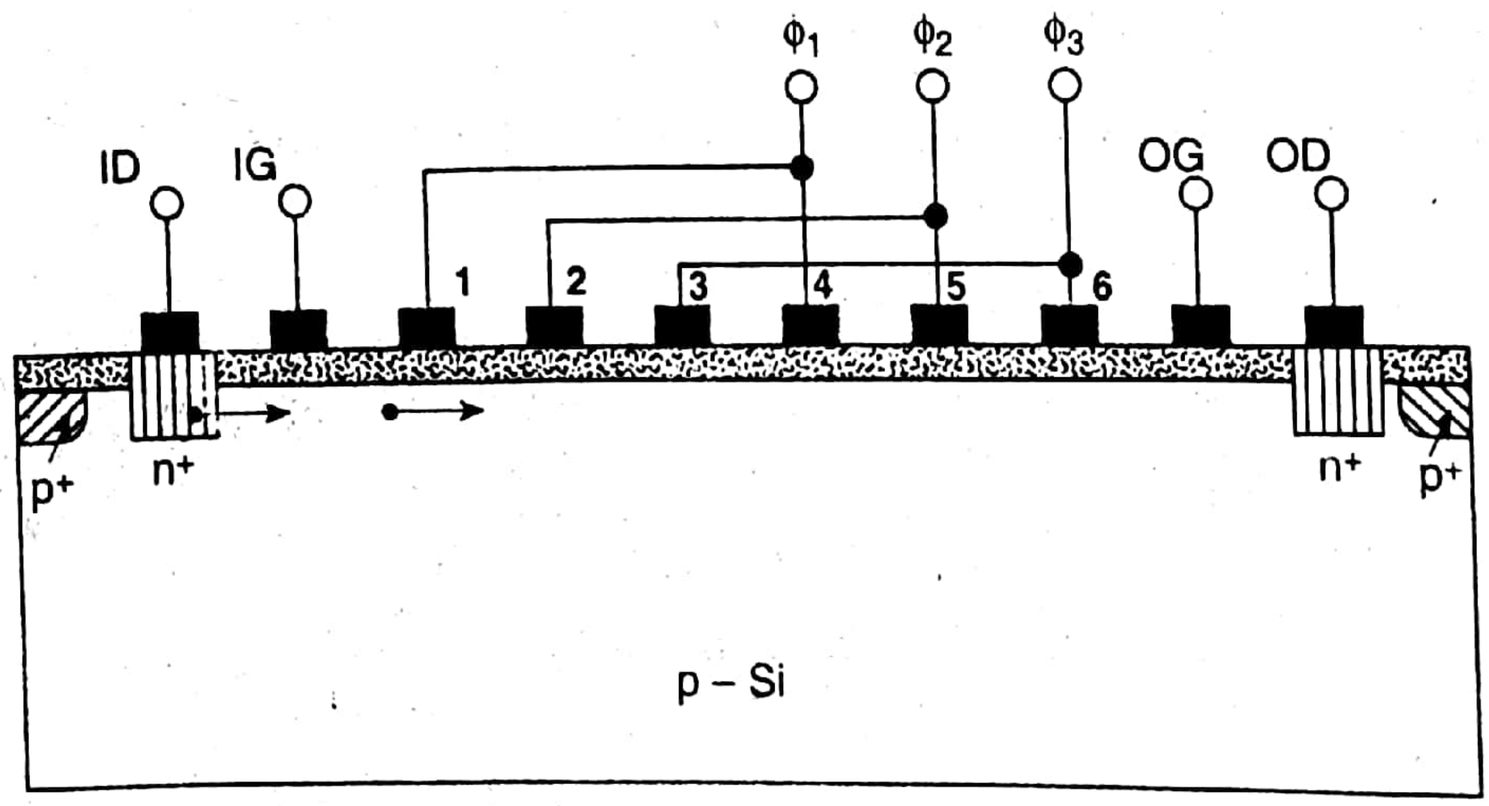


Fig. 8.23 IC using three-phase charge-coupled devices (CCDs)

A three-phase, n-channel CCD with three-electrodes per stage together with its input and output structures is shown in Fig. 18.23. The six MOS capacitors connected to the ϕ_1, ϕ_2 and ϕ_3 clock lines form the main body of the CCD. The input diode (ID), input gate (IG), output diode (OD), and output gate (OG) are elements that inject and detect charge packets to and from the main CCD. In designing a CCD, the capacitors must be *physically close* together so that the depletion layers overlap

strongly and the surface potential has a smooth transition at the boundaries between neighbouring electrodes. Typically, the oxide thickness is between 1000Å and 2000 Å, and the spacing between alluminium gate is $\lesssim 2.5\ \mu$m.

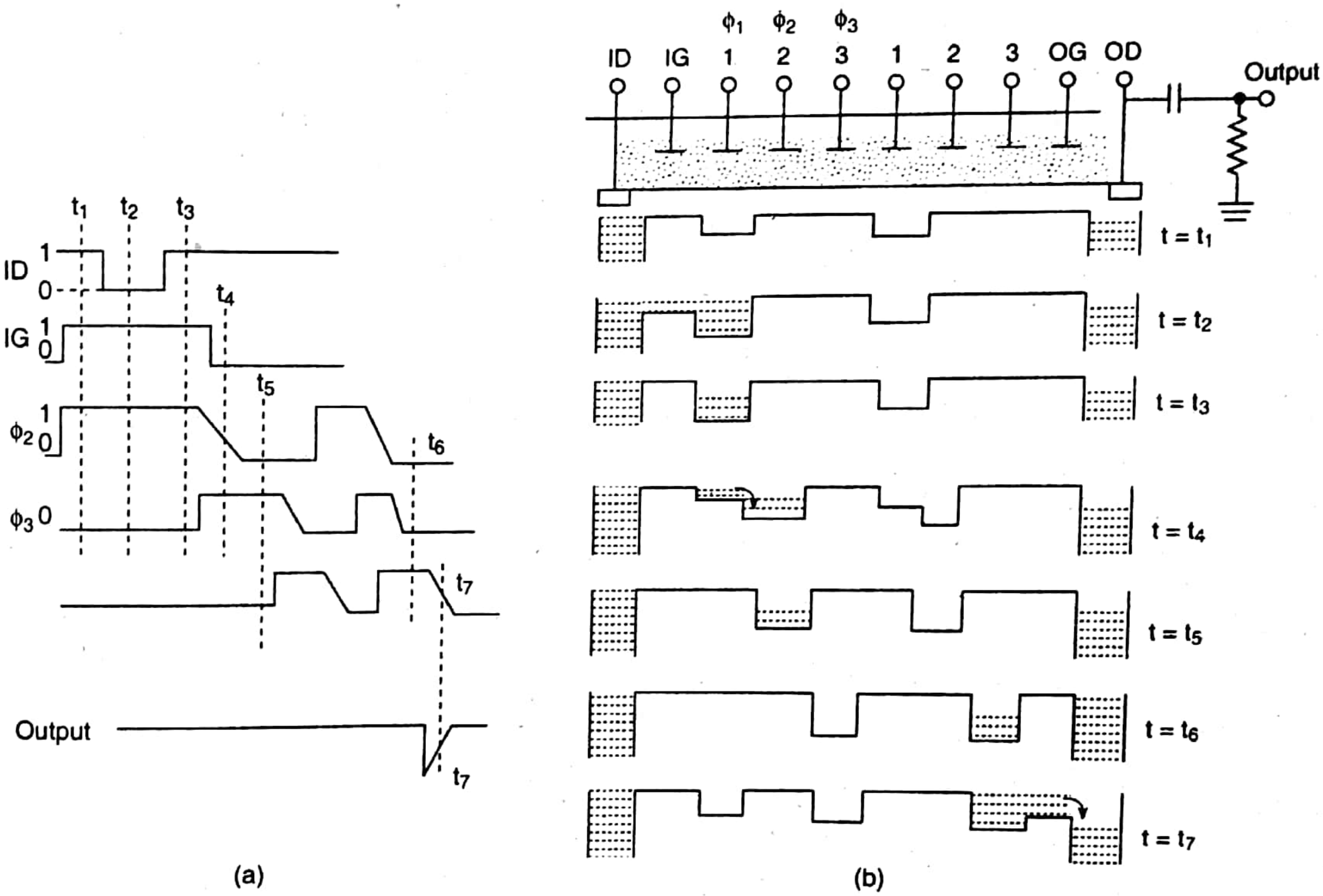


Fig. 8.24 Charge-coupled device, (a) clock pulses, (b) potential wells

Fig. 8.24(a) shows the clock waveforms and output signal for the CCD, and Fig. 8.24(b) illustrates the corresponding potential wells and charge distributions. In practice, the driving clock pulses with special features shown in Fig. 8.24(a) are designed to achieve better efficiency in charge transfer, as explained in the following paragraph.

The input electrode (ID) injects electrons across the n^+p junction. The injected current can be varied (or controlled) by varying the input gate (IG) voltage. Charge transfer from electrode 1 to electrode 2 and so on can take place only if the voltage applied to the electrodes is positive and the voltage at the next electrode exceeds that at the preceding one. Clock pulses are usually 10–20 V in magnitude.

At $t = t_1$, the clock line ϕ_1 is at a high voltage whereas ϕ_2 and ϕ_3 are at 0 V. ID and OD are biased with high positive voltages to prevent inversion of the surface under the IG and OG. Thus the surfaces under IG and OG are in deep depletion, and ID and OD cannot supply electrons into the main CCD chain. In other words, all potential wells under the CCD chain are *empty*. The potential well under ϕ_1 will be deeper than those under ϕ_2 and ϕ_3.

At $t = t_2$, the voltage of ID is lowered so that electrons can flow (inject) to the potential well under the first ϕ_1 electrode through IG. At the end of injection, the surface potentials under IG and ϕ_1 electrodes will be the same as the ID potential. Electrons will now be stored under IG and the first ϕ_1 electrode (No. 1).

At $t = t_3$, the voltage of ID is returned to a high value; electrons under IG and the excess electrons under gate 1 will be taken out of the device through ID head, creating a well-defined charge packet under electrode 1.

At $t = t_4$, the voltage applied to ϕ_1 is returning to the low value while the ϕ_2 electrodes have high voltage applied to them. The electrons stored under ϕ_1 are then transferred to the ϕ_2 electrode (No. 2) because the surface potential under ϕ_2 is higher. This process is called *charge transfer*. The voltage on ϕ_1 should have a *slowly* falling edge, because the charge carriers require a *finite time* to transport across the width of an electrode.

At $t = t_5$, the charge transfer process is completed and the original charge packet is now stored under the first ϕ_2 electrode (No. 2). This process is repeated and at $t = t_6$, the injected charge packet is stored under the second ϕ_3 electrode (No. 6).

At $t = t_7$, the voltage of ϕ_3 electrodes is returning to low value and pushing the electrons to the output diode (OD), thereby giving an output signal proportional to the size of the charge packet at the output terminal.

For analog and memory devices, the charge packets are introduced by applying suitable voltages to a $p - n$ junction at the input of the CCD. For optical imaging applications (such as digital camera), the charge packets are formed as a result of electron-hole pair generation caused by photons (light) incident on the semiconductor substrate. The magnitude of the output signal will be proportional the light insensity.

Charge storage in a CCD is limited by thermal electron-hole pair generation. The carriers so produced are trapped in the potential wells and, in time, may change the output giving a false information. This phenomenon is known as *dark current effect* and sets the lower frequency limit ($\sim$ kHz) and the maximum storage time ($< 100\mu$s) for CCDs. This is minimized at lower temperatures.

Charge transfer from one electrode to the next takes place by the drift mechanism whereas the process terminates as a diffusion of carriers. The fact that diffusion speed is sustantially lower than the drift velocity limits the operating speed of a CCD. However, the maximum frequency that can be obtained is $\sim$ 1 GHz.

Two-phase clocking, instead of three as described above, is possible if nonplanar electrodes are used and each CCD cell has two instead of three electrodes.

CCD video cameras have been produced with the size of a small cigarette pack, weighs as little as 200 gm (mainly due to mechanical attachements), and draws a mere 2W of power. The principle is based on solid-state imaging arrays. These imagers are simple in design and to make, small in size, light in weight, dissipate little power, have high sensitivity, and are able to operate in visible, infrared, and ultraviolet light. Those are already in use present an image as an array of several hundred thousand *pixels* a short form of *picture element*.