

3

Power Switching Devices and their Static Electrical Characteristics

There is a vast proliferation of power switching semiconductor devices, each offering various features, attributes, and limitations. The principal device families of concern in the power switching semiconductor range are the diode, transistor, and thyristor. Each family category has numerous different members. The basic characteristics of the three families and a range of their members will be analysed.

3.1 Power diodes

The diode is the simplest semiconductor device, comprising one pn junction. In attempts to improve both static and dynamic diode electrical properties for different application conditions, numerous diode types have evolved.

3.1.1 The pn fast-recovery diode

The doping concentration on each side of the junction influences the avalanche breakdown voltage, the contact potential, and the series resistance of the diode. The junction diode normally has the p-side highly doped compared with the n-side, and the lightly doped n-region determines many of the properties of the device. The n-region gives the device its high-voltage breakdown and under reverse bias, the scl penetrates deeply into the n-side. The lower the n-type concentration and the wider the n-side, the higher will be the reverse voltage rating and also, the higher the forward resistance. These n-region requirements can lead to thermal I^2R problems in silicon. Larger junction areas help reduce the thermal instability problem.

It is usual to terminate the lightly doped n-region with a heavily doped n⁺ layer to simplify ohmic contact and to reduce the access resistance to the scl. For better n-region width control, n-type silicon is epitaxially grown on an n⁺ substrate. The p⁺ anode is diffused or implanted into the epitaxial region, forming an epitaxial diode.

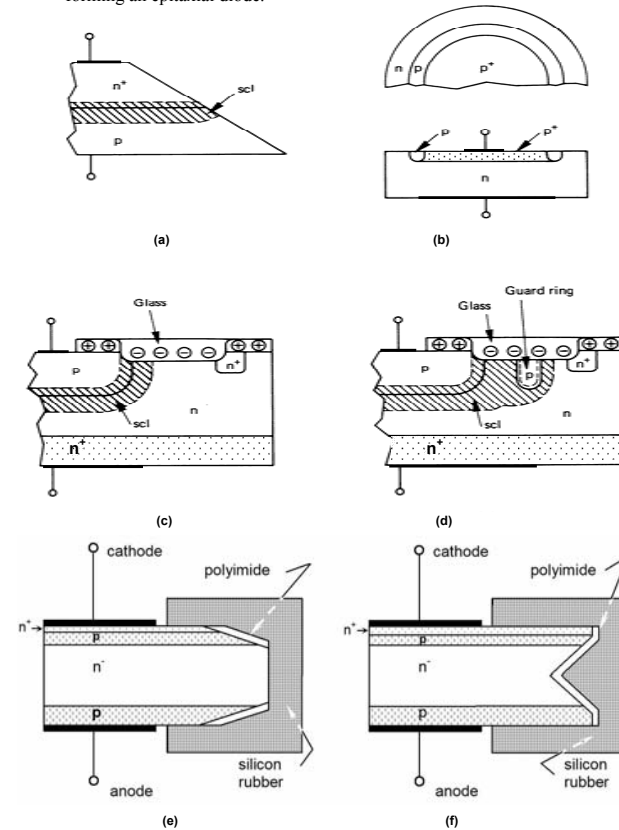


Figure 3.1. To prevent edge breakdown under junction reverse bias: (a) reduction of the space charge region near the bevel; (b) p-type guard ring; (c) glass guard ring; (d) glass plus p-type guard ring; (e) double negative bevel; and (f) double positive bevel angle.

In devices specifically designed for high reverse bias applications, care must be taken to avoid premature breakdown across the edge of the die or where the junction surfaces. Premature edge breakdown is reduced by *beveling* the edge as shown in figure 3.1a, or by diffusing a *guard ring* as shown in figure 3.1b, which isolates the junction from the edge of the wafer. The scl electric field is lower at the bevelled edge than it is in the main body of the device. In the case of a lightly doped p-type guard ring, the scl is wider in the p-ring, because of its lower concentration, than in the p^+ region. The maximum electric field is therefore lower at the pn-ring junction for a given reverse bias voltage. Negatively charged glass film techniques are also employed to widen the scl near the surface, as shown in figures 3.1c and 3.1d. Multiple guard rings are sometimes employed for very high breakdown voltage devices. Similar techniques are extendable to devices other than diodes, such as thyristors. Field control beveling on more complex junction structures is achieved with double-negative or double-positive beveling as shown in parts e and f of figure 3.1. The beveling is accomplished by grinding, followed by etching of the bevel surface to restore the silicon crystalline mechanical and structure quality. The processed area is passivated with a thin layer of polyimide, which is covered in silicon rubber. Negative bevels tend to be more stable electrically with ageing.

The foregoing discussion is directly applicable to the rectifier diode, but other considerations are also important if fast switching properties are required. The turn-on and reverse recovery time of a junction are minimised by reducing the amount of stored charge in the neutral regions and by minimising carrier lifetimes. *Lifetime killing* is achieved by adding gold or platinum, which is an efficient recombination centre. Electron and proton irradiation are preferred non-invasive lifetime control methods. Irradiation gives the lowest forward recovery voltage and the lowest reverse leakage current. The improved switching times must be traded off against increased leakage current and on-stage voltage. Switching times are also improved by minimising the length (thickness) of the n-region.

3.1.2 The p-i-n diode

The transient performance of diodes tends to deteriorate as the thickness of the silicon wafer is increased in attaining higher reverse voltage ratings. Gold lifetime killing only aggravates the adverse effects incurred with increased thickness. The p-i-n diode allows a much thinner wafer than its conventional pn counterpart, thus facilitating improved switching properties.

The p-i-n diode is a pn junction with a doping profile tailored so that an intrinsic layer, the i-region, is sandwiched between the p-layer and the n-layer, as shown in figure 3.2. In practice, the idealised *i-region* is approximated by a high resistivity n-layer referred to as a *v-layer*. Because of the low doping in the v-layer, the scl will penetrate deeply and most of the reverse bias potential will be supported across this region.

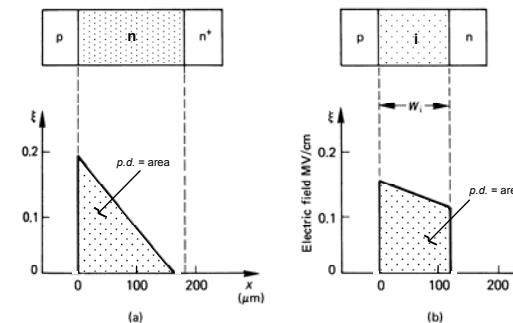


Figure 3.2. Cross-section and electric field distribution of: (a) a pn diode and (b) a p-i-n diode.

The power p-i-n diode can be fabricated by using either the epitaxial process or the diffusion of p and n-regions into a high-resistivity semiconductor substrate. The i-region width W_i specifies the reverse voltage breakdown of the p-i-n diode, which is the area under the electric field in figure 3.2b, viz.,

$$V_b \approx \xi_p W \approx 25 W_i \text{ (in } \mu\text{m)} \quad (\text{V}) \quad (3.1)$$

The thickness W_i , along with the distribution of any gold within it, determines the nature of the reverse and forward-conducting characteristics. These characteristics are more efficient in fast p-i-n diodes than in the traditional pn structures.

3.1.3 The power Zener diode

Zener diodes are pn diodes used extensively as voltage reference sources and voltage clamps. The diode reverse breakdown voltage is used as the reference or clamping voltage level.

The leakage current in a good pn diode remains small up to the reverse breakdown point where the characteristic has a sharp bend. Such a characteristic is called *hard*. Premature breakdown at weak spots in the junction area or periphery cause high leakage currents before final breakdown, and such diodes are said to have *soft* breakdown characteristics.

Zener diodes are especially made to operate in the breakdown range. Above a few volts, the breakdown mechanism is avalanche multiplication rather than Zener and the breakdown reference voltage V_Z is obtained by proper selection of the pn junction doping levels. Once in breakdown V_Z remains almost constant provided the manufacturer's power rating, $P = V_Z I$, is not exceeded. Where the breakdown

mechanism is due to the Zener effect, the temperature coefficient is negative, about -0.1 per cent/K, changing to positive, +0.1 per cent/K, after about 4.5V when the avalanche multiplication mechanism predominates.

Zener diodes require a hard breakdown characteristic not involving any local hot spots. They are available in a voltage range from a few volts to about 280V and with power dissipations ranging from 250mW to 75W, with heat sinking. Transient suppressing Zener diodes can absorb up to 50kW, provided energy limits and number of cycles are not exceeded, as shown in figure 10.19.

Practically, Zener diodes are difficult to make, less than ideal in application, and should be avoided if possible. The basic $I-V$ characteristics, and electrical circuit symbol for the different types of diodes, are shown in figure 3.3.

3.1.4 The Schottky barrier diode

The Schottky diode is a metal-semiconductor device which offers low on-state voltages, but in silicon is presently restricted to applications imposing a reverse bias of less than 400V. At lower voltages, less than 40V, devices of up to 300A are available and the maximum junction operating temperature is 175°C, which is higher than for conventional silicon pn junction devices.

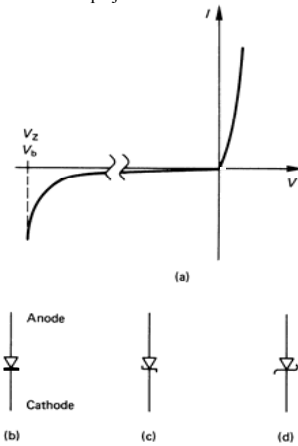


Figure 3.3. Diodes: (a) static $I-V$ characteristic; (b) symbol for a rectifier diode; (c) voltage reference or Zener diode; and (d) Schottky barrier diode.

The Schottky diode is formed by a metal (such as chromium, platinum, tungsten or molybdenum) in homogeneous contact with a substrate piece of n-type silicon, as shown in figure 3.4a. The contact is characterised by a potential barrier Φ_b

which determines the forward and reverse properties of the Schottky diode.

In forward conduction, electrons are emitted from the negative potential n-type silicon to the positive potential metal, passing over the barrier potential. Unlike the bipolar pn diode, only electrons are carriers, hence the Schottky barrier diode is a unipolar device. The forward on-state voltage drop is dominated by and proportional to the barrier potential Φ_b , while unfortunately the reverse leakage current is approximately inversely related. Thus a Schottky diode with a very low forward voltage drop will have very high reverse leakage current relative to the pn diode counterpart, as shown in figure 3.5.

Chromium provides the lowest forward voltage drop but is limited to an operating temperature of 125°C and has a high leakage current. Platinum allows operating temperatures to 175°C with a leakage current several orders of magnitude lower than chromium. The trade-off is a higher forward voltage.

A guard ring is used to improve device robustness, but its function is to act like a Zener diode and thus protect the Schottky barrier under excessive reverse bias. An optimally designed epitaxial layer, as shown in figure 3.4b, is also employed which reduces the field at the less than perfect metal-semiconductor interface and allows the whole interface to go safely into reverse bias breakdown.

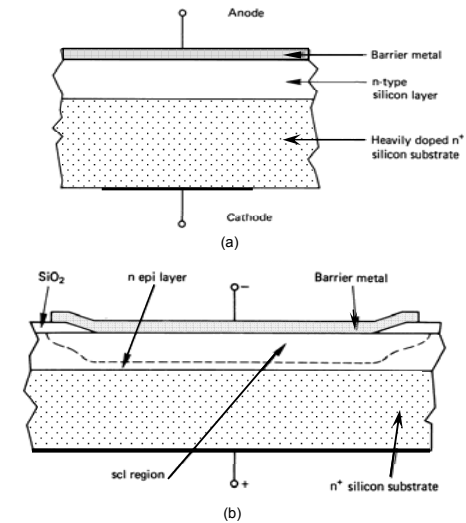


Figure 3.4. The Schottky barrier diode: (a) the basic structure and (b) the space charge layer region extending into the epi-substrate region under reverse bias.

There are a number of important differences between Schottky barrier and pn junction diodes.

- In a pn diode, the reverse bias leakage current is the result of minority carriers diffusing into the scl and being swept across it. This current level is highly temperature-sensitive. In the Schottky-barrier case, reverse current is the result of majority carriers that overcome the barrier. A much higher value results at room temperature, but is not temperature-dependent.
- The forward current is mostly injected from the n-type semiconductor into the metal and very little excess minority charge is able to accumulate in the semiconductor. Since minimal minority carrier recombination occurs, the Schottky barrier diode is able to switch rapidly from forward conduction to reverse blocking.
- Since under forward bias, barrier injection comes only from the semiconductor, and there is little recombination in the scl; thus the device can be well represented by the ideal diode equation (2.6).
- The majority electrons injected over the barrier into the metal have much higher energy than the other metal electrons which are in thermal equilibrium. Those injected electrons are therefore called *hot*, and the diode in some applications is referred to as a *hot electron diode*.

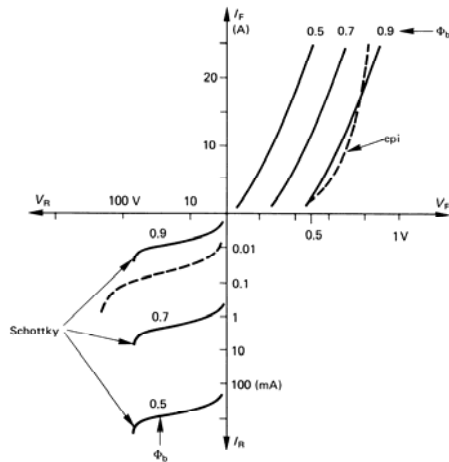


Figure 3.5. Schottky and epi diode I-V characteristics with different Schottky barrier potentials.

An important point arising from this brief consideration of the Schottky barrier diode is the importance of the connection of an n-type region to aluminium metallization that occurs in unipolar and bipolar semiconductor devices. A practical method of forming aluminium ohmic contacts on n-type materials is by doping the semiconductor very heavily, above the degeneracy level. Thus, in the contact region, if a barrier exists, the scl width is small enough to allow carriers to tunnel through the barrier in both directions. On the other hand, aluminium makes a good ohmic contact on p-type silicon since the required p⁺ surface layer is formed during the heat treatment of the contact after the aluminium is deposited. An ohmic contact acts as a virtual sink for minority carriers, because it has an enormous supply of majority carriers.

3.1.5 The silicon carbide Schottky barrier diode

Silicon carbide Schottky diodes are attractive for high voltages because the field breakdown of silicon carbide is eight times that of silicon. Additionally the wide band gap allows higher operating temperatures. Both nickel and titanium can be used as Schottky metals. Boron atoms (a dose of $1 \times 10^{15} / \text{cm}^2$ at 30keV) are implanted to form the edge termination that spreads any field crowding at the edge of the metal contact, as shown in figure 3.6. The lower barrier height of titanium produces a lower forward voltage device, but with a higher reverse leakage current, than when nickel is used as the barrier metal.

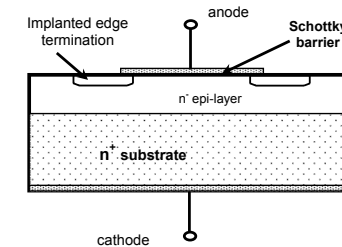


Figure 3.6. The silicon carbide Schottky barrier diode structure.

3.2 Power switching transistors

Two types of transistor are extensively used in power switching circuits, namely the power metal oxide semiconductor field effect transistor (MOSFET) and the insulated gate bipolar transistor (IGBT). The IGBT has a bipolar junction transistor (BJT) output stage and a MOSFET input stage, in an integrated *Darlington* pair configuration. Many of the IGBT power handling properties are associated with the limitations of the BJT. Thus some attention to the BJT's electrical characteristics is necessary, even though it is virtually obsolete as a discrete power-switching device.

- The **BJT** consists of a pnp or npn single-crystal silicon structure. It operates by the injection and collection of minority carriers, both electrons and holes, and is therefore termed a *bipolar transistor*.
- The **MOSFET** depends on the voltage control of a depletion width and is a majority carrier device. It is therefore a *unipolar transistor*.
- The **IGBT** has the desirable voltage input drive characteristics of the MOSFET but the power switching disadvantages of the minority carrier mechanisms of the BJT.

3.2.1 The bipolar npn power switching transistor (BJT)

As a discrete electrical device, the high-voltage, power-switching bipolar junction transistor, BJT, is virtually obsolete. The BJT has one unique redeeming electrical characteristic, viz.; it can conduct hundreds of amperes with an extremely low on-state voltage of less than 100mV, when saturated. Although superseded, its basic electrical operating characteristics are fundamental to the operation of most other power switching devices. Specifically, the MOSFET has a parasitic npn BJT, as shown in figure 3.14, that can cause false turn-on other than for the fact that understanding of BJT characteristics allows circumvention of the problem. The fundamental operation of thyristors (SCR, GTO, and GCT) relies totally on BJT characteristics and electrical mechanisms. The IGBT has two parasitic BJTs, as shown in figure 3.16, that form an undesirable pnp-npn SCR structure. Understanding of BJT gain mechanisms allows virtual deactivation of the SCR.

The first bipolar transistors were mainly pnp, and were fabricated by alloying techniques and employed germanium semiconductor materials. Most transistors are now npn, made of silicon, and utilise selective diffusion and oxide masking.

A typical high-voltage *triple-diffused* transistor doping profile is shown in figure 3.7a. The n-collector region is the initial high-resistivity silicon material and the collector n^+ diffusion is performed first, usually into both sides. One n^+ diffusion is lapped off and the p-base and n^+ emitter diffusions are sequentially performed.

A *planar epitaxial* structure is often used for transistors with voltage ratings of less than 1000V. The basic structure and processing steps are shown in figure 3.7b. The n-type collector region is an epitaxial layer grown on an n substrate. The base and emitter are sequentially diffused into the epitaxy. Ion implantation is also used. This approach allows greater control on the depth of the n-type collector region, which is particularly important in specifying device switching and high-voltage properties. Also, the parasitic series collector resistance of the substrate is minimised without compromising the pellet's mechanical strength as a result of a possible reduction in thickness.

3.2.1i - BJT gain

Figure 3.8 shows an npn bipolar transistor connected in the *common emitter* configuration. In this configuration, injection of electrons from the lower n^+p junction into the centre p-region supplies minority carrier electrons to participate in the reverse current through the upper np junction.

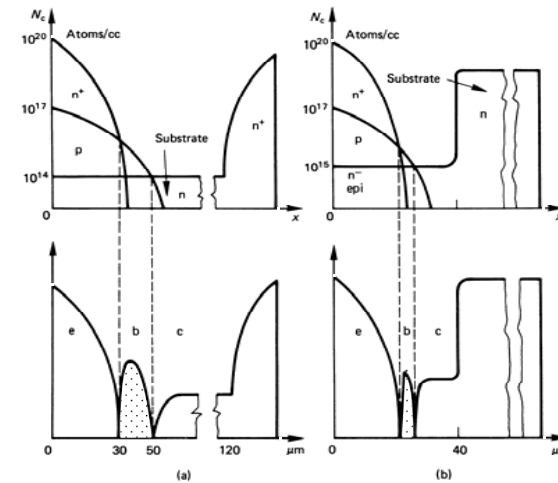


Figure 3.7. Impurity profile in two types of npn transistors: (a) a planar triple-diffused npn transistor obtained by three consecutive diffusions into a uniformly doped n-type substrate and (b) planar epitaxial npn transistor, obtained after two diffusions into n-type epitaxial layer which is first grown on a low-resistivity n-type silicon substrate.

The n^+ region which serves as the source of injected electrons is called the *emitter* and forms the emitter junction with the *p-base*, while the n-region into which electrons are swept by the reverse bias np junction is called the *collector* and, with the p-base, forms the collector junction.

To have a 'good' npn transistor almost all the electrons injected by the emitter into the base should be collected. Thus the p-base region should be narrow and the electron minority carrier lifetime should be long to ensure that the average electron injected at the emitter will diffuse to the collector scd without recombining in the base. The average lifetime of electrons in the p-base increases as the p-base concentration decreases, that is as the hole concentration decreases. The fraction of electrons which reach the collector is called the *base transport factor*, b_t . Electrons lost to recombination in the p-base must be re-supplied through the base contact. It is also required that the emitter junction carrier flow should be composed almost entirely of electrons injected into the base, rather than holes crossing from the base region to the emitter. Any such holes must be provided by the base current, which is minimised by doping the base region lightly compared with the emitter such that an n^+p emitter results. Such a junction is said to have a high *injection efficiency*, γ . A low lattice defect density also increases the injection efficiency.

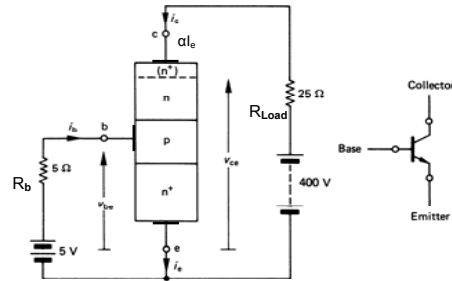


Figure 3.8. Common emitter junction bias conditions for an npn transistor and the npn bipolar junction transistor circuit symbol.

The relationship between collector and emitter current is

$$\frac{i_c}{i_e} = b_f \gamma = \alpha \tag{3.2}$$

The factor α is called the *current transfer ratio*. Since base current is necessary, α is less than 1, but close to 1, if the BJT

- has a good base transport factor, $b_f \approx 1$ (narrow base width and with long minority carrier lifetimes) and
- a high emitter injection efficiency, $\gamma \approx 1$ (high emitter doping relative to the base concentration).

In the common emitter configuration shown in figure 3.8 the ratio between the base current i_b and the collector current i_c , is of practical importance. Since the base current is the difference between the emitter and the collector current

$$\frac{i_c}{i_b} = \frac{i_c}{i_e - i_c} = \frac{i_c / i_b}{1 - i_c / i_b} \tag{3.3}$$

$$= \frac{\alpha}{1 - \alpha} = \beta \tag{3.4}$$

The factor β , relating the collector current to the base current, is defined as the base-to-collector *current amplification factor*. If α is near unity, β is large, implying the base current is small compared with the collector current.

3.2.1ii - BJT operating states

In power switching applications, a transistor is controlled in two states which can be referred to as the *off-state* or *cut-off state* and the conduction *on-state*. Ideally the transistor should appear as a short circuit when on and an open circuit when in

the off-state. Furthermore, the transition time between these two states is ideally zero. In reality, transistors only approximate these requirements.

The typical BJT collector output characteristics are shown in figure 3.9 which illustrates the various BJT operating regions. The *saturated on-state* shown in figure 3.9 occurs when both the collector and emitter junctions are forward biased. Consequently, the collector emitter voltage $V_{ce(sat)}$ is less than the base to emitter voltage $V_{be(sat)}$. The voltage breakdown phenomenon is of particular importance to the high-voltage, power-switching BJT, and is due to the characteristics of the device structure and geometry.

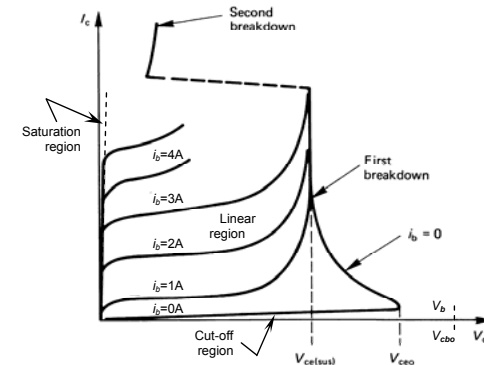


Figure 3.9. Output characteristics of a common emitter connected transistor showing its operating regions and the voltage breakdown range.

3.2.1iii - BJT maximum voltage - first and second breakdown

The collector junction supports the off-state voltage and in so doing develops a wide scl. This scl increases in width with increased reverse bias, penetrating into the base. It is unusual that a correctly designed high-voltage power switching BJT would break down as a result of punch-through of the collector scl through the base to the emitter scl. Because of the profile of the diffused base, collector junction voltage breakdown is usually due to the avalanche multiplication mechanism, created by the high electric field at the collector junction. In the common emitter configuration shown in figure 3.9, the transistor usually breaks down gradually, but before the collector junction avalanches at V_b . This occurs because the avalanche-generated holes in the collector scl are swept by the high field into the base. The emitter injects electrons in order to maintain base neutrality. This emitter junction current in turn causes more collector current, creating more avalanche pairs and causing a regenerative action. This voltage dependant avalanche effect is modelled by

$$M = \frac{1}{1 - (V_{ce}/V_b)^m} \quad (3.5)$$

Thus the gain mechanisms of the transistor cause collector to emitter breakdown - *first breakdown*, at voltage V_{ceo} , to occur before collector to base avalanche breakdown, at voltage V_{cbo} , which from $\alpha M = 1$ are related according to

$$\begin{aligned} V_{ceo} &= V_{cbo}(1 - \alpha)^{1/m} \approx V_{cbo} / \beta^{1/m} \\ &= V_b / \beta^{1/m} \end{aligned} \quad (V) \quad (3.6)$$

where the avalanche breakdown voltage V_b is given by equation (2.3);

$m \approx 6$ for a silicon p⁺n collector junction; and

$m \approx 4$ for a silicon n⁺p collector junction.

Contradictory device properties are that the higher the forward gain, the lower the breakdown voltage. A much higher collector emitter breakdown voltage level can be attained if the base emitter junction is reverse biased in the off-state.

First breakdown need not be catastrophic provided junction temperature limits are not exceeded. If local hot spots occur because of non-uniform current density distribution as a result of crystal faults, doping fluctuation, etc., *second breakdown* occurs. Silicon crystal melting and irreparable damage results, the collector voltage falls and the current increases rapidly as shown in figure 3.9.

3.2.2 The metal oxide semiconductor field effect transistor (MOSFET)

The basic low-power lateral structure of the metal oxide semiconductor field effect transistor (MOSFET) is illustrated in figure 3.10a. The n⁺ *source* and *drain* regions are diffused or implanted into the relatively lightly doped p-type substrate, and a thin silicon dioxide layer insulates the aluminium *gate* from the silicon surface. No lateral current flows from the drain to source without a conducting n-channel between them, since the drain-to-source path comprises two series, oppositely-directed pn junctions.

When a positive gate voltage is applied with respect to the source as shown in figure 3.10b, positive charges are created on the metal gate. In response, negative charges are induced in the underlying silicon, by the formation of a depletion region and a thin surface region containing mobile electrons. Effectively the positive gate potential *inverts* the p-channel, forming an electron-enhanced low-resistance *n-channel*, allowing current to flow freely in either direction between the drain and source. The inversion channel is essentially devoid of the thermal properties associated with the typical BJT.

An important parameter in mos transistors is the *threshold voltage* V_{TH} , which is the minimum positive gate voltage to induce the n-conducting channel. With zero gate voltage the structure is normally off. The device is considered to operate in the *enhancement mode* since the application of a positive gate voltage in excess of V_{TH} induces an n-conducting channel. The typical output characteristics of the MOSFET are shown in figure 3.10c.

3.2.2i - MOSFET structure and characteristics

The conventional horizontal structure in figure 3.10a has severe limitations associated with increasing die area that make it uneconomical for consideration as a viable high-current structure. A planar vertical n-channel dmos structure like those shown in figure 3.11 is used to overcome the inherent poor area utilisation of the basic mos structure. The peripheral p floating field guard is not shown.

The dmos structure is a vertical current flow device. An n⁺ epitaxial layer is grown on an n⁺ substrate. A series of p body regions are next diffused into the epitaxial layer. Then n⁺ source regions are diffused within the p body regions and a polycrystalline silicon gate is embedded in the silicon dioxide insulating layer. Source and gate metallization are deposited on the top surface of the die and the drain contact made to the bottom surface. Cell density is inversely related to voltage rating and varies from 200,000 to 1,000,000 cells per cm².

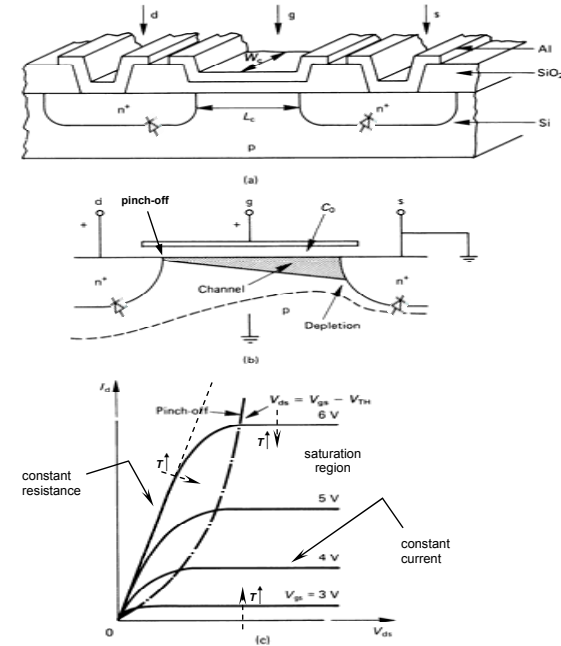


Figure 3.10. Enhancement-type n-channel mos transistor: (a) device cross-section; (b) induced n-channel near pinch-off; and (c) drain I-V characteristics as a function of gate voltage, showing the pinch-off locus and effects of increased temperature.

With a positive gate voltage, the device turns on and majority carriers flow laterally from the source to the drain region below the gate and vertically to the drain contact. Current can also flow freely in the reverse direction in the channel since the channel is bipolar conducting once enhanced.

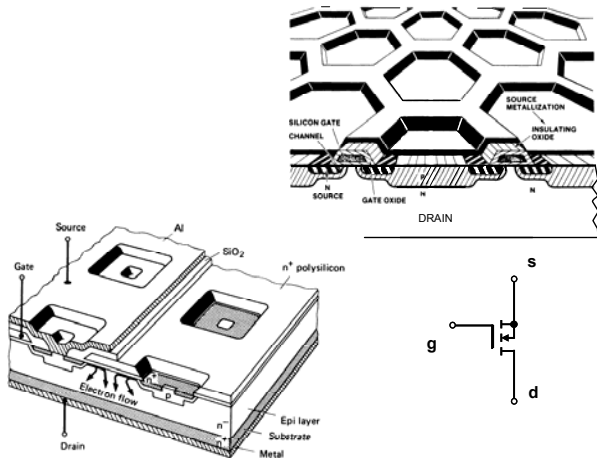


Figure 3.11. Two designs for the n-channel MOSFET and its circuit symbol (courtesy of Infineon and International Rectifier).

The obtainable drain-to-source breakdown voltage is not limited by the gate geometry. The scl associated with voltage blocking penetrates mostly in the n-type epitaxial layer. Thickness and doping concentration of this layer are thus decisive in specifying the blocking capability of the power MOSFET.

The basic drain current versus drain to source voltage static operating characteristics of the power MOSFET are illustrated in figure 3.10c. For a given gate voltage, there are two main operating regions on the drain current-voltage characteristic.

- The first is a *constant resistance* region, where an increase in drain to source voltage results in a proportional increase in drain current. (In practice, the effective resistance increases at higher drain currents.)
- At a certain drain current level, for a given gate voltage, a channel pinch-off effect occurs and the operating characteristic moves into a *constant current* region.

3.2.2ii - MOSFET drain current

When the power MOSFET is used as a switch, it is controlled in the on-condition, such that it is forced to operate in the resistive region. This ensures that the voltage

drop across the device is low so that the drain current is essentially defined by the load and the device power dissipation is minimal. Thus for switching applications, the on-resistance $R_{ds(on)}$ is an important characteristic because it will specify the on-state power loss for a given drain current. The lower $R_{ds(on)}$ is, the higher the current-handling capabilities of the device; thus $R_{ds(on)}$ is one important figure of merit of a power MOSFET.

A quadratic MOSFET model allows the inversion layer charge between the source and the drain to vary. For power MOSFETs that have short channels, the drain current I_d is related to the channel dimensions and the gate voltage V_{gs} according to

at low current, above pinch-off

$$I_d = \frac{1}{2} \mu \frac{W}{L_c} C_o (V_{gs} - V_{TH})^2 \tag{A} \tag{3.7}$$

if $V_{ds} \geq V_{gs} - V_{TH}$ for n-channel MOSFETs, as shown to the left of the pinch-off locus in figure 3.10c.

at high current after electron velocity saturation, the quadratic model is invalid and

$$I_d = \frac{1}{2} v_{sat} W_c C_o (V_{gs} - V_{TH}) \tag{A} \tag{3.8}$$

where C_o is the capacitance per unit area of the gate oxide (ϵ/ϵ_0)

W_c is the width of the channel

v_{sat} is the saturation velocity of electrons in silicon, (5.0×10^6 cm/s)

L_c is the effective channel length

μ is the conducting channel carrier mobility, ($300 \text{ cm}^2/\text{V}\cdot\text{s}$).

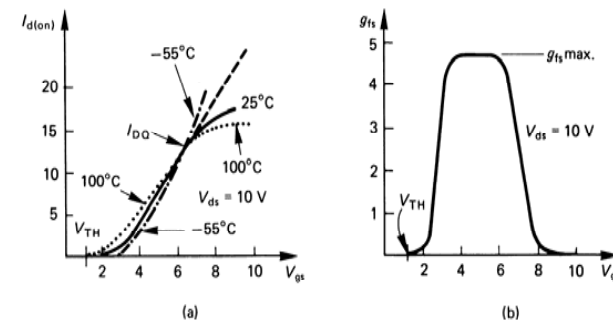


Figure 3.12. MOSFET gate voltage characteristics: (a) transfer characteristics of gate voltage versus drain current and (b) transconductance characteristics of gate voltage versus transconductance, g_{fs} .

Figure 3.12a shows that drain current exhibits both a positive and negative temperature coefficient with the drain current I_{DQ} being the boundary condition. If the drain current is greater than I_{DQ} there is a possibility of destruction by over-

current at low temperatures, while if the drain current is less than I_{DQ} , over-current can produce thermal runaway and destruction. Operation with a gate voltage corresponding to I_{DQ} avoids the need for any gate drive temperature compensation. At high gate voltages, the on-resistance of the resistive region and the drain current in the constant current region, become somewhat independent of the gate voltage. This phenomenon is best illustrated in the I_d versus V_{ds} characteristic by the curve cramping at high gate voltages in figure 3.10c.

3.2.2iii - MOSFET transconductance

Inspection of the static drain source characteristics of figure 3.10c reveals that as the gate voltage increases from zero, initially the drain current does not increase significantly. Only when a certain threshold gate voltage, V_{TH} , has been reached, does the drain current start to increase noticeably. This is more clearly illustrated in figure 3.12b which shows the characteristics of drain current I_d and small signal transconductance g_{fs} versus gate voltage, at a fixed drain voltage. It will be seen from these characteristics that no conduction occurs until V_{gs} reaches the threshold level, V_{TH} , after which the I_d versus V_{gs} characteristic becomes linear, the slope being the transconductance g_{fs} .

The amplification factor, forward transconductance, g_{fs} , is defined as

$$g_{fs} \triangleq \left. \frac{\partial I_d}{\partial V_{gs}} \right|_{V_{ds}=\text{constant}}$$

Differentiating equations (3.7) and (3.8) gives

at low current

$$g_{fs} = \mu \frac{W}{L_c} C_a (V_{gs} - V_{TH}) \quad (\text{mho}) \quad (3.9)$$

at high current

$$g_{fs} = \frac{1}{2} v_{sat} W C_a \quad (\text{mho}) \quad (3.10)$$

At high electric fields, that is high currents, the carrier velocity v_{sat} saturates. Inherent in the MOSFET structure are voltage-dependent capacitances and on-state resistance.

A typical minimum threshold voltage is about 2V and exhibits temperature dependence of approximately -10mV per K ($\alpha = 0.5$ per cent/K), as shown in figure 3.13. At high gate voltages, the drain current becomes constant as the transconductance falls to zero, implying the upper limit of forward drain current. The temperature variation of transconductance is small, typically -0.2 per cent/K, which results in extremely stable switching characteristics. The typical temperature coefficient for the gain of a bipolar transistor, the MOSFET equivalent to g_{fs} is +0.8 per cent/K. The temperature dependence of the MOSFET forward conductance is approximated by

$$g_{fs}(T) \approx g_{fs}(25^\circ\text{C}) \times \left(\frac{T}{300} \right)^{-2.3} \quad (\text{mho}) \quad (3.11)$$

since temperature effects are dominated by mobility variation with temperature.

3.2.2iv - MOSFET on-state resistance

In the fully on-state the drain-source conduction characteristics of the MOSFET can be considered as purely resistive. The on-resistance $R_{ds(on)}$ is the sum of the epitaxial region resistance, the channel resistance, which is modulated by the gate source voltage, and the lead and connection resistance. One reason for the wide proliferation of special gate geometries is to produce extremely short, reproducible channels, in order to reduce $R_{ds(on)}$. In high-voltage devices, the on-resistance is dominated by the resistance of the epitaxial drain region when the device is fully enhanced. For high-voltage n-channel devices the on-state resistance is approximated by

$$R_{ds(on)} = 8.3 \times 10^{-7} \times V_b^{2.5} / A \quad (\Omega) \quad (3.12)$$

where V_b is the breakdown voltage in volts

A is the die area in mm^2 .

A p-channel device with the same V_b as an n-channel device has an $R_{ds(on)}$ two to three times larger.

For low-voltage devices

$$R_{ds(on)} = L_c D / 34 A \quad (\Omega) \quad (3.13)$$

where D is the distance between cells in μm .

The factor $1/g_{fs}$ of $R_{ds(on)}$ is added to give the total $R_{ds(on)}$. On-state drain-source loss can therefore be based on $I_d^2 R_{ds(on)}$. On-resistance $R_{ds(on)}$ increases with temperature and approximately doubles over the range 25°C to 200°C , having a positive temperature coefficient of approximately +0.7 per cent/K above 25°C , as shown in figure 3.13. The temperature dependence of the on-state resistance is approximated by

$$R_{ds(on)}(T) = R_{ds(on)}(25^\circ\text{C}) \times \left(\frac{T}{300} \right)^{2.3} \quad (\Omega) \quad (3.14)$$

where the temperature T is in degrees Kelvin. This relationship closely follows the mobility charge dependence with temperature.

Since $R_{ds(on)}$ increases with temperature, current is automatically diverted away from a hot spot. Thus unlike the bipolar transistor, second breakdown cannot occur within the MOSFET. The breakdown voltage V_b has a positive temperature coefficient of typically 0.1 per cent/K as shown by $V_{(BR)DSS}$ in figure 3.13.

3.2.2v - MOSFET p-channel device

P-channel MOSFETs are very similar to n-channel devices except that the n and p regions are interchanged. In p-channel devices the on-resistance, for a given die area, will be approximately twice that of a comparable n-channel device. The reason for this is that in the n-channel device the majority carriers are electrons but in the p-channel device, the majority carriers are holes which have lower mobility. If the area of a p-channel device is increased to produce an equal $R_{ds(on)}$, then the various capacitances of the p-channel device will be larger, and the device costs will be greater.

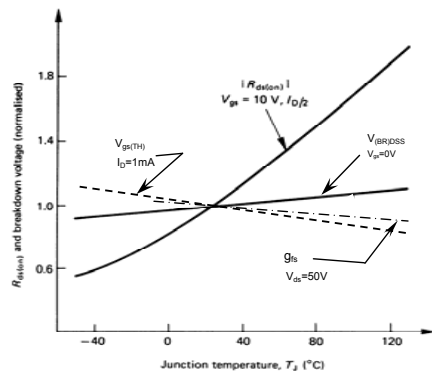


Figure 3.13. Normalised drain-source on-resistance, transconductance, gate threshold voltage, and breakdown voltage versus junction temperature.

Example 3.1: Properties of an n-channel MOSFET cell

A silicon n-channel MOSFET cell has a threshold voltage of $V_{TH} = 2V$, $W_c = 10\mu m$, $L_c = 1\mu m$, and an oxide thickness of $t_{ox} = 50nm$. The device is biased with $V_{GS} = 10V$ and $V_{DS} = 100V$.

- i. Assuming a quadratic model and a surface carrier mobility of $300\text{ cm}^2/V\text{-s}$, calculate the drain current and transconductance.
- ii. Assuming carrier velocity saturation ($5 \times 10^6\text{ cm/s}$), calculate the drain current and transconductance.

Solution

i. The MOSFET is biased in saturation since $V_{ds} > V_{gs} - V_{TH}$. Therefore, from equation (3.7) the drain current equals:

$$I_d = \frac{1}{2} \mu C_a \frac{W_c}{L_c} (V_{gs} - V_{TH})^2$$

$$= \frac{300 \times 10^{-4}}{2} \times \frac{3.85 \times 8.85 \times 10^{-12}}{50 \times 10^{-9}} \times \frac{10}{1} \times (10 - 2)^2 = 6.5\text{ mA}$$

From equation (3.9), the transconductance equals:

$$g_{fs} = \mu C_a \frac{W_c}{L_c} (V_{gs} - V_{TH})$$

$$= 300 \times 10^{-4} \times \frac{3.85 \times 8.85 \times 10^{-14}}{50 \times 10^{-9}} \times \frac{10}{1} \times (10 - 2) = 1.64\text{ mho}$$

ii. When the electron velocity saturates, the drain current is given by equation 3.8

$$I_d = \frac{1}{2} v_{sat} W_c C_a (V_{gs} - V_{TH})$$

$$= \frac{1}{2} \times 5 \times 10^4 \times 10^{-5} \times \frac{3.85 \times 8.85 \times 10^{-12}}{50 \times 10^{-9}} \times (10 - 2) = 136\text{ mA}$$

The transconductance is given by equation 3.10

$$g_{fs} = \frac{1}{2} v_{sat} W_c C_a$$

$$= \frac{1}{2} \times 5 \times 10^4 \times 10^{-5} \times \frac{3.85 \times 8.85 \times 10^{-12}}{50 \times 10^{-9}} = 16.1\text{ mho}$$

3.2.vi - MOSFET parasitic BJT

Figure 3.14 shows the MOSFET equivalent circuit based on its structure and features. The parasitic npn transistor shown in figure 3.14b is key to device operation and limitations.

Capacitance exists within the structure from the gate to the source, C_{gs} , the gate to the drain, C_{gd} , and from the drain to the source, C_{ds} . The capacitance C_{gs} varies little with voltage; however C_{ds} and C_{gd} vary significantly with voltage. Obviously these capacitances influence the switching intervals, an aspect considered in chapter 4.4.2.

The emitter of the parasitic npn transistor is the source of the MOSFET, the base is the p-type body and the collector is the drain region. In the construction of the MOSFET, the emitter and base of the npn transistor are purposely shorted out by the source metallization to disable the parasitic device by reducing its injection efficiency. However, this short circuit cannot be perfect and R_{be} models the lateral p-body resistance, while C_{ob} is essentially C_{ds} . The npn transistor has a collector-emitter breakdown voltage, between V_{cbo} and V_{ceo} . If an external dv/dt is applied between the drain and source as shown in figure 3.14b, enough displacement current could flow through C_{ob} to generate a voltage drop across R_{be} sufficient to turn on the parasitic bipolar device, causing MOSFET failure in second breakdown.

When the drain to source voltage is negative, current can flow from the source to drain through R_{be} and the base to collector junction of the parasitic npn transistor within the structure, the dashed line shown in figure 3.14b. This is termed the *body diode*, inherent in the MOSFET structure.

3.2.vii - MOSFET on-state resistance reduction

Most power switching devices have a *vertical structure*, where the gate and source of the MOSFET (or emitter in the case of the IGBT) are on one surface of the substrate, while the drain (or collector) is on the other substrate surface. The principal current flows vertically through the substrate but the conductive channel is lateral due to the *planar gate structure*, as shown in figure 3.11. The structure resistance components between the drain and source are:

- the drift region;
- the JFET region;

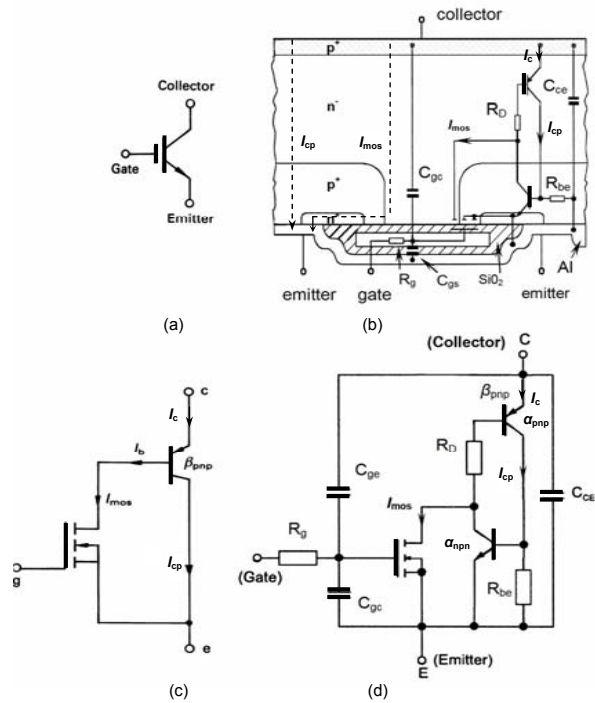


Figure 3.16. Insulated gate bipolar transistor (IGBT): (a) circuit symbol; (b) physical structure showing current paths; (c) normal operation equivalent circuit; and (d) high current latching equivalent circuit.

3.2.3ii - IGBT in the on-state

The p⁺ substrate conductively modulates the n⁻ region with minority carriers, which whilst conducting the main collector current, produces a low on-state voltage at the expense of a 0.6 to 0.8V offset in the output voltage characteristics due to the collector pn junction. From figure 3.16c, the IGBT collector current is approximated by

$$I_c = I_{mos}(1 + \beta_{pnp}) \tag{3.15}$$

3.2.3iii - IGBT at turn-off

Excess p-stored charge that remains after the high voltage scl has been established must recombine in the externally inaccessible n⁻ drift region. This storage charge produces a tail current.

The operational mechanisms are those of any minority carrier device and result in slower switching times than the majority carrier MOSFET. On-state voltage and switching characteristics can be significantly improved by using the trench gate technique used on the MOSFET, as considered in section 3.2.1 and shown in figure 3.15b. A less stable structure improvement involves using wider trenches, judiciously spaced, so that accumulate holes under the trench enhance emitter injection of electrons. This injection enhancement reduces the on-state voltage without degrading the switching performance.

Further performance enhancement is gained by using the punch through, PT-IGBT, structure shown in figure 3.17a, which incorporates an n⁺ buffer region. The conventional non-punch through NPT-IGBT structure is shown in figure 3.17b. Both collector structures can have the same emitter structure, whether a lateral gate as shown, or the MOSFET trench gate in figure 3.15b.

Figure 3.17 shows the electric field in the off-state, where the PT-IGBT develops a field as in the pin diode in figure 3.2b, which allows a thinner wafer. The NPT-IGBT requires a thicker wafer (about 200µm for a 1200V device) which results in a larger substrate resistance and a slower switching device.

- The PT-IGBT has n⁺ and n⁻ layers formed by epitaxial growth on a p⁺ substrate. The electric field plot in figure 3.17b shows that the off-state voltage scl consumes the n⁻ substrate and is rapidly reduced to zero in the n⁺ buffer.
- The NPT-IGBT has a lightly doped n⁻ substrate with the p-regions (p wells and p collector) formed by ion implantation. The electric field distribution in figure 3.17b shows that the n⁻ drift region has to be wide enough to support all the off-state voltage, without punch through to the p collector implant.

3.2.3iv - IGBT latch-up

The equivalent circuit in figure 3.16d shows non-ideal components associated with the ideal MOSFET. The parasitic npn bipolar junction transistor (the n⁺ emitter/ p⁺ well/ n⁻ drift region) and the pnp transistor (p⁺ collector/ n⁻ drift/ p⁺ well) are the BJT e-b-c) couple together to form an SCR thyristor structure, as considered in section 3.3. Latching of this parasitic SCR can occur:

- in the on-state if the current density exceeds a critical level, which adversely decreases with increased temperature or
- during the turn-off voltage rise when the hole current increases in sensitive regions of the structure due to the charge movement associated with the scl widening.

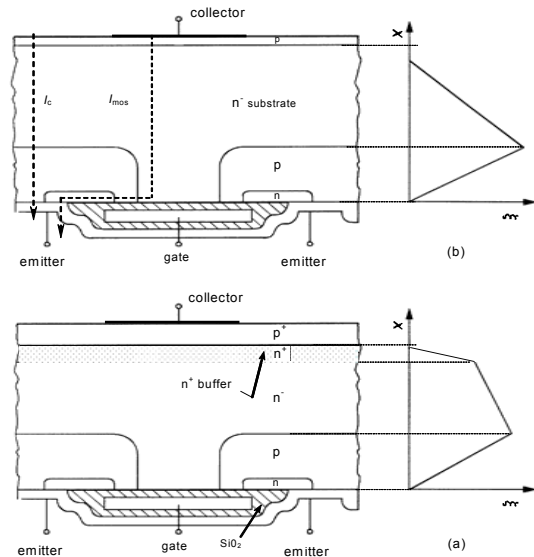


Figure 3.17. Insulated gate bipolar transistor structures and electric field profile: (a) fieldstop PT-IGBT and (b) conventional NPT-IGBT.

1 - IGBT on-state SCR static latch-up is related to the temperature dependant transistor gains which are related to the BJT base transport factor b_i and emitter injection efficiency γ , defined for the BJT in equation (3.2)

$$\alpha_{npn} + \alpha_{pnp} = b_{npn}\gamma_{npn} + b_{pnp}\gamma_{pnp} = 1 \quad (3.16)$$

To avoid loss of control and possible IGBT failure, the factors in equation (3.16), which is valid for on-state latch-up, are judiciously adjusted in the device design.

Common to both device types is the gate structure, hence the base-emitter junction of the npn parasitic BJT have the same properties. In each structure, the shorting resistor R_{be} decreases the injection efficiency of the npn BJT emitter. This resistance is minimized by highly doping the p^+ wells directly below the n-emitters and by shortening the length of the n-emitter. The gain α_{npn} in equation (3.16) is decreased since the injection efficiency γ_{npn} is lowered.

Reduction of the pnp BJT gain of the PT-IGBT and NPT-IGBT is achieved with different techniques.

- For the NPT-IGBT, the emitter injection efficiency of holes from the p^+ zone into the n^- drift region is high because of the large difference in

doping concentrations at the junction. This yields a high injection efficiency γ_{pnp} . The base transport factor $b_{i,pnp}$ is already low because of the large width of the n^- drift region, and is further reduced by lifetime killing of minority carriers in the n^- drift region by using gold doping or electron beam radiation.

- For the PT-IGBT, the p^+ emitting junction at the collector is a well-controlled shallow implant thus reducing the injection efficiency γ_{pnp} . Charge carrier lifetime killing in the n^- drift region to reduce the base transport factor b_{pnp} , is therefore not necessary.

2 - IGBT turn-off SCR dynamic latch-up can occur while the collector voltage is rising, before the collector current decreases. Equation (3.16) is modified by equation (3.5) to account for voltage avalanche multiplication effects.

$$M_{npn}\alpha_{npn} + M_{pnp}\alpha_{pnp} = 1 \quad (3.17)$$

$$\text{where } M = \frac{1}{1 - (v_{ce}/V_b)^n}$$

This dynamic latch-up mode is adversely affected by increased temperature and current magnitude during the voltage rise time at turn-off.

Since $v_{ce} \ll V_b$, $M \rightarrow 1$, and the multiplication effect is not significant in the on-state static latch-up analysis. IGBTs are designed and rated so that the latch-up current is 10 to 15 times the rated current.

PT IGBT and NPT IGBT comparison

Generally, faster switching speed is traded for lower on-state losses.

Table 3.1 PT versus NPT IGBTs

IGBT TYPE	PT IGBT	NPT IGBT
conduction loss (same switching speed)	Lower $V_{ce(sat)}$ Decreases slightly with temperature A slight positive temperature coefficient at high current densities allows parallel connection.	Higher $V_{ce(sat)}$ Increases with temperature Suitable for parallel connection
switching speed (same on-state loss)	Faster switching due to high gain and reduced minority carrier lifetime	
short circuit rating		more rugged due to wider base and low pnp gain
turn-on switching loss	Largely unaffected by temperature	Largely unaffected by temperature
turn-off switching loss	Loss increases with temperature but lower than NPT devices to start with	Virtually constant with temperature

3.2.4 The junction field effect transistor (JFET)

The field effect for a FET may be created in two ways:

- A voltage signal controls charge indirectly using a capacitive effect as in the MOSFET, section 3.2.2.
- In a junction FET (JFET), the voltage dependant scl width of a junction is used to control the effective cross-sectional area of a conducting channel. If the zero bias voltage cuts off the channel then the JFET is normally off, otherwise if a reverse bias is needed to cut-off the channel, the JFET is termed normally on.

The electrical properties of SiC make the JFET a viable possibility as a power switch. Two normally on JFET structures are shown in figure 3.18, where it is seen how the scl layer decreases the channel width as the source to gate voltage reverse bias increases. In SiC, the channel has a positive temperature coefficient, $R_{on} \propto T^{2.6}$, hence parallel connection is viable. Natural current saturation with a positive temperature coefficient means lengthy short-circuit currents of over a millisecond can be sustained. Although the channel is bidirectional, in the biased off-state an integral fast, robust pn body diode is inherent as seen in figure 3.18b.

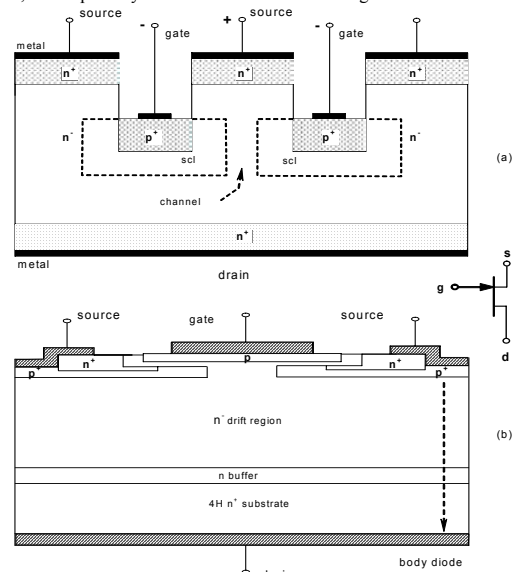


Figure 3.18. Cross-section of the SiC vertical junction field effect transistor: (a) trench gate with channel shown and (b) variation incorporating a pn body diode.

3.3 Thyristors

The name thyristor is a generic term for a bipolar semiconductor device which comprises four semiconductor layers and operates as a switch having a latched on-state and a stable off-state. Numerous members of the thyristor family exist. The simplest structurally is the silicon-controlled rectifier (SCR) while the most complicated is the triac.

3.3.1 The silicon-controlled rectifier (SCR)

The basic SCR structure and doping profile in figure 3.19 depicts the SCR as three pn junctions J1, J2, and J3 in series. The contact electrode to the outer p-layer is called the anode and that to the outer n-layer is termed the cathode. With a gate contact to the inner p-region, the resultant three-terminal thyristor device is technically called the silicon-controlled rectifier (SCR).

A low concentration n-type silicon wafer is chosen as the starting material. A single diffusion process is then used to form simultaneously the p1 and p2 layers. Finally, an n-type layer, n1, is diffused selectively into one side of the wafer to form the cathode. The masked-out areas are used for the gate contact to the p1 region. To prevent premature breakdown at the surface edge, bevelling is used as in figure 3.1, to ensure that breakdown will occur uniformly in the bulk.

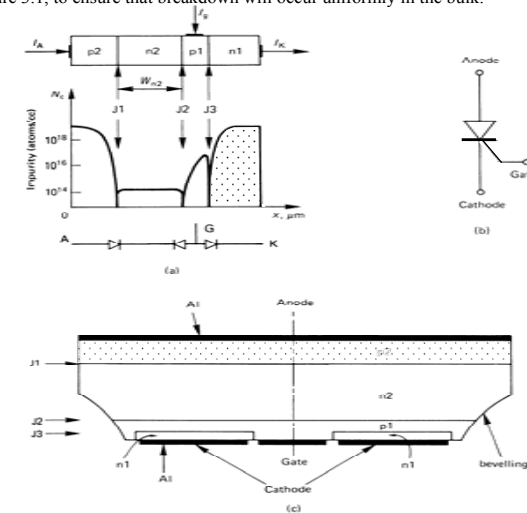


Figure 3.19. The silicon-controlled rectifier, SCR: (a) net impurity density profile; (b) circuit symbol; and (c) cross-sectional view.

A number of observations can be made about the doping profile of the SCR which relate to its electrical characteristics.

The anode and cathode would both be expected to be good emitters of minority carriers into the n2 and p1 regions respectively because of their relative high concentrations with respect to their injected regions.

The n2 region is very wide, typically hundreds of micrometres, and low concentration, typically less than $10^{14}/\text{cc}$. Even though the hole lifetime may be very long, $100\mu\text{s}$, the base transport factor for hole minority carriers, $b_{h,n2}$ is low. The low-concentration provides high forward and reverse blocking capability and the associated reverse-biased scl's penetrate deeply into the n2 region. Gold lifetime killing or electron irradiation, most effective in the n2 region, is employed to improve the switching speed by increasing the number of carrier recombination centres.

The two-transistor model of the SCR shown in figure 3.20 can be used to represent the p2-n2-p1-n1 structure and explain its characteristics. Transistor T_1 is an npn BJT formed from regions n2-p1-n1 while T_2 is a pnp BJT formed from SCR regions p2-n2-p1.

The application of a positive voltage between anode and cathode does not result in conduction because the SCR central junction J2 is reverse-biased and *blocking*. Both equivalent circuit transistors have forward-biased emitter junctions and with reverse-biased collector junctions, both BJT's can be considered to be cut off.

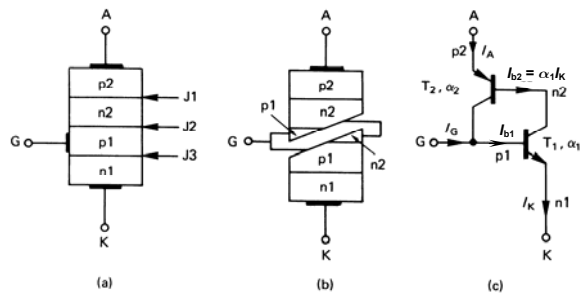


Figure 3.20. Cross-section of the SCR showing its model derivation: (a) schematic of the SCR cross-section; (b) the division of the SCR into two transistors; and (c) the npn-pnp two-transistor model of the basic SCR.

3.3.1i - SCR turn-on

It is evident from figure 3.20c that the collector current of the npn transistor provides the base current for the pnp transistor. Also, the collector current of the pnp transistor along with any gate current I_G supplies the base drive for the npn transistor. Thus a *regenerative* situation occurs when the loop gain exceeds unity.

The base current of the pnp transistor T_2 with dc current gain α_2 is

$$I_{b2} = (1 - \alpha_2) I_A - I_{co2}$$

which is supplied by the collector of the npn transistor. The current I_{co} is the collector reverse bias leakage current. The collector current of the npn transistor T_1 with a dc current gain of α_1 is given by

$$I_{c1} = \alpha_1 I_K + I_{co1}$$

By equating I_{b2} and I_{c1}

$$(1 - \alpha_2) I_A - I_{co2} = \alpha_1 I_K + I_{co1}$$

Since $I_K = I_A + I_G$

$$I_A = \frac{\alpha_1 I_G + I_{co1} + I_{co2}}{1 - (\alpha_1 + \alpha_2)} = \frac{\alpha_1 I_G + I_{co1} + I_{co2}}{1 - G_I} \tag{A} \tag{3.18}$$

where $\alpha_1 + \alpha_2$ is called the *loop gain*, G_I . At high voltages, to account for avalanche multiplication effects, the gains are replaced by $M\alpha$, where M is the avalanche multiplication coefficient in equation 3.15. Hence, G_I becomes $M_1\alpha_1 + M_2\alpha_2$. By inspection of equation (3.18) it can be seen that a large anode current results when $G_I \rightarrow 1$, whence the circuit regenerates and each transistor drives its counterpart into saturation. All junctions are forward-biased and the total device voltage is approximately that of a single pn junction, with the anode current limited by the external circuit. The n2-p1-n1 device acts like a saturated transistor and provides a remote contact to the n2 region. Therefore the device behaves essentially like a p-i-n diode (p2-i-n1), where the voltage drop across the i-region is inversely proportional to the recombination rate. Typical SCR static I - V characteristics are shown in figure 3.21.

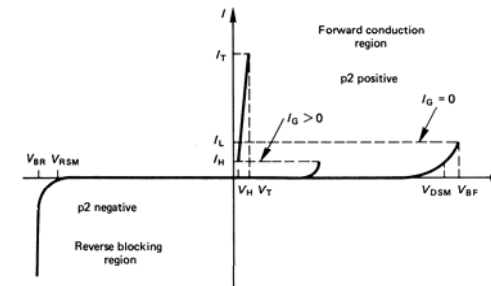


Figure 3.21. The silicon-controlled rectifier static I - V characteristics.

At low current levels, α_1 and α_2 are small because of carrier recombination effects, but increase rapidly as current increases. The conventional gate turn-on mechanism is based on these current gain properties. External gate current starts the regeneration action and the subsequent increase in anode current causes the

gains to increase, thus ensuring a high loop gain, whence the gate current can be removed. The $I-V$ characteristics in figure 3.21 show this property, where a minimum anode current I_L is necessary for the loop gain to increase sufficiently to enable the SCR to latch on by the regeneration mechanism.

The SCR can be brought into conduction by a number of mechanisms other than via the gate (other than the light activated SCR used in high-voltage dc converters).

- If the anode-cathode voltage causes avalanche multiplication of the central junction, the increased current is sufficient to start the regenerative action. The forward anode-cathode breakover voltage V_{BF} is dependent on the central junction J2 avalanche voltage and the loop gain according to

$$V_{BF} = V_b (1 - \alpha_1 - \alpha_2)^{1/m} \quad (V) \quad (3.19)$$

where the avalanche breakdown voltage, at room temperature, for a typical SCR p⁺n central junction is given by (equation (2.3))

$$V_b = 5.34 \times 10^{13} \times N_D^{-2/4} \quad (V) \quad (3.20)$$

where N_D is the concentration of the high resistivity n2 region when $10^{13} < N_D < 5 \times 10^{14}$ /cc.

- Turn-on can also be induced by means of an anode-to-cathode *applied* dv/dt where the peak ramp voltage is less than V_{BF} . The increasing voltage is supported by the central blocking junction J2. The associated scl width increases and a charging or displacement current flows according to $i = d(Cv)/dt$. The charging current flows across both the anode and cathode junctions, causing hole and electron injection respectively. The same mechanism occurs at the cathode if gate current is applied; hence if the terminal dv/dt is large enough, SCR turn-on occurs.
- The forward SCR leakage current, which is the reverse-biased pn junction J2 leakage current, doubles approximately with every 8K temperature rise. At elevated temperatures, the thermally generated leakage current can be sufficient to increase the SCR loop gain such that turn-on occurs.

3.3.1ii - SCR cathode shorts

All SCR turn-on mechanisms are highly temperature-dependent. A structural modification commonly used to reduce device temperature sensitivity and to increase dv/dt rating is the introduction of *cathode shorts*. A cross-sectional structure schematic and two-transistor equivalent of the cathode shorting technique are shown in figure 3.22. It will be seen that the cathode metallization overlaps the p1 region, which is the gate contact region. The technique is based on some of the anode forward-blocking current being shunted from the cathode junction via the cathode short. The cathode electron injection efficiency is effectively reduced, thereby decreasing α_1 which results in an increase in the forward voltage-blocking rating V_{BF} and dv/dt capability. The holding and latching currents are also increased.

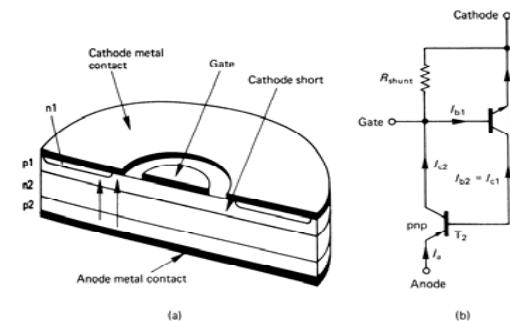


Figure 3.22. Shorted cathode SCR: (a) SCR cross-section showing some anode current flowing through cathode shorts and (b) the SCR two-transistor equivalent circuit SCR with cathode shorts.

The cathode-anode, reverse breakdown voltage V_{BR} is shown in figure 3.21. The anode p2⁺n2 junction J1 characterises SCR reverse blocking properties and V_{BR} is given by (equation (3.6))

$$V_{BR} = V_b (1 - \alpha_2)^{1/m}$$

If a very high resistivity n2 region, N_{Dn2} , is used and breakdown is due to punch-through to J2, then the terminal breakdown voltage will be approximated by (equation (2.2))

$$V_{PF} = 7.67 \times 10^{-16} N_{Dn2} W_{n2}^2$$

where W_{n2} is the width of the n2 region. This relationship is valid for both forward and reverse SCR voltage breakdown arising from punch-through.

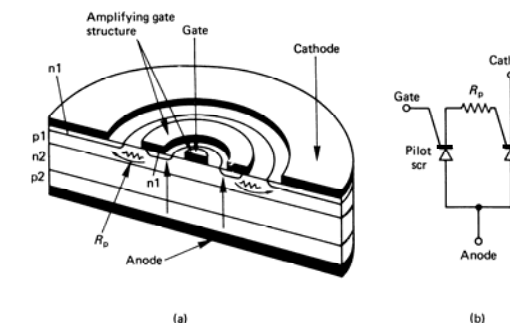


Figure 3.23. The amplifying gate SCR: (a) cross-section of the structure and (b) two-SCR equivalent circuit.

3.3.1iii - SCR amplifying gate

At SCR turn-on, only a small peripheral region of the cathode along the gate region conducts initially. The conducting area spreads at about 50m/s, eventually encompassing the whole cathode area. If at turn-on a very large anode current is required, that is a high initial di/dt , a long gate-cathode perimeter is necessary in order to avoid excessively high initial cathode current densities. The usual method employed to effectively enlarge the SCR initial turn-on area is to fabricate an integrated *amplifying gate*, as shown in figure 3.23. A small gate current is used to initiate the *pilot SCR*, which turns on very rapidly because of its small area. The cathode current of this pilot SCR provides a much larger gate current to the main SCR section than the original triggering current. Once the main device is fully on, the pilot device turns off if the gate current is removed.

An important property of the SCR is that once latched on, the gate condition is of little importance. The regenerative action holds the device on and SCR turn-off can only be achieved by reducing the anode current externally to a level below which the loop gain is significantly less than unity.

3.3.2 The asymmetrical silicon-controlled rectifier (ASCR)

The doping profiles and cross-sectional views comparing the asymmetrical SCR and SCR are shown in figure 3.24. In each case the electric field ξ within the p1n2 junction reverse-bias scl is shown and because the n2 region is lightly doped, the scl extends deeply into it. The scl applied reverse-bias voltage is mathematically equal to the integral of the electric field, ξ (area under the curve).

If, in the conventional SCR, the scl edge reaches the p2⁺ layer, then punch-through has occurred and the SCR turns on. To prevent such a condition and to allow for manufacturing tolerances, the n2⁻ region is kept thick with the unfortunate consequence that on-state losses, which are proportional to n2 layer thickness, are high.

In the case of the ASCR, a much thinner n2⁻ region is possible since a highly doped n layer adjacent to the p2⁺ anode is utilised as an *electric field stopper*. The penalty paid for this layer construction is that in the reverse voltage blocking mode, the n2p2⁺ junction avalanches at a low voltage of a few tens of volts. Thus the ASCR does not have any usable repetitive reverse-blocking ability, hence the name asymmetrical SCR. By sacrificing reverse-blocking ability, significant improvements in lower on-state voltage, higher forward-blocking voltage, and faster turn-off characteristics are attained.

3.3.3 The reverse-conducting thyristor (RCT)

The RCT is electrically equivalent to an SCR in anti-parallel with a diode, but both are integrated into the same wafer. The reason for integrating the SCR and diode is to minimise the interconnecting lead inductance. The circuit symbol, cross-sectional wafer view, and typical doping profile are shown in figure 3.25.

Since no reverse voltage will be applied to the RCT there is only the cathode-side deep p-diffused layer. This and the ASCR n-region type field stopper result in low forward voltage characteristics. As in the ASCR case, the highly n-type doped anode end of the wide n-region also allows higher forward voltages to be blocked. Both anode and cathode shorts can be employed to improve thermal and dv/dt properties. As shown in figure 3.25a, an amplifying gate can be used to improve initial di/dt capability.

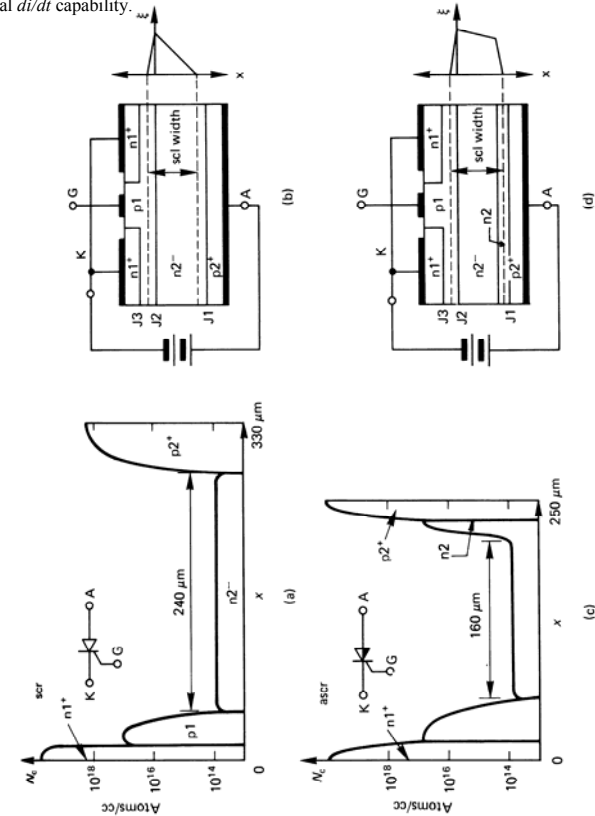


Figure 3.24. Doping profile and cross-section, including the electric field of J2 in the forward biased off-state for: (a) and (b) the conventional SCR; (c) and (d) the asymmetrical SCR.

The integral anti-parallel diode comprises an outer ring and is isolated from the central SCR section by a diffused guard ring, or a groove, or by irradiation lifetime control techniques. The guard ring is particularly important in that it must confine the carriers associated with the reverse-blocking diode to that region so that these carriers do not represent a forward displacement current in the SCR section. If the carriers were to spill over, the device dv/dt rating would be reduced - possibly resulting in false turn-on.

Gold or irradiation lifetime killing can be employed to reduce the turn-off time without significantly increasing the on-state voltage.

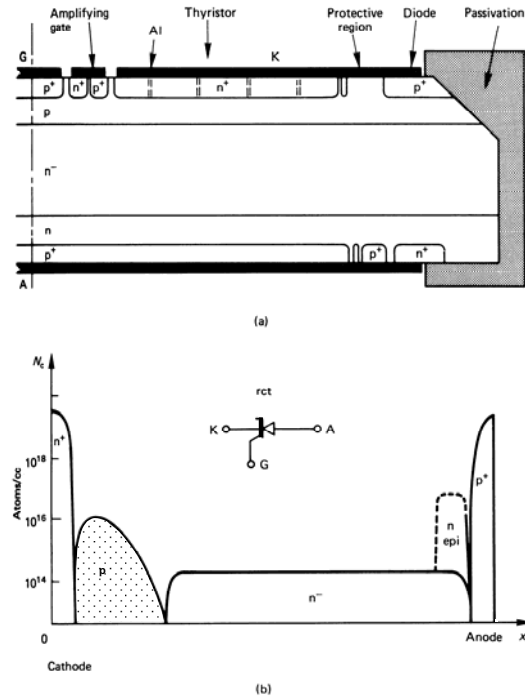


Figure 3.25. Reverse conducting thyristor with an amplifying gate structure: (a) cross-section of the structure and (b) typical doping profile of the SCR section.

3.3.4 The bi-directional-conducting thyristor (BCT)

Two anti-parallel connected SCRs can be integrated into one silicon wafer, as shown in figure 3.26. As a result of integrated symmetry, both devices have near identical electrical properties. The mechanical feature different to the triac, is that there are two gates – one on each side of the wafer. Also, unlike the triac, the two SCR sections are physically separated in the wafer to minimise carrier diffusion interaction. The equivalent circuit comprises two SCRs connected in anti-parallel. As such, one device turning off and supporting a negative voltage, represents a positive dv/dt impressed across the complementary device, tending to turn it on. Also, any charge carriers which diffusion from the SCR previously on, exasperate the dv/dt stress on the off SCR.

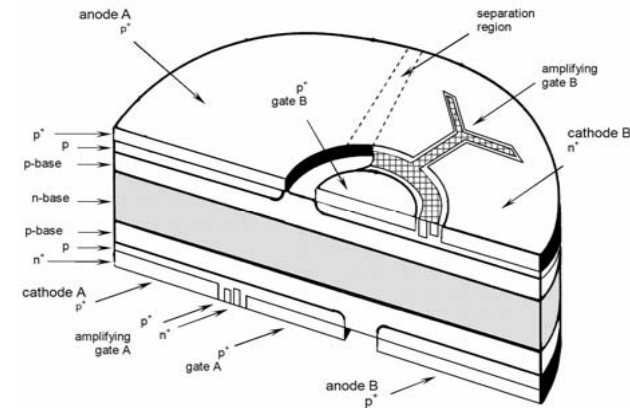


Figure 3.26. Cross-section structure of the bidirectional conducting phase-control SCR with an amplifying gate structure.

The two central amplifying gate structures are as for the RCT, in figure 3.25a. A separation of a few minority carrier lateral diffusion lengths, along with an increased density of cathode shorts along the separating edge of each cathode and in the amplifying gate region close to the anode of the complementary SCR, enhance the physical separation. The amplifying gate fingers are angled away from the separation regions to minimise the shorting effect of the complementary SCR anode emitter shorting.

The on-state voltage of each SCR is fine tuned, match for on-state loss, using electron irradiation.

3.3.5 The gate turn-off thyristor (GTO)

The gate turn-off thyristor is an SCR that is turned on by forward-biasing the cathode junction and turned off by reverse-biasing the same junction, thereby preventing the cathode from injecting electrons into the p1 region. Other than its controlled turn-off properties, the GTO's characteristics are similar to the conventional SCR. The basic structure and circuit symbol are shown in figure 3.27.

3.3.5i - GTO turn-off mechanism

In the on-state, due to the high injection efficiency of junctions J1 and J3, the central p-base is flooded with electrons emitted from the n-cathode and the central n-base is flooded with holes emitted from the p-anode. If a reverse gate current flows from the cathode to the gate, with a driving voltage tending to reverse bias the gate-cathode junction – then p-base holes are extracted from the gate, suppressing the cathode junction from injecting electrons. Eventually the cathode junction is cut-off and the pnp transistor section, now without base current turns off, thereby turning off the GTO.

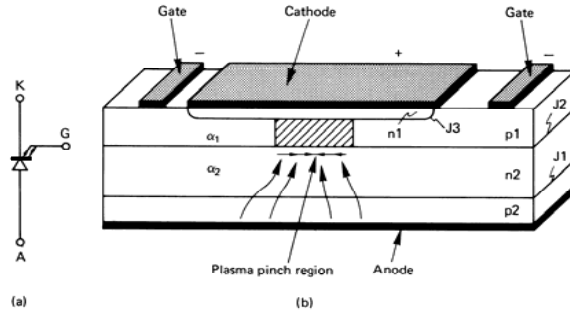


Figure 3.27. The gate turn-off thyristor:

(a) circuit symbol and (b) the basic structure along an interdigitated finger showing plasma focussing in the p1 region at the cathode junction at turn-off.

The turn-off mechanism can be analyzed by considering the two-transistor equivalent circuit model for the SCR shown in figure 3.20c. The reverse gate current I_{GQ} flows from the gate and is the reverse base current of the npn transistor T_1 . The base current for transistor T_1 is $I_B = \alpha_2 I_A - I_{GQ}$, where $I_{GQ} = -I_G$. The reverse base current in terms of the gain of T_1 is $I_{RB} = (1 - \alpha_1) I_K$. The GTO as a three terminal device must satisfy $I_A = I_K + I_{GQ}$ and to turn-off the GTO, $I_B < I_{RB}$. These conditions yield

$$(\alpha_1 + \alpha_2 - 1)I_A < \alpha_2 I_{GQ}$$

The turn-off gain of the GTO, β_Q , is defined as the ratio of anode current I_A to reverse gate current I_{GQ} , that is

$$\beta_Q \equiv I_A / I_{GQ} < \alpha_1 / (\alpha_1 + \alpha_2 - 1) \quad (3.21)$$

Thus for high turn-off gain it is important to make α_1 for the npn section as close to unity as possible, while α_2 of the pnp section should be small. A turn-off current gain of 5 is typical.

During the turn-off process, the conducting plasma is squeezed to the centre of the cathode finger, since the lateral p1 region resistance causes this region to be last in changing from forward to reverse bias. This region has the least reverse bias and for reliable GTO operation, the final area of the squeezed plasma must be large enough to prevent an excessive current density. Device failure would be imminent because of overheating.

The doping profile is characterised by a low p1 region sheet resistance and an inter-digitated cathode region to ensure even distribution of the reverse bias across the cathode junction at turn-off. Both turn-off and temperature properties are enhanced by using an anode shorting and defocusing technique as shown in figure 3.28a, but at the expense of any reverse-blocking capability and increased on-state voltage.

The shown two-level cathode and gate metallization used on large-area devices allow a flat metal plate for the cathode connection. As with the conventional SCR, a reverse conducting diode structure can be integrated, as shown in figure 3.28b.

3.3.6 The gate commutated thyristor (GCT)

GTO frequency limitations and the need for an external parallel connected capacitive turn-off snubber (to limit re-applied dv/dt), have motivated its enhancement, resulting in the gate commutated thyristor, GCT. As shown in figure 3.28c, a number of processing and structural variations to the basic GTO result in a more robust and versatile high power switch.

- *n-type buffer*
An n-type buffer layer allows a thinner n-drift region. A 40% thinner silicon wafer, for the same blocking voltage, reduces switching losses and the on-state voltage. An integral reverse conducting diode is also possible, as with the conventional SCR and GTO.
- *transparent emitter*
A thin lightly doped anode p-emitter is used instead of the normal GTO anode shorts. Some electrons pass through the layer as if the anode were shorted and recombine at the anode contact metal interface, without causing hole emission into the n-base. Effectively, a reduced emitter injection efficiency is achieved without anode shorts. Consequently, gate current triggering requirements are an order of magnitude lower than for the conventional GTO.
- *low inductance*
A low inductance gate structure, contact, and wafer assembly allow the full anode current to be shunted from the gate in less than 1 μ s, before the anode voltage rises at turn-off.

The four different trigger modes of the triac are illustrated in figure 3.31 and the turn-on mechanism for each mode is as follows.

(a) M2 positive, I_g positive (Mode I)

The main terminal M2 is positive with respect to M1 and gate current forward-biases the p2-n2 junction, J3. The active main SCR section is p1-n1-p2-n2. Turn-on is that for a conventional SCR, as shown in figure 3.31a.

(b) M2 positive, I_g negative (Mode II)

In figure 3.31b, M2 is positive with respect to M1 but negative gate voltage is applied. Junction J4 is now forward-biased and electrons are injected from n3 into p2. A lateral current flows in p2 towards the n3 gate and the auxiliary SCR section p1-n1-p2-n3 turns on as the gain of the n3-p2-n1 transistor section increases. Current flow in this auxiliary SCR results in a current flow across J3 into n2, hence piloting the SCR p1-n1-p2-n2 into conduction.

(c) M1 positive, I_g negative (Mode III)

Figure 3.31c shows the bias condition with M2 negative with respect to M1 and the gate negative with respect to M1 such that J4 is forward-biased and electrons are injected from n3 into the p2 region. The potential in n1 is lowered, causing holes to be injected from p2 into the n1 layer which provide base current for the p2-n1-p1 transistor section. This brings the p2-n1-p1-n4 SCR into conduction.

(d) M1 positive, I_g positive (Mode IV)

When M2 is negatively biased with respect to M1 and the gate is positively biased such that J3 is forward-biased, as in figure 3.31d, electrons are injected from n2 to p2 and diffused to n1. This increases the forward bias of J2 and eventually the SCR section p2-n1-p1-n4 comes into full conduction.

The various turn-on mechanisms are highly reliant on the judicious lateral separation of the various contacts and regions. The main advantage of the triac lies in the fact that two anti-parallel SCR's in the one silicon structure can be triggered into conduction from the one gate. Because of the need for extra structure layers, hence processing steps, some conventional SCR characteristics are sacrificed and poor device area utilisation results. Two anti-parallel SCR's therefore tend to be more robust than a triac but unlike the BCT device in section 3.3.4, only one gate drive circuit is needed.

Reading list

Baliga, B. J., *Modern Power Devices*, John Wiley-Interscience, 1987.

Blicher, A., *Thyristor Physics*, Springer, New York, 1976.

Gentry, F. E. et al., *Semiconductor Controlled Rectifiers: Principles and Application of pnpn Devices*, Prentice-Hall, 1964.

Ghandhi, S. K., *Semiconductor Power Devices*, John Wiley-Interscience, New York, 1977.

Graham, D. R. et al., *SCR Manual* General Electric Company, 6th edition, 1979.

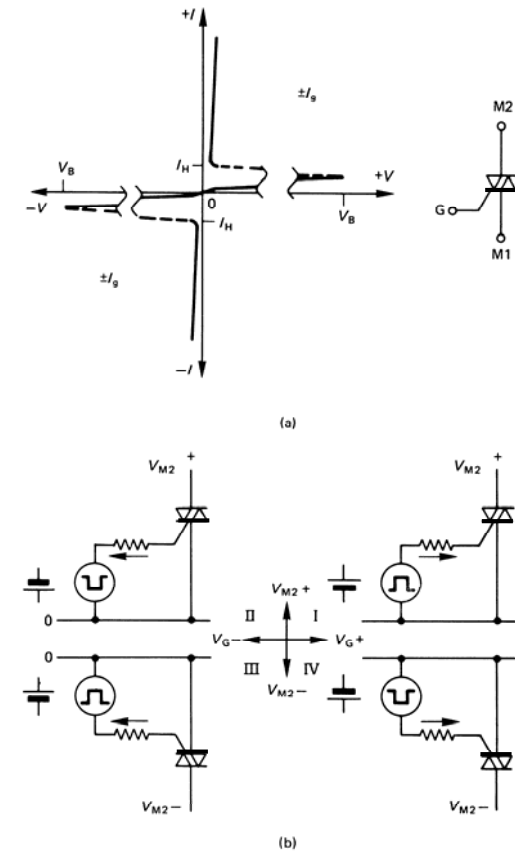


Figure 3.30. The triac: (a) I-V characteristics and circuit symbol and (b) its four firing modes.

