Schottky Barrier Diodes (SBD)—As Detector and Mixer

We know that for metal contacts with semiconductor to be ohmic, the semiconductor has to be doped close to degenerate level (i.e. n^{++}, p^{++} making its conductivity close to that of metal by doping density $\geq 10^{19}$/cc). If the semiconductor doping is just below degenerate level than normally with metal, it makes a Schottky barrier rectifying contact diode. The ON-OFF or OFF-ON switching time is very small (in pico seconds range), being a majority carrier device as in the 'n'-type devices electrons enter the metal quickly. Therefore it is used as μW detector diode (see Sect. 4.12.2).

Let ϕ_{m}, ϕ_{s} be the work function energy (in joules or eV) of metal and semiconductor. Work function is the difference between fermi level and its energy levels of electron in free state in vacuum (i.e. zero level) in the metal or semiconductor (Fig. 6.35). Then the difference of work functions between metal and semiconductor contact will be:

$$W = (\phi_{\mathrm{m}} - \phi_{\mathrm{s}}) \tag{6.52}$$

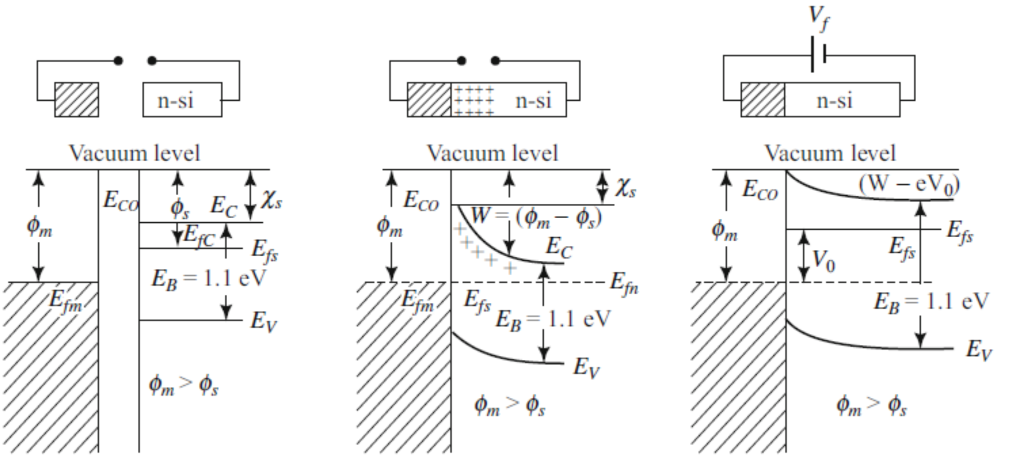
Following three conditions are there for having a Schottky barrier rectifying diode property.

(a) Doping density is low ($<10^{17}$–10^{18}/cc).
(b) Metal of type $\phi_{\mathrm{m}} > \phi_{\mathrm{s}}$ (i.e. $W = +\mathrm{ve}$) for n-type semiconductor (Fig 6.35).
(c) Metal of type $\phi_{\mathrm{m}} < \phi_{\mathrm{s}}$ (i.e. $W = -\mathrm{ve}$) for p-type semiconductor (Fig 6.36).

Reverse of any of the above conditions makes it ohmic contact, e.g. $n > 10^{18}$/cc or $\phi_{\mathrm{m}} < \phi_{\mathrm{s}}$ (in n-type) or $\phi_{\mathrm{m}} < \phi_{\mathrm{s}}$ (in p-type).

Work functions of platinum, palladium, nickel, gold are ≥ 5.1 eV and therefore are more suitable for n-type Schottky, while molybdenum, silver, aluminium, titanium, and tungsten have $\phi_{\mathrm{m}} \approx 4.2$–$4.6$ can be used for p-type Schottky diode (Tables 6.6 and 6.7).

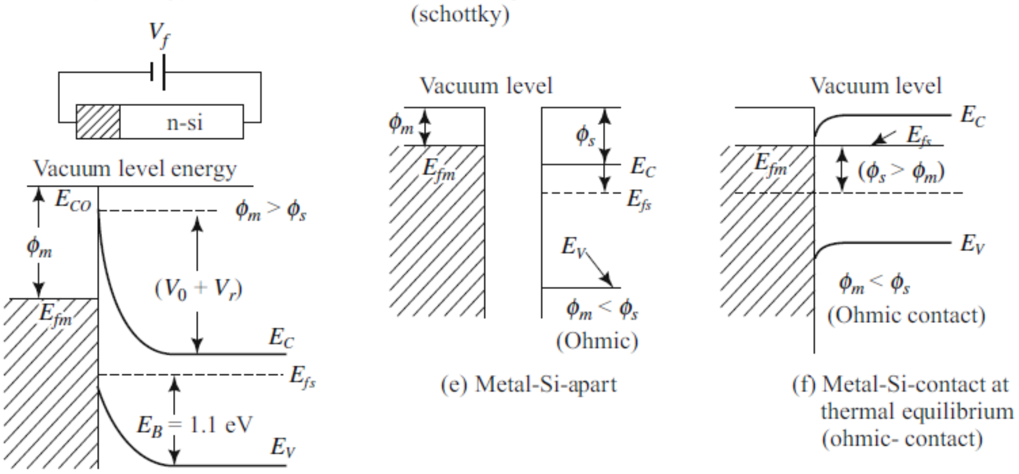
(i) **n-type semiconductor in contact with metal of type $\phi_{\mathrm{m}} > \phi_{\mathrm{s}}$, ($E_{fm} < E_{fs}$) making Schottky contact:**



(a) Metal-Si-apart (schottky)

(b) Metal-Si-contact at thermal equilibrium (schottky)

(c) Forward bias (schottky)



(e) Metal-Si-apart

(f) Metal-Si-contact at thermal equilibrium (ohmic- contact)

(d) Reverse bias (schottky)

Fig. 6.35 Energy bond diagram for *n*-type semiconductor in contact with metals of two type (i) $\phi_m > \phi_s$ for Schottky barrier diode rectifier **a** for apart, **b** in contact, **c** forward bias (V_f), and **d** reverse (V_f) bias cases (ii) $\phi_m < \phi_s$ ohmic **e** apart **f** in contact. Here E_c. E_v, E_{fs} are conduction band level, valence band level, and fermi level of the *n*-semiconductor

When the metal and *n*-semiconductor are in contact (Fig. 6.35) such that work function ϕ_m of metal (energy needed to remove an electron from solid out to a point in the vacuum) is higher than ϕ_s in semiconductor, i.e. fermi energy level of electrons in semiconductor is higher than in metal, as a result electrons will flow from *n*-semiconductor to the metal, till the fermi-level equalisation takes place. This process makes the semiconductor +vely charged as it gets depleted of electrons, generating a potential barrier of $eV_0 = (\phi_m - \chi_s)$ Joules or V_0 electron volts, with a field which opposes further flow of electrons. Here χ_s is the affinity of electron in semiconductor.

Under forward bias (V_f), the potential barrier height reduces to $eV_0 = (\phi_m - \chi_s - eV_f)$ Joules (Fig. 6.35c) as a result more electrons gets injected from semiconductor into the metal. In reverse bias case, this barrier height increases (Fig. 6.35d) and electron flow almost stops. Further increase of reverse bias leads to breakdown.

Table 6.6 Electron affinity (χ) and work function of some semiconductors

Semiconductor	Electron affinity (χ_s)	[a]Work function of the semiconductors (ϕ_s) $\phi_s = [\chi_s + (E_c - E_f)]$	
		For *n*-type (ϕ_{sn})	For *p*-type ϕ_{sp}
Ge (eV)	4.13	4.23	4.66
Si (eV)	4.01	4.11	5.01
GaAs (eV)	4.07	4.17	5.27

[a]Here it has been assumed that $(E_c - E_f) - (E_c - 0.1)$ eV for *p*-type close to $\approx$0.1 eV for *n*-type close to degenerate

Table 6.7 Metals suitable for Schottky diodes

Suitable for *n*-type semiconductor metals with ($\phi_m > \phi_s$)		Suitable for *p*-type semiconductor metals with ($\phi_m < \phi_s$)	
Metals	ϕ_m (eV)	Metals	ϕ_m (eV)
Au	5.10	Ag	4.26
Ni	5.15	Al	4.28
Pd	5.12	Ti	4.33
Pt	5.65	Cr	4.50
		Mo	4.60
		W	4.55

The *I–V* characteristic of the diode action is shown in Fig. 6.37, and it acts as Schottky barrier rectifying diode, with turn-on voltage of 0.3 V or so, i.e. just half of normal diode as barrier potential is in semiconductor only.

(ii) **n-type semiconductor in contact with metal of type $\phi_m < \phi_s$, $(E_{fm} > E_{fs})$ making ohmic contact:**

With such metal, the band diagram without and with contact with semiconductor will be as given in Fig. 6.35e, f, respectively. The free electron negative charge in metal being at higher level ($E_{fm} > E_{fs}$) will flow to the semiconductor after contacting, increasing the −ve charge over there (see Fig. 6.35f), but no depletion region or barrier potential is formed. Therefore full current flows in both the directions (whether forward or reversed biased) and hence behaves as an ohmic contact.

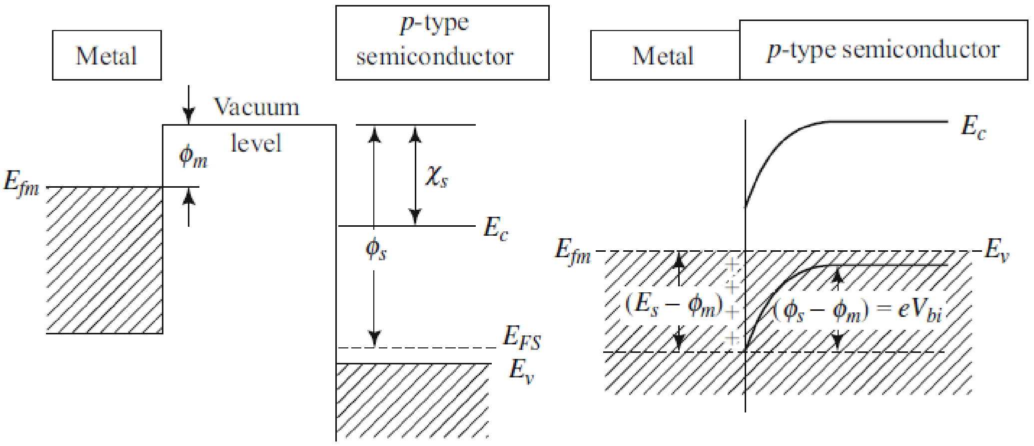
(iii) **p-type semiconductor in contact with metal of type $\phi_m < \phi_s$, $(E_{fm} > E_{fs})$ making Schottky diode:**

Such a contact can be well explained by diagram Fig. 6.36a, b. The electron has higher energy in metal, therefore flows from metal to semiconductor, until fermi level is same thought. The excess −ve charge in semiconductor makes an electric field across the junction and hence a potential barrier, therefore it acts like a diode with rectifying behaviour.

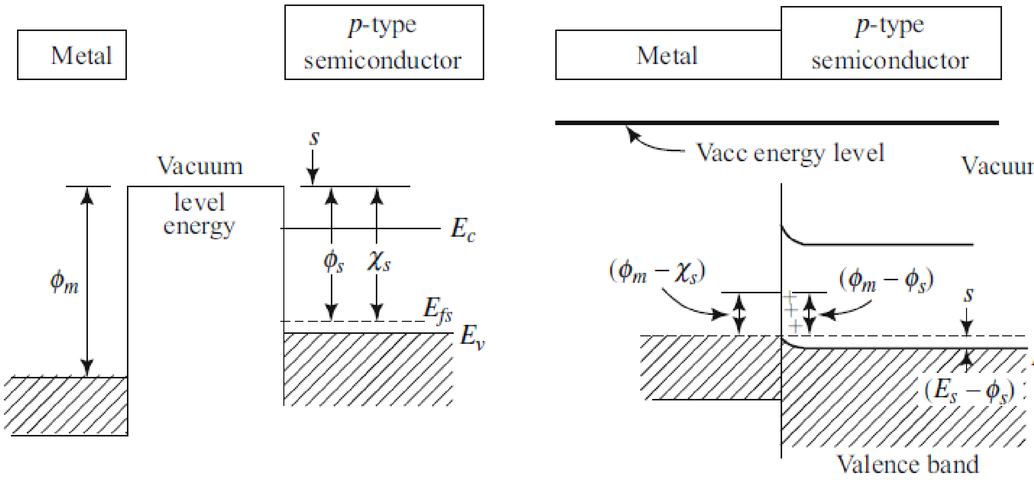
(iv) **p-type semiconductor in contact with metal of type $\phi_m > \phi_s$, $(E_{fm} < E)$ making ohmic contact:**

Such a contact is explained in Fig. 6.36c, d for metal and semiconductor separate and then in contact, respectively. Here the fermi level (E_{fs}) in semiconductor has higher energy level than of metal and therefore electrons flow from semiconductor to metal, making the semiconductor as positively charged, but no depletion or potential barrier. Therefore it behaves as ohmic contact, in forward as well as in reverse bias.

The Schottky barrier diode is sometimes called the hot electron diode because of electrons in semiconductor having higher energy level than in metal. In the forward current, it is all majority



(a) $\phi_m < \phi_s$ (separate) (b) $\phi_m < \phi_s$ (in contact Schottky diode)



(c) $\phi_m > \phi_s$ (separate) (d) $\phi_m > \phi_s$ in contact (ohmic)

Fig. 6.36 Energy-level diagrams *p*-type semiconductors in contact with metal: **a** and **b** are for $\phi_m < \phi_s$. Contact (b) acts as a Schottky barrier diode rectifier. In **c** and **d** the contact is ohmic

carriers with minority initially absent. Therefore it has very short reverse recovery time as the storage capacitance is almost nil. Therefore Schottky diode switches from non-conducting stable to conductivity state very fast (in less than 100 p/s), whereas ordinary pn junction switching time is large (around 100 n/s).

Figure 6.38 gives the construction of Schottky diodes (a) the point contact type and (b) planar technology type. The latter has a n^+–Si-substrate, upon which a thin epitaxial *n*-layer of 2–3 μ thickness is grown. Schottky contact is from n-surface to Si-ohmic contact from n^+ is taken through a window opened and gold evaporated for Au–Al contact.

The second one can be manufactured in large scale easily, but can be used up to 100 GHz only due to larger metal contact area capacitance of 0.3–0.5 pF. Two more advantages of this type are:

(a) Lower forward resistance (<0.5 Ω)
(b) Lower noise generation (<4 dB)

The point contact type can be used for frequencies up to 1000 GHz due to very low shunt capacitance of the contact of 0.01 pF, but has high series resistance of 2–5 Ω.

The limitations of SBD are (a) low reverse breakdown voltage (<100 V) and (b) high

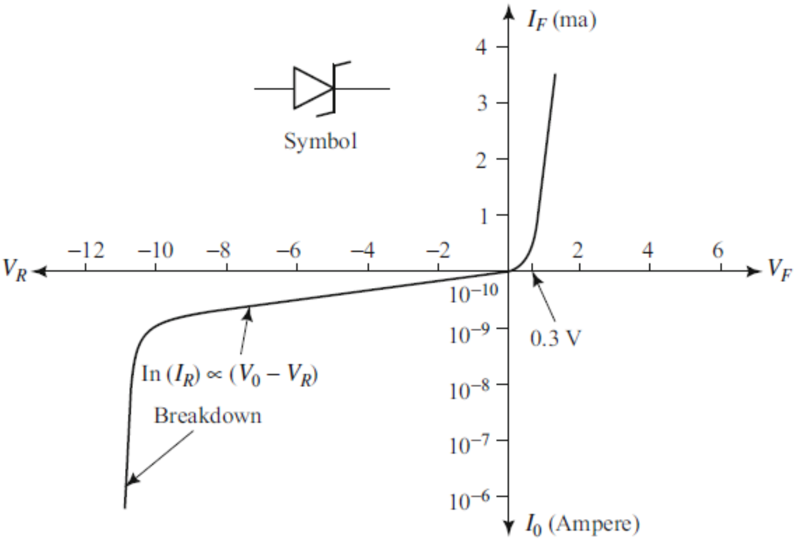


Fig. 6.37 *I–V* characteristic of a typical *n*-type Schottky barrier diode

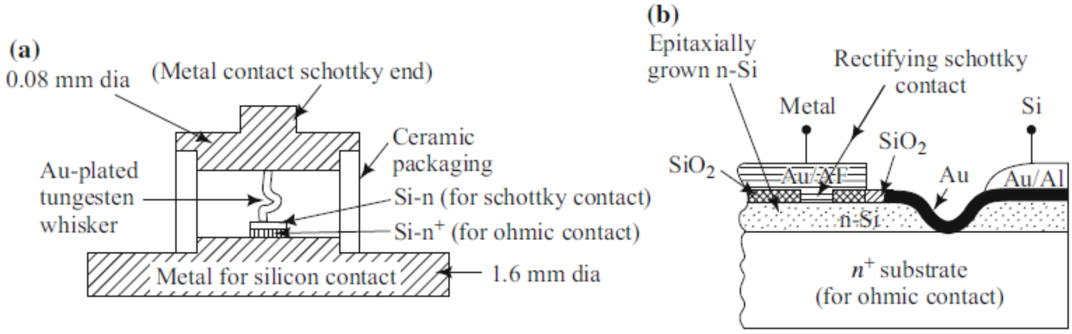


Fig. 6.38 Construction of Schottky barrier diode (SBD) **a** point contact Schottky **b** planar technology Schottky

reverse leakage current which increases with temperature causing thermal instability. The reverse bias leakage current is of the order of 10^{-6} A/cm^2 when compared to 10^{-11} cm^2 in conventional pn junctions.

For more properties in a circuit as a detector, see Sect. 4.12.2.

6.12.1 Application

The SBDs are used in the following circuits as microwave devices:

(i) Mixer in CW-RADAR

(ii) Microwave power detection as in Fig. 6.39 due to very small switching time $\approx 10^{-9}$ s. The rectified (detected) output is dc, and it goes to VSWR meter, etc.

6.13 PIN Diode for Switching/Controlling Microwave Power, Phase Shifting, Modulating etc.

PIN diode is very useful control element used at microwave frequencies as a (a) switch, (b) attenuator, (c) phase shifter, (d) power limiters,

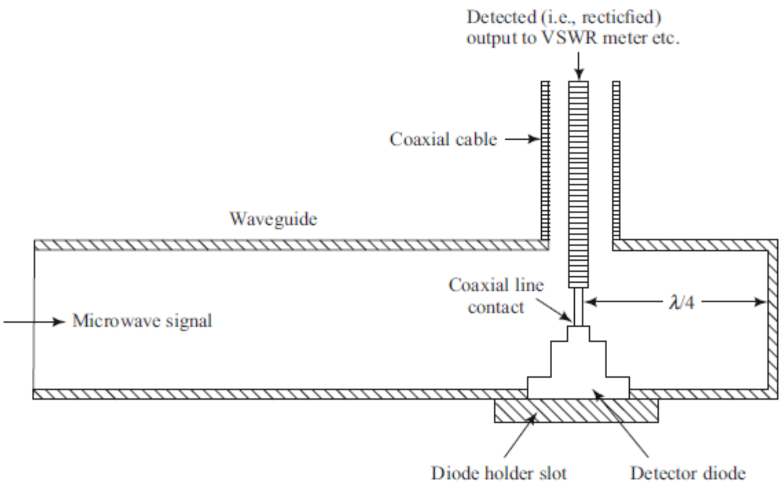


Fig. 6.39 Schottky diode detector mount assembly in a waveguide

(e) amplitude modulating element. This is because of its four important properties:

1. It can control very large microwave power of kW range, just by changing its bias by a small voltage.
2. It does not behave as a rectifier at microwave frequencies.
3. Breakdown voltage is very large generally over 500 V, as a result, even large μW power in its positive cycle cannot forward bias it.
4. Capacitance is very small.

Now we will discuss its structure, characteristics, and working:

(i) **Structure**: PIN Diode consists of heavily doped *p*- and *n*-regions separated by a high-resistivity *i*-region (≈1000 Ω cm) (Fig. 6.40), nearly intrinsic. In fact, this so-called i-region is high-resistivity *p*-layer (called π-type) or high-resistivity *n*-type (called ν-type). The reverse bias resistance being very high, most of the reverse bias voltage is across it, fully depleting the region. Therefore the reverse breakdown voltage will be very high over 1000 V or so and the capacitance very small (0.2 pF or so). Therefore:

(i) Length L of the *i*-region is kept large (100 μm or so)
(ii) Doping level is kept very low (10^{12}–10^{13}), i.e. π or ν-type
(iii) Device capacitance is kept quite small (<1 pF by keeping small area). Moreover it remains constant with voltage, as whole of *i*-region gets depleted with a very small reverse bias voltage itself and higher reverse voltage has no effect.

$$C_s = (\varepsilon_s A/L) \tag{6.53}$$

(iv) For $L = 200$ μm, transit time of charge across *i*-region is approximately:

$$\begin{aligned}\tau &= L/v_s \\ &= (200 \times 10^{-4}/1.3 \times 10^{-2}) = 1.4 \times 10^{-9}\,\mathrm{s}\end{aligned} \tag{6.54}$$

Figure 6.40 gives the (a) systematic diagram, (b) impurity concentration, (c) space charge, and (d) electric field distribution in fully depleted PIN diode. Figure 6.41 gives the equivalent circuit and RV and IV characteristics of the three biasing regions (before A, AB and after B). Figure 6.42 gives the actual PIN device and diode structures. It also gives the V_B *vs* doping level, forward bias resistance *vs* current and transit time *vs* length of the I region of the PIN diode.

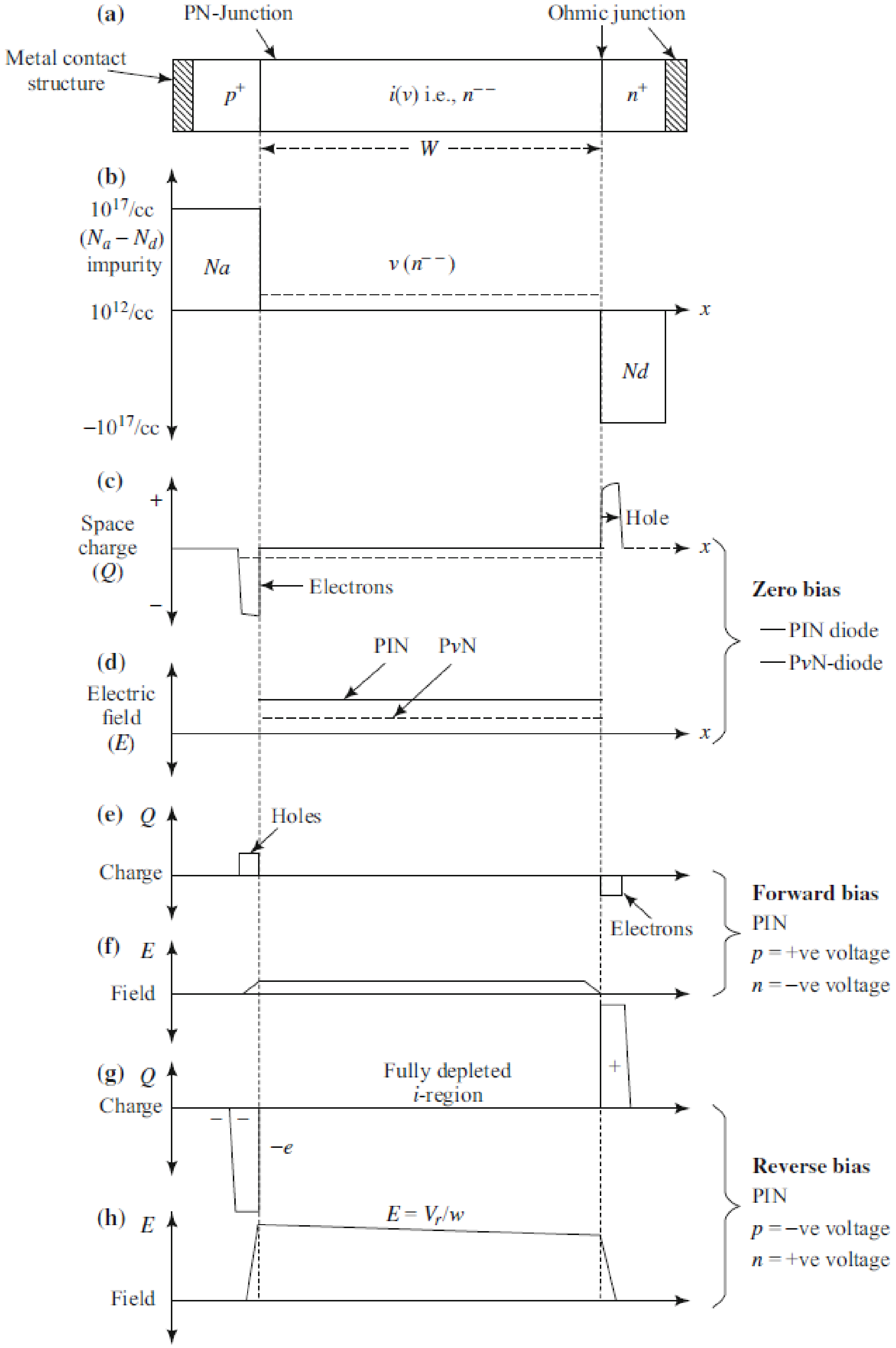


Fig. 6.40 PIN diodes **a** schematic diagram, **b** impurity distribution, **c** and **d** space charge and electric field for zero bias case, **e** and **f** space charge and electric field for forward bias case, **g** and **h** space charge and electric field for reverse bias case

(i) Characteristic

Because of the following three important properties of PIN diode, it is used as microwave switch;

1. **Can control/switch large μW power** of kW level or so, by a small change of bias from forward to reverse. This is by allowing the μW signal to flow through a very low resistance of 0.1 Ω or so in forward bias and stop μW signal in reverse bias resistance of 10 kΩ or so.
2. **Does not behave as a rectifier at microwave frequencies**, like other diodes like Schottky diodes, etc. At low frequencies up to 100 MHz, PIN diode behaves like other pn-diodes as rectifiers, but at higher frequencies the rectification property decreases due to
   (a) Large transit time across the *i*-region.
   (b) Large switching/recombination time (τ_{sw}) from ON to OFF, causing the carrier storage taking place in the *i*-region, which as a result acts as a variable resistance as a function of both voltage and current. This switching time/carrier re-combination time is $\tau_{sw} \approx 10^{-4}$ s for silicon and 10^{-9} s for GaAs from ON (forward bias) to OFF (rev. bias), which is ≫ than μW time period.
   (c) Large carrier lifetime τ_{sw}, which in turn leads to large diffusion length which is the length after moving, it recombines

$$L_d = \sqrt[2]{(D_{Si} \cdot \tau_{sw})} \approx 400\text{–}600\ \mu\text{m}$$

3. **Very high breakdown voltage V_{bd} of over 1000 V or so**, as V_{bd} depends on the doping and on W. For large V_{bd}, W had to be large but less than the diffusion length L_d for recombination to take place within *i*-layer thickness W, otherwise forward bias voltage drop across *i*-region as well as forward bias resistance will become high. For $W = 100$ μm, typical values of V_{bd} and L_d are as:

Semiconductor	Doping/cc	L_d (μm)	V_{bd} (V)
GaAs	10^{12}–10^{13}	10–20	150–250
Si	10^{12}–10^{13}	400–600	1000–6000

GaAs PIN diode has the disadvantages of (a) low V_{bd} and (b) three times higher thermal resistance, but has the advantages of smaller switching/recombination time < 10 ns, still Si-PIN diode is preferred specially due to high V_{bd}.

Therefore PIN diode is safe even at large μW power, as in its positive cycle also it cannot forward bias it, when kept at a reverse bias of 500 V. For handling still larger power, series–parallel connection can be made.

4. **Capacitance is very low due to** (a) large W and (b) by keeping/device area small. As a result the forward bias device impedance is also small.

(ii) Working Mechanism

1. At zero bias: The diffusion of holes and electrons across the junction causes;
   (a) Space charge region: The thickness of which is inversely proportional to doping density.
   (b) Fixed +ve space charge in the *n*-region.
   (c) Fixed −ve space charge in the *p*-region
   (d) No depletion region in ideal *i*-region

   ——

   (e) High resistance of PIN diode 10 kΩ or so Figs. 6.40c and 6.41.
2. At reverse bias: The space charge regions of p and n-layers become thicker and denser (Fig. 6.41b), with reverse resistance remaining high ≈10 kΩ and constant (Fig. 6.40g, h) along with uniform and high electric field, falling to zero in p^+, n^+ regions. This high resistance is equivalent to open circuit.
3. At forward bias: The hole and electron carriers get injected from both the junctions (n^+i and p^+i), respectively, into the *i*-region and this:
   (a) reduces the thickness as well as density of the space charge regions of n and p layers.

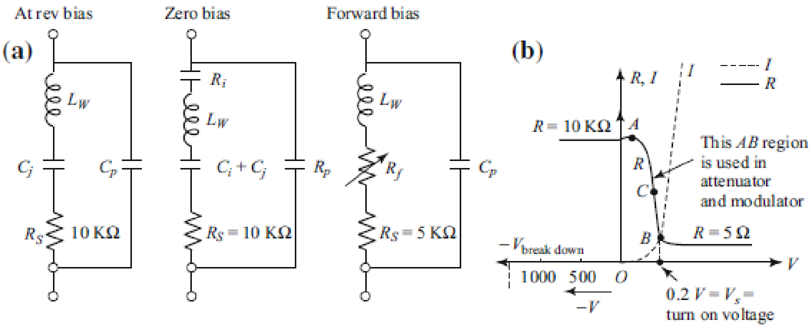


Fig. 6.41 PIN diodes a equivalent circuit of diode in package C_p = package cap; C_j = junction cap; C_{is}, R_i = Unswept intrinsic portion cap and res; R_s = series res; R_f = forward res; L_w = bonding Au-wire inductance and b R–V and I–V characteristic

(b) raises the carrier concentration in the *i*-layer above equilibrium.
(c) reduces the resistance of *i*-layers.
(d) causes fall of over all resistance of the diode, (Fig. 6.40e, f).
(e) causes virtual short circuit, as the switching time/carrier lifetime ($\approx 10^{-4}$–10^{-9} s) in this-*i*-region is ≫ period of microwave frequency. From high µW power to low power switching time is 40–1 ns.

6.13.1 PIN Diode Application in Circuits (as Switch, Attenuator, Phase Shifter, Limiter, and AM Unit)

PIN diodes are used in the circuit as (a) switch, (b) attenuator, (c) phase shifter, (d) power limiter, (e) amplitude modulating elements.

The property of short and open in forward and reverse bias is used as switch, and the switch property is used as phase shifter and limiter. The variation in resistance with bias voltage (Figs. 6.41b and 6.42) is used in attenuator and modulator.

(a) **PIN Diode as a Microwave Power Switch**: PIN diode is used in the series configuration or shunt configuration as in Fig. 6.43. The series configuration is more suitable for coaxial line while shunt one for waveguides. With the dc forward bias, the
  (i) series circuit transmission is ON.
  (ii) shunt circuit transmission is OFF.

In I-region impurity concentration has to be quite low, width $L > 20$ µm so that breakdown voltages remain large >500 V. This is because the microwave signal should not be able to forward bias it even in its +ve cycle, when it is reverse biased, i.e. OFF. The capacitances C_1, C_2 and RFC_1, RFC_2 have to be large to pass and stop the signal, respectively.

(b) **PIN Diode as Attenuator**: If we use *A*–*B* portion of the *I*–*V* characteristic in Fig. 6.41b, where the forward resistance decreases with bias along the points *A*, *C*, *B*, then PIN diode can function as attenuator using the same circuit of switch (Fig. 6.43). In this attenuator, the forward bias value will control the attenuation.
At point A-bias the series (shunt) configuration has max (nil) attenuation
At point B-bias the series (shunt) configuration has max (nil) attenuation

Fig. 6.42 PIN diode **a** construction of the planar diode device with measurements in mil and **b** a typical diode in package. **c** Breakdown voltage of PIN diode as a function of doping in I-layer for different width (W_I). Corresponding to the doping, the resistivity (ρ) in Si is also given at the top (from references 12 and 13, Chaturvedi et al.). **d** PIN diode as a variable resistance 0.1 Ω – 10 kΩ versus forward bias current 1 µA –100 mA, with centre point C at 50 Ω making Smith chart analyses. **e** Carrier transit time t_{tt} as a function of L and V_{BD}

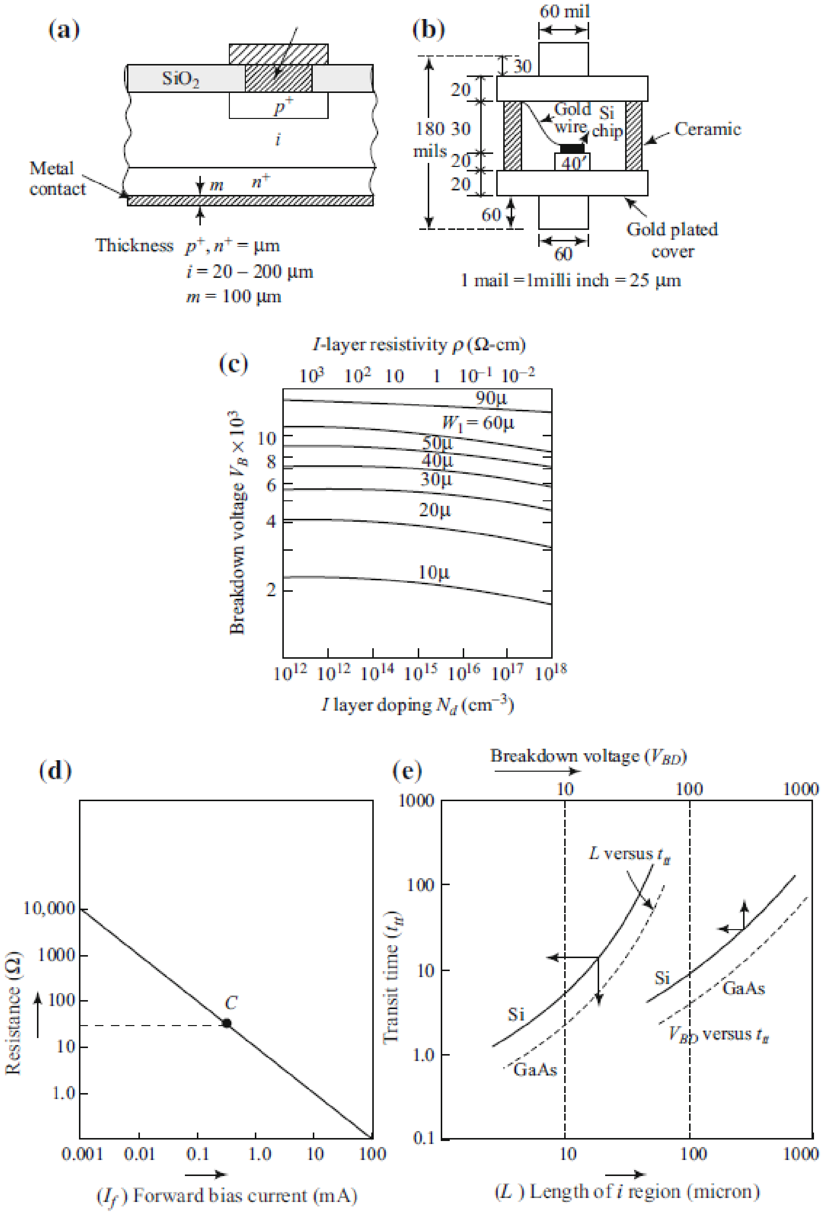
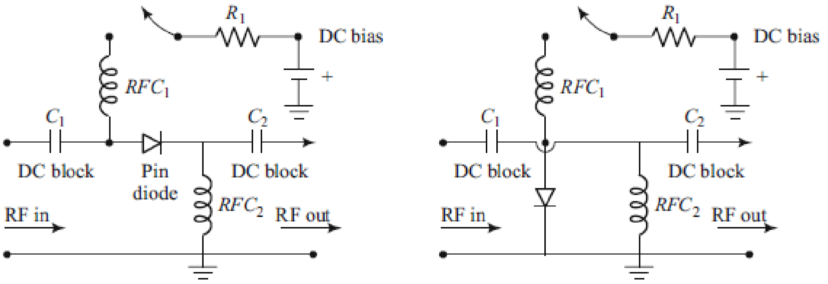


Fig. 6.43 PIN diode as a switch



(a) Series configuration ON-with forward bias OFF-with reverse

(b) Shunt configuration ON-with reverse bias OFF-with forward bias

At point C-bias the series (shunt) configuration has medium (medium) attenuation

(c) **PIN Diode as Phase Shifter**: The switch property is indirectly used for introducing a phase shift in the signals (Fig. 6.44). The signal is made to travel a length of 2 $(l_1 + l_2)$ at the port 2 when diode is OFF (to and fro path), while only length $2l_1$ when diode is ON i.e. short at diode. The phase shift $2l_2$ can be changed by moving the short plunger. The phase shift is

$$\phi = 2\pi \cdot l_2/\lambda \tag{6.55}$$

l_2 has to be fraction of λ. These phase shifters are used in phased array radars.

(d) **PIN Diode as Power Limiter**: PIN diodes are also used as microwave power limiters and act as short beyond a power limit (which is set $p_{\max}$ as per requirement). Here we make use of the diode property, i.e. beyond V_1 (the junction turn-on voltage) acts as short (Very low resistance) and microwave voltage above V_1 gets shorted (Fig. 6.45). Therefore external bias on diode is not required and microwave signal voltage acts as bias.

For increasing the power handling capacity, $p_{\max}$, shorting current can be increased by having more number of diodes in shunt and can handle up to 100 kW.

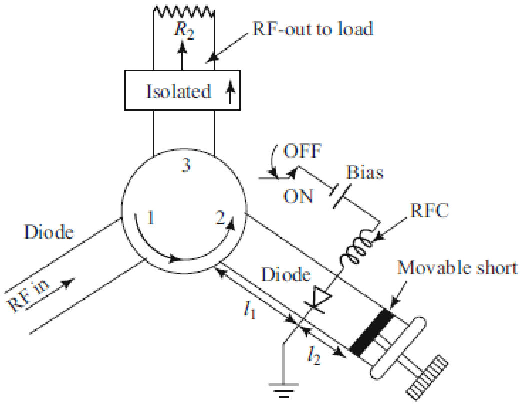


Fig. 6.44 Phase shifting using PIN diode as switch

Thus shorting the input power beyond $p_{\max}$ will not allow power to reach the output end beyond $p_{\max}$, for protection of microwave system.

(e) **PIN Diode as Amplitude Modulator**: PIN diode can be employed as an amplitude modulator as in Fig. 6.46, where the amplitude modulating (AM) signal of low frequency (f_m) and amplitude lower than the microwave carrier signal (t) are mixed. The diode is kept at a very low reverse (or zero) bias and in series with the modulating signal (f_m).

The f_m signal in its +ve cycle (point B) will make the diode forward biased and hence of very low resistance across PQ (Fig. 6.45a, b), allowing very low power to reach the output end. During the −ve cycle of f_m, (points A_1, A_2), more power will reach the output and hence we get AM output. The modulation could be by a square wave also. In this reference, the Gunn diode power supply could be referred (see Fig. 8.11 also).

6.14 Varactor Diode as a Variable Capacitor

The term varactor (or varicap, as it is so called) was coined from its property of variable reactance (or capacitor) of a pn junction with reverse bias, due to variation of its depletion layer width (acting as a dielectric) (Fig. 6.47).

It is a pn junction having I–V characteristic just like other pn junction diodes, but because of special impurity doping (Figs. 6.48 and 6.49) profile (abrupt or hyperabrupt), its properties differ as:

(a) Its capacitance vary in a nonlinear manner $\left(C\alpha 1/V_R^n\right)$ (Fig. 6.50) with the reverse bias voltage.

(b) It is fast enough to follow microwave frequency.

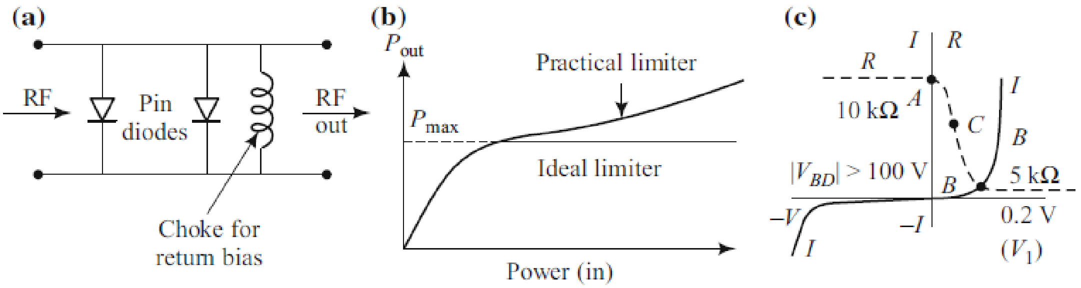


Fig. 6.45 a PIN diode limiter and b input–output characteristic of PIN-limiter c *I*–*V* and *R*–*V* characteristic and of PIN diode

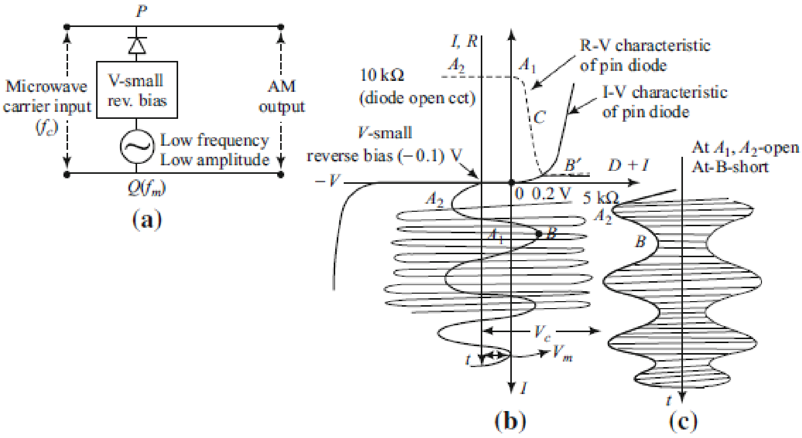


Fig. 6.46 PIN diode as amplitude modulator **a** the circuit in principle, **b** mix of *dc* with f_m, f_c, **c** modulated outputs

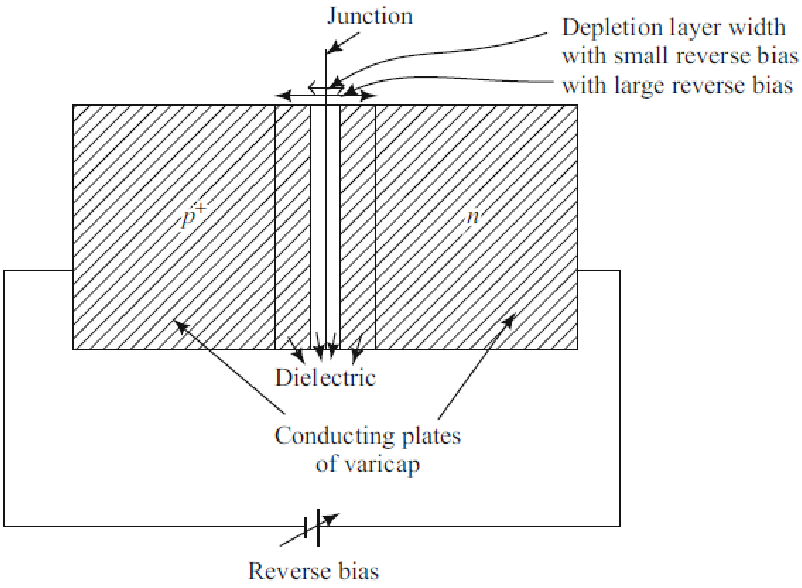


Fig. 6.47 Increase of reverse bias increases the depletion width and hence reduces capacitance ($C = \varepsilon A/d$)

(c) The junction capacitance variation depends on the type of impurity profile linear or abrupt or hyperabrupt. The hyperabrupt gives best variation of C (Fig. 6.50). i.e. gradual change with voltage.

(d) It has negligible power loss, as the equivalent series resistance is very small (Fig. 6.51).

In an ordinary pn junction diode, the C–V curve does not show large variation but in varactor diode it does, as the depletion width vary as:

$$\begin{array}{l} W\alpha(V_0+V_s)^m \\ C\alpha\frac{1}{(V_0+V_R)^m} \end{array} \left(\begin{array}{l} m=\frac{1}{3}\text{ for linearly graded Junction} \\ m=\frac{1}{2}\text{ for abrupt Junction} \\ m=2\text{ for hyper abrutpt Junction} \end{array} \right) \tag{6.56}$$

6.14.1 The Device Structure

The three types of impurity profiles (linear, abrupt, and hyperabrupt) are given in Figs. 6.48 and 6.49 with their C–V characteristic in

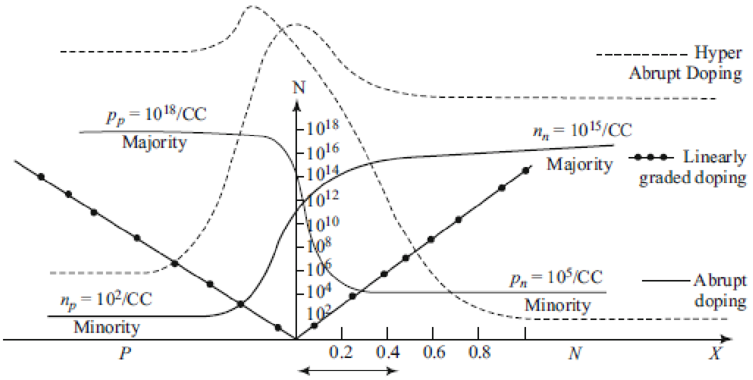


Fig. 6.48 Doping density and carrier density of a linearly graded/abrupt/hyperabrupt pn junction diode around the junction

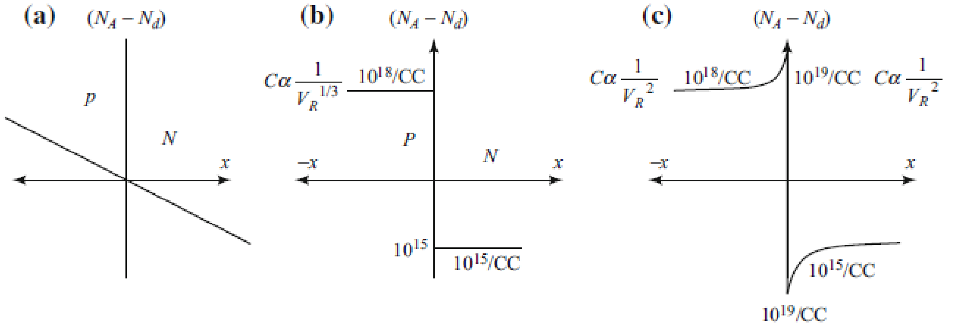


Fig. 6.49 Ideal doping profile of pn junction electrons (Jn) show by $N_a - N_d$ around the junction, **a** both sided linearly graded Jn (ordinary diode), **b** both sided abrupt Jn (varactor diode), **c** both sided hyperabrupt Jn (best varactor diode) with gradual variation of C with V_o

Fig. 6.50 *C–V* characteristic of a typical hyperabrupt varactor diode and ordinary diode

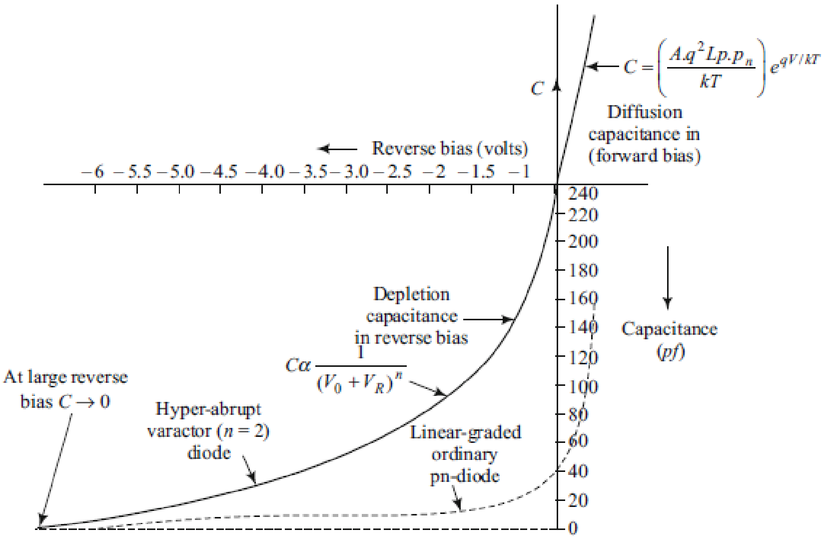


Fig. 6.51 Varactor diode **a** full equivalent circuit, **b** simplified equivalent circuit, **c** symbols used, **d** packaged device with mesa-structure (trapezium type)

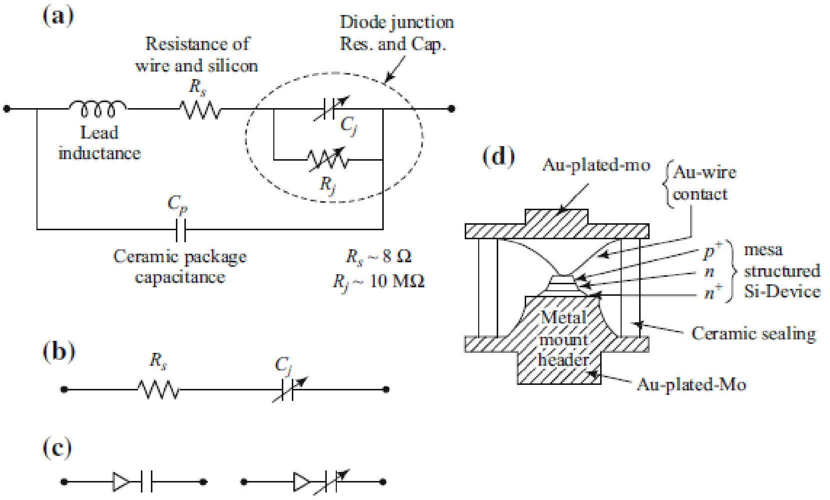


Fig. 6.50. In the abrupt and hyperabrupt junction, the change of in purity from N_D to N_A takes place in a very small distance of 0.4 µm, while in linear graded it happens in 2–10 µm region. As a result of this, depletion region in hyperabrupt junction is very thin leading to higher capacitance value as well as gradual variation of *C* with reverse bias as shown in Fig. 6.50. The abrupt junction diode is fabricated (Fig. 6.51d) by starting with n^+ substrate over which epitaxial growth of *n*-layer is done and the np^+ is diffused. For hyperabrupt diode, on the n^+ substrate, two epitaxial layers of *n*-Si and p^+-Si are deposited, such that at the junction, the doping of *n* and p^+ is very high (Fig. 6.49c).

Individual Si devices are then given a mesa-structure (Trapezium Type) before dicing from the water. The device is then mounted on the header by T.C. bonding and then encapsulated with gold wire bonded on the top (p^+) (Fig. 6.51).

6.14.2 Characteristic

(a) **At zero bias** the pn junction has:
- Built-in voltage due to diffusion of charges $V_0 = \frac{kT}{\mathrm{e}} \ln\left(\frac{N_\mathrm{A} N_\mathrm{d}}{n_i^2}\right)$
- Diffusion capacitance $C_\mathrm{d} = \frac{A \cdot \mathrm{e}^2 \cdot L_\mathrm{p} \cdot p_n}{\mathrm{kT}} \cdot \mathrm{e}^{\frac{\mathrm{eV}}{\mathrm{kT}}}$

This C_d becomes very large, while V_0 depends on doping levels N_d, N_A on the two sides.

(b) **In the reverse bias** the variation of capacitance (C_j) is quite large in hyperabrupt junction diode Eq. (6.56) (Fig. 6.50).

$$C_\mathrm{j} \alpha \frac{1}{(V_0 + V_\mathrm{R})^2} \tag{6.57}$$

(c) **For large reverse** bias $V_\mathrm{R} \gg V_0$, therefore above equation reduces to an approximate equation as:

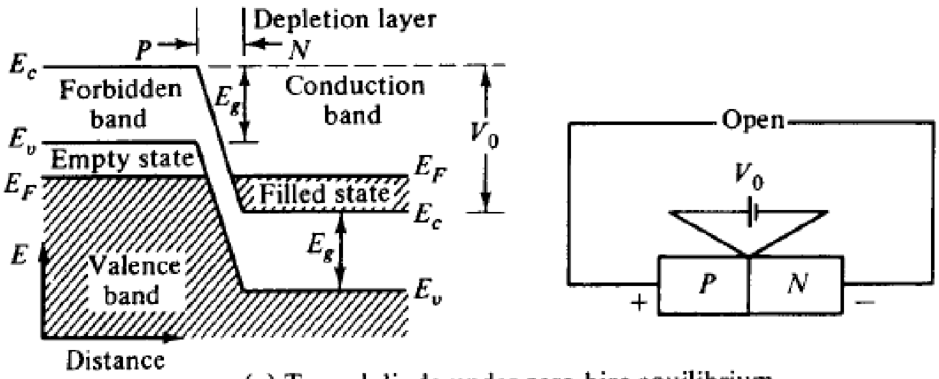
$$C_\mathrm{j} \alpha \frac{1}{V_\mathrm{R}^2} \tag{6.58}$$

MICROWAVE TUNNEL DIODES

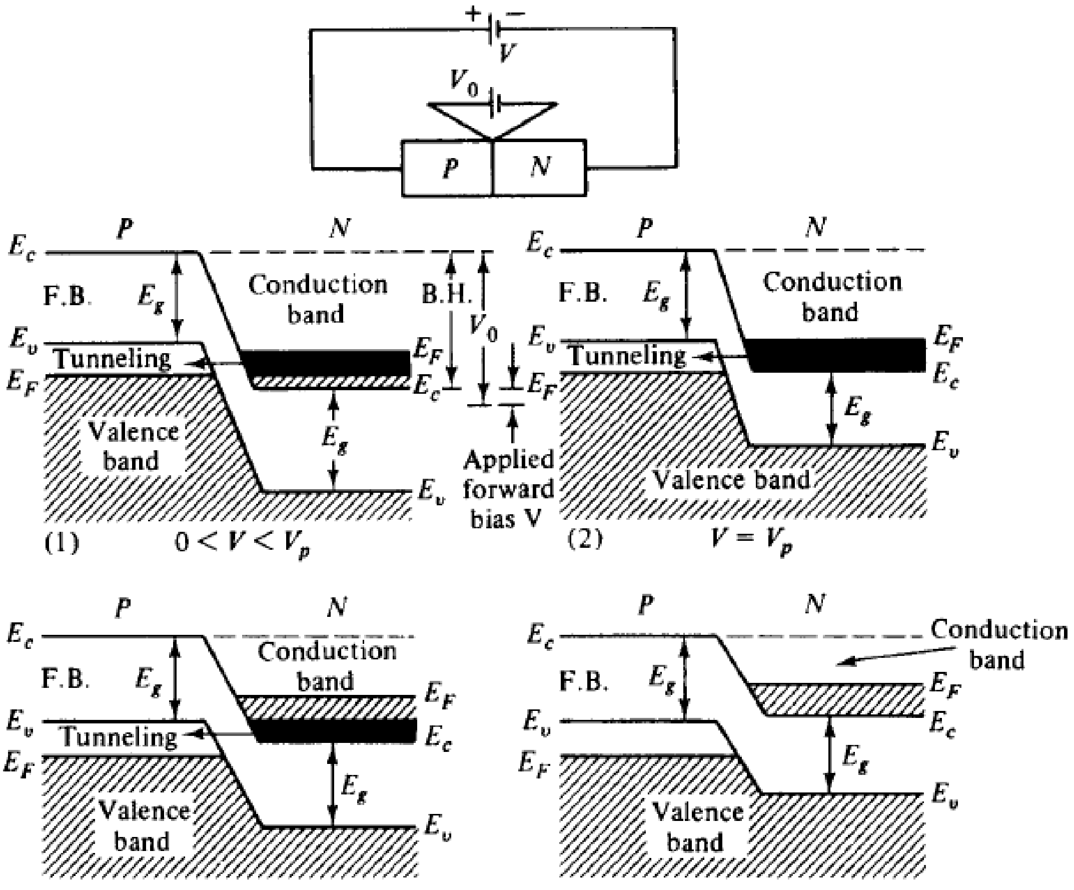
After the publication of Esaki's classic paper on tunnel diodes in 1958, the potential of tunnel diodes for microwave applications was quickly established. Prior to 1958 the anomalous characteristics of some *p-n* junctions were observed by many scientists, but the irregularities were rejected immediately because they did not follow the "classic" diode equation. Esaki, however, described this anomalous phenomenon by applying a quantum tunneling theory. The tunneling phenomenon is a majority carrier effect. The tunneling time of carriers through the potential energy barrier is not governed by the classic transit-time concept—that the transit time is equal to the barrier width divided by the carrier velocity—but rather by the quantum transition probability per unit time. Tunnel diodes are useful in many circuit applications in microwave amplification, microwave oscillation, and binary memory because of their low cost, light weight, high speed, low-power operation, low noise, and high peak-current to valley-current ratio.

5-3-1 Principles of Operation

The tunnel diode is a negative-resistance semiconductor *p-n* junction diode. The negative resistance is created by the tunnel effect of electrons in the *p-n* junction. The doping of both the *p* and *n* regions of the tunnel diode is very high—impurity concentrations of 10^{19} to 10^{20} atoms/cm^3 are used—and the depletion-layer barrier at the junction is very thin, on the order of 100 Å or 10^{-6} cm. Classically, it is possible for those particles to pass over the barrier if and only if they have an energy equal to or greater than the height of the potential barrier. Quantum mechanically, however, if the barrier is less than 3 Å there is an appreciable probability that particles will tunnel through the potential barrier even though they do not have enough kinetic energy to pass over the same barrier. In addition to the barrier thinness, there must also be filled energy states on the side from which particles will tunnel and allowed empty states on the other side into which particles penetrate through at the same energy level. In order to understand the tunnel effects fully, let us analyze the energy-band pictures of a heavily doped *p-n* diode. Figure 5-3-1 shows energy-band diagrams of a tunnel diode.



(a) Tunnel diode under zero-bias equilibrium



(b) Tunnel diode with applied forward bias

E_F is the Fermi level representing the energy state with 50% probability of being filled if no forbidden band exists

V_0 is the potential barrier of the junction

E_g is the energy required to break a covalent bond, which is 0.72 eV for germanium and 1.10 eV for silicon

E_c is the lowest energy in the conduction band

E_v is the maximum energy in the valence band

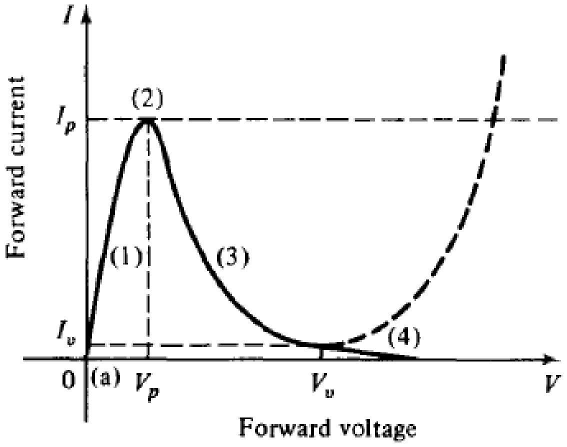
V is the applied forward bias

F.B. stands for the forbidden band

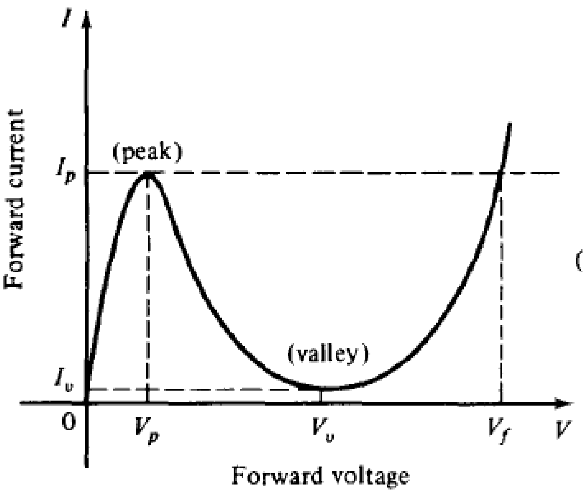
B.H. represents the barrier height

Figure 5-3-1 Energy-band diagrams of tunnel diode.

Under open-circuit conditions or at zero-bias equilibrium, the upper levels of electron energy of both the p type and n type are lined up at the same Fermi level as shown in Fig. 5-3-1(a). Since there are no filled states on one side of the junction that are at the same energy level as empty allowed states on the other side, there is no flow of charge in either direction across the junction and the current is zero, as shown at point (a) of the volt-ampere characteristic curve of a tunnel diode in Fig. 5-3-2.



(a) Tunnelling current (solid curve) and injection current (dashed curve)



(b) I–V characteristic of tunnel diode

Figure 5-3-2 Ampere-voltage characteristics of tunnel diode.

In ordinary diodes the Fermi level exists in the forbidden band. Since the tunnel diode is heavily doped, the Fermi level exists in the valence band in p-type and in the conduction band in n-type semiconductors. When the tunnel diode is forward-biased by a voltage between zero and the value that would produce peak tunneling current $I_p(0 < V < V_p)$, the energy diagram is shown in part (1) of Fig. 5-3-1(b). Accordingly, the potential barrier is decreased by the magnitude of the applied forward-bias voltage. A difference in Fermi levels in both sides is created. Since there are filled states in the conduction band of the n type at the same energy level as allowed empty states in the valence band of the p type, the electrons tunnel through the barrier from the n type to the p type, giving rise to a forward tunneling current from the p type to the n type as shown in sector (1) of Fig. 5-3-2(a). As the forward bias is increased to V_p, the picture of the energy band is as shown in part (2) of Fig. 5-3-1(b). A maximum number of electrons can tunnel through the barrier from the filled states in the n type to the empty states in the p type, giving rise to the peak current I_p in Fig. 5-3-2(a). If the bias voltage is further increased, the condition shown in part (3) of Fig. 5-3-1(b) is reached. The tunneling current decreases as shown in sector (3) of Fig. 5-3-2(a). Finally, at a very large bias voltage, the band structure of part (4) of Fig. 5-3-1(b) is obtained. Since there are now no allowed empty states in the p type at the same energy level as filled states in the n type, no electrons can tunnel through the barrier and the tunneling current drops to zero as shown at point (4) of Fig. 5-3-2(a).

When the forward-bias voltage V is increased above the valley voltage V_v, the ordinary injection current I at the p-n junction starts to flow. This injection current is increased exponentially with the forward voltage as indicated by the dashed curve of Fig. 5-3-2(a). The total current, given by the sum of the tunneling current and the injection current, results in the volt-ampere characteristic of the tunnel diode as shown in Fig. 5-3-2(b). It can be seen from the figure that the total current reaches a minimum value I_v (or valley current) somewhere in the region where the tunnel-diode characteristic meets the ordinary p-n diode characteristic. The ratio of peak current to valley current (I_p/I_v) can theoretically reach 50 to 100. In practice, however, this ratio is about 15.

5-3-2 Microwave Characteristics

The tunnel diode is useful in microwave oscillators and amplifiers because the diode exhibits a negative-resistance characteristic in the region between peak current I_p and valley current I_v. The I-V characteristic of a tunnel diode with the load line is shown in Fig. 5-3-3.

Here the *abc* load line intersects the characteristic curve in three points. Points a and c are stable points, and point b is unstable. If the voltage and current vary about b, the final values of I and V would be given by point a or c, but not by b. Since the tunnel diode has two stable states for this load line, the circuit is called *bistable*, and it can be utilized as a binary device in switching circuits. However, mi-

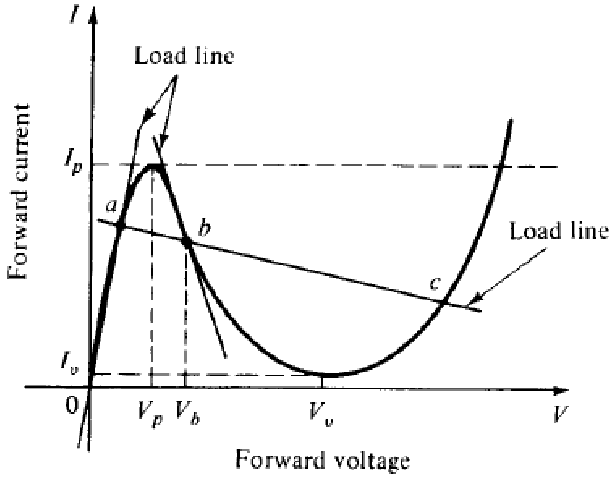


Figure 5-3-3 I-V characteristic of tunnel diode with load line.

crowave oscillation or amplification generated by the tunnel diode is our major concern in this section. The second load line intersects the *I-V* curve at point *b* only. This point is stable and shows a dynamic negative conductance that enables the tunnel diode to function as a microwave amplifier or oscillator. The circuit with a load line crossing point *b* in the negative-resistance region is called *astable*. Another load line crossing point *a* in the positive-resistance region indicates a *monostable* circuit. The negative conductance in Fig. 5-3-3 is given by

$$-g = \left.\frac{\partial i}{\partial v}\right|_{V_b} = \frac{1}{-R_n} \qquad (5\text{-}3\text{-}1)$$

where R_n is the magnitude of negative resistance.

For a small variation of the forward voltage about V_b, the negative resistance is constant and the diode circuit behavior is stable. A small-signal equivalent circuit for the tunnel diode operated in the negative-resistance region is shown in Fig. 5-3-4. Here R_s and L_s denote the inductance and resistance of the packaging circuit of a tunnel diode. The junction capacitance C of the diode is usually measured at the valley point; R_n is the negative resistance of the diode. Typical values of these parameters for a tunnel diode having a peak current I_p of 10 mA are

$$-R_n = -30\ \Omega \qquad R_s = 1\ \Omega \qquad L_s = 5\ \text{nH} \qquad C = 20\ \text{pF}$$

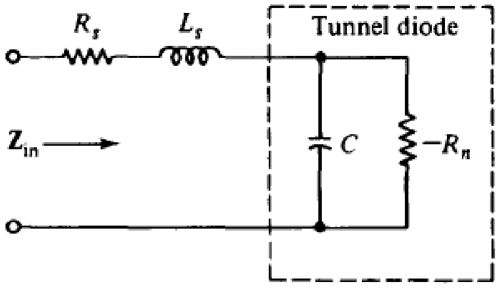


Figure 5-3-4 Equivalent circuit of tunnel diode.

The input impedance $\mathbf{Z}_{\text{in}}$ of the equivalent circuit shown in Fig. 5-3-4 is given by

$$\mathbf{Z}_{\text{in}} = R_s + j\omega L_s + \frac{R_n[j/(\omega C)]}{-R_n - j/(\omega C)}$$

$$\mathbf{Z}_{\text{in}} = R_s - \frac{R_n}{1 + (\omega R_n C)^2} + j\left[\omega L_s - \frac{\omega R_n^2 C}{1 + (\omega R_n C)^2}\right] \qquad (5\text{-}3\text{-}2)$$

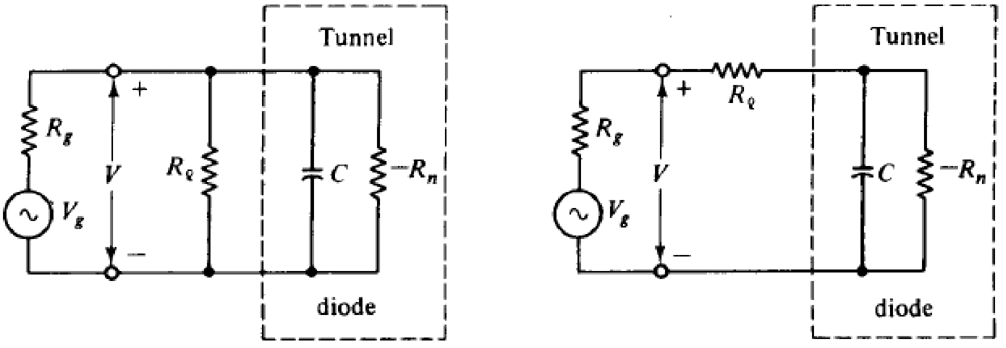
For the resistive cutoff frequency, the real part of the input impedance $\mathbf{Z}_{\text{in}}$ must be zero. Consequently, from Eq. (5-3-2) the resistive cutoff frequency is given by

$$f_c = \frac{1}{2\pi R_n C}\sqrt{\frac{R_n}{R_s} - 1} \qquad (5\text{-}3\text{-}3)$$

For the self-resonance frequency, the imaginary part of the input impedance must be zero. Thus,

$$f_r = \frac{1}{2\pi R_n C}\sqrt{\frac{R_n^2 C}{L_s} - 1} \qquad (5\text{-}3\text{-}4)$$

The tunnel diode can be connected either in parallel or in series with a resistive load as an amplifier; its equivalent circuits are shown in Fig. 5-3-5.



(a) Parallel loading

(b) Series loading

Figure 5-3-5 Equivalent circuits of tunnel diodes.

A tunnel diode can be connected to a microwave circulator to make a negative-resistance amplifier as shown in Fig. 5-3-6. A microwave circulator is a multiport junction in which the power may flow only from port 1 to port 2, port 2 to port 3, and so on in the direction shown. Although the number of ports is not restricted, microwave circulators with four ports are most commonly used. If the circulator is perfect and has a positive real characteristic impedance R_0, an amplifier with infinite gain can be built by selecting a negative-resistance tunnel diode whose input impedance has a real part equal to $-R_0$ and an imaginary part equal to zero. The reflection coefficient from Fig. 5-3-6 is infinite. In general, the reflection coefficient is given by

$$\Gamma = \frac{-R_n - R_0}{-R_n + R_0} \tag{5-3-11}$$

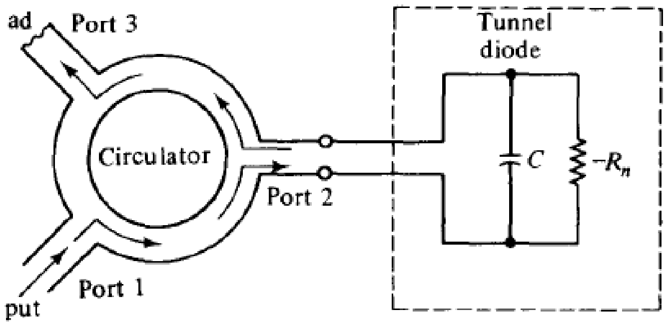


Figure 5-3-6 Tunnel diode connected to circulator.

Avalanche Transit-Time Devices

Avalanche transit-time diode oscillators rely on the effect of voltage breakdown across a reverse-biased p-n junction to produce a supply of holes and electrons. Ever since the development of modern semiconductor device theory scientists have speculated on whether it is possible to make a two-terminal negative-resistance device. The tunnel diode was the first such device to be realized in practice. Its operation depends on the properties of a forward-biased p-n junction in which both the p and n regions are heavily doped. The other two devices are the transferred electron devices and the avalanche transit-time devices. In this chapter the latter type is discussed.

The transferred electron devices or the Gunn oscillators operate simply by the application of a dc voltage to a bulk semiconductor. There are no p-n junctions in this device. Its frequency is a function of the load and of the natural frequency of the circuit. The avalanche diode oscillator uses carrier impact ionization and drift in the high-field region of a semiconductor junction to produce a negative resistance at microwave frequencies. The device was originally proposed in a theoretical paper by Read [1] in which he analyzed the negative-resistance properties of an idealized n^+-p-i-p^+ diode. Two distinct modes of avalanche oscillator have been observed. One is the IMPATT mode, which stands for *impact ionization avalanche transit-time* operation. In this mode the typical dc-to-RF conversion efficiency is 5 to 10%, and frequencies are as high as 100 GHz with silicon diodes. The other mode is the TRAPATT mode, which represents *trapped plasma avalanche triggered transit* operation. Its typical conversion efficiency is from 20 to 60%.

Another type of active microwave device is the BARITT (*barrier injected transit-time*) diode [2]. It has long drift regions similar to those of IMPATT diodes. The carriers traversing the drift regions of BARITT diodes, however, are generated by minority carrier injection from forward-biased junctions rather than being extracted from the plasma of an avalanche region. Several different structures have been operated as BARITT diodes, such as p-n-p, p-n-v-p, p-n-metal, and metal-n-metal. BARITT diodes have low noise figures of 15 dB, but their bandwidth is relatively narrow with low output power.

READ DIODE

The basic operating principle of IMPATT diodes can be most easily understood by reference to the first proposed avalanche diode, the Read diode [1]. The theory of this device was presented by Read in 1958, but the first experimental Read diode was reported by Lee et al. in 1965 [3]. A mode of the original Read diode with a doping profile and a dc electric field distribution that exists when a large reverse bias is applied across the diode is shown in Fig. 8-1-1.

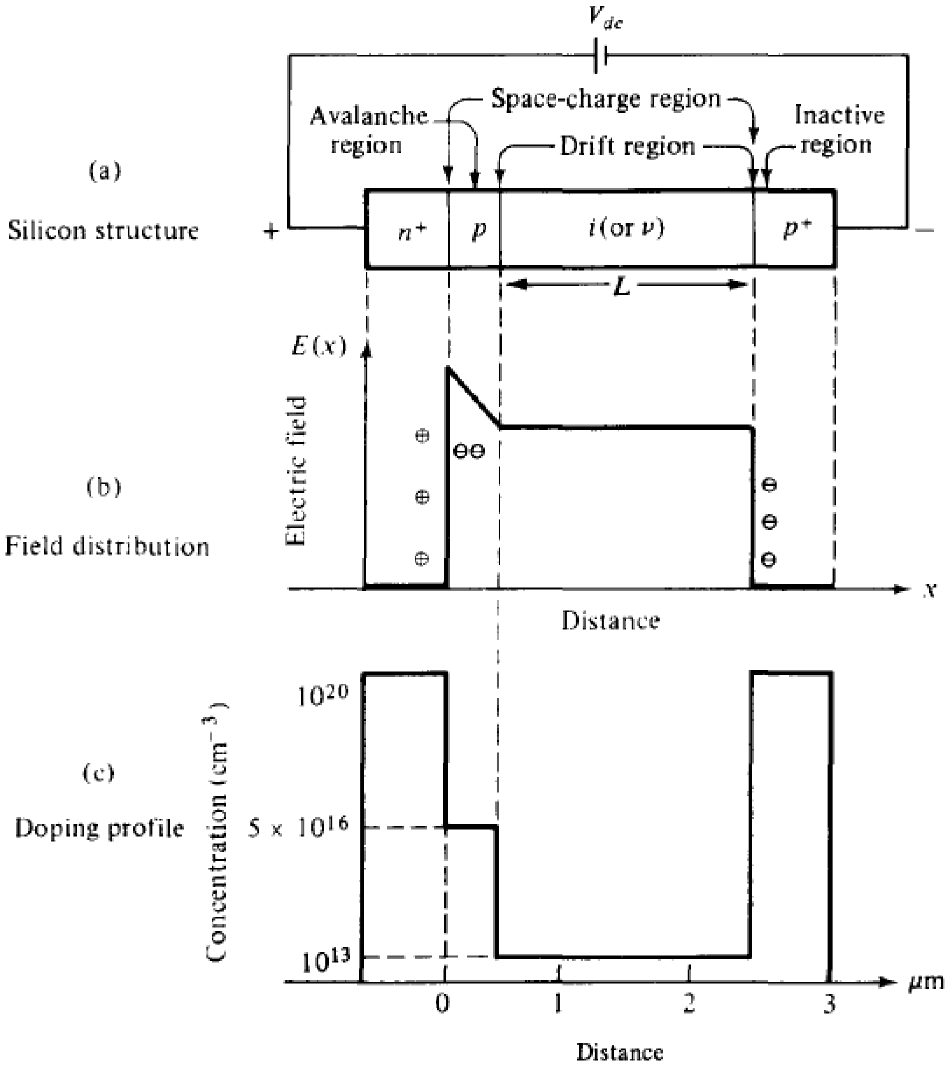


Figure 8-1-1 Read diode.

The Read diode is an n^+-p-i-p^+ structure, where the superscript plus sign denotes very high doping and the i or v refers to intrinsic material. The device consists essentially of two regions. One is the thin p region at which avalanche multiplication occurs. This region is also called the high-field region or the avalanche region. The other is the i or v region through which the generated holes must drift in moving to the p^+ contact. This region is also called the intrinsic region or the drift region. The p region is very thin. The space between the n^+-p junction and the i-p^+ junction is called the space-charge region. Similar devices can be built in the p^+-n-i-n^+ structure, in which electrons generated from avalanche multiplication drift through the i region.

The Read diode oscillator consists of an n^+-p-i-p^+ diode biased in reverse and mounted in a microwave cavity. The impedance of the cavity is mainly inductive and is matched to the mainly capacitive impedance of the diode to form a resonant circuit. The device can produce a negative ac resistance that, in turn, delivers power from the dc bias to the oscillation.

Avalanche Multiplication

When the reverse-biased voltage is well above the punchthrough or breakdown voltage, the space-charge region always extends from the n^+-p junction through the p and i regions to the i-p^+ junction. The fixed charges in the various regions are shown in Fig. 8-1-1(b). A positive charge gives a rising field in moving from left to right. The maximum field, which occurs at the n^+-p junction, is about several hundred kilovolts per centimeter. Carriers (holes) moving in the high field near the n^+-p junction acquire energy to knock valence electrons into the conduction band, thus producing hole-electron pairs. The rate of pair production, or avalanche multiplication, is a sensitive nonlinear function of the field. By proper doping, the field can be given a relatively sharp peak so that avalanche multiplication is confined to a very narrow region at the n^+-p junction. The electrons move into the n^+ region and the holes drift through the space-charge region to the p^+ region with a constant velocity v_d of about 10^7 cm/s for silicon. The field throughout the space-charge region is above about 5 kV/cm. The transit time of a hole across the drift i-region L is given by

$$\tau = \frac{L}{v_d} \tag{8-1-1}$$

and the avalanche multiplication factor is

$$M = \frac{1}{1 - (V/V_b)^n} \tag{8-1-1a}$$

where V = applied voltage
V_b = avalanche breakdown voltage
n = 3–6 for silicon is a numerical factor depending on the doping of p^+-n or n^+-p junction

Carrier Current $I_o(t)$ and External Current $I_e(t)$

As described previously, the Read diode is mounted in a microwave resonant circuit. An ac voltage can be maintained at a given frequency in the circuit, and the total field across the diode is the sum of the dc and ac fields. This total field causes breakdown at the n^+-p junction during the positive half of the ac voltage cycle if the field is above the breakdown voltage, and the carrier current (or the hole current in this case) $I_0(t)$ generated at the n^+-p junction by the avalanche multiplication grows exponentially with time while the field is above the critical value. During the negative half cycle, when the field is below the breakdown voltage, the carrier current $I_0(t)$ decays exponentially to a small steady-state value. The carrier current $I_0(t)$ is the current at the junction only and is in the form of a pulse of very short duration as shown in Fig. 8-1-3(d). Therefore the carrier current $I_0(t)$ reaches its maximum in the middle of the ac voltage cycle, or one-quarter of a cycle later than the voltage. Under the influence of the electric field the generated holes are injected into the space-charge region toward the negative terminal. As the injected holes traverse the drift space, they induce a current $I_e(t)$ in the external circuit as shown in Fig. 8-1-3(d).

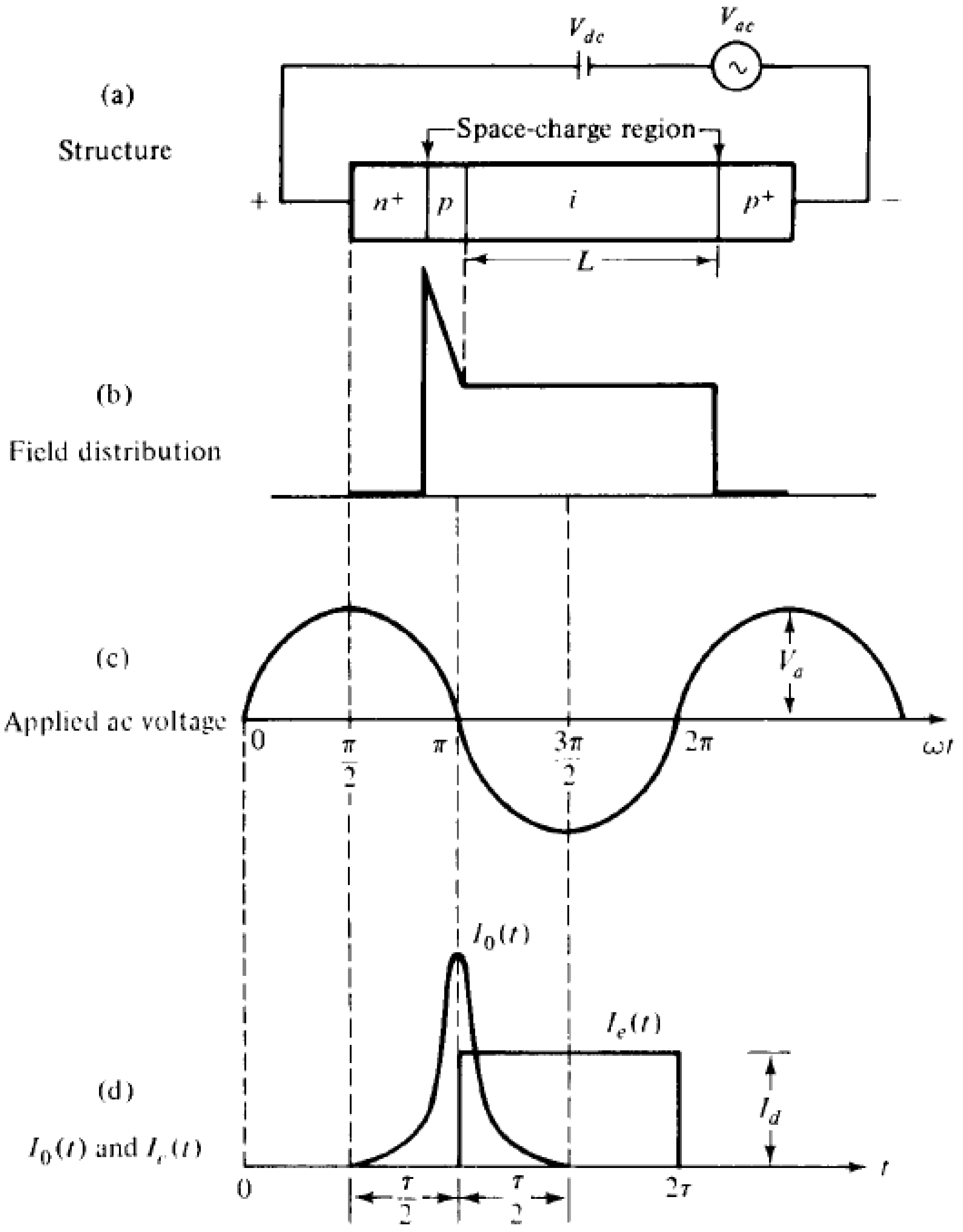


Figure 8-1-3 Field, voltage, and currents in Read diode. (*After Read [1]; reprinted by permission of the Bell System, AT&T Co.*)

$$f = \frac{1}{2\tau} = \frac{v_d}{2L}$$

Since the applied ac voltage and the external current $I_e(t)$ are out of phase by 180°, negative conductance occurs and the Read diode can be used for microwave oscillation and amplification. For example, taking $v_d = 10^7$ cm/s for silicon, the optimum operating frequency for a Read diode with an i-region length of 2.5 μm is 20 GHz.

IMPATT DIODES

A theoretical Read diode made of an n^+-p-i-p^+ or p^+-n-i-n^+ structure has been analyzed. Its basic physical mechanism is the interaction of the impact ionization avalanche and the transit time of charge carriers. Hence the Read-type diodes are called IMPATT diodes. These diodes exhibit a differential negative resistance by two effects:

1. The impact ionization avalanche effect, which causes the carrier current $I_0(t)$ and the ac voltage to be out of phase by 90°
2. The transit-time effect, which further delays the external current $I_e(t)$ relative to the ac voltage by 90°

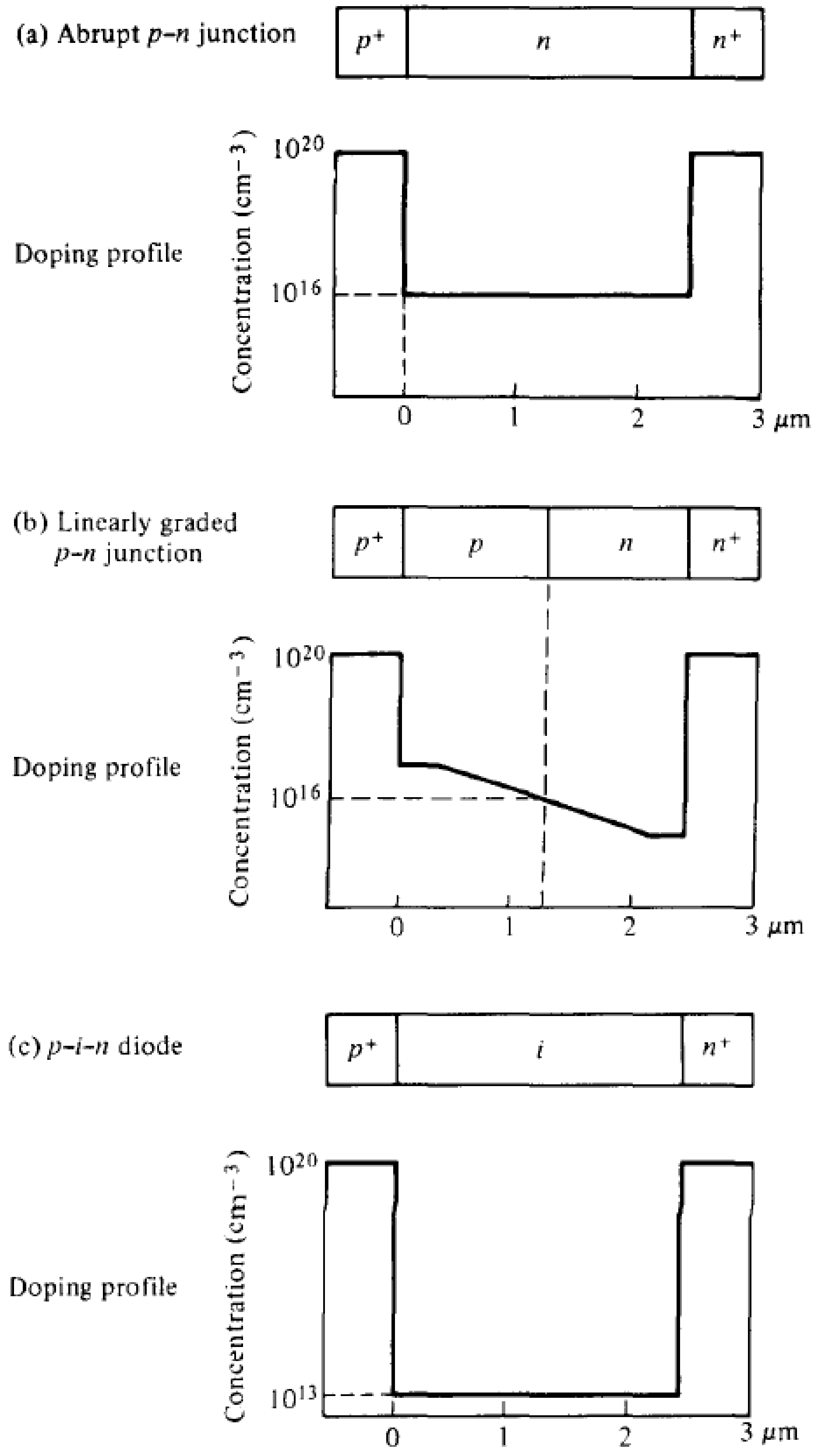
The first IMPATT operation as reported by Johnston et al. [4] in 1965, however, was obtained from a simple p-n junction. The first real Read-type IMPATT diode was reported by Lee et al. [3], as described previously. From the small-signal theory developed by Gilden [5] it has been confirmed that a negative resistance of the IMPATT diode can be obtained from a junction diode with any doping profile. Many IMPATT diodes consist of a high doping avalanching region followed by a drift region where the field is low enough that the carriers can traverse through it without avalanching. The Read diode is the basic type in the IMPATT diode family. The others are the one-sided abrupt p-n junction, the linearly graded p-n junction (or double-drift region), and the p-i-n diode, all of which are shown in Fig. 8-2-1. The principle of operation of these devices, however, is essentially similar to the mechanism described for the Read diode.

8-2-2 Negative Resistance

Small-signal analysis of a Read diode results in the following expression for the real part of the diode terminal impedance [5]:

$$R = R_s + \frac{2L^2}{v_d \epsilon_s A} \frac{1}{1 - \omega^2/\omega_r^2} \frac{1 - \cos\theta}{\theta} \tag{8-2-1}$$

where R_s = passive resistance of the inactive region
v_d = carrier drift velocity
L = length of the drift space-charge region
A = diode cross section
e_s = semiconductor dielectric permittivity


(a) Abrupt p–n junction
p+
n
n+
Doping profile
Concentration (cm−3)
10^20
10^16
0
1
2
3 μm
(b) Linearly graded p–n junction
p+
p
n
n+
Doping profile
Concentration (cm−3)
10^20
10^16
0
1
2
3 μm
(c) p–i–n diode
p+
i
n+
Doping profile
Concentration (cm−3)
10^20
10^13
0
1
2
3 μm

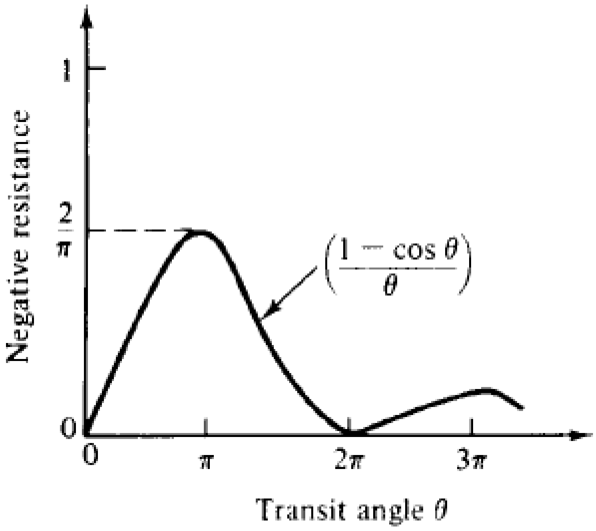
$$f = \frac{1}{2\tau} = \frac{v_d}{2L}$$

Moreover, θ is the transit angle, given by

$$\theta = \omega\tau = \omega \frac{L}{v_d}$$

and ω_r is the avalanche resonant frequency, defined by

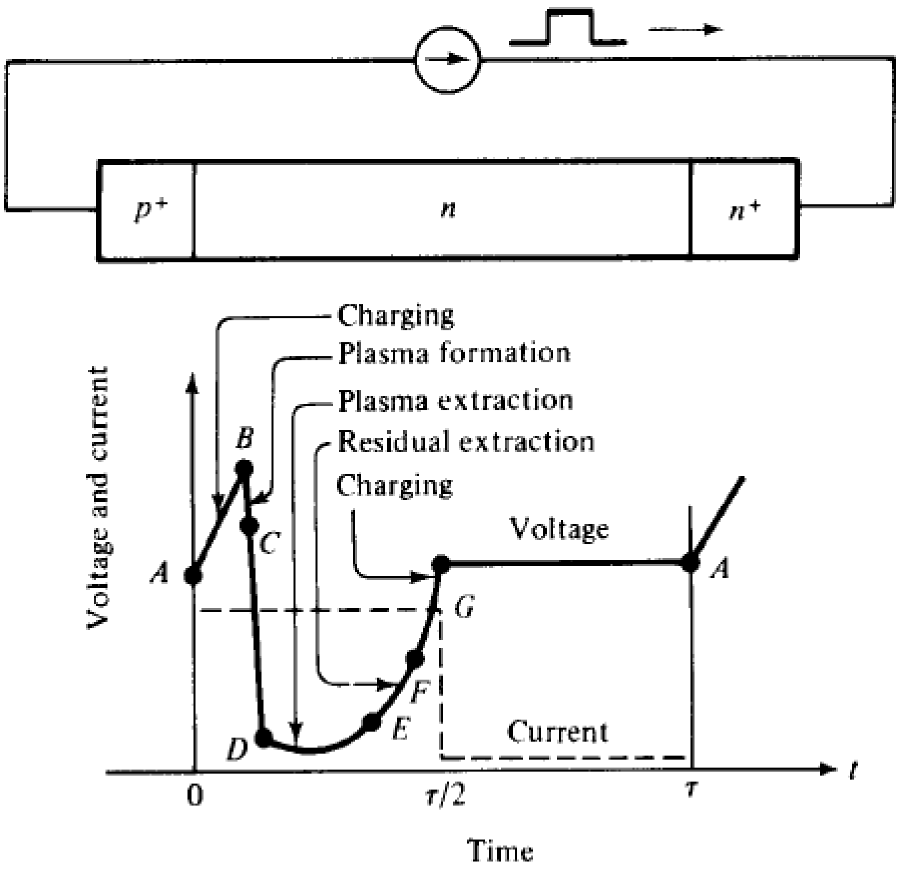
$$\omega_r \equiv \left(\frac{2\alpha' v_d I_0}{\epsilon_s A} \right)^{1/2}$$



TRAPATT DIODES

The abbreviation TRAPATT stands for *trapped plasma avalanche triggered transit* mode, a mode first reported by Prager et al. [7]. It is a high-efficiency microwave generator capable of operating from several hundred megahertz to several gigahertz. The basic operation of the oscillator is a semiconductor p-n junction diode reverse-biased to current densities well in excess of those encountered in normal avalanche operation. High-peak-power diodes are typically silicon n^+-p-p^+ (or p^+-n-n^+) structures with the n-type depletion region width varying from 2.5 to 12.5 μm. The doping of the depletion region is generally such that the diodes are well "punched through" at breakdown; that is, the dc electric field in the depletion region just prior to breakdown is well above the saturated drift-velocity level. The device's p^+ region is kept as thin as possible at 2.5 to 7.5 μm. The TRAPATT diode's diameter ranges from as small as 50 μm for CW operation to 750 μm at lower frequency for high-peak-power devices.

Principles of Operation



At the instant of time at point A, the diode current is turned on. Since the only charge carriers present are those caused by the thermal generation, the diode initially charges up like a linear capacitor, driving the magnitude of the electric field above the breakdown voltage. When a sufficient number of carriers is generated, the particle current exceeds the external current and the electric field is depressed throughout the depletion region, causing the voltage to decrease. This portion of the cycle is shown by the curve from point B to point C. During this time interval the electric field is sufficiently large for the avalanche to continue, and a dense plasma of electrons and holes is created. As some of the electrons and holes drift out of the ends of the depletion layer, the field is further depressed and "traps" the remaining plasma. The voltage decreases to point D. A long time is required to remove the plasma because the total plasma charge is large compared to the charge per unit time in the external current. At point E the plasma is removed, but a residual charge of electrons remains in one end of the depletion layer and a residual charge of holes in the other end. As the residual charge is removed, the voltage increases from point E to point F. At point F all the charge that was generated internally has been removed. This charge must be greater than or equal to that supplied by the external current; otherwise the voltage will exceed that at point A. From point F to point G the diode charges up again like a fixed capacitor. At point G the diode current goes to zero for half a period and the voltage remains constant at V_A until the current comes back on and the cycle repeats. The electric field can be expressed as

$$E(x, t) = E_m - \frac{qN_A}{\epsilon_s}x + \frac{Jt}{\epsilon_s} \qquad (8\text{-}3\text{-}2)$$

where N_A is the doping concentration of the n region and x is the distance.

Thus the value of t at which the electric field reaches E_m at a given distance x into the depletion region is obtained by setting $E(x, t) = E_m$, yielding

$$t = \frac{qN_A}{J}x \qquad (8\text{-}3\text{-}3)$$

Differentiation of Eq. (8-3-3) with respect to time t results in

$$v_z \equiv \frac{dx}{dt} = \frac{J}{qN_A} \qquad (8\text{-}3\text{-}4)$$

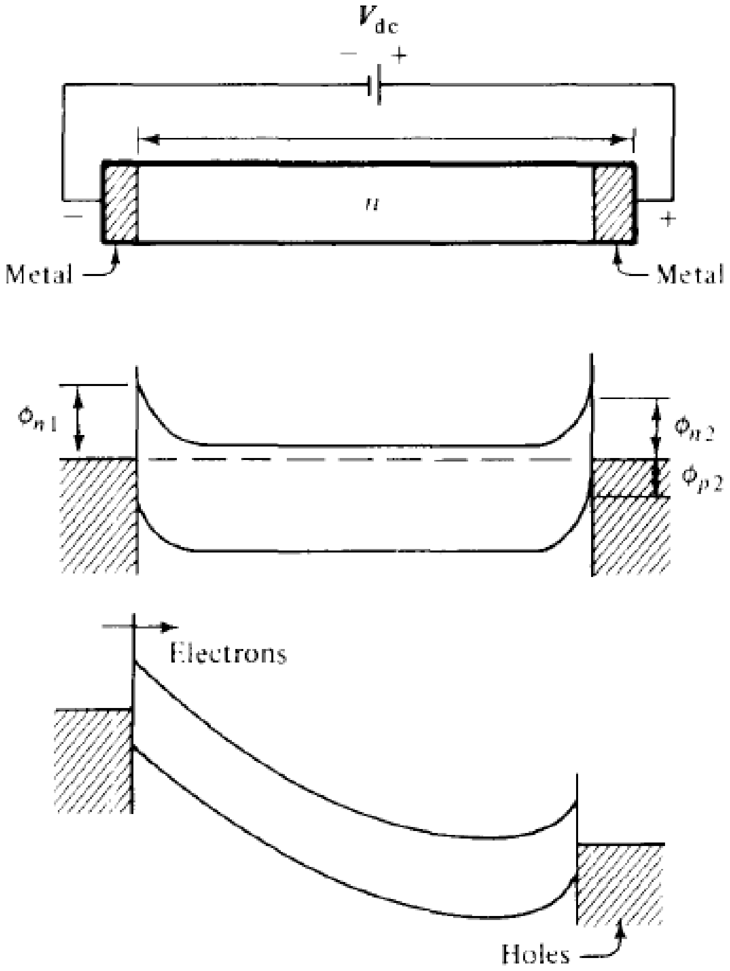
where v_z is the avalanche-zone velocity.

BARITT DIODES

BARITT diodes, meaning *barrier injected transit-time* diodes, are the latest addition to the family of active microwave diodes. They have long drift regions similar to those of IMPATT diodes. The carriers traversing the drift regions of BARITT diodes, however, are generated by minority carrier injection from forward-biased junctions instead of being extracted from the plasma of an avalanche region.

Several different structures have been operated as BARITT diodes, including *p-n-p*, *p-n-v-p*, *p-n*-metal, and metal-*n*-metal. For a *p-n-v-p* BARITT diode, the forward-biased *p-n* junction emits holes into the ν region. These holes drift with saturation velocity through the ν region and are collected at the *p* contact. The diode exhibits a negative resistance for transit angles between π and 2π. The optimum transit angle is approximately 1.6π.

Such diodes are much less noisy than IMPATT diodes. Noise figures are as low as 15 dB at C-band frequencies with silicon BARITT amplifiers. The major disadvantages of BARITT diodes are relatively narrow bandwidth and power outputs limited to a few milliwatts.



(a) M-*n*-M diode

(b) Energy band diagram in thermal equilibrium

(c) Energy band under bias condition

Figure 8-4-1 M-*n*-M diode. (*After D. J. Coleman and S. M. Sze [13]; reprinted by permission of the Bell System, AT&T Co.*)

Principles of Operation

A crystal n-type silicon wafer with 11 Ω-cm resistivity and 4×10^{14} per cubic centimeter doping is made of a 10-μm thin slice. Then the n-type silicon wafer is sandwiched between two PtSi Schottky barrier contacts of about 0.1 μm thickness. A schematic diagram of a metal-n-metal structure is shown in Fig. 8-4-1(a).

The energy-band diagram at thermal equilibrium is shown in Fig. 8-4-1(b), where ϕ_{n1} and ϕ_{n2} are the barrier heights for the metal-semiconductor contacts, respectively. For the PtSi–Si–PtSi structure mentioned previously, $\phi_{n1} = \phi_{n2} =$ 0.85 eV. The hole barrier height ϕ_{p2} for the forward-biased contact is about 0.15 eV. Figure 8-4-1(c) shows the energy-band diagram when a voltage is applied. The mechanisms responsible for the microwave oscillations are derived from:

1. The rapid increase of the carrier injection process caused by the decreasing potential barrier of the forward-biased metal-semiconductor contact
2. An apparent $3\pi/2$ transit angle of the injected carrier that traverses the semiconductor depletion region

The rapid increase in terminal current with applied voltage (above 30 V) as shown in Fig. 8-4-2 is caused by thermionic hole injection into the semiconductor as the depletion layer of the reverse-biased contact reaches through the entire device thickness. The critical voltage is approximately given by

$$V_c = \frac{qNL^2}{2\epsilon_s} \tag{8-4-1}$$

where N = doping concentration
L = semiconductor thickness
ϵ_s = semiconductor dielectric permittivity

The current-voltage characteristics of the silicon MSM structure (PtSi–Si–PtSi) were measured at 77° K and 300° K. The device parameters are $L = 10$ μm, $N = 4 \times 10^{14}$ cm^3, $\phi_{n1} = \phi_{n2} = 0.85$ eV, and area $= 5 \times 10^{-4}$ cm^2.

The current increase is not due to avalanche multiplication, as is apparent from the magnitude of the critical voltage and its negative temperature coefficient. At 77° K the rapid increase is stopped at a current of about 10^{-5} A. This saturated current is expected in accordance with the thermionic emission theory of hole injection from the forward-biased contact with a hole barrier height (ϕ_{p2}) of about 0.15 eV.