9 Solid state microwave devices

9.1 INTRODUCTION

The devices considered so far in this book are passive linear devices. In order to make useful systems we also require active devices which can convert d.c. into r.f. energy (oscillators and amplifiers) and non-linear devices for mixing, detection and switching. There are two main groups of devices which can perform these functions: those which depend upon the movement of electrons in semiconductor materials (considered in this chapter) and those in which the electrons move in a vacuum or a gas discharge (considered in Chapter 10).

This book is about electromagnetic waves and their applications so we shall concentrate upon the terminal properties of the different kinds of semiconductor device and the way in which they are put to use. For information about the inner workings of the devices the reader must refer to other books. (e.g. Bar-Lev, 1979).

There are two main classes of semiconductor device: diodes (with two terminals) and transistors (with three). Devices designed to work at high frequencies differ in some respects from their low-frequency counterparts. High-frequency diodes and transistors are dicussed in Sections 9.3 and 9.4. In the rest of the chapter we shall see how their properties are employed to make a variety of useful subsystems.

9.2 SEMICONDUCTOR MATERIALS

The commonest semiconductor material is silicon. This element has four valence electrons and it forms crystals in which each atom has four nearest neighbours to which it is bound by covalent bonds. These bonds each involve two electrons: one from each atom.

In an isolated atom the electrons which surround the nucleus cannot have arbitrary energies. Instead there is a set of permitted energy states

which they may occupy. According to the Pauli exclusion principle each state may only accommodate two electrons. When electrons move from one state to another they absorb or emit energy in the form of packets of electromagnetic radiation known as photons. The frequency of the photon emitted when an electron moves between two states with energies E_1 and E_2 is given by

$$f = (E_1 - E_2)/h, \tag{9.1}$$

where h is Planck's constant which has the value 6.625×10^{-34} J s. The frequencies associated with transitions in isolated atoms normally fall in the optical region of the electromagnetic spectrum. They account for the characteristic colours of, for example, sodium and neon discharge lamps.

When atoms are assembled together to form a crystalline solid the individual energy levels are perturbed and form corresponding energy bands each containing $2N$ closely spaced energy states, where N is the number of atoms in the crystal. The energy differences between these levels are very small so that thermal energy is sufficient to produce transitions between them. The typical thermal energy available is given by

$$E = kT, \tag{9.2}$$

where T is the absolute temperature and k is Boltzmann's constant which has the numerical value 1.380×10^{-23} J K^{-1}. Normally it is only the highest energy levels which play any part in determining the physical properties of the crystal.

In silicon there are two energy bands which are of interest. These are known as the conduction band and the valence band. They are shown diagrammatically in Fig. 9.1. At the absolute zero of temperature the valence band would be full of electrons and the conduction band empty. The two bands are separated by a forbidden region whose width is 1.11 electron volts. (One electron volt (eV) is the change in the energy of an electron when it moves through a potential difference of one volt). At room temperature a tiny proportion of the electrons in the valence band can aquire enough energy to excite them into states in the conduction

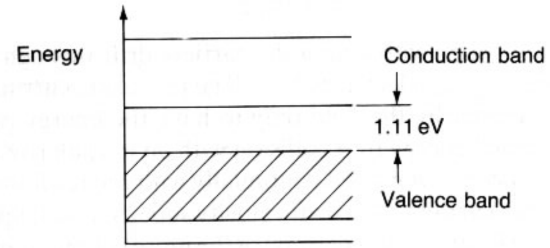

Fig. 9.1 Band diagram for silicon.

band. They are then free to change to other energy states near by, that is they are free to move through the crystal and form a conduction current. At the same time the states vacated at the top of the valence band allow some redistribution of the electrons there. This appears in experiments to be a motion of positively charged particles which are known as 'holes'. In pure silicon the result of thermal excitation is always to produce mobile charge carriers as electron–hole pairs. The thermal production of electron–hole pairs is balanced by recombination to produce a dynamic steady state. The number of charge carriers is only a tiny fraction of the total number of electrons in the valence band so the electrical conductivity of silicon is low. It is referred to as an intrinsic semiconducting material.

The properties of a semiconductor material can be altered by adding tiny quantities of other elements when the crystal is being grown or by subsequent diffusion or ion implantation. Usually these are chosen to have three or five valence electrons. Their presence produces additional holes or additional conduction electrons respectively so that there are no longer equal numbers of the two types of charge carrier. At room temperature virtually all the energy states associated with the impurity atoms are ionized. Material in which holes are the dominant charge carriers is known as p-type and that in which electrons dominate as n-type. The carrier densities and, hence, the electrical conductivity can be controlled by the impurity concentration. Material which has high conductivity is distinguished by the symbols p^+ and n^+.

Besides silicon there are other semiconductor materials. Germanium which was used to make the first bipolar transistors is still used for special purposes. More interesting at high frequencies are the intermetallic compounds, especially gallium arsenide (GaAs). Gallium has three valence electrons and arsenic has five so that together they can form crystals having the same structure as those of silicon. The intermetallic compounds add considerably to the range of semiconductor materials available. They all have different physical properties so it is possible to select the material which is best for a particular purpose.

The most important property as far as high-frequency devices are concerned is the mobility of the charge carriers. This is defined by the equation

$$\mu = v_d/E, \tag{9.3}$$

where v_d is the velocity with which the carriers drift through the material under the influence of an electric field E. The process occurring is that each electron is accelerated by the field only to have the energy gained turned into random thermal energy by a collision with an irregularity in the crystal structure. The process is repeated continually with the result that the whole assembly of charge carriers has a drift velocity superimposed upon a random thermal motion. The properties of some of the more common semiconductor materials are given in Table 9.1.

Table 9.1 Properties of semiconductor materials

	Si	Ge	GaAs	InSb
Energy gap (eV)	1.11	0.67	1.4	0.18
Gap type	Ind.	Ind.	Dir.	Dir.
Electron mobility ($cm^2 V^{-1} s^{-1}$)	1350	3900	8500	80 000
Hole mobility ($cm^2 V^{-1} s^{-1}$)	480	1900	450	150

Table 9.1 shows that GaAs has a much higher electron mobility than silicon. For this reason devices for high-speed and high-frequency operation are usually made from GaAs. Superficially InSb (indium antimonide) would be an even better choice. The reason why it and GaAs are not more generally used is that the technology needed to make devices is more difficult than that for silicon.

One other important property of semiconductor materials is shown in Table 9.1 as the gap type. This affects the way in which energy is released if an electron and a hole recombine. In direct band gap materials the energy is released as a photon whilst in indirect band gap materials it is released as thermal energy. That is why silicon and germanium cannot be used to make light-emitting diodes and semiconductor lasers.

9.3 DIODES

A semiconductor junction diode is formed by creating adjacent regions of p-type and n-type material as shown schematically in Fig. 9.2(a). These regions are characterized by having holes and electrons respectively as the dominant carrier types. It is found that there is a region at the junction where the carrier densities are much lower than in the bulk material on either side. This region is known as the depletion layer. If the diode is biased so that the p-type material is positive with respect to the n-type

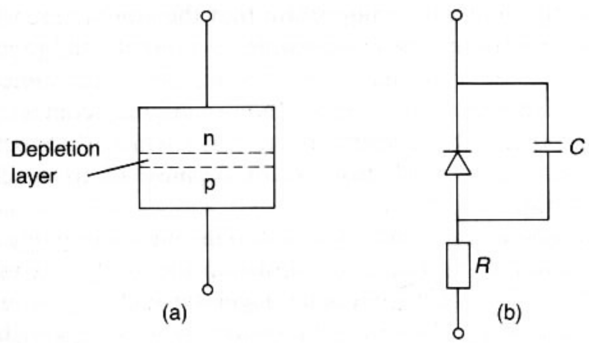

Fig. 9.2 p–n junction diode: (a) schematic diagram, and (b) equivalent circuit.

material majority charge carriers are injected across the junction and the diode conducts. If the opposite bias is applied the majority carrier cannot cross the junction and the diode carries only a very small current.

The I–V characteristic of a semiconductor diode is given by

$$I = I_s\left[\exp\left(\frac{V}{\eta V_T}\right) - 1\right], \tag{9.4}$$

where I_s is the reverse-bias saturation current, V_T ($= kT/e$) is the volt equivalent of temperature (25 mV at room temperature) and η is a constant whose value depends upon the material from which the diode is made.

The type of diode generally used at low frequencies is the silicon p–n junction shown in Fig. 9.2(a). This diode can be represented by the equivalent circuit of Fig. 9.2(b) in which the capacitance represents both the capacitance between the p and n regions and the storage of charge in the depletion region which separates them. The resistor represents the bulk resistance of the p and n regions. At high frequencies this kind of diode is unsatisfactory for two reasons. The first is the large junction capacitance which effectively short circuits the diode. The second is the slow speed of reaction caused by the low mobility of the charges and by storage of charge in the depletion layer. The capacitance of a reverse-biased junction varies with the bias voltage because the width of the depletion layer changes. This effect enables diodes to be used as voltage-variable capacitors. Diodes made for this purpose are known as varactor diodes, they are used to make electrically tunable circuits.

The problem of slow response can be overcome by using a metal–semiconductor junction in place of a p–n junction. Such a diode, known as a Schottky barrier diode, has a much faster switching time than a p–n junction, typically around 10 picoseconds. Figure 9.3 shows two examples of diodes of this type. In the point-contact diode (Fig. 9.3(a)) a metal whisker makes contact with the surface of a piece of semiconductor material. The usual combinations are p-type silicon with tungsten and n-type germanium with titanium. It important that the contact area between the two materials is kept as small as possible to minimize the junction capacitance. The other contact is made via metallization of the semiconductor. It is important that this is an 'ohmic' (non-rectifying) contact. The point-contact diode traces its ancestry to the 'cat's whisker' detectors used in early radio sets. It is now obsolescent because of its mechanical and electrical fragility.

Developments in semiconductor materials and technology have led to the form of Schottky barrier diode shown in Fig. 9.3(b). A metal contact makes contact with a semiconductor layer through a hole etched in an insulating oxide layer. The semiconductor layer is backed by a heavily doped, high-conductivity, layer of the same type of semiconductor and

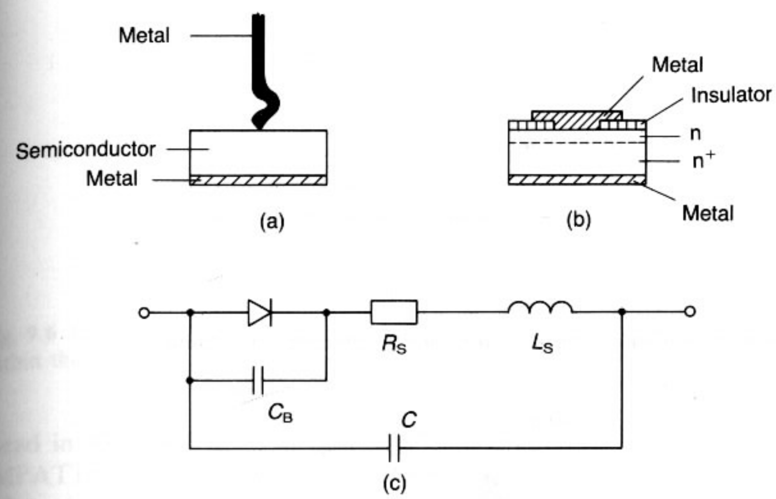

Fig. 9.3 Schottky barrier diodes: (a) point-contact diode, (b) modern form of Schottky barrier diode, and (c) equivalent circuit.

this, in turn, makes an ohmic contact with a metal base. The diameter of the contact between the metal and the semiconductor can be as small as 2 µm. The metal is commonly nickel, titanium or gold. The semiconductor is p-type or n-type silicon or n-type gallium arsenide. The latter has higher carrier mobilities than silicon and is therefore used to make diodes for millimetre wave frequencies (above 30 GHz).

The diodes shown in Figs. 9.3(a) and (b) can be represented by the equivalent circuit of Fig. 9.3(c) in which C_B is the capacitance of the junction, R_S the series resistance and L_S and C are the inductance and capacitance associated with the packaging of the diode. The I–V characteristic is given by the diode equation (9.4). For a silicon junction diode η is around 1.3 whilst for a silicon Schottky barrier diode it is typically 1.03 giving a faster turn on as the forward voltage rises. V in (9.4) is the voltage across the junction which is less than the terminal voltage by the voltage drop in R_S. At low currents the difference is negligible. The I–V characteristic curve (Fig. 9.4) has the same features as those of other semiconductor diodes with low forward resistance, high reverse resistance and breakdown at some reverse voltage.

The tunnel diode, first described by Esaki in 1958, is a p–n junction diode in which the two regions are very heavily doped and the depletion layer between them is very thin. Classically this layer is a barrier to the movement of charge across the junction but quantum theory shows that an appreciable number of electrons can 'tunnel' through it. Under conditions of zero bias the maximum electron energies on the two sides of the junction are equal and no net current flows. As forward bias is applied it is found

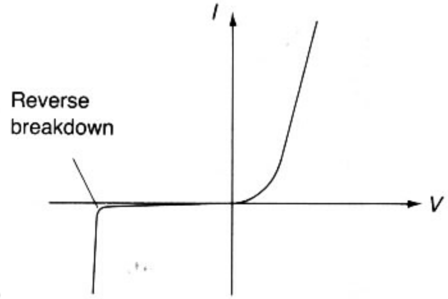

Fig. 9.4 Typical characteristic curve for a semiconductor diode.

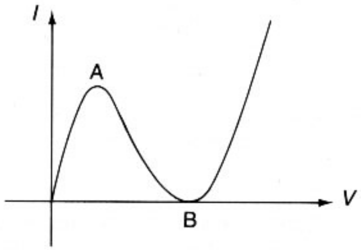

Fig. 9.5 Characteristic curve for a tunnel diode.

that the current increases to a maximum and then falls off again. For still higher forward bias the ordinary injection current of a p–n junction starts to flow and the diode conducts. Figure 9.5 shows the resulting I–V characteristic. The importance of this curve is that it has a region A–B in which the dynamic resistance is negative. This negative resistance can be used to generate oscillations as we shall see in Section 9.7.

The reverse breakdown condition of a semiconductor diode shown in Fig. 9.4 can be caused by tunnelling but is more commonly the result of avalanche breakdown. Ordinarily the charge carriers moving through a semiconductor material are accelerated by the applied electric field only to have their velocity randomized after a short while by collisions with lattice defects. The result is that a net drift velocity is superimposed upon the random thermal velocities. However, if the electric field is strong enough, the charge carriers may gain sufficient energy to cause ionization when they make collisions. The result is the creation of an additional pair of charge carriers (electron and hole) which can in their turn make ionizing collisions. This phenomenon is like an uncontrolled chain reaction producing a rapid increase in the number of charge carriers and in the current flowing through the device. Avalanche effects are used to produce negative resistance in IMPATT and TRAPATT diodes.

The IMPATT (IMPact Avalanche Transit Time) diode was proposed by

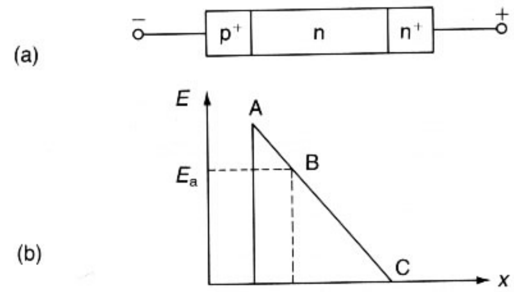

Fig. 9.6 IMPATT diode: (a) schematic diagram, and (b) electric field distribution within the device.

Read in 1958 and first demonstrated experimentally in 1965. A typical IMPATT idode has the structure shown in Fig. 9.6(a). The layers are, from left to right, a heavily doped p region, a long lightly doped n region and a heavily doped n region. When this diode is strongly reverse biased the region A–C is depleted of free charges and contains a strong electric field since most of the external potential difference appears across it. The resulting field distribution is shown in Fig. 9.6(b). If the external field is strong enough an avalanche builds up in A–B and the electrons produced move to B and drift across B–C. Now avalanching depends critically upon the applied field so, if a d.c. bias is applied to a level just below that at which an avalanche will develop then the superposition of a small a.c. signal causes avalanching for part of each a.c. cycle. Pulses of current are then injected into B–C. It is a property of semiconductor materials that, at high field strengths, the drift velocity tends to a constant value. The time taken for the current pulses to drift across B–C is then independent of the external field. If the length B–C is chosen correctly the current pulses emerge at the terminals of the device in antiphase with the applied a.c. signal. The device is then exhibiting a negative dynamic resistance and it may be used as the basis of an oscillator. The TRAPATT (TRApped Plasma Avalanche Triggered Transit) diode which has a similar structure to the IMPATT diode also shows negative dynamic resistance. Details of its mode of operation are given by Liao (1980).

The Gunn diode or transferred electron device depends for its operation upon the special properties of gallium arsenide. This material has two energy bands which can contain conduction electrons. One of these bands is at a lower energy level than the other and has the greater value of electron mobility. If a sample of GaAs has a voltage across it some of the electrons gain energy and are transferred from the lower energy band to the higher one. They then have lower mobilities and contribute less to the current flow through the material. This state is unstable and space-charge instability domains form which travel through the material producing an

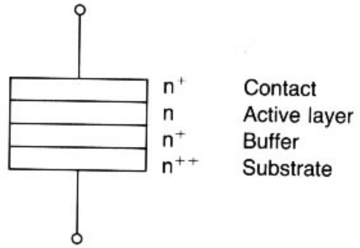

Fig. 9.7 Schematic diagram of a Gunn diode.

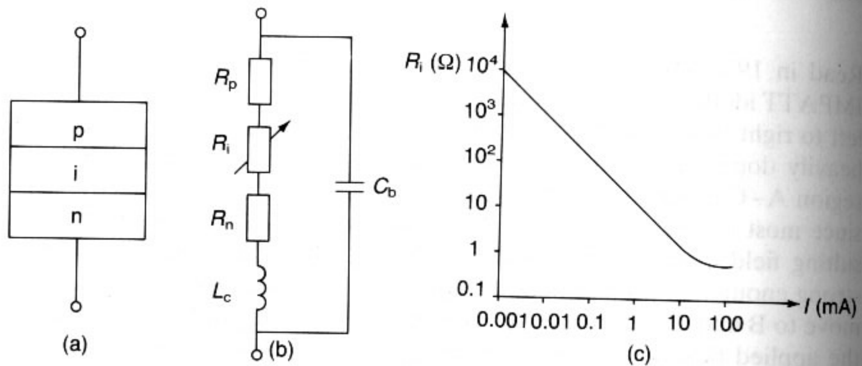

Fig. 9.8 PIN diode: (a) schematic diagram, (b) equivalent circuit, and (c) variation of R_i with current.

a.c. current which is superimposed upon the d.c. bias current. This effect was discovered experimentally by Gunn in 1963. It had been predicted by Ridley, Watkins and Hilsum a few years earlier but, at that time, no GaAs of sufficient quality was available for its experimental demonstration. Gunn diodes do not depend upon p–n junctions for their operation. A typical device structure is shown in Fig. 9.7.

The final example of a diode with microwave applications is the PIN diode shown in Fig. 9.8(a). In this device a layer of intrinsic semiconductor material is sandwiched between p-type and n-type layers. When the device is reverse biased it behaves as a capacitor and no current flows. When it is forward biased charge carriers are injected into the intrinsic region at both junctions so that the region conducts. It is found that the resistance of the intrinsic region can be varied from close to zero to almost infinity by varying the bias current. The PIN diode can therefore be used as a variable attenuator or a switch by connecting it across a transmission line or waveguide. The equivalent circuit of a forward-biased PIN diode is shown in Fig. 9.8(b). Here R_p, R_i and R_n are the resistances of the layers of the

diode, L_c is the inductance of the conductors and C_b is the parasitic capacitance of the encapsualtion. The variation of R_i with bias current is shown in Fig. 9.8(c).

9.4 TRANSISTORS

Bipolar and field-effect transistors are familiar circuit elements at low frequencies. High-frequency versions of both these devices exist but at frequencies above 4 GHz the field-effect transistor is much better than its bipolar rival.

Figure 9.9(a) shows the arrangement of a typical n–p–n silicon planar bipolar transistor. At high frequencies it can be represented by the small-signal hybrid-pi equivalent circuit shown in Fig. 9.9(b). The derivation of this model and the physical significance of the components are discussed by Bar-Lev (1979). It is shown in that book that the common-emitter short-circuit current gain obeys the equation

$$\frac{\beta}{\beta_0} = \frac{1}{1 + j(\omega/\omega_\beta)}, \tag{9.5}$$

where β_0 is the low-frequency value of β and the corner frequency is given by

$$\omega_\beta = \frac{1}{\beta_0(C_e + C_c + C_{\text{diff}})r_e}. \tag{9.6}$$

The bandwidth of the transistor can be defined by the frequency at which $|\beta| = 1$. If this frequency is large compared with ω_β then it is given with sufficient accuracy by

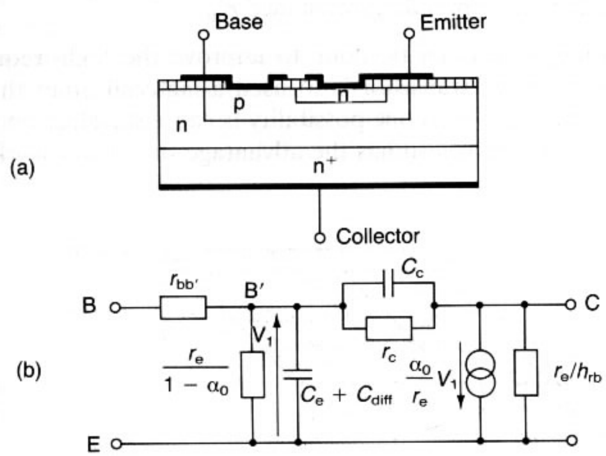

Fig. 9.9 Bipolar transistor: (a) arrangement of a silicon planar n–p–n transistor, and (b) the hybrid pi equivalent circuit.

$$\frac{1}{\omega_T} = \frac{1}{\beta_0\omega_\beta} = r_e(C_e + C_c + C_{diff}). \tag{9.7}$$

It can also be shown that the power gain of the transistor working into a matched load is

$$G = \frac{\omega_T}{4\omega^2 r_{bb'} C_c}. \tag{9.8}$$

To obtain the best high-frequency performance we must therefore minimize $r_{bb'}$ and C_c and maximize ω_T.

The base spreading resistance $r_{bb'}$ is the resistance of the thin base layer between the base terminal and the active part of the transistor. C_c represents the capacitance of the reverse-biased collector–base junction. Equation (9.7) does not include all the factors affecting the bandwidth of a transistor at high frequencies. Other factors include the transit times across the collector junction and the base and the effect of the series resistance of the collector on the charging time of C_c.

Thus, in order to make a high-frequency transistor, a number of steps can be taken.

1. The geometry of the transistor is made as small as possible to minimize the junction capacitances.
2. The thickness of the emitter is made as small as possible and the emitter–base junction is operated at the highest possible forward bias to minimize r_e.
3. The base is made as thin as possible and the doping graded to provide an internal electric field to minimize the base transit time.
4. The transistor is made in the interdigital form shown in Fig. 9.10 to minimize the base spreading resistance $r_{bb'}$.

The other thing which can be done to improve the high-frequency performance of a bipolar transistor is to used a material other than silicon. Gallium arsenide (GaAs) is one possibility because the electron mobility is higher than in silicon. It also has the advantage that it can work at higher

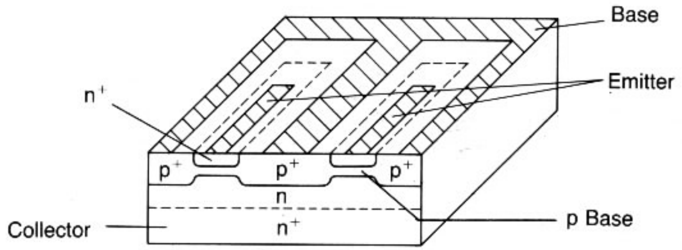

Fig. 9.10 Interdigital structure used for high-frequency transistors.

temperatures so allowing more power to be handled within the small sizes which must be used at high frequencies. Recent research has concentrated on the development of heterojunction transistors in which the junctions are between different semiconductor materials (e.g. gallium aluminium arsenide (GaAlAs) and GaAs). Bipolar transistors with cut-off frequencies of 40 GHz have been made in this way.

Gallium arsenide is used to make microwave field-effect transistors. These are akin to the low-frequency junction field-effect transistor (JFET) but employ a metal semiconductor Schottky junction in place of a p–n junction. This device which is known as a MESFET (MEtal Semiconductor FET) has the arrangement shown in Fig. 9.11. The active layer of high-conductivity n-type GaAs less than a micrometre in thickness is produced by ion implantation or epitaxial growth upon a semi-insulating substrate. A high-resistivity buffer layer separates the two. The source and drain contacts are ohmic (non-rectifying) whilst that of the gate forms a Schottky diode. The Schottky junction produces a region in the active layer which is depleted of conduction electrons and whose thickness is controlled by the reverse bias applied to the junction. Thus the thickness of the conducting channel, and therefore its resistance, is controlled by the source–gate voltage. The separation of the source and the drain is typically a few micrometres whilst the lengths of the electrodes are perhaps a hundred micrometres.

A simple equivalent circuit of the MESFET can be constructed by considering Fig. 9.11 with the result shown in Fig. 9.12(a). In this circuit R_I takes account of the potential difference between the source and the active region of the channel under the gate. R_{DS} is the channel resistance and the transconductance g_m represents the small-signal effect of the gate-channel voltage on the current flowing between source and drain. This equivalent circuit is not capable of representing the transistor correctly at high frequencies. For that purpose a much more complicated circuit is used which includes the parasitic capacitances and inductances associated with the packaging of the device with the result shown in Fig. 9.12(b). Circuits like this one are used to represent MESFETs in computer-aided-design programs for microwave circuits.

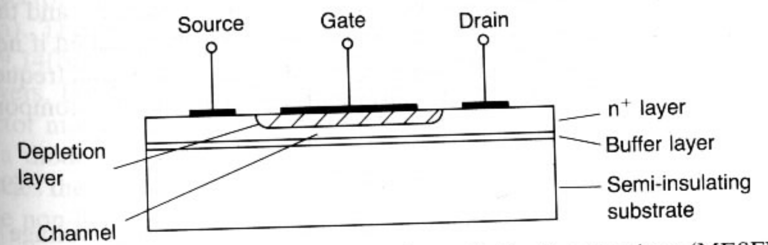

Fig. 9.11 Arrangement of a metal semiconductor field-effect transistor (MESFET).

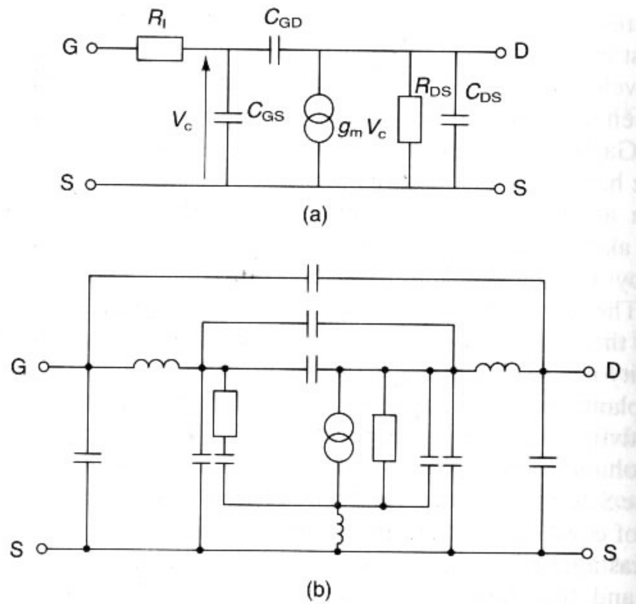

Fig. 9.12 Equivalent circuits for MESFETs: (a) basic equivalent circuit for the device, and (b) equivalent circuit which includes the parasitic effects associated with packaging.

9.5 DETECTORS AND MIXERS

The non-linear properties of diodes find application in detectors and mixers. The diode equation (9.4) can be expanded as a power series to give

$$I = I_s(\alpha V + \tfrac{1}{2}\alpha^2 V^2 + \cdots), \tag{9.9}$$

where $\alpha = 1/\eta V_T$. For small signals the higher-order terms can be neglected. If the voltage applied to the diode varies with time as $V_0 \cos \omega t$ the current is

$$I = I_s[\alpha V_0 \cos \omega t + \tfrac{1}{2}\alpha^2 V_0^2(\tfrac{1}{2} + \tfrac{1}{2} \cos 2\omega t)] \tag{9.10}$$

if terms beyond the second are neglected. The equivalent circuit of the diode (Fig. 9.3(c)) has shunt capacitance and series inductance and therefore acts as a low-pass filter. External components can be added if necessary to adjust the cut-off frequency so that it lies below the signal frequency ω. The current flowing in the external circuit is then just the d.c. component of (9.10), namely

$$I = \tfrac{1}{4} I_s \alpha V_0^2. \tag{9.11}$$

There is thus a square-law relationship between the applied voltage and the current flow in the diode, and measuring instruments such as VSWR

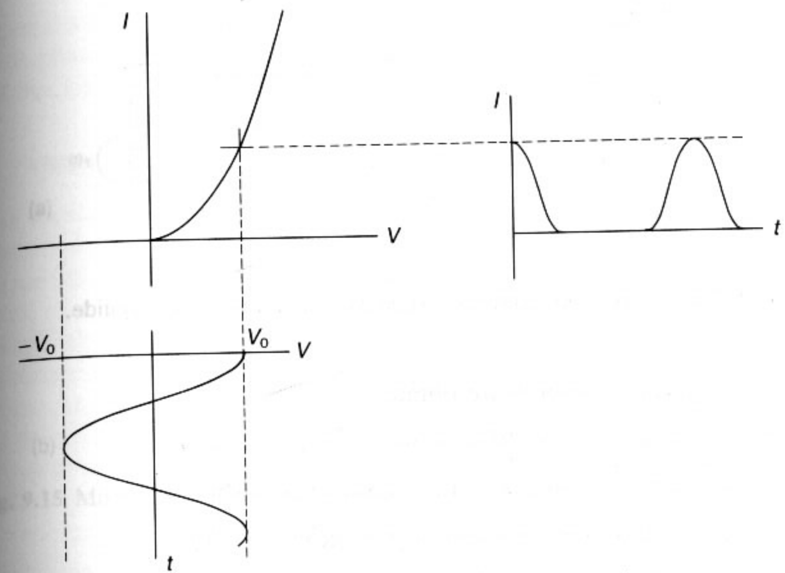

Fig. 9.13 Use of a diode as a detector showing how current flows only during the positive half cycles of the voltage waveform.

meters are calibrated on this basis. The rectifying operation of a diode can be represented graphically as shown in Fig. 9.13. The graph below the diode characteristic curve shows the variation of the signal voltage with time. When V is positive the diode conducts whereas in the negative half cycle virtually no current flows. The resulting current waveform is shown on the right of the diagram. The time average current flow is evidently non-zero.

In microwave measuring systems the signal is commonly square-wave modulated in order to improve the signal-to-noise ratio of the system. Provided that the modulating frequency is below the cut-off of the low-pass filter in the diode circuit the output current waveform is a square wave at the modulating frequency.

Detector diodes are often mounted in coaxial form as shown in Fig. 9.14(a). They can also be mounted directly in a waveguide as shown in Fig. 9.14(b) and this is a common configuration in detectors for radar receivers. For laboratory instrumentation it is more usual to use a coaxial detector mount attached to a waveguide–coaxial line transformer.

If a diode is subjected simultaneously to two signals at different frequencies the result is the production of signals at other frequencies because of the non-linearity of the diode. Suppose that the input signal is

$$V = V_1 \cos \omega_1 t + V_2 \cos \omega_2 t \tag{9.12}$$

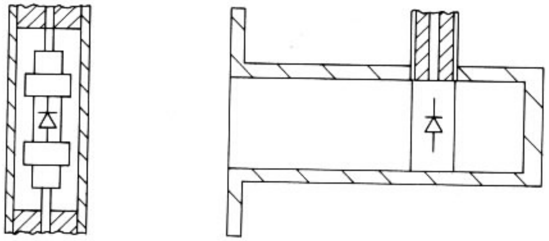

Fig. 9.14 Microwave diode mounts: (a) coaxial line, and (b) waveguide.

then, substituting into (9.9) we obtain

$$\begin{aligned} I &= I_s[\alpha V_1 \cos \omega_1 t + \alpha V_2 \cos \omega_2 t + \tfrac{1}{2}\alpha^2 V_1^2 \cos^2 \omega_1 t \\ &\quad + \tfrac{1}{2}\alpha^2 V_2^2 \cos^2 \omega_2 t + \alpha^2 V_1 V_2 \cos \omega_1 t \cos \omega_2 t] \\ &= I_s[\alpha V_1 \cos \omega_1 t + \alpha V_2 \cos \omega_2 t + \tfrac{1}{4}\alpha^2 (V_1^2 + V_2^2) \\ &\quad + \tfrac{1}{4}\alpha^2 V_1^2 \cos 2\omega_1 t + \tfrac{1}{4}\alpha^2 V_2^2 \cos 2\omega_2 t \\ &\quad + \tfrac{1}{2}\alpha^2 V_1 V_2 \cos (\omega_1 - \omega_2)t + \tfrac{1}{2}\alpha^2 V_1 V_2 \cos (\omega_1 + \omega_2)t] \end{aligned} \tag{9.13}$$

which contains signals at frequencies ω_1, ω_2, $2\omega_1$, $2\omega_2$, $(\omega_1 - \omega_2)$ and $(\omega_1 + \omega_2)$. If the higher-order terms are taken into account as well the resulting frequencies are

$$\omega = (m\omega_1 + n\omega_2), \tag{9.14}$$

where $m, n = \pm 0, 1, 2$, etc. These frequencies are known as intermodulation products. In an amplifier they are undesirable because they produce distortion of the amplified signal but they can also be used for frequency changing and modulation. The simplest form of mixer circuit is the single-ended mixer shown in Fig. 9.15(a). The signals to be mixed are fed into the main arm and the side arm of a directional coupler and the output applied to a diode. A filter is then used to select whichever of the intermodulation products is required. Typical applications include conversion of a received r.f. signal to an intermediate frequency (i.f.) for further amplification and generation of a frequency-modulated carrier.

An improved form of mixer is the balanced mixer shown in Fig. 9.15(b). This is commonly used for the conversion of r.f. to i.f. In that case ω_1 is the frequency of the incoming r.f. signal and ω_2 that of the local oscillator. The local oscillator will have some noise output at the intermediate frequency and this limits the sensitivity of the receiver. If a hybrid junction such as a magic tee is used as the 3 dB coupler then the contributions of the local oscillator noise to the outputs of the two diodes are in antiphase so that they cancel each other out.

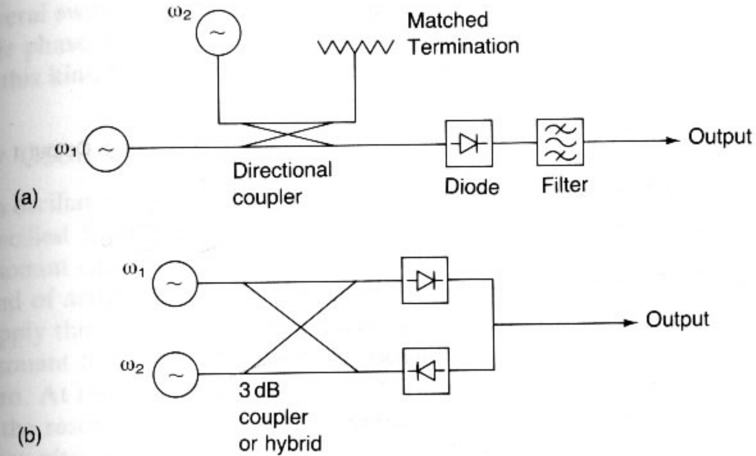

Fig. 9.15 Mixer circuits: (a) single-ended mixer, and (b) balanced mixer.

The circuits shown in Fig. 9.15 can be realized in microstrip, waveguide and other types of transmission line such as co-planar waveguide. For further information on mixers and their applications consult texts on telecommunications (Dunlop and Smith, 1984).

It is, of course, important that a diode used as a mixer or detector should be matched as far as possible to the transmission line in which it is mounted. The equivalent circuit of an encapsulated diode shown in Fig. 9.3(c) shows that a diode has both series and parallel resonances. At low frequencies the diode impedance is the forward resistance R_s. This is generally rather small. At the parallel resonant frequency the impedance is real and quite close to R_s. The series resonance on the other hand produces a high impedance which is generally closer to that of the transmission line.

9.6 SWITCHES

The PIN diode described in Section 9.3 can be used as a swtich. In its simplest form this device is a single diode which is arranged either in parallel or in series in a transmission line as shown in Fig. 9.16(a) and (b). In microstrip the series connection is easiest to realise whilst shunt connection must be used in waveguide. Because of the parasitic inductance and capacitance associated with the encapsulation of a diode the best results are obtained with unencapsulated diodes connected directly into a microstrip circuit. To obtain greater isolation it is possible to use several diodes. Figure 9.16(c) shows how three diodes could be arranged in waveguide so that their 'on' admittances add to give greater isolation. A switch of this kind can be used to provide square-wave modulation at high frequencies.

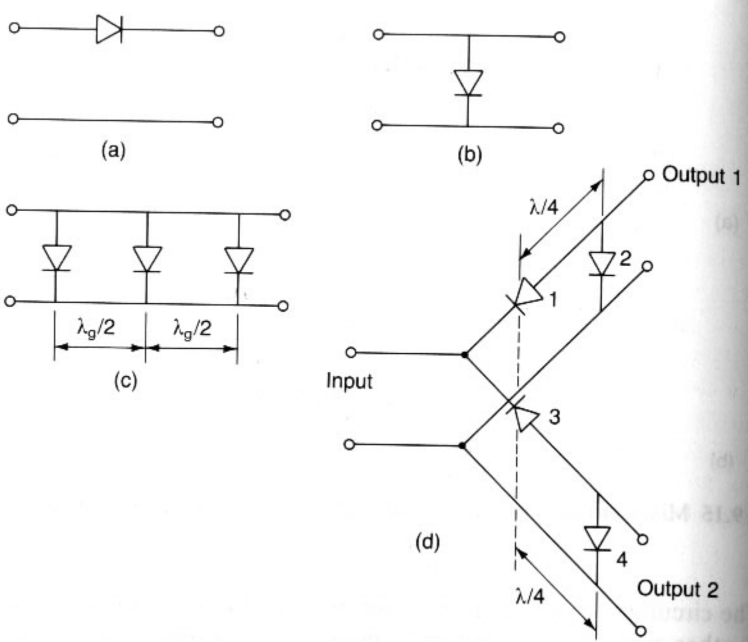

Fig. 9.16 PIN diode switching circuits: (a) series connection, (b) shunt connection, (c) use of several diodes to obtain greater isolation, and (d) arrangement of diodes to switch the input signal to either of two output ports.

Two or more switches can be combined to enable an input signal to be switched to one of two, or more, output lines. Figure 9.16(d) shows such an arrangement. The use of series and parallel diodes in pairs increases the isolation produced when a switch is open. When diodes 1 and 4 are 'on' and 2 and 3 are 'off' the output is directed to output port 1. Reversing the situation switches the output to port 2. The method of making the d.c. connections to the diodes is not shown. It requires a combination of capacitors and chokes to block d.c. and a.c. paths respectively.

Switches like that shown in Fig. 9.16(d) can be used to make switchable phase shifters. One method is to use pairs of switches to select alternative transmission lines with different lengths as shown in Fig. 9.17. By cascading

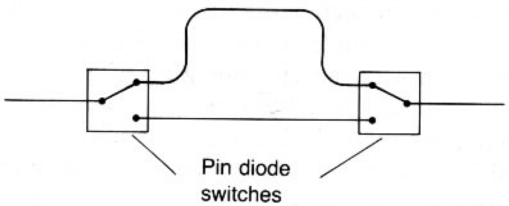

Fig. 9.17 Switchable phase shifter using PIN diode switches.

several switchable phase shifters it is possible to make an electrically variable phase shifter whose phase can be varied in small steps. Phase shifters of this kind find application in circuits for feeding phased array antennas.

9.7 OSCILLATORS

An oscillator is a subsystem which converts d.c. energy into r.f. energy at a specified frequency. The frequency is commonly selected by a parallel resonant circuit as shown in Fig. 9.18. In order to make an oscillator some kind of active device must be connected in parallel with the resonator to supply the r.f. output power. At all frequencies except those close to the resonant frequency the resonator has a high admittance and the output is zero. At resonance the resistance R represents the losses in the resonator. If the resonator were excited by a voltage pulse the oscillations would die away after a few cycles because of these losses. Even if the resonator were lossless so that the oscillations continued undamped it would not be possible to extract any power from it without causing damping. It follows that the condition for continuous oscillations to occur with useful output power is for the impedance presented at the terminals A–B to be a negative resistance at the desired frequency. There are a number of ways in which this negative resistance can be achieved.

In Section 9.3 we saw that tunnel diodes and IMPATT diodes exhibit negative resistance characteristics. Tunnel diodes are very fragile and are not in common use. IMPATT diodes, on the other hand, are valuable because they are able to produce powers of a few watts of continuous-wave (c.w.) power and several tens of watts of pulsed power at microwave frequencies. Gunn diodes can operate in a number of modes but the one normally used is the accumulation layer mode. In this mode the transit time of charge carriers through the device is important and it can be thought of as having negative dynamic resistance.

The performance of an oscillator of this kind depends critically upon the Q factor of the resonator. A low Q factor gives poor frequency stability and a higher noise figure. Oscillators commonly make use of waveguide, coaxial line or microstrip resonators. The unloaded Q factor of a metal waveguide resonator is typically 500 to 2000. Figure 9.19 shows the arrangement of a waveguide oscillator. The diode is placed approximately half a

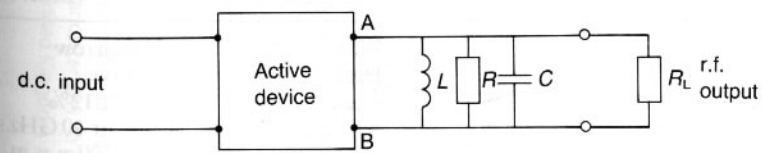

Fig. 9.18 General arrangement of an oscillator.

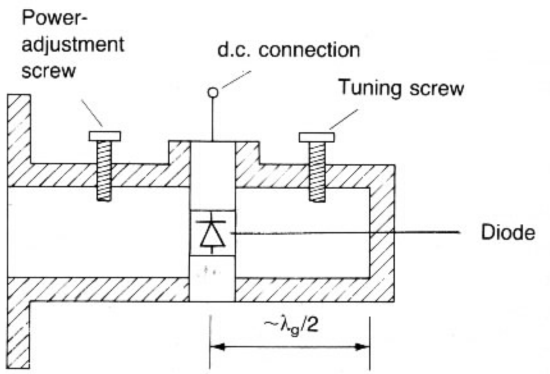

Fig. 9.19 Arrangement of a typical waveguide oscillator.

guide wavelength from a short circuit where the shunt impedance of the resonator is low and matched to that of the diode.

If a coaxial-line resonator is used the configuration is as shown in Fig. 9.20(a). At the resonant frequency the line is close to half a wavelength long, the difference being accounted for by the loading effect of the diode and its packaging. The unloaded Q factor is quite low (20 to 50) so the oscillator has poor stability and noise figure. This kind of arrangement can also include a varactor diode to allow the frequency to be changed electrically (Fig. 9.20(b)). Its main practical use is for testing diodes because of the ease with which matching can be achieved using coaxial line techniques.

Microstrip circuits tend to have high losses and therefore low Q factors. They are, however, simple, compact and cheap to manufacture. The frequency stability can be improved by using a separate high-Q resonator such as a YIG or dielectric resonator. Tuning can be achieved with YIG or with a varactor diode. YIG tuning provides bandwidths in excess of an octave and good temperature stability at the price of a relatively slow rate at which the frequency can be swept (the 'slew rate'). Varactor-tuned oscillators have narrower bandwidths and poorer temperature stability but higher slew rates. Table 9.2 shows a comparison between these two methods of tuning.

Table 9.2 Characteristics of voltage-tunable microwave oscillators

	YIG	Varactor
Bandwidth	Wide	Narrow
Q factor	High	Low
Linearity	<±1%	<±12%
Slew rate	<1 MHz s^{-1}	1 to 10 GHz s^{-1}
Temperature frequency sensitivity	<100 p.p.m. K^{-1}	<300 p.p.m. K^{-1}

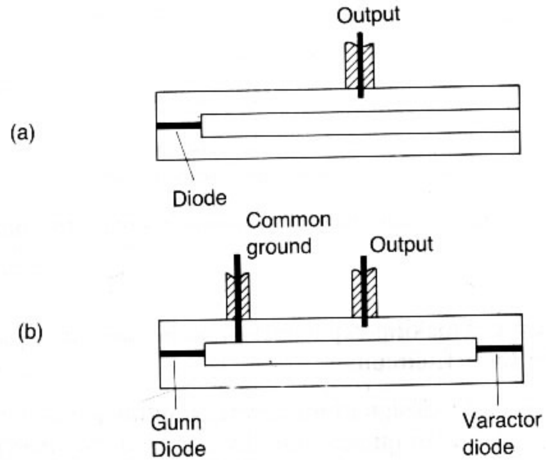

Fig. 9.20 Arrangements of coaxial-line oscillators: (a) fixed-frequency, and (b) varactor-diode tuned oscillators.

An alternative approach to oscillator design employs a three terminal active device, normally a MESFET. At frequencies up to 5 GHz good power output can be obtained using silicon bipolar junction transistors. At higher frequencies gallium arsenide FETs are used. Transistor oscillators tend to be noisier than those which use diodes but they have the advantage that they can be tuned over multioctave bands. This is useful in test equipment applications. The tuning ranges of diode oscillators are limited to around an octave by the transit time of electrons through the device.

9.8 AMPLIFIERS

The oscillators considered in the previous section are sources of microwave power whose frequency is controlled by the resonant circuit incorporated in them. An amplifier, in contrast, produces an output whose power is taken from the d.c. power supply but whose amplitude and frequency are controlled by the input signal.

Transistor amplifiers at microwave frequencies have configurations very like those familiar from low-frequency practice as shown in Fig. 9.21. There are, however, some differences.

1. The parasitic reactances of the devices are much more important than at low frequencies.
2. Networks must be designed to match the input and output to the external transmission lines.

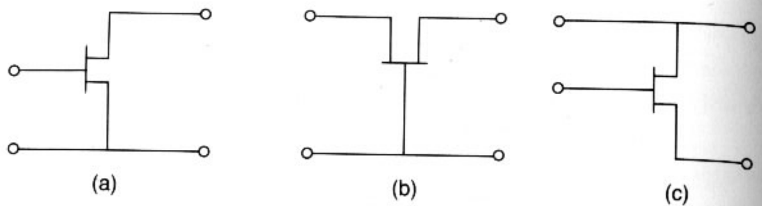

Fig. 9.21 FET amplifier configurations: (a) common source, (b) common gate, and (c) common drain.

3. Chokes and d.c. blocking capacitors must be used to separate the bias circuits from the r.f. circuits.

These factors make the design of microwave amplifiers much more difficult than the design of low-frequency amplifiers (Pengelly, 1986). Computer-aided design methods are normally employed for this purpose and the circuits realized using microstrip technology.

Microwave transistor amplifiers can readily be made with bandwidths of an octave. For a single stage the gain is low (a few decibels) and the power-added efficiency around 20 to 30%. As at low frequencies, it is possible to get increased bandwidth by using feedback to trade gain for bandwidth. Amplifiers have been made in this way with bandwidths of 3.5 octaves for use in electronic counter-measures (ECM) systems (Section 12.7). A second way of getting very wide bandwidth is to use a distributed amplifier circuit. Figure 9.22 shows the arrangement of such an amplifier with the bias connection omitted. Essentially it consists of a pair of transmission lines coupled together at intervals by transistors. The parasitic capacitances of the gates and drains modify the characteristics of the transmission lines but do not otherwise limit the performance of the devices. Multi-octave bandwidths can be achieved using this technique.

The power output available from a single transistor goes down as the frequency goes up. This is because transistors for high-frequency operation are smaller and so less able to dissipate heat. Powers of several watts at 10 GHz can be achieved in single devices. To obtain higher powers the outputs from several transistors can be combined using power dividers or

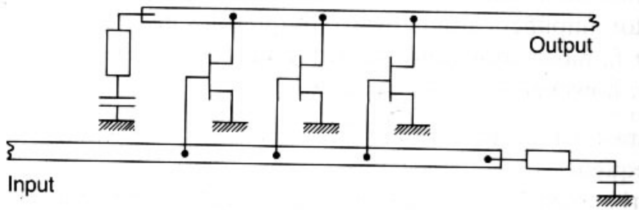

Fig. 9.22 Arrangement of a distributed amplifier.

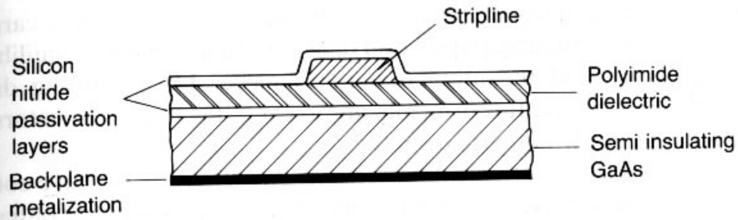

Fig. 9.23 Arrangement of stripline used in GaAs monolithic microwave integrated circuits (MMICs).

3 dB hybrid junctions operated in reverse. The insertion losses in these devices reduce the overall efficiency of the amplifier. A typical conversion efficiency for a microwave amplifier at 8 GHz is 15%.

9.9 MONOLITHIC MICROWAVE INTEGRATED CIRCUITS

Since 1960 the electronics industry has been dramatically altered by the development of silicon integrated circuits. The development of monolithic integrated circuits for use at microwave frequencies was much more difficult and circuits of this kind have only been available commercially since 1986. Monolithic microwave integrated circuits (MMICs) are constructed on gallium arsenide substrates using a miniaturized form of microstrip circuit. Figure 9.23 shows the construction of a stripline for a GaAs MMIC, the strip thickness is typically a few micrometres and its width a few tens of micrometres. The transistors may have gate lengths down to a few tenths of a micrometre and gate widths of a few hundred micrometres. One interesting effect of this miniaturization of circuits is that the dimensions are small compared with the free-space wavelength at microwave frequencies. It is therefore sometimes possible to use lumped-element circuit designs in MMIC technology.

A wide range of circuits is now available including amplifiers, oscillators, phase shifters, switches, mixers, and even complete receivers. Power outputs of up to a watt have been achieved at frequencies up to 18 GHz. It is to be expected that MMIC technology will have an effect on microwave engineering comparable with that of silicon technology at lower frequencies.

9.10 LEDS AND LASER DIODES

The devices discussed so far in this chapter have been for use at microwave frequencies. Semiconductor devices can also be used as sources at optical frequencies.

We have already noted that in direct band gap materials such as GaAs a photon is emitted when an electron and a hole recombine. When a p–n junction diode is forward biased majority carriers are injected across the junction into a region where they swell the population of minority carriers. The majority and minority population densities are then out of equilibrium with each other and recombination takes place. Light-emitting diodes (LEDs) which work on this principle are in everyday use as indicators and as elements in alpha-numeric displays.

For optical fibre communication systems it is desirable to have sources which are more intense than LEDs and which emit coherent radiation. Such a device is the laser. Here we shall discuss the laser (light amplification by stimulated emission of radiation) diode as an illustration of the principles upon which all lasers work.

When light interacts with matter one or more of three processes occur.

1. Absorption: a photon is absorbed and an electron excited to a higher energy state.
2. Spontaneous emission: an electron moves to a lower energy state and a photon is emitted.
3. Stimulated emission: a photon whose energy matches that of a transition stimulates an electron to undergo that transition emitting another photon of the same frequency and phase as the first.

Normally, higher energy states are less densely populated than lower energy states and a photon is much more likely to be absorbed than to cause stimulated emission. The population distribution is determined by considerations of thermal equilibrium. It also depends upon the average time an electron will spend in a higher state before spontaneous emission occurs. In order for stimulated emission of radiation to become the dominant process two conditions must be satisfied:

1. there must be a high-energy state with a relatively long lifetime to which electrons can be excited, and
2. some way must be found of exciting enough electrons into this state to make stimulated emission more probable than absorption.

These conditions are satisfied for all lasers. They differ only in the mechanisms by which the second condition (population inversion) is brought about.

The final thing needed to make a laser work is some form of optical resonator to ensure a high density of photons at the frequency required for laser action within the device.

Figure 9.24 shows the arrangement of a semiconductor injection laser. The junction is arranged to be in a certain orientation with respect to the crystal structure of the semiconductor. This makes it possible for two opposing faces A–A to be cleaved accurately flat and parallel forming a Fabry–Pérot resonator. The p and n regions of the diode are heavily doped so that when a large forward current is passed through the diode large numbers of electrons are injected into the p region and large numbers

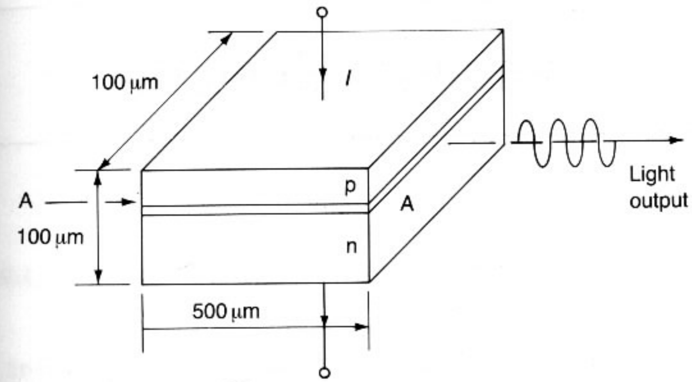

Fig. 9.24 Arrangement of a semiconductor injection laser.

of holes are injected into the n region. The diode is made from a direct band gap material so that the energy of recombination appears as photons. These travel through the diode stimulating more emission at the same frequency and phase. The light is partially reflected at the faces A–A ensuring a high intensity of electromagnetic radiation within the diode. The remainder of the light is emitted as an intense coherent beam which may be used to carry a signal along an optical fibre. Laser diodes are made in a variety of intermetallic compounds. The output power is limited by the maximum current which can be passed through the diode without destroying it. Continuous output powers of up to 70 W have been obtained with driving currents up to 250 A.

9.11 CONCLUSION

In this chapter we have considered the types of solid state device which are used to generate or amplify signals at microwave and optical frequencies. The bipolar transistor which is the commonest device at low frequencies is only usable up to about 4 GHz. Above that frequency gallium arsenide MESFETs are used in amplifiers and oscillators. Several types of diode exist having negative resistance characteristics which can be used to sustain the oscillations in a resonant circuit. It is easiest to make high Q circuits using hollow metal waveguides and that technology is commonly used for oscillators when high stability is required. Microstrip circuits have much lower Q factors but good stability can be achieved by using them in conjunction with high Q YIG or dielectric resonators. Electronically tuned oscillators can be made using YIG or varactor diodes.

Infrared and optical signals can be generated using light-emitting diodes and lasers. The semiconductor injection laser was used to explain the principles of laser action.