Q.2 a. What is a Carrier Wave? Draw and explain the block diagram of typical radio transmitter. (7)

Answer:

Transmitter

Unless the message arriving from the information source is electrical in nature, it will be unsuitable for immediate transmission. Even then, a lot of work must be done to make such a message suitable. This may be demonstrated in *single-sideband modulation* (see Chapter 4), where it is necessary to convert the incoming sound signals into electrical variations, to restrict the range of the audio frequencies and then to *compress* their amplitude range. All this is done before any *modulation*. In wire telephony no processing may be required, but in long-distance communications, a transmitter is required to process, and possibly encode, the incoming information so as to make it suitable for transmission and subsequent reception.

Eventually, in a transmitter, the information modulates the *carrier*, i.e., is superimposed on a high-frequency sine wave. The actual method of modulation varies from one system to another. Modulation may be *high level* or *low level*, and the system itself may be *amplitude modulation, frequency modulation, pulse modulation* or any variation or combination of these, depending on the requirements. Figure 1-2 shows a high-level amplitude-modulated broadcast transmitter of a type that will be discussed in detail in Chapter 6.

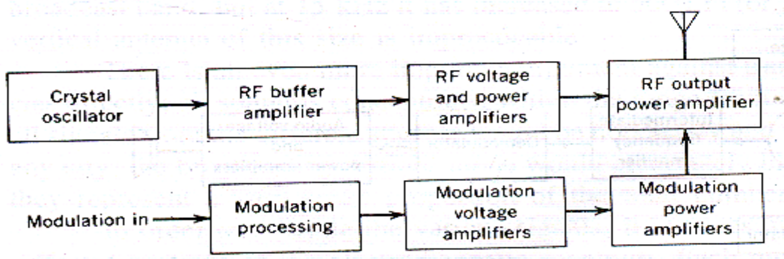


FIGURE 1-2 Block diagram of typical radio transmitter.

b. What is shot noise and explain its calculations? (5)

Answer:

Shot Noise

Thermal agitation is by no means the only source of noise in receivers. The most important of all the other sources is the *shot effect*, which leads to shot noise in all amplifying devices and virtually all active devices. *It is caused by random variations in the arrival of electrons* (or holes) *at the output electrode* of an amplifying device and appears as a randomly varying noise current superimposed on the output. When amplified, it is supposed to sound as though a shower of lead shot were falling on a metal sheet. Hence the name *shot* noise.

Although the average output current of a device is governed by the various bias voltages, at any instant of time there may be more or fewer electrons arriving at the output electrode. In bipolar transistors, this is mainly a result of the random drift of the discrete current carriers across the junctions. The paths taken are random and therefore unequal, so that although the average collector current is constant, minute variations nevertheless occur. Shot noise behaves in a similar manner to thermal agitation noise, apart from the fact that it has a different source.

Many variables are involved in the generation of this noise in the various amplifying devices, and so it is customary to use approximate equations for it. In addition, shot-noise *current* is a little difficult to add to thermal-noise *voltage* in calculations, so that for all devices with the exception of the diode, shot-noise formulas used are generally simplified. For a diode, the formula is exactly

$$i_n = \sqrt{2ei_p\,\delta f} \qquad \textbf{(2-3)}$$

where i_n = rms shot-noise current

e = charge of an electron = 1.6×10^{-19}C

i_p = direct diode current

δf = bandwidth of system

Note: It may be shown that, for a vacuum tube diode, Equation (2-3) applies only under so-called temperature-limited conditions, under which the "virtual cathode" has not been formed.

In all other instances not only is the formula simplified but it is not even a formula for shot-noise current. The most convenient method of dealing with shot noise

c. An amplifier operating over the frequency range from 18 to 20 MHZ has a 10 KΩ input resistor.
What is the rms noise voltage at the input to this amplifier if the ambient temperature is $27^\circ C$? (4)

Answer:

The rms noise voltage at the input to the amplifier is given by:

$V_n = \sqrt{4kT\delta fR}$

$= \sqrt{4X1.38X10^{-23}X(27+273)X(20-18)X10^6X10^4}$

$= \sqrt{4X1.38X3X2X10^{-11}}$ = 1.82 X 10^{-5}

= 18.2 μV

Q.3 a. Define Amplitude Modulation. Derive an expression for its Amplitude Modulated Wave by assuming the modulating signal voltage $e_m = E_m \sin\omega_m t$ and the carrier signal voltage as $e_c = E_c \sin\omega_c t$. (8)

Answer:

In amplitude modulation, the amplitude of a *carrier* signal is varied by the *modulating voltage*, whose frequency is invariably lower than that of the carrier. In practice, the carrier may be high-frequency (HF) while the modulation is audio. Formally, *AM* is defined as a system of modulation in which the *amplitude of the carrier is made proportional to the instantaneous amplitude of the modulating voltage.*

Let the carrier voltage and the modulating voltage, v_c and v_m, respectively, be represented by

$$v_c = V_c \sin \omega_c t \tag{3-1}$$

$$v_m = V_m \sin \omega_m t \tag{3-2}$$

Note that phase angle has been ignored in both expressions since it is unchanged by the *amplitude* modulation process. Its inclusion here would merely complicate the proceedings, without affecting the result. However, it will certainly not be possible to ignore phase angle when we deal with frequency and phase modulation in Chapter 4.

From the definition of AM, you can see that the (maximum) amplitude V_c of the unmodulated carrier will have to be made proportional to the *instantaneous* modulating voltage $V_m \sin \omega_m t$ when the carrier is amplitude-modulated.

3-1.1 Frequency Spectrum of the AM Wave

We shall show mathematically that the frequencies present in the AM wave are the carrier frequency and the first pair of sideband frequencies, where a sideband frequency is defined as

$$f_{SB} = f_c \pm nf_m \tag{3-3}$$

and in the first pair $n = 1$.

When a carrier is amplitude-modulated, the proportionality constant is made equal to unity, and the instantaneous modulating voltage variations are *superimposed* onto the carrier amplitude. Thus when there is temporarily no modulation, the amplitude of the carrier is equal to its unmodulated value. When modulation is present, the amplitude of the carrier is varied by its instantaneous value. The situation is illustrated in Figure 3-1, which shows how the maximum amplitude of the amplitude-modulated voltage is made to vary in accordance with modulating voltage changes. Figure 3-1 also shows that something unusual (distortion) will occur if V_m is greater than V_c (this distortion is a result of overdriving the amplifier stage). This, and the fact that the ratio V_m/V_c often occurs, leads to the definition of the *modulation index* given in Equation (3-4).

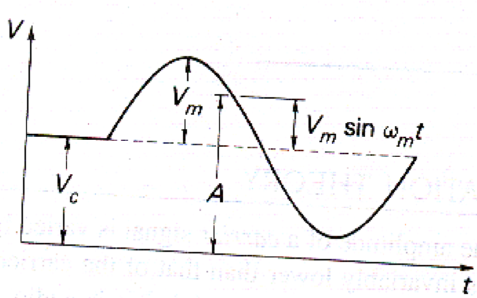


FIGURE 3-1 Amplitude of AM wave.

$$m = \frac{V_m}{V_c} \tag{3-4}$$

The modulation index is a number lying between 0 and 1, and it is very often expressed as a percentage and called the *percentage modulation*.

From Figure 3-1 and Equation (3-4) it is possible to write an equation for the amplitude of the amplitude-modulated voltage. We have

$$\begin{aligned} A &= V_c + v_m = V_c + V_m \sin \omega_m t = V_c = mV_c \sin \omega_m t \\ &= V_c\,(1 + m \sin \omega_m t) \end{aligned} \tag{3-5}$$

The instantaneous voltage of the resulting amplitude-modulated wave is

$$v = A \sin \theta = A \sin \omega_c t = V_c\,(1 + m \sin \omega_m t) \sin \omega_c t \tag{3-6}$$

Equation (3-6) may be expanded, by means of the trigonometrical relation $\sin x \sin y = ½\,[\cos (x - y) - \cos (x + y)]$, to give

$$v = V_c \sin \omega_c t + \frac{mV_c}{2} \cos (\omega_c - \omega_m)t - \frac{mV_c}{2} \cos (\omega_c + \omega_m)t \tag{3-7}$$

It has thus been shown that the equation of an amplitude-modulated wave contains three terms. The first term is identical to Equation (3-1) and represents the unmodulated carrier. It is apparent that the process of amplitude modulation has the effect of adding to the unmodulated wave, rather than changing it. The two additional terms produced are the two sidebands outlined. The frequency of the lower sideband (LSB) is $f_c - f_m$, and the frequency of the upper sideband (USB) is $f_c + f_m$. The very important conclusion to be made at this stage is that the bandwidth required for amplitude modulation is twice the frequency of the modulating signal. In modulation by several sine waves simultaneously, as in the AM broadcasting service, *the bandwidth required is twice the highest modulating frequency*.

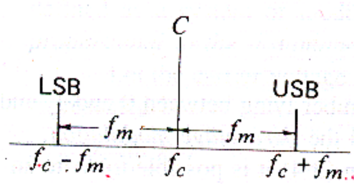


FIGURE 3-2 Frequency spectrum of AM wave.

b. **With the help of a block diagram, explain the working of Third Method used for generating SSB signal. (8)**

Answer:

The third method of generating SSB was developed by Weaver as a means of retaining the advantages of the phase-shift method, such as its ability to generate SSB at any frequency and use low audio frequencies, without the associated disadvantage of an AF phase-shift network required to operate over a large range of audio frequencies. The third method is in direct competition with the filter method, but is very complex and not often used commercially.

From the block diagram of Figure 4-6, we see that the latter part of this circuit is identical to that of the phase-shift method, but the way in which appropriate voltages are fed to the last two balanced modulators at points C and F has been changed. Instead of trying to phase-shift the whole range of audio frequencies, this method combines them with an AF carrier f_0, which is a fixed frequency in the middle of the audio band, 1650 Hz. A phase shift is then applied to this frequency only, and after the resulting voltages have been applied to the first pair of balanced modulators, the low-pass filters whose cutoff frequency is f_0 ensure that the input to the last pair of balanced modulators results in the proper eventual sideband suppression.

The proof of this method is unduly complex, and therefore not given here. It may be shown that all lower sideband signals will be canceled for the configuration of

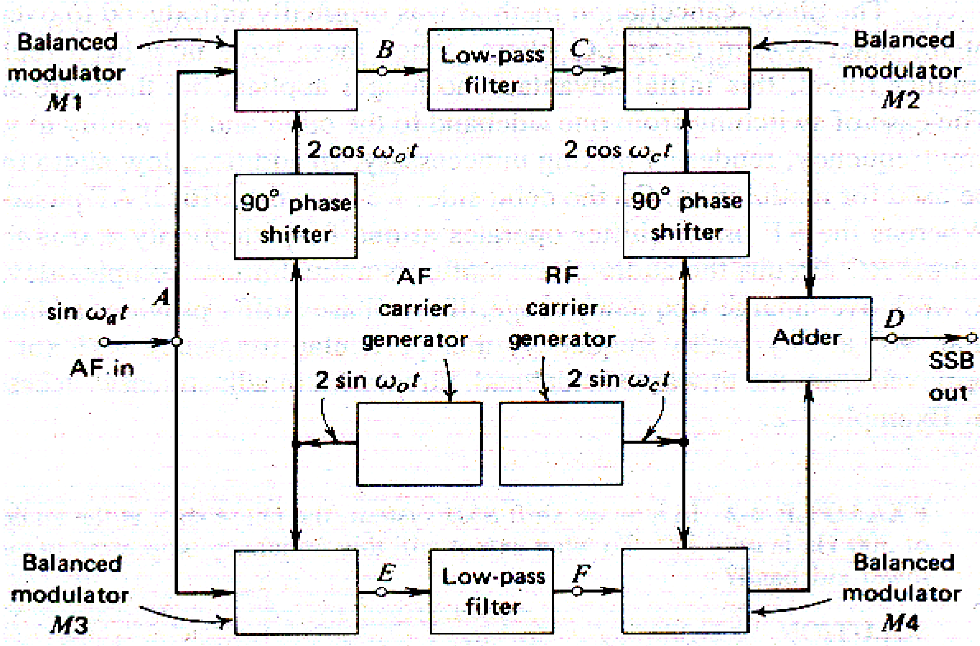


FIGURE 4-6 Third method of SSB generation.

Figure 4-6, regardless of whether audio frequencies are above or below f_0. If a lower sideband signal is required, the phase of the carrier voltage applied to M_1 may be changed by 180°.

Q.4 a. Draw the circuit diagram of basic Reactance Modulator and explain its working by deriving an expression for its Capacitive Reactance. (8)

Answer:

Basic reactance modulator Provided that certain simple conditions are met, the impedance z, as seen at the input terminals A–A of Figure 5-11, is almost entirely reactive. The circuit shown is the basic circuit of a FET reactance modulator, which behaves as a three-terminal reactance that may be connected across the tank circuit of the oscillator to be frequency-modulated. It can be made inductive or capacitive by a simple component change. The value of this reactance is proportional to the transconductance of the device, which can be made to depend on the gate bias and its variations. Note that an FET is used in the explanation here for simplicity only. Identical reasoning would apply to a bipolar transistor or a vacuum tube, or indeed to any other amplifying device.

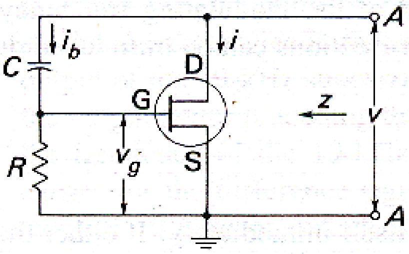


FIGURE 5-11 Basic reactance modulator.

Theory of reactance modulators In order to determine z, a voltage v is applied to the terminals A–A between which the impedance is to be measured, and the resulting current i is calculated. The applied voltage is then divided by this current, giving the impedance seen when looking into the terminals. In order for this impedance to be a pure reactance (it is capacitive here), two requirements must be fulfilled. The first is that the bias network current i_b must be negligible compared to the drain current. The impedance of the bias network must be large enough to be ignored. The second requirement is that the drain-to-gate impedance (X_C here) must be greater than the gate-to-source impedance (R in this case), preferably by more than 5:1. The following analysis may then be applied:

$$v_g = i_b R = \frac{Rv}{R - jX_C} \tag{5-15}$$

The FET drain current is

$$i = g_m v_g = \frac{g_m R v}{R - jX_C} \tag{5-16}$$

Therefore, the impedance seen at the terminals A–A is

$$z = \frac{v}{i} = v \div \frac{g_m R v}{R - jX_C} = \frac{R - jX_C}{g_m R} = \frac{1}{g_m}\left(1 - \frac{jX_C}{R}\right) \tag{5-17}$$

If $X_C \gg R$ in Equation (5-17), the equation will reduce to

$$z = -j\frac{X_C}{g_m R} \tag{5-18}$$

This impedance is quite clearly a capacitive reactance, which may be written as

$$X_{eq} = \frac{X_C}{g_m R} = \frac{1}{2\pi f g_m RC} = \frac{1}{2\pi f C_{eq}} \tag{5-19}$$

From Equation (5-19) it is seen that under such conditions the input impedance of the device at A–A is a pure reactance and is given by

$$C_{eq} = g_m RC \tag{5-20}$$

The following should be noted from Equation (5-20):

1. This equivalent capacitance depends on the device transconductance and can therefore be varied with bias voltage.
2. The capacitance can be originally adjusted to any value, within reason, by varying the components R and C.
3. The expression $g_m RC$ has the correct dimensions of capacitance; R, measured in ohms, and g_m, measured in siemens (s), cancel each other's dimensions, leaving C as required.
4. It was stated earlier that the gate-to-drain impedance must be much larger than the gate-to-source impedance. This is illustrated by Equation (5-17). If X_C/R had not been much greater than unity, z would have had a resistive component as well.

If R is not much less than X_C (in the particular reactance modulator treated), the gate voltage will no longer be exactly 90° out of phase with the applied voltage v, nor will the drain current i. Thus, the input impedance will no longer be purely reactive. As shown in Equation (5-17), the resistive component for this particular FET reactance modulator will be $1/g_m$. This component contains g_m, it will vary with the applied modulating voltage. This variable resistance (like the variable reactance) will appear directly across the tank circuit of the master oscillator, varying its Q and therefore its output voltage. A certain amount of amplitude modulation will be created. This applies to all the forms of reactance modulator. If the situation is unavoidable, the oscillator being modulated must be followed by an amplitude limiter.

The gate-to-drain impedance is, in practice, made five to ten times the gate-to-source impedance. Let $X_C = nR$ (at the carrier frequency) in the capacitive RC reactance FET so far discussed. Then

$$X_C = \frac{1}{\omega C} = nR$$

$$C = \frac{1}{\omega nR} = \frac{1}{2\pi fnR} \tag{5-21}$$

Substituting Equation (5-21) into (5-20) gives

$$C_{eq} = g_m RC = \frac{g_m R}{2\pi fnR}$$

$$C_{eq} = \frac{g_m}{2\pi fn} \tag{5-22}$$

Equation (5-22) is a very useful formula. In practical situations the frequency of operation and the ratio of X_C to R are the usual starting data from which other calculations are made.

b. Compare the differences between Frequency and Phase modulations. (4)

Answer:

Frequency and phase modulation From the purely theoretical point of view, the difference between FM and PM is quite simple—the modulation index is defined differently in each system. However, this is not nearly as obvious as the difference between AM and FM, and it must be developed further. First the similarity will be stressed.

In phase modulation, the phase deviation is proportional to the amplitude of the modulating signal and therefore independent of its frequency. Also, since the phase-modulated vector sometimes leads and sometimes lags the reference carrier vector, its instantaneous angular velocity must be continually changing between the limits imposed by ϕ_m; thus some form of frequency change must be taking place. In frequency modulation, the frequency deviation is proportional to the amplitude of the modulating voltage. Also, if we take a reference vector, rotating with a constant angular velocity which corresponds to the carrier frequency, then the FM vector will have a phase lead or lag with respect to the reference, since its frequency oscillates between $f_c - \delta$ and $f_c + \delta$. Therefore FM must be a form of PM. With this close similarity of the two forms of angle modulation established, it now remains to explain the difference.

If we consider FM as a form of phase modulation, we must determine what causes the phase change in FM. The larger the frequency deviation, the larger the phase deviation, so that the latter depends at least to a certain extent on the amplitude of the modulation, just as in PM. The difference is shown by comparing the definition of PM, which states in part that the modulation index is proportional to the modulating voltage *only,* with that of the FM, which states that the modulation index *is also inversely proportional to the modulation frequency*. This means that under identical conditions *FM and PM are indistinguishable for a single modulating frequency*. When the modulating frequency is changed the PM modulation index will remain constant, whereas the FM modulation index will increase as modulation frequency is reduced, and vice versa. This is best illustrated with an example.

The practical effect of all these considerations is that if an FM transmission were received on a PM receiver, the bass frequencies would have considerably more deviation (*of phase*) than a PM transmitter would have given them. Since the output of a PM receiver would be proportional to phase deviation (or modulation index), the signal would appear unduly bass-boosted. Phase modulation received by an FM system would appear to be *lacking in bass*. This deficiency could be *corrected by bass boosting the modulating signal prior to phase modulation*. This is the practical difference between phase and frequency modulation.

c. A Frequency Modulated Wave is represented by the voltage equation as

$v = 12\sin\left(6X10^8 t + 5\sin 1250t\right)$. **Find:** **(4)**

(i) Carrier Frequency

(ii) Modulating Frequency

Answer:

By comparing the given equation $v = 12\sin\left(6X10^8 t + 5\sin 1250t\right)$ with the Mathematical equation for Frequency Modulated Wave as

$v = A\sin\left(\omega_c t + m_f \sin\omega_m t\right)$,

We get,

$$\omega_c = 2\pi f_c = 6X10^8$$ and

$$\omega_m = 2\pi f_m = 1250$$

Therefore, (i) The Carrier Frequency

$$f_c = \frac{6X10^8}{2\pi} = 95.5MHz$$

(ii)The Modulating Frequency

$$fm = \frac{1250}{2\pi} = 199Hz$$

Q.5 a. Draw the block diagram of FM Superheterodyne Receiver and explain the function of each block. (8)

Answer:

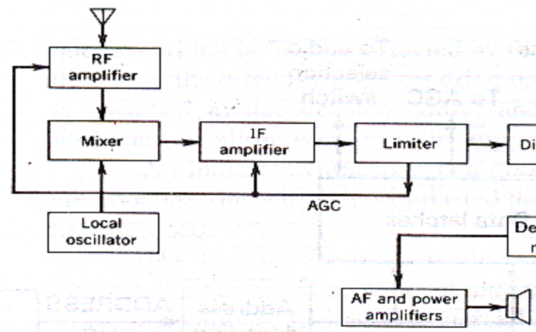


FIGURE 6-28 FM receiver block diagram.

6-4 FM RECEIVERS

The FM receiver is a superheterodyne receiver, and the block diagram of Figure 6-28 shows just how similar it is to an AM receiver. The basic differences are as follows:

1. Generally much higher operating frequencies in FM
2. Need for limiting and de-emphasis in FM
3. Totally different methods of demodulation
4. Different methods of obtaining AGC

6-4.1 Common Circuits—Comparison with AM Receivers

A number of sections of the FM receiver correspond exactly to those of other receivers already discussed. The same criteria apply in the selection of the intermediate frequency, and IF amplifiers are basically similar. A number of concepts have very similar meanings so that only the differences and special applications need be pointed out.

RF amplifiers An RF amplifier is always used in an FM receiver. Its main purpose is to reduce the noise figure, which could otherwise be a problem because of the large bandwidths needed for FM. It is also required to match the input impedance of the receiver to that of the antenna. To meet the second requirement, grounded gate (or base) or cascode amplifiers are employed. Both types have the property of low input impedance and matching the antenna, while neither requires neutralization. This is because the input electrode is grounded on either type of amplifier, effectively isolating input from output. A typical FET grounded-gate RF amplifier is shown in Figure 6-29. It has all the good points mentioned and the added features of low distortion and simple operation.

Oscillators and mixers The oscillator circuit takes any of the usual forms, with the Colpitts and Clapp predominant, being suited to VHF operation. Tracking is not nor-

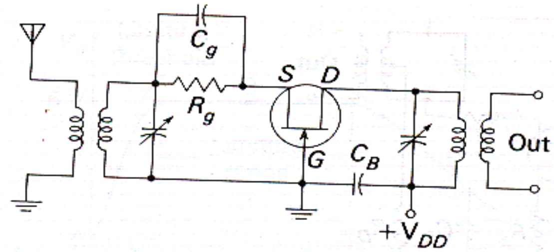


FIGURE 6-29 Grounded-gate FET RF amplifier.

mally much of a problem in FM broadcast receivers. This is because the tuning frequency range is only 1.25:1, much less than in AM broadcasting.

A very satisfactory arrangement for the front end of an FM receiver consists of FETs for the RF amplifier and mixer, and a bipolar transistor oscillator. As implied by this statement, separately excited oscillators are normally used, with an arrangement as shown in Figure 6-6.

Intermediate frequency and IF amplifiers Again, the types and operation do not differ much from their AM counterparts. It is worth noting, however, that the intermediate frequency and the bandwidth required are far higher than in AM broadcast receivers. Typical figures for receivers operating in the 88- to 108-MHz band are an IF of 10.7 MHz and a bandwidth of 200 kHz. As a consequence of the large bandwidth, gain per stage may be low. Two IF amplifier stages are often provided, in which case the shrinkage of bandwidth as stages are cascaded must be taken into account.

6-4.2 Amplitude Limiting

In order to make full use of the advantages offered by FM, a demodulator must be preceded by an amplitude limiter, as discussed in Chapter 5, on the grounds that any amplitude changes in the signal fed to the FM demodulator are spurious. (This does not apply to a receiver with a ratio detector which, as is shown in Section 6-4.4, provides a fair amount of limiting.) They must therefore be removed if distortion is to be avoided. The point is significant, since most FM demodulators react to amplitude changes as well as frequency changes. The limiter is a form of clipping device, a circuit whose output tends to remain constant despite changes in the input signal. Most limiters behave in this fashion, provided that the input voltage remains within a certain range. The common type of limiter uses two separate electrical effects to provide a relatively constant output. There are leak-type bias and early (collector) saturation.

b. Draw the block diagram of Pilot-Carrier Single Sideband Receiver and describe its working. (8)

Answer:

Pilot-carrier receiver As shown in Figure 6-46, in block form, a pilot-carrier receiver is a fairly straightforward communications receiver with trimmings. It uses double conversion, and AFC based on the pilot carrier. AFC is needed to ensure good frequency stability, which must be at least 1 part of 10^7 (long-term) for long-distance telephone and telegraph communications. Note also the use of one local crystal oscillator, with multiplication by 9, rather than two separate oscillators; this also improves stability.

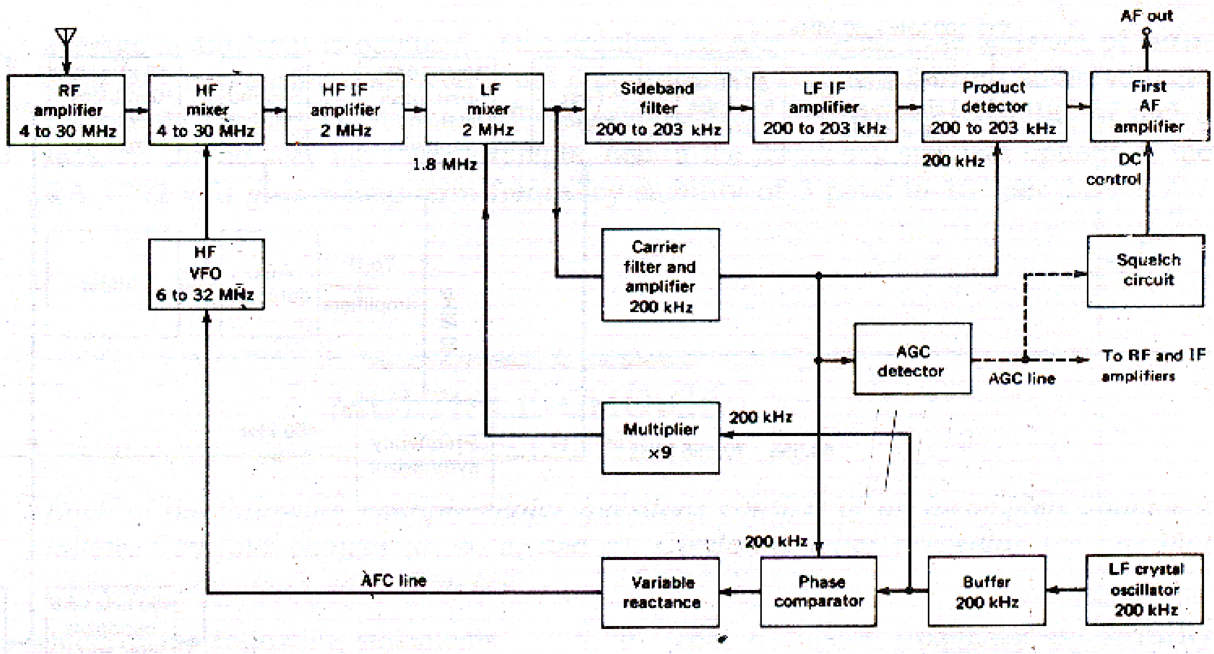


FIGURE 6-46 Block diagram of pilot-carrier single-sideband receiver.

The output of the second mixer contains two components—the wanted sideband and the weak carrier. They are separated by filters, the sideband going to the product detector, and the carrier to AGC and AFC circuits via an extremely narrowband filter and amplifier. The output of the carrier amplifier is fed, together with the buffered output of the crystal oscillator, to a *phase comparator*. This is almost identical to the phase discriminator and works in a similar fashion. The output depends on the *phase difference* between the two applied signals, which is zero or a positive or negative dc voltage, just as in the discriminator. The phase difference between the two inputs to the phase-sensitive circuit can be zero only if the frequency difference is zero. Excellent frequency stability is obtainable. The output of the phase comparator actuates a varactor diode connected across the tank circuit of the VFO and pulls it into frequency as required.

Because a pilot carrier is transmitted, automatic gain control is not much of a problem, although that part of the circuit may look complicated. The output of the carrier filter and amplifier is a carrier whose amplitude varies with the strength of the input signal, so that it may be used for AGC after rectification. Automatic gain control is also applied to the squelch circuit, as explained in Section 6-3.2 It should also be mentioned that receivers of this type often have AGC with two different time constants. This is helpful in telegraphy reception, and in coping to a certain extent with signal-strength variations caused by fading.

Suppressed-carrier receiver A typical block diagram is shown in Figure 6-47. This is actually a much simplified version of the receiver of Figure 6-18, which is capable of receiving all forms of AM but has here been shown in the ISB mode. The receiver has a number of very interesting features, of which the first is the fixed-frequency RF amplifier. This may be wideband, covering the entire 100-kHZ to 30-MHz receiving range; or, optionally, a set of filters may be used, each covering a portion of this range. The second very interesting feature is the very high first intermediate frequency, 40.455 MHz. Such high frequencies have been made possible by the advent of VHF crystal bandpass filters. They are increasingly used by SSB receivers, for a number of reasons. One, clearly, is to provide image frequency rejections much higher than previously available. Another reason is to facilitate receiver tuning. In the RA 1792, which is typical of high-quality professional receivers, a variety of tuning methods are available, such as push-button selection, or even automatic selection of a series of wanted preset channels stored in the microprocessor memory. However, an important method is the orthodox continuous tuning method, which utilizes a tuning knob. Since receivers of this type are capable of remote tuning, the knob actually adjusts the voltage applied to a varactor diode across the VFO in an indirect frequency synthesizer. There is a limit to the tuning range. If the first IF is high, the resulting range (70.455 MHz ÷ 40.555 MHz = 1.74:1) can be covered in a single sweep, with a much lower first IF it cannot be tuned so easily.

It will be seen that this is, nonetheless, a double-conversion superheterodyne receiver, up to the low-frequency IF stages. After this the main differences are due to the presence of the two independent sidebands, which are separated at this point with mechanical filters. If just a single upper and a single lower sideband are transmitted, the USB filter will have a bandpass of 455.25 to 458 kHz, and the LSB filter 452 to 454.75 kHz. Since the carrier is not transmitted, it is necessary to obtain AGC by rectifying part of the combined audio signal. From this a dc voltage proportional to the average audio level is obtained. This requires an AGC circuit time constant of sufficient length to ensure that AGC is not proportional to the instantaneous audio voltage. Because of the presence of the frequency synthesizer, the frequency stability of such a receiver can be very high. For example, one of the frequency standard options of the RA 1792 will give a long-term frequency stability of 3 parts in 10^9 per day.

Q.6a. Explain the following related to antennas:- (8)

(i) **Directive Gain**
(ii) **Antenna Efficiency**
(iii)**Bandwidth**
(iv) Beamwidth

Answer: (i) Directive Gain

Directive gain *Directive gain* is defined as the ratio of the power density in a particular direction of one antenna to the power density that would be radiated by an omnidirectional antenna (isotropic antenna). The power density of both types of antenna is measured at a specified distance, and a comparative ratio is established.

The gain of a Hertzian dipole with respect to an isotropic antenna = 1.5:1 power (1.5 (10 $\log_{10}$) = 1.76 dB).

The gain of a half-wave dipole compared to the isotropic antenna = 1.64:1 power (1.64 (10 $\log_{10}$) = 2.15 dB).

The wire antennas discussed in the preceding section have gains that vary from 1.64 (2.15 dB) for a half-wave dipole to 7.1 (8.51 dB) for an eight-wave dipole. These figures are for resonant antennas in free space. Similar nonresonant antennas have gains of 3.2 (5.05 dB) and 17.4 (12.4 dB) respectively. Two sets of characteristics can be obtained from the previous information:

1. The longer the antenna, the higher the directive gain.
2. Nonresonant antennas have higher directive gain than resonant antennas.

(ii)

Antenna losses and efficiency In addition to the energy radiated by an antenna, power losses must be accounted for. Antenna losses can be caused by ground resistance, corona effects, imperfect dialectric near the antenna, energy loss due to eddy currents induced into nearby metallic objects, and I^2R losses in the antenna itself. We can combine these losses and represent them as shown in Equation (9-5).

$$P\text{in} = P_d + P_{rad} \tag{9-5}$$

where P_{in} = power delivered to the feed point

P_d = power lost

P_{rad} = power actually radiated

Converting Equation (9-5) to I^2R terms, we may state the equation as follows.

$$I^2R_{in} = I^2R_d + I^2R_{rad}$$
$$R_{in} = R_d + R_{rad}$$

From this expression we can now develop an equation for calculating antenna efficiency.

$$n = \frac{R_{rad}}{R_{rad} + R_d} \times 100\% \tag{9-6}$$

R_d = antenna resistance

R_{rad} = antenna radiation resistance

Low- and medium-frequency antennas are least efficient because of difficulties in achieving the proper physical (resonant) length. These antennas can approach efficiencies of only 75 to 95 percent. Antennas at higher frequencies can easily achieve values approaching 100 percent. Radiation resistance values may vary from a few ohms to several hundred ohms depending on the choice of feed points and physical and electrical characteristics.

(iii)

Bandwidth The term bandwidth refers to the range of frequencies the antenna will radiate effectively; i.e., the antenna will *perform satisfactorily* throughout this range of frequencies. When the antenna power drops to ½ (3 dB), the upper and lower extremities of these frequencies have been reached and the antenna no longer *performs satisfactorily*.

Antennas that operate over a wide frequency range and still maintain satisfactory performance must have compensating circuits switched into the system to maintain impedance matching, thus ensuring no deterioration of the transmitted signals.

(iv)

Beamwidth The *beamwidth* of an antenna is described as the angles created by comparing the half-power points (3 dB) on the main radiation lobe to its maximum power point. In Figure 9-9, as an example, the *beam angle* is 30°, which is the sum of the two angles created at the points where the *field strength* drops to 0.707 (field strength is measured in μ/V/m) of the maximum voltage at the center of the lobe. (These points are known as the half-power points.)

b. Explain the characteristics, radiation pattern and applications of the following antennas:- (8)
(i) Yagi-Uda Antenna
(ii) Rhombic Antenna

Answer:

The Yagi-Uda antenna A Yagi-Uda antenna is an array consisting of a driven element and one or more parasitic elements. They are arranged collinearly and close together, as shown in Figure 9-25, together with the optical equivalent and the radiation pattern.

Since it is relatively unidirectional, as the radiation pattern shows, and has a moderate gain in the vicinity of 7 dB, the Yagi antenna is used as an HF transmitting antenna. It is also employed at higher frequencies, particularly as a VHF television receiving antenna. The back lobe of Figure 9-25*b* may be reduced, and thus the *front-to-back ratio* of the antenna improved, by bringing the radiators closer. However, this has the adverse effect of lowering the input impedance of the array, so that the separation shown, 0.1λ, is an optimum value.

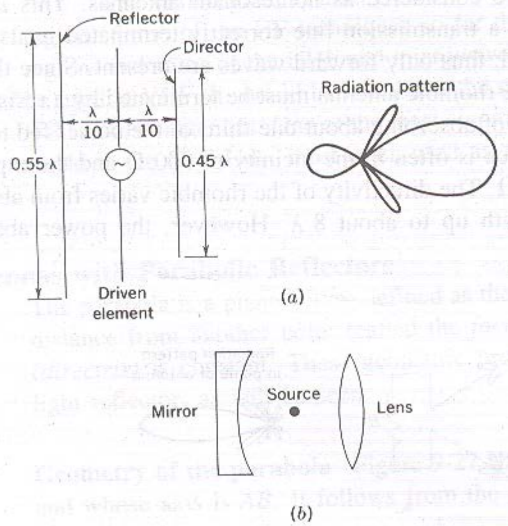


FIGURE 9-25 Yagi antenna. (*a*) Antenna and pattern; (*b*) optical equivalent.

The precise effect of the parasitic element depends on its distance and tuning, i.e., on the magnitude and phase of the current induced in it. As already mentioned, a parasitic element resonant at a lower frequency than the driven element (i.e., longer) will act as a mild reflector, and a shorter parasitic will act as a mild "director" of radiation. As a parasitic element is brought closer to the driven element, it will load the driven element more and reduce its input impedance. This is perhaps the main reason for the almost invariable use of a folded dipole as the driven element of such an array.

The Yagi antenna admittedly does not have high gain, but it is very compact, relatively broadband because of the folded dipole used and has quite a good unidirectional radiation pattern. As used in practice, it has one reflector and several directors which are either of equal length or decreasing slightly away from the driven element. Finally, it must be mentioned that the folded dipole, along with one or two other antennas, is sometimes called a *supergain antenna,* because of its good gain and beamwidth per unit area of array.

9-6.3 Nonresonant Antennas—The Rhombic

A major requirement for HF is the need for a multiband antenna capable of operating satisfactorily over most or all of the 3- to 30-MHz range, for either reception or transmission. One of the obvious solutions is to employ an array of nonresonant antennas, whose characteristics will not change too drastically over this frequency range.

A very interesting and widely used antenna array, especially for point-to-point communications, is shown in Figure 9-26. This is the *rhombic antenna,* which consists of nonresonant elements arranged differently from any previous arrays. It is a planar rhombus which may be thought of as a piece of parallel-wire transmission line bowed in the middle. The lengths of the (equal) radiators vary from 2 to 8 λ, and the radiation angle, ϕ, varies from 40 to 75°, being mostly determined by the leg length.

The four legs are considered as nonresonant antennas. This is achieved by treating the two sets as a transmission line correctly terminated in its characteristic impedance at the far end; thus only forward waves are present. Since the termination absorbs some power, the rhombic antenna must be terminated by a resistor which, for transmission, is capable of absorbing about one-third of the power fed to the antenna. The terminating resistance is often in the vicinity of 800 Ω and the input impedance varies from 650 to 700 Ω. The directivity of the rhombic varies from about 20 to 90°, increasing with leg length up to about 8 λ. However, the power absorbed by the

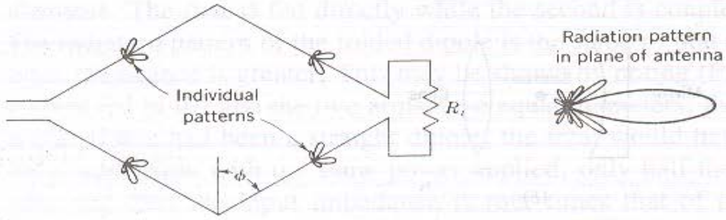


FIGURE 9-26 Rhombic antenna and radiation patterns.

termination must be taken into account, so that the *power gain* of this antenna ranges from about 15 to 60°. The radiation pattern is unidirectional as shown (Figure 9-26).

Because the rhombic is nonresonant, it does not have to be an integral number of half-wavelengths long. It is thus a broadband antenna, with a frequency range at least 4 : 1 for both input impedance and radiation pattern. The rhombic is ideally suited to HF transmission and reception and is a very popular antenna in commercial point-to-point communications.

Q.7 a. What is ionosphere? Explain its effects. (8)

Answer:

The ionosphere and its effects The ionosphere is the upper portion of the atmosphere, which absorbs large quantities of radiant energy from the sun, becoming heated and ionized. There are variations in the physical properties of the atmosphere, such as temperature, density and composition. Because of this and the different types of radiation received, the ionosphere tends to be stratified, rather than regular, in its distribution. The most important ionizing agents are ultraviolet and α, β, and γ radiation from the sun, as well as cosmic rays and meteors. The overall result, as shown in Figure 8-13, is a range of four main layers, D, E, F_1 and F_2, in ascending order. The last two combine at night to form one single layer.

The D layer is the lowest, existing at an average height of 70 km, with an average thickness of 10 km. The degree of its ionization depends on the altitude of the sun above the horizon, and thus it disappears at night. It is the least important layer from the point of view of HF propagation. It reflects some VLF and LF waves and absorbs MF and HF waves to a certain extent.

The E layer is next in height, existing at about 100 km, with a thickness of perhaps 25 km. Like the D layer, it all but disappears at night; the reason for these disappearances is the recombination of the ions into molecules. This is due to the absence of the sun (at night), when radiation is consequently no longer received. The main effects of the *E layer* are to aid MF surface-wave propagation a little and to reflect some HF waves in daytime.

The E_s layer is a thin layer of very high ionization density, sometimes making an appearance with the *E layer*. It is also called the *sporadic E layer;* when it does occur, it often persists during the night also. On the whole, it does not have an impor-

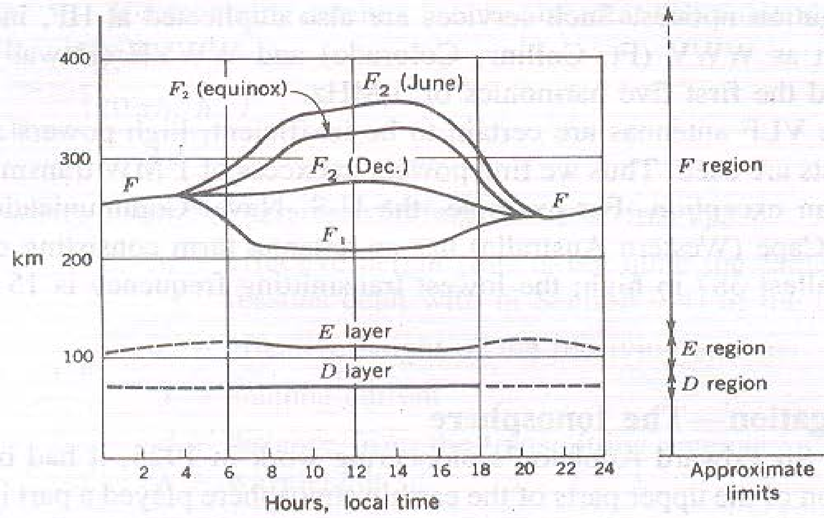


FIGURE 8-13 Ionospheric layers and their regular variations. ***(F. R. East, "The Properties of the Ionosphere Which Affect HF Transmission")***

tant part in long-distance propagation, but it sometimes permits unexpectedly good reception. Its causes are not well understood.

The F_1 layer, as shown in Figure 8-13, exists at a height of 180 km in daytime and combines with the F_2 layer at night, its daytime thickness is about 20 km. Although some HF waves are reflected from it, most pass through to be reflected from the F_2 layer. Thus the main effect of the F_1 layer is to provide more absorption for HF waves. Note that the absorption effect of this and any other layer is doubled, because HF waves are absorbed on the way up and also on the way down.

The F_2 layer is by far the most important reflecting medium for high-frequency radio waves. Its approximate thickness can be up to 200 km, and its height ranges from 250 to 400 km in daytime. At night it falls to a height of about 300 km, where it combines with the F_1 layer. Its height and ionization density vary tremendously, as Figure 8-13 shows. They depend on the time of day, the average ambient temperature and the sunspot cycle (see also the following sections dealing with the normal and abnormal ionospheric variations). It is most noticeable that the *F* layer persists at night, unlike the others. This arises from a combination of reasons; the first is that since this is the topmost layer, it is also the most highly ionized, and hence there is some chance for the ionization to remain at night, to some extent at least. The other main reason is that although ionization density is high in this layer, the *actual air density* is not, and thus most of the molecules in it are ionized. Furthermore, this low actual density gives the molecules a large *mean free path* (the statistical average distance a molecule travels before colliding with another molecule). This low molecular collision rate in turn means that, in this layer, ionization does not disappear as soon as the sun sets. Finally, it must be mentioned that the reasons for better HF reception at night are the combination of the F_1 and F_2 layers into one *F* layer, and the virtual disappearance of the other two layers, which were causing noticeable absorption during the day.

Reflection mechanism Electromagnetic waves returned to earth by one of the layers of the ionosphere appear to have been reflected. In actual fact the mechanism involved is refraction, and the situation is identical to that described in Figure 8-6. As the ionization density increases for a wave approaching the given layer at an angle, so the refractive index of the layer is reduced. (Alternatively, this may be interpreted as an increase in the conductivity of the layer, and therefore a reduction in its electrical density or dielectric constant.) Hence the incident wave is gradually bent farther and farther away from the normal, as in Figure 8-6.

If the rate of change of refractive index per unit height (measured in wavelengths) is sufficient, the refracted ray will eventually become parallel to the layer. It will then be bent downward, finally emerging from the ionized layer at an angle equal to the angle of incidence. Some absorption has taken place, but the wave has been returned by the ionosphere (well over the horizon if an appropriate angle of incidence was used).

Terms and definitions The terminology that has grown up around the ionosphere and sky-wave propagation includes several names and expressions whose meanings are not obvious. The most important of these terms will now be explained.

The virtual height of an ionospheric layer is best understood with the aid of Figure 8-14. This figure shows that as the wave is refracted, it is bent down gradually rather than sharply. However, below the ionized layer, the incident and refracted rays follow paths that are exactly the same as they would have been if *reflection* had taken place from a surface located at a greater height, called the *virtual height* of this layer. If the virtual height of a layer is known, it is then quite simple to calculate the angle of incidence required for the wave to return to ground at a selected spot.

The critical frequency (f_c) for a given layer is the highest frequency that will be returned down to earth by that layer after having been beamed straight up at it. It is important to realize that there is such a maximum, and it is also necessary to know its value under a given set of conditions, since this value changes with these conditions. It was mentioned earlier that a wave will be bent downward provided that the rate of change of ionization density is sufficient, and that this rate of ionization is measured

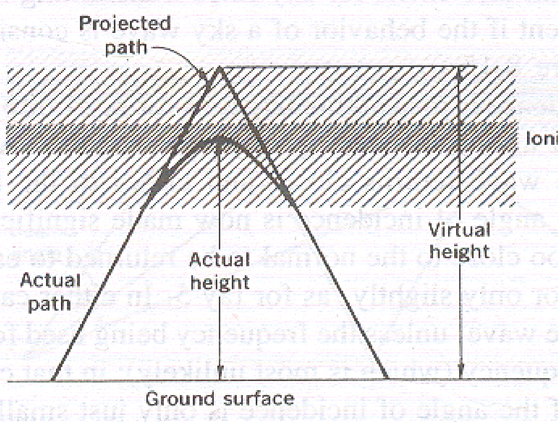


FIGURE 8-14 Actual and virtual heights of an ionized layer.

per unit wavelength. It also follows that the closer to being vertical the incident ray, the more it must be bent to be returned to earth by a layer. The result of these two effects is twofold. First, the higher the frequency, the shorter the wavelength, and the less likely it is that the change in ionization density will be sufficient for refraction. Second, the closer to vertical a given incident ray, the less likely it is to be returned to ground. Either way, this means that a maximum frequency must exist, above which rays go through the ionosphere. When the angle of incidence is normal, the name given to this maximum frequency is *critical frequency;* its value in practice ranges from 5 to 12 MHz for the F_2 layer.

The maximum usable frequency, or *MUF,* is also a limiting frequency, but this time for some specific angle of incidence other than the normal. In fact, if the angle of incidence (between the incident ray and the normal) is θ, it follows that

$$MUF = \frac{critical\ frequency}{\cos\theta}$$

$$= f_c \sec\theta \qquad \textbf{(8-12)}$$

This is the so-called *secant law,* and it is very useful in making preliminary calculations for a specific MUF. Strictly speaking, it applies only to a flat earth and a flat reflecting layer. However, the angle of incidence is not of prime importance, since it is determined by the distance between the points that are to be joined by a sky-wave link. Thus MUF is defined in terms of two such points, rather than in terms of the angle of incidence at the ionosphere, it is defined at the highest frequency that can be used for sky-wave communication between two given points on earth. It follows that there is a different value of MUF for each pair of points on the globe. Normal values of MUF may range from 8 to 35 MHz, but after unusual solar activity they may rise to as high as 50 MHz. The highest working frequency between a given pair of points is naturally made less than the MUF, but it is not very much less for reasons that will be seen.

The skip distance is the shortest distance from a transmitter, measured along the surface of the earth, at which a sky wave of fixed frequency (more than f_c) will be returned to earth. That there should be a minimum distance may come as a shock. One expects there to be a maximum distance, as limited by the curvature of the earth, but nevertheless a definite minimum also exists for any fixed transmitting frequency. The reason for this becomes apparent if the behavior of a sky wave is considered with the aid of a sketch, such as Figure 8-15.

When the angle of incidence is made quite large, as for ray 1 of Figure 8-15, the sky wave returns to ground at a long distance from the transmitter. As this angle is slowly reduced, naturally the wave returns closer and closer to the transmitter, as shown by rays 2 and 3. If the angle of incidence is now made significantly less than that of ray 3, the ray will be too close to the normal to be returned to earth. It may be bent noticeably, as for ray 4, or only slightly, as for ray 5. In either case the bending will be insufficient to return the wave, unless the frequency being used for communication is less than the critical frequency (which is most unlikely); in that case everything is returned to earth. Finally, if the angle of incidence is only just smaller than that of ray 3, the wave may be returned, but at a distance farther than the return point of ray 3; a ray such as this is ray 6 of Figure 8-15. This upper ray is bent back very gradually,

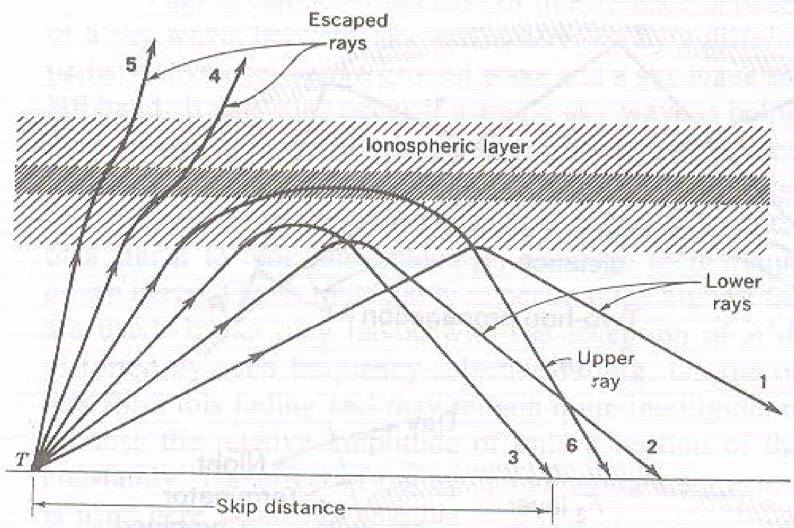


FIGURE 8-15 Effects of ionosphere on rays of varying incidence.

because ion density is changing very slowly at this angle. It thus returns to earth at a considerable distance from the transmitter and is weakened by its passage.

Ray 3 is incident at an angle which results in its being returned as close to the transmitter as a wave of this frequency can be. Accordingly, the distance is the *skip distance*. It thus follows that any higher frequency beamed up at the angle of ray 3 will not be returned to ground. It is seen that the frequency which makes a given distance correspond to the skip distance is the MUF for that pair of points.

At the skip distance, only the normal, or lower, ray can reach the destination, whereas at greater distances the upper ray can be received as well, causing interference. This is a reason why frequencies not much below the MUF are used for transmission. Another reason is the lack of directionality of high-frequency antennas, which is discussed in Section 9-6. If the frequency used is low enough, it is possible to receive lower rays by two different paths after either one or two hops, as shown in Figure 8-16, the result of this is interference once again.

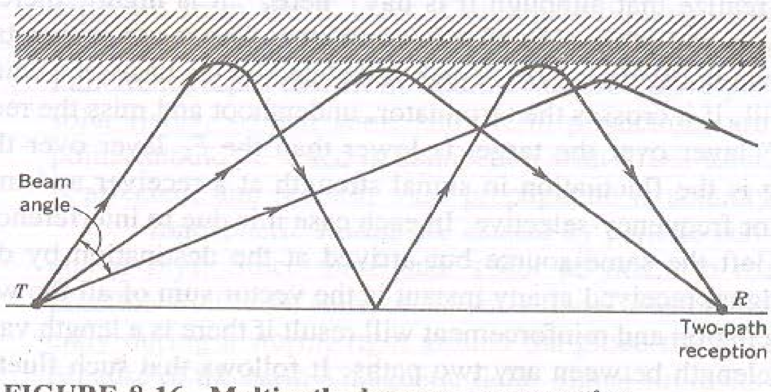


FIGURE 8-16 Multipath sky-wave propagation.

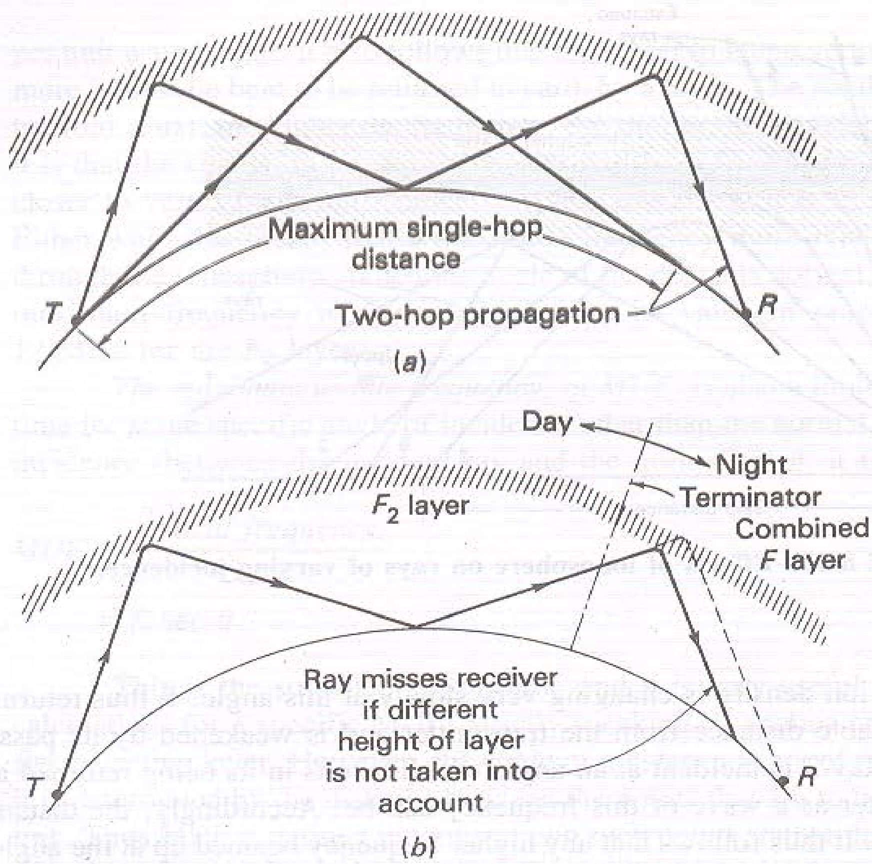


FIGURE 8-17 Long-distance sky-wave transmission paths. *(a)* North-south; *(b)* east-west.

The *transmission path* is limited by the skip distance at one end and the curvature of the earth at the other. The longest single-hop distance is obtained when the ray is transmitted tangentially to the surface of the earth, as shown in Figure 8-17. For the F_2 layer, this corresponds to a maximum practical distance of about 4000 km. Since the semicircumference of the earth is just over 20,000 km, multiple-hop paths are often required, and Figure 8-17 shows such a situation. No unusual problems arise with multihop north-south paths. However, care must be taken when planning long east-west paths to realize that although it is day "here," it is night "there," if "there" happens to be on the other side of the terminator. The result of not taking this into account is shown in Figure 8-17*b*. A path calculated on the basis of a constant height of the F_2 layer will, if it crosses the terminator, undershoot and miss the receiving area as shown—the *F* layer over the target is lower than the F_2 layer over the transmitter.

Fading is the fluctuation in signal strength at a receiver and may be rapid or slow, general or frequency-selective. In each case it is due to interference between two waves which left the same source but arrived at the destination by different paths. Because the signal received at any instant is the vector sum of all the waves received, alternate cancellation and reinforcement will result if there is a length variation as large as a half-wavelength between any two paths. It follows that such fluctuation is more likely with smaller wavelengths, i.e., at higher frequencies.

Fading can occur because of interference between the lower and the upper rays of a sky wave; between sky waves arriving by a different number of hops or different paths; or even between a ground wave and a sky wave especially at the lower end of the HF band. It may also occur if a single sky wave is being received, because of fluctuations of height or density in the layer reflecting the wave. One of the more successful means of combating fading is to use space or frequency diversity (see Section 6-3.2).

Because fading is frequency-selective, it is quite possible for adjacent portions of a signal to fade independently, although their frequency separation is only a few dozen hertz. This is most likely to occur at the highest frequencies for which sky waves are used. It can play havoc with the reception of AM signals, which are seriously distorted by such frequency-selective fading. On the other hand, SSB signals suffer less from this fading and may remain quite intelligible under these conditions. This is because the relative amplitude of only a portion of the received signal is changing constantly. The effect of fading on radiotelegraphy is to introduce errors, and diversity is used here wherever possible.

b. A rectangular waveguide measures 3 X 4.5 cm internally and has a 9-GHz signal propagated in it. Calculate the following characteristics for $TM_{1,1}$ mode
(i) Cutoff wavelength
(ii) Guide Wavelength
(iii)Group Velocity
(iv) Phase velocity **(8)**

Answer:

The free-space wavelength given by

$$\lambda = \frac{v_c}{f} = \frac{3X10^{10}}{9X10^{9}} = 3.33cm.$$

(i) The Cutoff Wavelength for mode $TM_{1,1}$ is given by:

$$\lambda_o = \frac{2}{\sqrt{(m/a)^2 + (n/b)^2}} = \frac{2}{\sqrt{(1/4.5)^2 + (1/3)^2}} = \frac{2}{0.4} = 5cm$$

(ii) The Guide Wavelength for mode $TM_{1,1}$ is given by:

$$\lambda_p = \frac{\lambda}{\rho} \text{ , where } \rho = \sqrt{1 - \left(\frac{\lambda}{\lambda_o}\right)^2} = \sqrt{1 - \left(\frac{3.33}{5}\right)^2} = 0.746.$$

Therefore, $\lambda_p = \frac{\lambda}{\rho} = \frac{3.33}{0.746} = 4.6cm.$

(iii) The Group Velocity for mode $TM_{1,1}$ is given by:

$$v_g = v_c \rho = 3X10^8 X 0.746 = 2.24X10^8 m/s.$$

(iv) The Phase Velocity for mode $TM_{1,1}$ is given by:

$$v_p = \frac{v_c}{\rho} = \frac{3X10^8}{0.746} = 4.02X10^8 m/s.$$

Q.8 a. What is Pulse Code Modulation? Describe its principle with the help of suitable diagram. Also list out its advantages and applications. (8)

Answer:

Pulse-Code Modulation (PCM)

Pulse-code modulation is just as different from the forms of pulse modulation so far studied as they were from AM or FM. PAM and PTM differed from AM and FM because, unlike in those two continuous forms of modulation, the signal was sampled and sent in pulse form. Like AM and FM, they were forms of *analog* communication—in all these forms a signal is sent which has a characteristic that is infinitely variable and proportional to the modulating voltage. In common with the other forms of pulse modulation, PCM also uses the sampling technique, but it differs from the others in that it is a *digital* process. That is, instead of sending a pulse train capable of continuously varying one of the parameters, the PCM generator produces a series of numbers, or digits (hence the name *digital* process). Each one of these digits, almost always in binary code, represents the *approximate amplitude* of the signal sample at that instant. The approximation can be made as close as desired, but it is always just that, an *approximation*.

Principles of PCM In PCM, the total amplitude range which the signal may occupy is divided into a number of standard levels, as shown in Figure 13-8. Since these levels are transmitted in a binary code, the actual number of levels is a power of 2; 16 levels are shown here for simplicity, but practical systems use as many as 128. By a process called *quantizing*, the level actually sent at any sampling time is the nearest standard (or *quantum*) level. As shown in Figure 13-8, should the signal amplitude be 6.8 V at any time, it is not sent as a 6.8-V pulse, as it might have been in PAM, nor as a 6.8-μs-wide pulse as in PWM, but simply as the digit 7, because 7 V is the standard amplitude nearest to 6.8 V. Furthermore, the digit 7 is sent at that instant of time as a

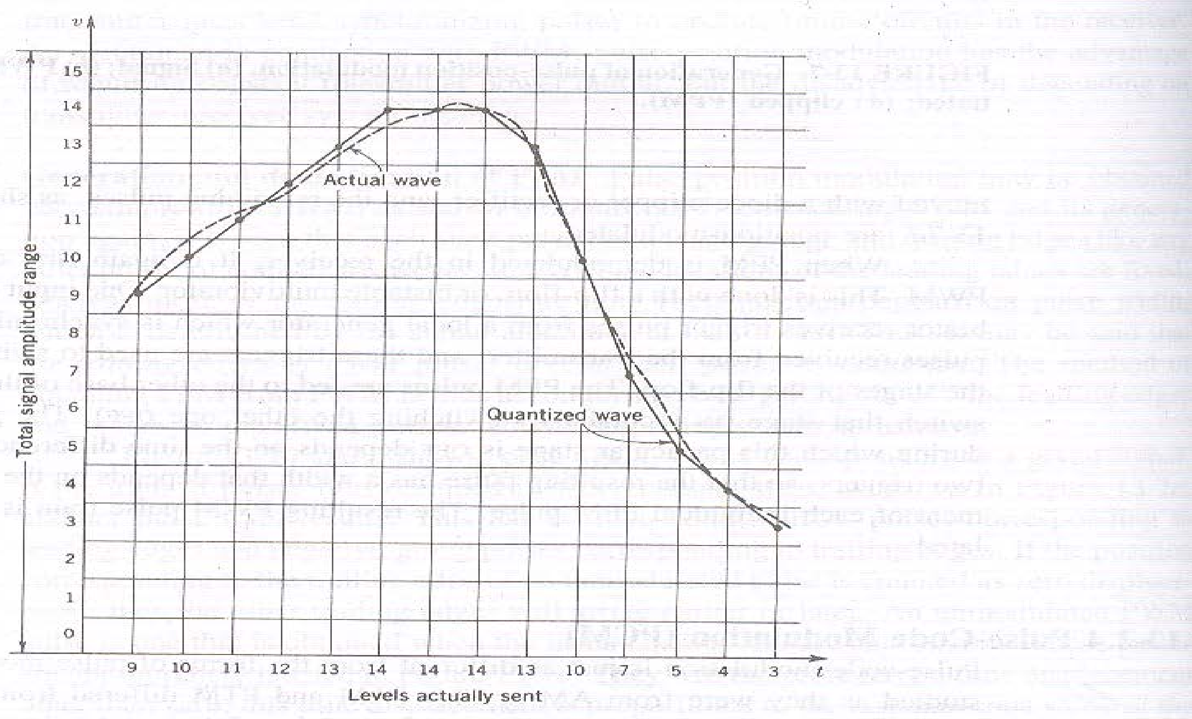


FIGURE 13-8 Quantization of signal for pulse-code modulation.

series of pulses corresponding to the number 7. Since there are 16 levels (2^4), 4 binary places are required; the number becomes 0111, and could be sent as 0PPP, where P = pulse and 0 = no-pulse. Actually, it is often sent as a binary number back-to-front, i.e., as 1110, or PPP0, to make demodulation easier.

As shown in Figure 13-8, the signal is continuously sampled, quantized, coded and sent, as each sample amplitude is converted to the nearest standard amplitude and into the corresponding back-to-front binary number. Provided sufficient quantizing levels are used, the result cannot be distinguished from that of analog transmission.

A supervisory or signaling bit is generally added to each code group representing a quantized sample. Hence each group of pulses denoting a sample, here called a *word*, is expressed by means of $n + 1$ bits, where 2^n is the chosen number of standard levels.

Advantages and applications of PCM A person may well ask, at this stage, "If PCM is so marvelous, why are any other modulation systems used?" There are three answers to this question, namely:

1. The other systems came first.
2. PCM requires very complex encoding and quantizing circuitry.
3. PCM requires a large bandwidth compared to analog systems.

PCM was invented by Alex H. Reeves in Great Britain, in 1937. When he patented it the following year, it was an astonishingly detailed and complete system. However, its very complexity prevented its immediate use—there were no really suitable electronic devices to implement it. By the end of World War II it was being used in a microwave relay system designed and operated by the U.S. Army Signal Corps. Soon after the war the system was adapted and evolved for commercial use by the Bell System. PCM then received a real boost from a very important paper of Oliver and others, in 1948, but it was some 10 years before it was actually used for telephony. Its first practical application in commercial telephony was in short-distance, medium-density work, in Great Britain and the United States in the early 1960s. Semiconductors and integration (it was not yet "large-scale" then) made its use practicable. Quite a number of new communication facilities built around the world have used PCM, and its use has grown very markedly during the 1980s.

As regards the second point, it is perfectly true that PCM requires much more complex modulating procedures than analog systems. However, multiplexing equipment is very much cheaper, and repeaters do not have to be placed so close together because PCM tolerates much worse signal-to-noise ratios. Especially because of very large-scale integration, the complexity of PCM is no longer a significant cost penalty.

Although the large bandwidth requirements still represent a problem, it is no longer as serious as it had earlier been, because of the advent of large-bandwidth fiber-optic systems. However, the large bandwidth requirements should be recognized. A typical first-level PCM system is the Bell T1 digital transmission system in use in North America. As described in greater detail in Section 15-1.2, it provides 24 PCM channels with time-division multiplexing. Each channel requires 8 bits per sample and thus 24 channels will need $24 \times 8 + 1 = 193$ bits—the extra 1 bit is an additional sync signal. With a sampling rate of 8000 per second, a total of $8000 \times 193 = 1{,}544{,}000$ bps will be sent by using this system. Work earlier in this chapter showed that the bandwidth in hertz would have to be at least half that figure, but the practical system in fact uses a bandwidth of 1.5 MHz as an optimum figure. It will be shown in Chapter 15 that 24 channels correspond to two groups, requiring a bandwidth of 96 kHz if frequency-division multiplex is used. PCM is seen to require 16 times as much bandwidth for the same number of channels. However, the situation in practice is not quite so bad, because economies of scale begin to appear when higher levels of digital multiplexing are used (see also Section 15-1.2).

The following considerations ensured that the main application of PCM for telephony was in 24-channel frames over wire pairs which previously had carried only one telephone conversation each. Their performance was not good enough to provide 24 FDM channels, but after a little modification 24 PCM channels could be carried over the one pair of wires. Since the mid-1970s the picture has changed dramatically. First, very large-scale integration reduced costs significantly. Then came the proliferation of digital systems, such as data transmissions, which were clearly advantaged by not having to be converted into analog prior to transmission and reconverted to digital after reception. Finally, fiber-optic systems became practical, with two effects. On the one hand, the current state of development of lasers and receiving diodes is such that digital operation is preferable to analog because of nonlinearities (see Chapter 18). Huge bandwidths, e.g., 565 Mbps per pair of fibers, have become available without attendant huge costs. The use of PCM in the broadband networks of advanced countries is increasing by leaps and bounds.

PCM also finds use in space communications. Indeed, the *Mariner IV* probe was an excellent example of the noise immunity of PCM, when, back in 1965, it transmitted the first pictures of Mars. Admittedly, each picture took 30 minutes to transmit, whereas it takes only 1/30 s in TV broadcasting. The *Mariner IV* transmitter was just over 200,000,000 km away, and the transmitting power was only 10 W. PCM was used; no other system would have done the job.

b. What is Frequency Shift Keying? Describe briefly. (8)

Answer:

Frequency-shift keying (FSK) It would be quite possible to transmit teletype by the ordinary ON-OFF keying of the transmitter. We could use amplitude modulation with pulses, ON corresponding to mark and OFF to space. Such a system has the inherent disadvantage that there is no real indication for the space. In addition, a system such as this would suffer from all the usual ailments of amplitude modulation, as a result of which it is never used for automatic telegraphy (it is, of course, widely used for manual Morse code CW operation). A system known as *frequency-shift keying* is generally used instead.

FSK is a system of frequency modulation. In it, the nominal unmodulated carrier frequency corresponds to the *mark* condition, and a *space* is represented by a downward frequency shift. The amount was 850 Hz in the original wideband FSK system designed for HF radio. For transmission by line or broadband systems, the current shift is 60 Hz, as laid down in CCITT Rec. R35. This is known as narrowband FSK, or frequency-modulated voice-frequency telegraph (FMVFT). FSK is still often used for HF radio transmissions, with a frequency shift that is commonly 170 Hz. As with other forms of FM, the main advantage of the wideband system is greater noise immunity, while the narrowband systems are used to conserve the allocated frequency spectrum. Note that FSK may be thought of as an FM system in which the carrier frequency is midway between the mark and space frequencies, and modulation is by a square wave. In practice, of course, only the fundamental frequency of the square wave is transmitted, and regeneration takes place in the receiver.

In the FSK generator, the frequency shift may be obtained by applying the varying dc output of the telegraph machine to a varactor diode in a crystal oscillator. At the receiving end, the signal is demultiplexed (if, as is common, FDM was used to send a number of telegraph or telex transmissions together) and applied to a standard phase discriminator. From the discriminator, signals of either polarity will be available. After some pulse shaping, they are applied to the receiving teletypewriter. If the telegraph transmission is by HF radio, the phase discriminator works at a (fairly low) intermediate frequency, although other methods are also possible for demodulation. An amplitude limiter is always used in the receiver, to take full advantage of the noise immunity of FSK.

Q.9 Write short note on <u>any TWO</u> the following: (2×8)
(i) Time Division Multiplexing
(ii) Coaxial Cables
(iii) Submarine Cables

Answer: (i) Time Division Multiplexing:

Time-Division Multiplex

The topic of TDM is an extension of pulse modulation, discussed in Chapter 13. It is covered here to permit the two major multiplexing methods to be compared. In time-division multiplex, use is made of the fact that narrow pulses with wide spaces between them are generated in any of the pulse modulation systems, so that the spaces can be used by signals from other sources. Moreover, although the spaces are relatively fixed in width, pulses may be made as narrow as desired, thus permitting the generation of high-level hierarchies.

The method of achieving TDM is best illustrated by describing the makeup of an actual system, and so a practical basic PCM system used in North America has been selected as the example. In somewhat simplified fashion, this may be described as a 24-channel system, having a sampling rate of 8000 samples per second, 8 bits (i.e., 256 sampling levels) per sample, and a pulse width of approximately 0.625 μs. This means that the sampling interval is $1/8000 = 0.000125$ s $= 125$ μs, and the period required for each pulse group is $8 \times 0.625 = 5$ μs. If there were no multiplexing and only one channel were sent, the transmission would consist of 8000 frames per second, each made up of furious activity during the first 5 μs and nothing at all during the remaining 120 μs. This would clearly be wasteful and would represent an unnecessarily complicated method of encoding a single channel, and so this system exploits the large spaces between the pulse groups. In fact, each 125-μs frame is used to provide 24 adjacent channel time slots, with the twenty-fifth slot assigned for synchronization. Each frame consists of 193 bits—24×8 for each channel, plus 1 for sync, and since there are 8000 frames per second, the bit rate is 1.544 Mbit/s.

Slow-speed TDM, as often used in radiotelemetry, is produced simply with rotating mechanical switches. A number of channels are fed simultaneously to the switch in the transmitter—one channel to each switch contact—while the output is taken from the moving rotor. This rotates slowly and remains in contact with each channel for a predetermined period, during which time the output of that channel is the only one passed on for transmission. There is a corresponding rotating switch in the receiver, synchronized to the one in the transmitter, which reverses the process to separate the received channels.

The high-speed TDM described here uses electronic switching and delay lines to accomplish the same result. Each sampling circuit, one per channel, simultaneously receives a trigger pulse which causes it to sample its signal, and each channel output is then fed to an adder. However, whereas the output of the first sampler goes straight to the adder, that of the second is delayed by 5 μs, with a delay line or delay circuit. The output of the third sampling circuit is similarly delayed but by 10 μs, and so on, until the twenty-fourth channel is delayed by 115 μs. In this way, each successive interval during the 125-μs frame is occupied by the transmission of a different channel, and the process is repeated 8000 times per second.

In the receiver, the output of the main detector is fed simultaneously to 24 AND gates. An *AND gate*, or coincidence circuit, is a simple device having one output and two or more input terminals, so arranged that an output is obtained only if all (in this case both) input signals are present. In this case each gate has two input terminals, and the second input to each gate is provided from a clock-synchronized gating generator, which is a monostable multivibrator providing rectangular pulses of 5 μs duration, 8000 times per second. Delay lines or circuits are used once again, with the gating pulse to the first gate not delayed at all, that to the second gate delayed by 5 μs and so forth. In this fashion each gate is open only during the appropriate time intervals, and the 24 channels are duly separated.

If transmission is by wire, the 1.544-Mbit/s pulse train is the signal sent, but if cable or radio communication is used, the pulse train either modulates the carrier or else is further multiplexed, with similar pulse trains, all combined together into a higher TDM hierarchical level.

(ii) Coaxial Cables

Coaxial Cables

A coaxial cable system consists of a tube carrying a number of coaxial cables of the type covered in Chapter 7, together with repeaters and other ancillary equipment. Separate cables are used for the two directions of transmission, and a pair of spare cables is also provided for protection in case of failure. The number of cables per tube may be as low as four in smaller systems or as high as 22 in major systems, as illustrated in Figure 15-4. The typical number of channels per cable varies from 600 in a 3-MHz system to 3600 in an 18-MHz system.

Since signals are attenuated as they travel along the cable (see Section 7-1.3), amplifying repeaters must be placed at suitable intervals along the route. The distance varies, being roughly inversely proportional to the bandwidth of the system. It may be as much as 10 km between repeaters for a small system, but in the L5 system of Figure 15-4, where bandwidths for all cables are nearly 58 MHz, repeaters are placed at 1.6-km intervals. Since there are repeaters, a dc supply must be fed to the cable to power them. In the L5 system, dc power-feeding stations are located 120 km apart, i.e., 75 repeaters apart. Assuming an 18-V drop across each repeater, and noting that repeaters are in series for direct current since otherwise the required currents would be too high, this means that the dc voltage applied at each station must be 1350 V. To minimize insulation problems, what is done in practice is to apply voltages of half that value, but of opposite polarities, at the two adjoining dc feed stations. A station at one end may thus feed +675 V to the cable, while the next station along feeds −675 V toward the first station and +675 V toward the next station down the cable.

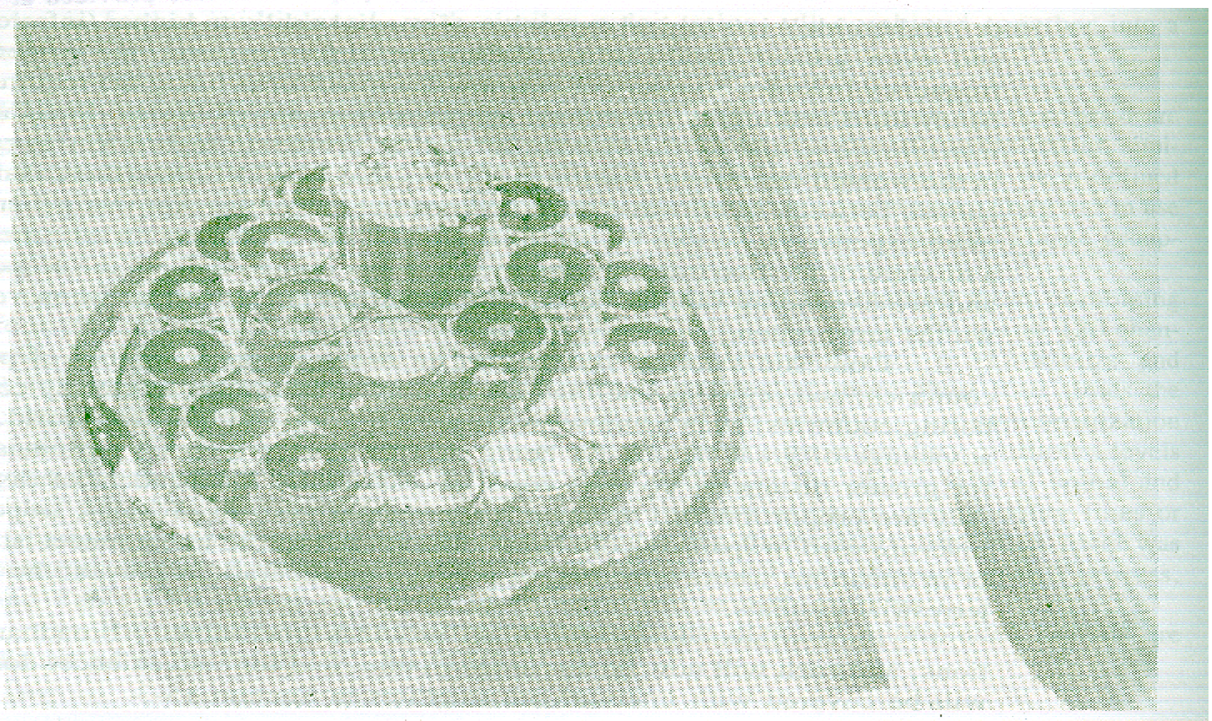


FIGURE 15-4 Coaxial cable used in the L5 system for carrying up to 108,000 simultaneous two-way telephone conversations. *(By permission of AT&T Long Lines.)*

Broadband systems must have excellent frequency and phase-delay responses to be of use. This cannot be achieved by cables and repeaters unaided, so that equalizers are also located along the cable, 60 km apart in the L5 system. It should be noted that there is need for two kinds of equalizers. The fixed type compensates for constant, known deviations in frequency and phase response which are inherent in each particular system. Adjustable equalizers, generally provided at the two ends of the system, are used to compensate for the variables and the unpredictable variations. Where adjustable equalizers are located in underground stations along the cable, they are normally adjustable in steps rather than continuously. In modern systems these adjustments may be made from the control stations at the ends, by sending appropriate signals down the cable. Finally, to ensure constant gain along the system, thus preventing excessive noise and intermodulation distortion, the gain of repeaters is regulated. This may be done by having adjustable-gain repeaters at intervals along the cable and altering their gain as required with suitable control signals.

Multiplexing and demultiplexing bays form the major portion of the terminal equipment. It is in these bays that FDM, as described in Section 15-1, takes place. Dc power feed equipment is also located at the terminals, as are interconnections to other systems, be they local or trunk. Surveillance equipment is also provided at terminal stations. It is here that system pilots are applied, and those that were applied at the other end are extracted. A distinction should be made between a supergroup—or even supermastergroup—pilot, as described in Section 15-1.1, and a system pilot. The latter belongs to the system and is used for end-to-end system regulation and monitoring. The supergroup pilot is applied at the point at which the supergroup is formed and extracted at the point at which it is broken up. It is used for regulating and monitoring that particular supergroup, which may traverse many different links. Although each is regulated, small, in-tolerance departures from correct response in the various links may be additive, resulting in a supergroup that is out of tolerance end to end. Finally, each terminal is provided with equipment which, should there be a cable failure, permits it to interrogate the repeaters in the link, so as to allow quick localization of the fault. Furthermore, to minimize the effects of outages, terminal stations may be provided with redundant and/or duplicated systems, allowing their staff to patch rapidly around any breaks.

Some students may wonder why communications systems tend to have more and more capacity. The answer is that long-distance telephony, telex and television transmissions in most countries have been increasing at high rates, for over two decades, while data transmission in developed countries is growing at very high annual rates of close to 50 percent. Coupled to this demand growth is the fact that a 10,800-channel system is decidedly cheaper to install and maintain than three 3600-channel systems. Such broadband links are manufactured by some of the world's most modern, efficient and reliable companies.

(iii) Submarine Cables

Submarine Cables

Submarine cables use principles very much like those of coaxial cables. Thus they are coaxial, have repeaters and equalizers and have dc power fed to them, with opposite polarities fed from opposite ends to reduce insulation problems. However, submarine cables use a single coaxial tube for both directions of transmission, with frequency techniques similar to those of microwave links to separate the two directions. The extent to which cables have spread out around the world, since *TAT-1* in 1956, is shown in Figure 15-6.

Cables such as the 48-circuit *TAT-1* and the 80-circuit *CANTAT-1* (1961) are often referred to as "first-generation" cables. They feature vacuum-tube repeaters, at intervals of 50 to 60 km. Second-generation cables, such as the *SAT-1* (1968) cable from Portugal to South Africa, have up to 360 circuits, with vacuum-tube repeaters at 18-km intervals. Vacuum tubes were used as late as 1968 because of their proven reliability. Submerged cable or repeater repair is perfectly feasible, but is a complex and costly process. It involves sending cableships to the affected area and dragging the sea bottom for the cable, while the interrupted circuits are restored via another cable or a satellite (at no small cost). It can therefore be appreciated that reliability is the keynote, and vacuum tubes had certainly established a reputation for that in submarine systems.

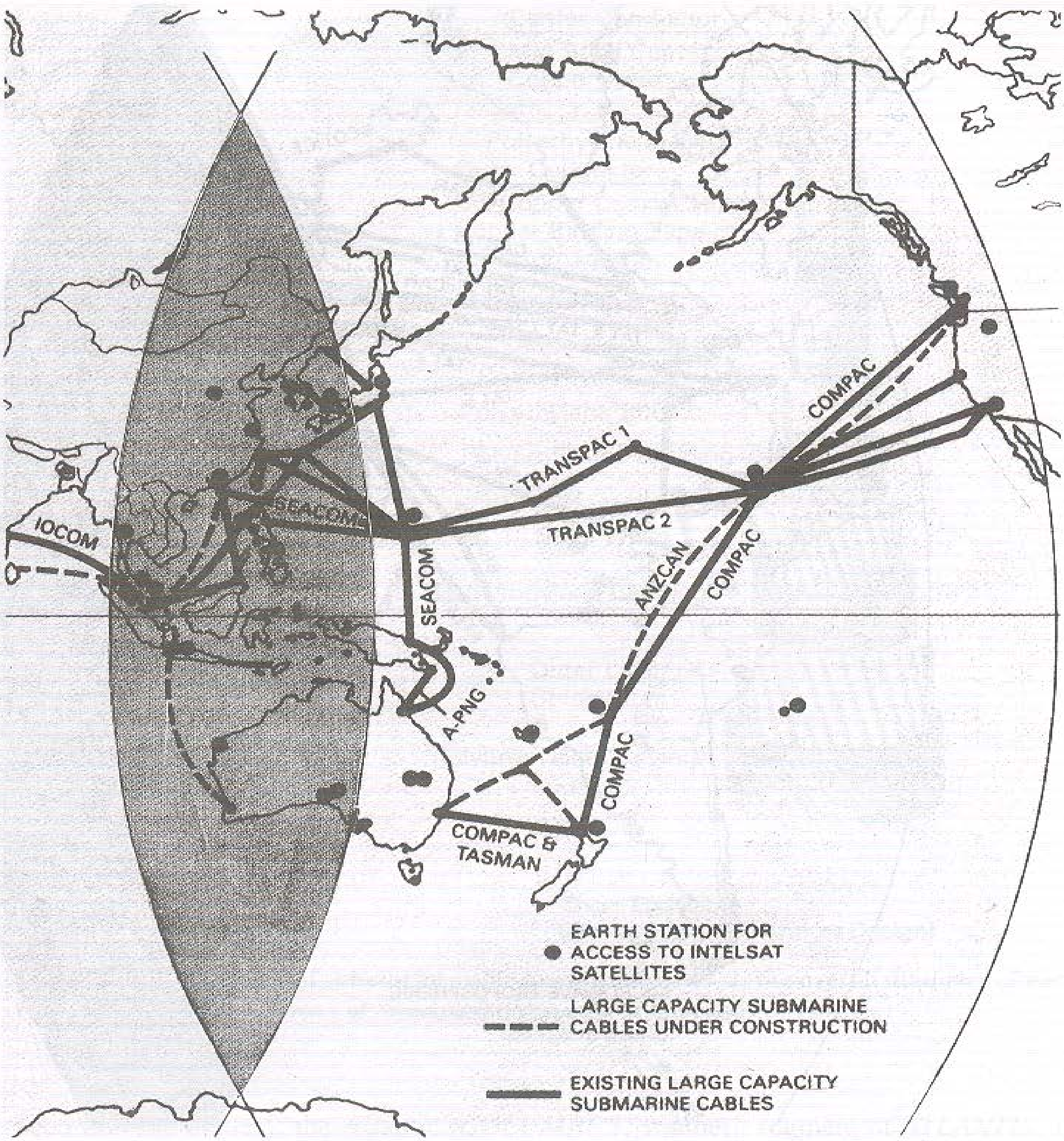
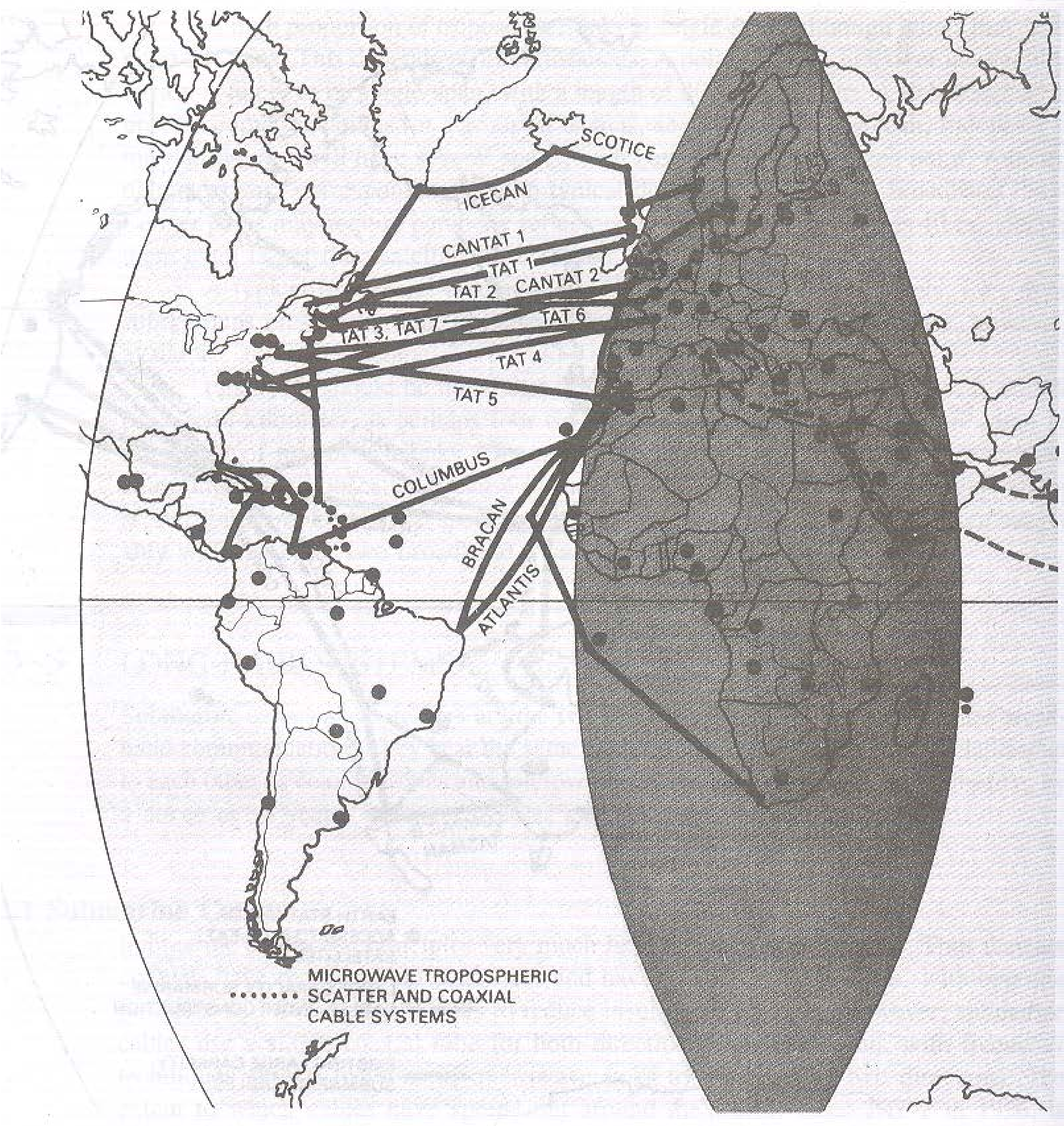


FIGURE 15-6 The world's major submarine cables and satellite earth stations. The curved lines indicate the coverage area limits of the satellites shown along the equator. (Map continues on p. 578.) *(Courtesy of Overseas Telecommunications Commission, Australia.)*

However, increased bandwidths mean reduced repeater gains and increased cable losses, and so repeaters must be placed closer together. For long cable segments, this results in unduly high dc voltages required at the two ends to accommodate the 70-V drop per vacuum tube repeater. Thus the third- and subsequent-generation cables have used transistor repeaters exclusively, with voltage drops of only 12 V per repeater. The *TASMAN* cable (1974, 480 circuits from Australia to New Zealand) and



(Map continued from p. 577.)

the *TAT-5* cable (1970, 845 circuits from the United States to Spain), both shown on Figure 15-6, are typical examples of third-generation cables.

CANTAT 2 is typical of fourth-generation cables. It was laid in 1974 and provides 1840 circuits between Canada and Great Britain. Figure 15-7 shows the cable, both lightweight and armored, used in *CANTAT 2*, and a repeater from the system is shown in Figure 15-8. The repeaters are, of course, all solid-state, with separations of about 11 km in practice. This is a very successful design, first used in 1971 for a cable between Spain and the Canary Islands and subsequently employed in the Mediterra-

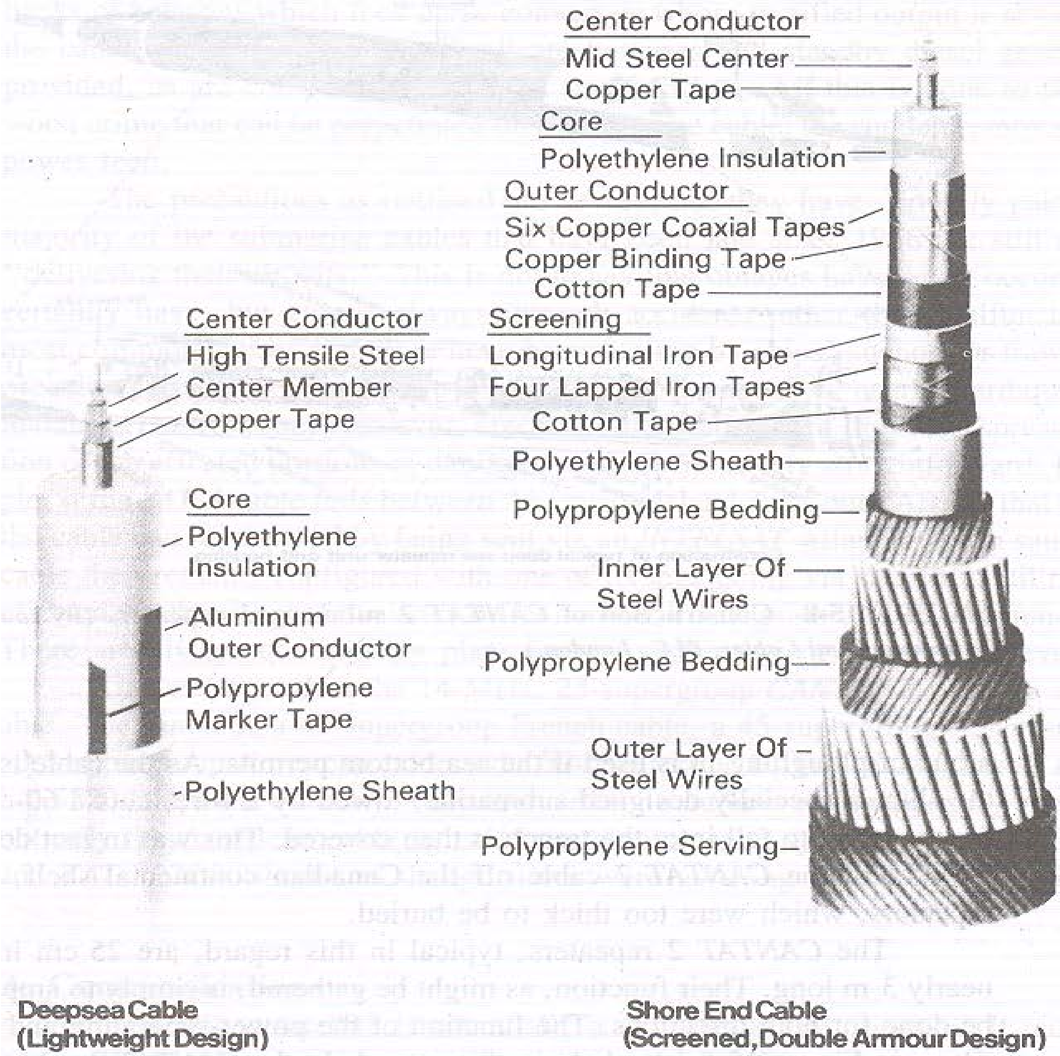
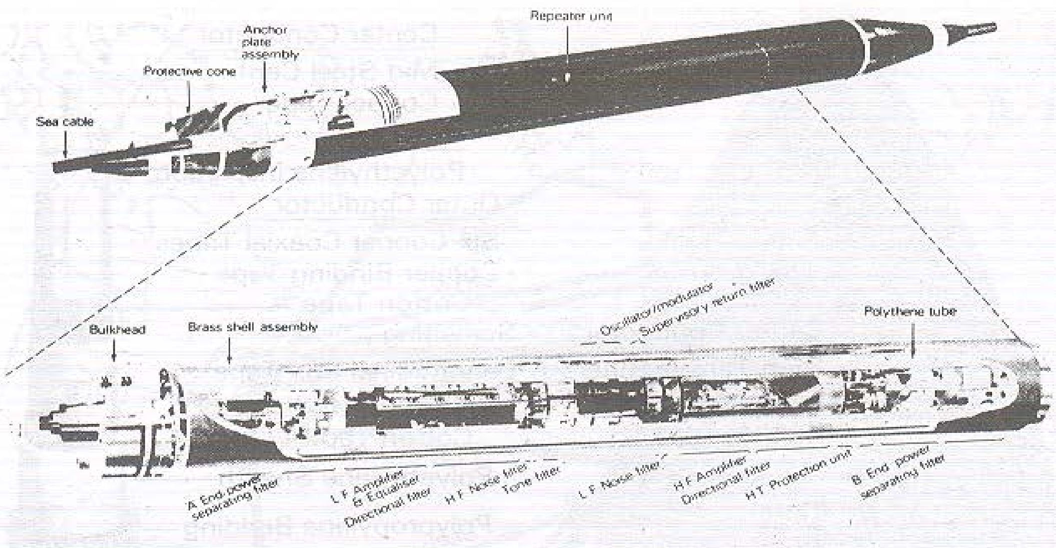


FIGURE 15-7 Display of submarine cable used in *CANTAT 2;* the overall diameter of each cable is 44.5 mm. ***(Courtesy of Standard Telephones and Cables, PLC, London.)***

nean (several cables), the Atlantic (*COLUMBUS*, southern segment of *ATLANTIS*, in 1982) and the Pacific (*ANZCAN*, 1984), as well as several shorter cables in Europe and southeast Asia. All these are shown in Figure 15-6, except the many Mediterranean cables, which are omitted for lack of space.

Cable is laid by cableships operating from the two ends separately and sometimes simultaneously, moving at typical speeds of about 8 knots (about 15 km/h)—the final splice is thus the midocean one. Lightweight cable is used for most of the length, including all deep sea portions. Sometimes, where great depths are involved, the cable is laid with sea parachutes, to slow its descent and therefore the rate of temperature change undergone by the cable and electronic components. The repeaters are rigid, and ingenious methods of bypassing shipboard sheaves have been developed. Armored cable is used for the shore ends as protection against trawlers, ships' anchors and tidal movements. In well-known fishing areas, particularly if they are shallow, the tech-



Construction of typical deep sea repeater unit and housing

FIGURE 15-8 Construction of *CANTAT 2* submerged repeater. ***(By courtesy of Standard Telephones and Cables, PLC, London.)***

nique of ploughing-in is used if the sea bottom permits. As the cable is paid out from the ship, a specially designed submarine, towed by a wire, cuts a 60-cm-deep trench for the cable to fall into; the trench is then covered. This was in fact done for the first 220 km of the *CANTAT 2* cable off the Canadian continental shelf, except for the repeaters, which were too thick to be buried.

The *CANTAT 2* repeaters, typical in this regard, are 25 cm in diameter and nearly 3 m long. Their function, as might be gathered, is simply to amplify. This must be done for both directions. The function of the power-separating and the directional filters in Figure 15-8 is to help in this regard. In the *CANTAT 2* cable, the 23 supergroups are accommodated in the frequency band 312 to 6012 kHz in one direction, and 8000 to 13,700 kHz in the other direction. Inquisitive students who perform the appropriate calculations will realize that the above figures correspond to 3-kHz circuits and 80-circuit supergroups. It will be recalled that submarine cables are expensive, and 3-kHz voice circuits are often used. Supervisory tones and cable and system pilots are assigned various portions of the nearly 14-MHz spectrum, leaving 940 kHz for separation between the two directions; this is quite adequate in practice.

Reliability is the keynote of a submarine cable project. This point cannot be stressed enough. Whether it is the cable itself, repeaters, equalizers, cable station terminal equipment or power feed equipment, everything is engineered for a long life and slight, predictable aging. All cable and repeater welding is done by specially trained personnel, and all welds are checked by x-ray. The electronic components are assembled and tested under dustfree, laboratory conditions. All the components are used at well below their maximum ratings, and key components are duplicated. The performance of the system is monitored by the cableship during laying, and from the terminals for the rest of the cable life. Power feed arrangements are complex, with main supplies rectified and regulated at the terminals and then used to float-charge the banks of batteries which feed dc/ac converters whose rectified output is actually fed to the cable at constant current. Duplicate batteries and standby diesel generators are provided, as are complicated interlock arrangements. All this is done to prevent the worst crime that can be perpetrated on a submarine cable: the sudden removal of the dc power feed.

The precautions as outlined are severe, but they have certainly paid off. The majority of the submarine cables that have been laid since 1956 are still operating, "delivering their circuits." This is not to say that outages have never occurred. They certainly have, but almost always through accidents rather than malfunctions. The most common causes of failure have been fouling by ships' anchors or trawlers, with occasional turbidity currents (undersea avalanches caused by nearby earthquakes) also making a contribution. However, since satellite stations are now widespread, restoration of the affected portions of damaged cables is relatively straightforward. For example, if the *SAT-1* cable fails between Ascension Island and South Africa, that portion of the cable can be restored by being sent via an *INTELSAT* Atlantic Ocean satellite. The cable then remains configured with one of its legs going via satellite until repairs are effected, so that most of the users suffer a minor interruption instead of a major outage. There are always contingency plans for the restoration of each leg of every cable.

Cables larger than the 14-MHz, 23-supergroup *CANTAT 2* type are also available. They include a 43-supergroup French cable, a 45-supergroup Japanese cable, a 50.8-supergroup American cable and a 69-supergroup British cable (capable of providing 5520 telephone circuits). They are used for a number of high-density applications, but only the American cable is used in intercontinental systems, for example, *TAT-6* and *TAT-7*. It is almost as though users were awaiting the advent of fiber optics.

TEXT-BOOK

I. Electronic Communication Systems, George Kennedy and Bernard Davis, Fourth Edition (1999), Tata McGraw Hill Publishing Company Ltd