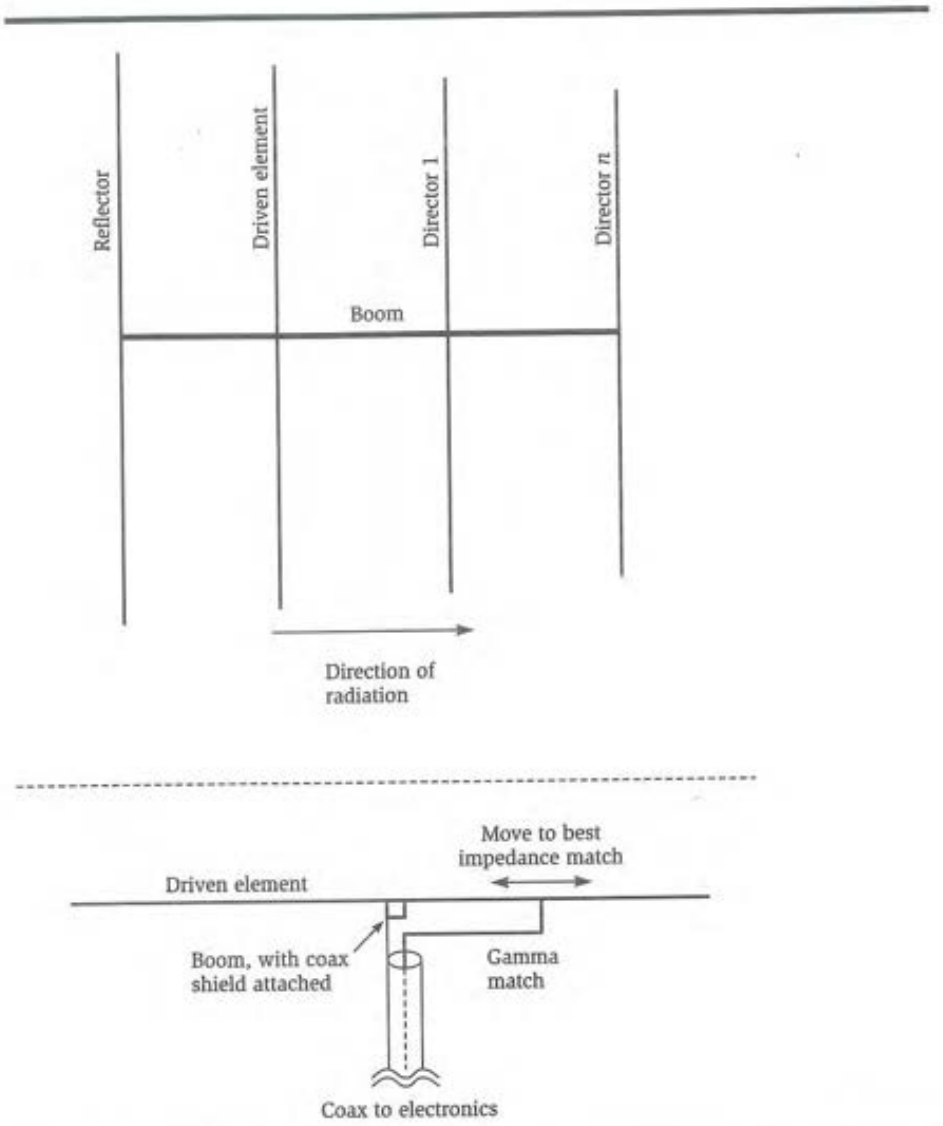


Figure 7.4 Yagi antenna.

higher, especially as the frequency increases. All else being equal, the more elements, the larger the structure, and the more directional and higher gain the antenna.

The driven element is essentially a dipole though the length is somewhat reduced owing to proximity effects of the other elements. A balun may be used

to feed a Yagi, but other feed methods are more common. Shown is a gamma match, which is used frequently, especially at lower frequencies. The coax shield is connected to the boom, which is electrically attached to the center of the driven element. The center conductor is extended along one side of the driven element to the point at which the impedance is correctly matched to that of the coax.

A two-element Yagi typically has a driven element and one reflector, located behind the driven element, where *front* is the direction of signal radiation or reception. The reflector is slightly longer than the driven element. As more elements are added, they are added in front of the driven element and are progressively shorter. A Yagi allows a number of degrees of design freedom in the spacing between elements and the length of each. Because of this, Yagi characteristics may be modified, resulting in trade-offs between gain, bandwidth, and pattern.[5]

Log Periodic Antennas

Figure 7.5[6] illustrates the other type of off-air receiving antenna commonly used in cable television, particularly at lower frequencies. This is the log periodic antenna. It can be a very broadband antenna with good pattern characteristics. The penalty is that it can get rather large. Each element of the log periodic

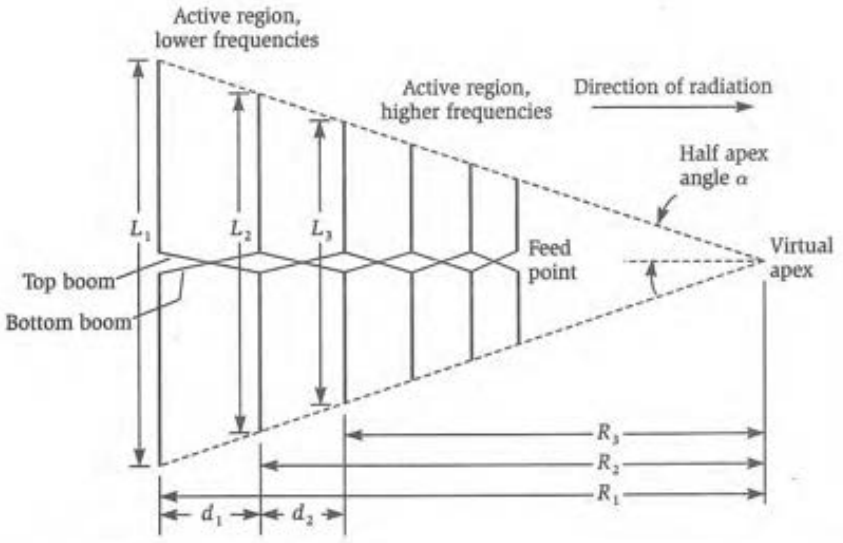


Figure 7.5 Log periodic antenna.

antenna is driven. It is driven from the front with a balanced feed. The length of each element, and the spacing between them, have a ratio relationship, τ, called the *scaling constant*:

$$L_2 = \tau L_1 \quad R_2 = \tau R_1 \quad d_2 = \tau d_1 \quad L_3 = \tau L_2 = \tau^2 L_1$$

and so on. A second constant, the spacing constant σ, further describes the antenna:

$$\sigma = \frac{d_n}{2L_n}$$

The log periodic antenna exhibits an active region for every frequency at which it operates. At lower frequencies, the active region is toward the rear of the antenna. It moves progressively forward as the frequency increases.

The log periodic antenna used for television work is usually constructed using two conducting booms, one above the other. The booms are fed at the front with a balanced signal, which propagates along the two booms to the active region. In order to feed that antenna with an unbalanced coax line, the coax is routed through one of the two booms, which then forms an infinite balun to convert from unbalanced to balanced.

Figure 7.6 illustrates a typical pattern for either a Yagi or log periodic antenna. As with all directional arrays, these antennas exhibit maximum radiation (sensitivity) in one direction. The antenna may be characterized by its *beamwidth*, the angle θ between the directions in which the power drops to one-half (−3 dB) of that in the direction of maximum radiation, or the *main beam* (sometimes called the main lobe). Ideally, the antenna would have no response in other directions, but as a practical matter, any antenna will have some response in other directions. Those responses are called *side lobes*, and an antenna will usually have several side lobes. The number of decibels that any side lobe is below the main beam is a figure of merit for the antenna. The ratio of sensitivity in the main beam to that from behind the antenna is called the front-to-back ratio. This is usually specified separately from the side lobe level. Between the side lobes, there is often an angle of minimum radiation, called a *null*.

Interference and Ghost Reduction

Directional antennas are selected for two characteristics in cable television receive applications: gain and directivity. The antennas exhibit gain, which increases the signal level at which a distant station is received, possibly improving the carrier-to-noise ratio. The directional characteristics of the antenna can be useful for reducing co-channel and ghosting interference. Co-channel interference results when a desired station is being interfered with by a more distant station operating on the same frequency. In cases of potential interference, the

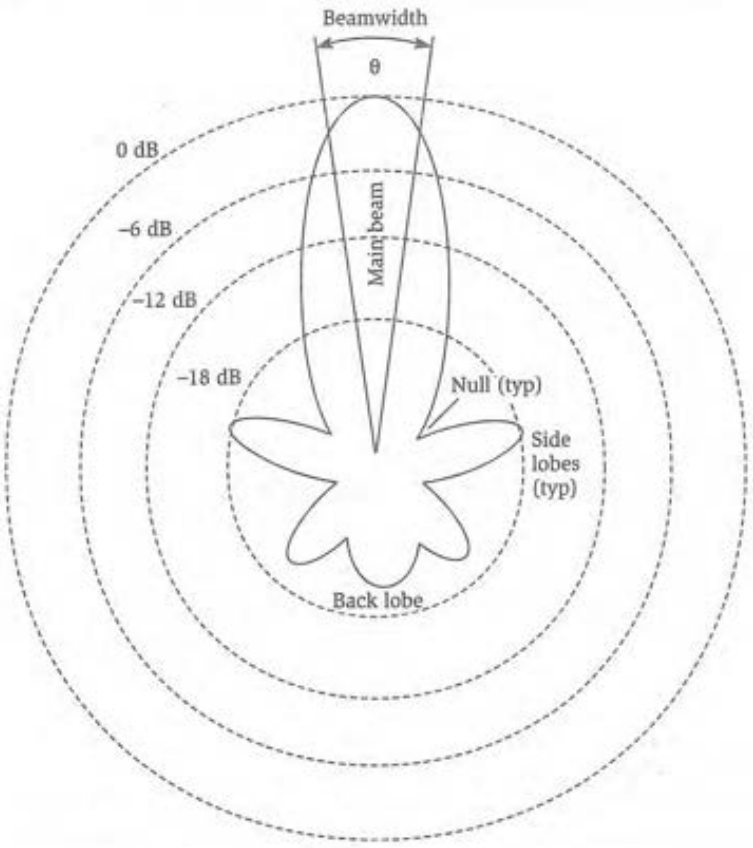


Figure 7.6 Typical antenna pattern for Yagi or log periodic antenna.

FCC assigns offsets to the stations involved, such that their picture carriers will be nominally offset from each other by either 10 or 20 kHz. It is known that such an offset will reduce the visibility of a beat. The visibility of the beat is also related to the level of the interfering signal relative to the level of the desired signal. It is often possible to reduce a co-channel beat by using an antenna with better directional characteristics so as to reject the interfering signal. Alternatively, it may be possible to rotate the receive antenna slightly so as to place the interfering signal in a null in the antenna pattern, as illustrated in Figure 7.6.

The interference may not be due to another station, but it may be due to the station being received. This can occur when a signal is reflected from some object such as a building, water tower, or mountain, and arrives at the antenna after the main signal. By the time the reflected signal arrives at the antenna, the

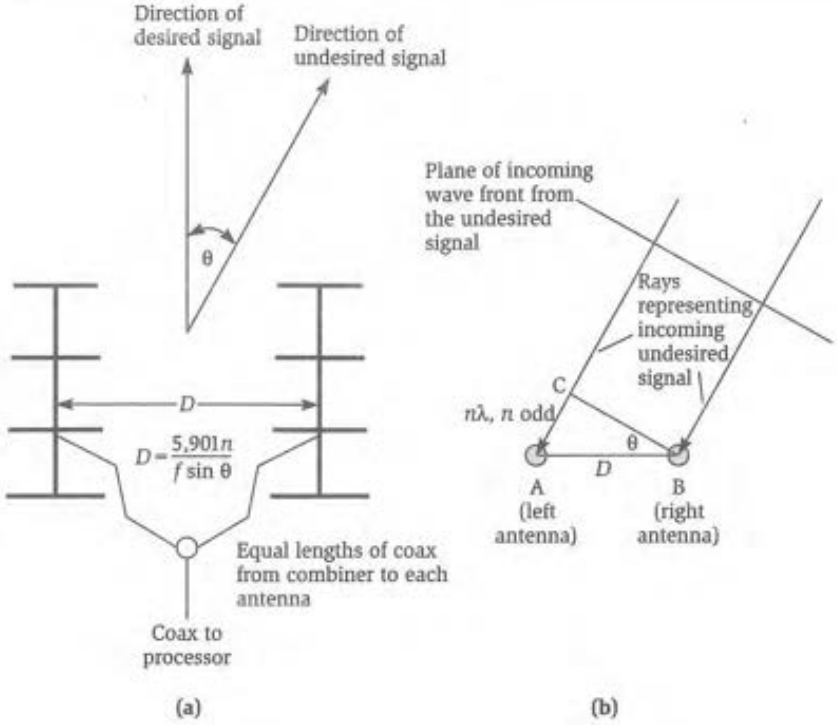


Figure 7.7 Use of multiple antennas to reject interference. (a) Antenna spacing. (b) Geometry showing why phase cancellation takes place.

electron beam in the picture tube has moved, so the reflected signal is seen as a ghost. Again, it may be possible to move the antenna slightly to put the reflected signal in a null, or it may be possible to use a more directive antenna to reject the interference.

Figure 7.7 illustrates another technique that is used to reject interference when that interference is coming from a significant angle with respect to the desired signal. Two antennas are mounted horizontally, spaced by a distance D, which varies with the angle of arrival of the interfering signal. As shown in Figure 7.7(a), the two antennas are fed with equal lengths of identical coax, which supply the signals to a combiner. The lengths of the two pieces of coax are not critical, so long as they are exactly the same.

Figure 7.7(b) shows how the cancellation takes place. The phase of the signal coming from the desired source is the same at the two antennas, so at the combiner, the signals from each antenna add. However, distance D is selected such that the signal from the undesired source will phase cancel when the signals from the two antennas are combined. This condition is satisfied when D is chosen

such that the undesired signal must travel one-half wavelength farther to the left antenna than to the right. The angle between the desired and undesired is θ, shown in right triangle ABC. The problem then reduces to defining D such that, given θ, side AC is an odd multiple of a wavelength at the frequency of reception (use the picture carrier frequency). This will cause the interfering signal to cancel. Substituting known relations, we can derive the formula for D:[7]

$$D = \frac{5{,}901n}{f \sin \theta} \tag{7.3}$$

where

D = distance between antennas (inches)
f = frequency (MHz)
θ = angle between desired and undesired signals
n = an odd integer

Another technique that has been used in the past to cancel an undesired signal is a bucking array: a second antenna is pointed in the direction of the undesired signal. Its received amplitude and phase are adjusted such that it cancels the undesired signal being received from the main antenna. This technique was once popular, but seems to be less so now owing to the critical need for matching the level and phase between two (usually) unlike antennas.

Yet another method that has been used for ghost cancellation requires the cooperation of the television station. The station radiates a circularly polarized signal, and the cable system receives it on a circularly polarized receiving antenna. It may be shown that upon reflection, the polarization of a circularly polarized wave changes the sense of polarization, from left hand to right hand or vice versa. A circularly polarized receiving antenna can then reject the undesired polarity.

In the 1980s, the FCC permitted television stations to transmit circularly polarized signals so as to aid in rejecting ghosting. A few stations have elected to do so, but more have not. One problem is that a new and expensive antenna is required. The antenna must be installed and fed carefully to ensure truly circular polarization. Another problem is that, if all else stays the same when a station goes to circular polarization, the radiated signal in the horizontal plane traditionally used for television transmission is halved. This means that people receiving the signal with a horizontally polarized antenna will suffer a 3-dB loss of signal level. To recover this would require another 3-dB gain in the antenna or a doubling of transmitter power. Neither is a simple or inexpensive proposition.

Offsetting these objections somewhat is the observation that a circularly polarized signal is somewhat easier to receive on rabbit ear antennas, which respond to both horizontal and vertical components. This benefit may be obtained even with elliptical polarization, which is circular polarization in which

the phase or amplitude of the vertical and horizontal components is not controlled carefully.

Yet another method that has been used to reduce reflection is the use of a ghost-canceling system. A training signal is transmitted in the vertical blanking interval of the television signal (line 19 has been designated by the FCC). This training signal is subjected to the same ghosting as is the video. A filter in the receiver is adjusted until its characteristics return the training signal to its known transmitted form. The same filter is applied to the video, removing the ghost. The system has been shown to be effective in reducing ghosts. Cooperation of the station is required to insert the ghost-canceling reference signal. This technique is simpler than going to circular polarization. Although it is being used in some places, the public has never bought a lot of television sets that have ghost-canceling circuits built in. Ghost-canceling hardware is available to the cable industry for headend use.

7.2.2 Propagation of Off-Air Signals

The best receiving antenna is of no use if signals are not present for it to receive. Signals propagate from the transmitting antenna to the receiving antenna in ways that are not totally predictable though some characteristics are sufficiently constant that we can make some statements about the propagation. Signals are reflected from natural and man-made structures so that the resulting signal at the receiving antenna comprises direct and reflected *rays*. (A ray is a path taken by a signal from one point to another.) Direct and reflected rays may add constructively (in phase) or destructively (out of phase) or in between (some arbitrary phase angle). If sufficient delay is present between the signals, a ghost is produced. If the differential delay between paths is short, a distinct ghost may not result though the video waveform will be distorted.

Free Space Propagation

In the far field, the power varies inversely with the square of the distance from the source. This may be intuitively comprehended by remembering that the power from the source is spreading over two dimensions perpendicular to the direction of propagation. If so, then the area over which the power is spread quadruples every time the distance doubles. The same power is spread over four times the area, so the power density is quartered.

Signal strength is normally measured in microvolts per meter, and the units are often expressed in decibel terms, where the unit of signal level is decibels with respect to 1 microvolt per meter (dBμV/m). A particularly useful expression describing propagation of a signal in free space is[8]

$$E = 105 + P_t + G - 20 \log d \qquad (7.4)$$

where

E = field strength (dBμV)/m)
P_t = power transmitted (dB above 1 kW, dBk)
G = gain of transmitting antenna (over isotropic) in the direction of concern
d = distance from transmitter (km)

Though useful in satellite work, free space propagation is more of a reference point in discussing terrestrial propagation. Reflections from the ground and other natural obstacles, and from man-made objects, will significantly modify the equation. At lower frequencies, reflections from the ionosphere will change the observed propagation. These ionospheric effects can sometimes extend well into the VHF TV band though they tend to show the greatest effect at lower frequencies.

Effects of Antenna Height

With the assumption that ground approximates a large flat surface, an approximation to field strength as a function of antenna height is given by Hall as[9]

$$E = 2E_d \sin\left(\frac{2\pi h_1 h_2}{\lambda d}\right) \tag{7.5}$$

where

h_1 = height of transmitting antenna
h_2 = height of receiving antenna
λ = wavelength
d = distance between transmitter and receiver

To get a feel for how receiving antenna height affects received signal level, consider Figure 7.8, which plots the relative strength of a signal from Hall's equation, for channels 2 and 40 (off-air), for a 1,500-foot transmitting antenna. At channel 2, the receive antenna is not high enough, even at 100 feet, to permit destructive interference between the direct ray and one reflected from the ground. At channel 40 (off-air, not cable), you can see clearly the effects of ground reflection.

We caution you against taking this illustration literally. It assumes a flat, perfectly conducting earth with no extraneous signal reflections. None of these conditions applies, except possibly to paths across open water. It does illustrate, however, that though increasing receive antenna height can be useful, there are times when you might expect a lower signal level as the antenna height is raised.

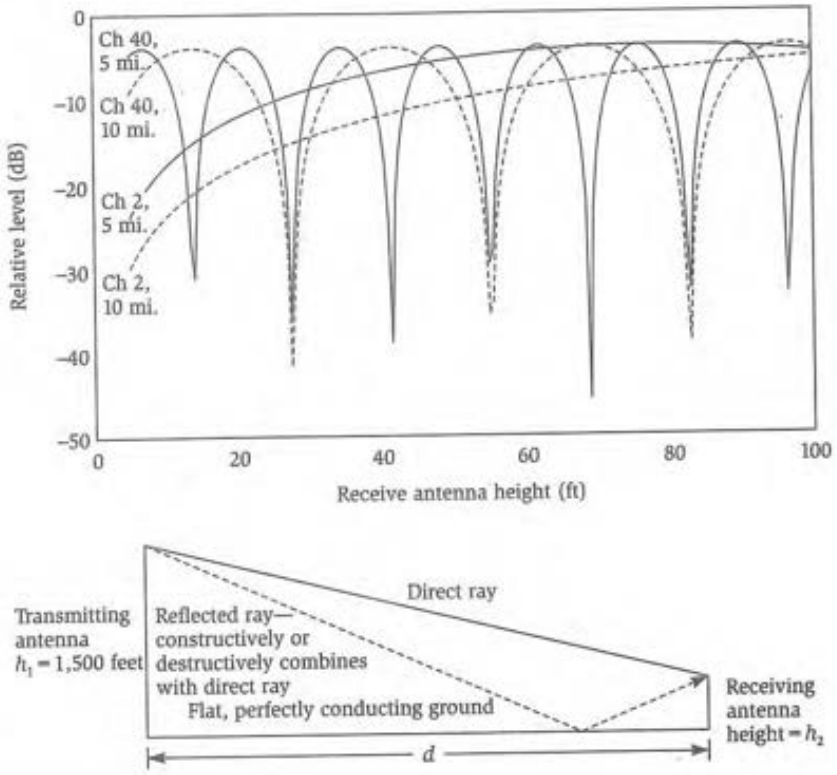


Figure 7.8 Estimated relative field strength versus receiving antenna height.

FCC Methodology for Prediction of Signal Strength

At television frequencies, radio waves are subject to shielding by the curvature of the earth though the effective radius of the earth for propagation purposes is larger than the physical radius. (See Chapter 14 for more information relating to the effective radius of the earth.) Also, one must account for reflection from the earth, which adds constructively or destructively with a direct signal. Thus, height of the transmitting and receiving antennas must be taken into account when predicting the strength of a received signal. For purposes of computing television signal strengths, the FCC uses a standard receiving antenna height of 30 feet (9 meters). The technical regulations for the broadcast industry, 47CFR section 73.699, feature graphs (Figures 9 through 11 of those rules) to be used in predicting the strength of a signal at various distances from a television station. All curves are based on a receive antenna height of 9 meters. For each frequency band, channels 2 through 6, and FM channels 7 through 13 and channels 14

through 69, two sets of curves are supplied. The first, F(50,50), is used to predict the field strength that is exceeded at 50% of potential receiver locations for at least 50% of the time. The second, F(50,10), is used to predict field strength exceeded at 50% of the potential receiver locations for at least 10% of the time. The curves are based on theory coupled with practical measurements. At higher frequencies, there may be modifications to the curves in the future.

Once a television station is on the air, it must make field strength measurements using procedures prescribed by the FCC to determine actual coverage and interference potential. These measurements are made using calibrated antennas and field strength meters not unlike the better selective voltmeters used in the cable industry. Note that a selective voltmeter by itself is not capable of measuring field strength. The old name used by the cable industry, field strength meter, is a misnomer. The selective voltmeter is not a field strength meter until it is combined with an antenna having known gain-versus-frequency characteristics, and certain computations are made (Equation (7.2)).

The FCC sets standards for the minimum field strength for two grades of service. Grade A service is generally expected to be provided to the principal community served by a television station. Grade B service is a lower field strength, but one that should result in satisfactory pictures if good outside antennas are used. Of course, further improved service may be obtained by use of better antennas and lower noise receiving systems.

To aid in planning of cable reception, computer programs are available that use the FCC's database of station coverage to predict the received signal level at any geographic coordinates. The programs are useful, but in the final analysis, the only way to know with certainty what signals exist where is to go look for them. The old cable industry axiom is still true: signals are where you find them.

7.3 Satellite Reception

Communications satellites used in the cable television service are in geosynchronous orbit, about 23,500 miles above the equator. At this altitude and above the equator, the orbital period of a satellite is 24 hours, equal to the time the earth requires to rotate one complete revolution. If a satellite is launched to this orbit, moving in the same easterly direction as the earth, it appears to stay stationary over one point on the equator. Because of slight irregularities in the earth's shape, and because insertion into an exact orbit is impossible, it is necessary to track the position of the satellite, and command small corrections as it drifts out of its assigned slot. Tracking stations operated by or for the satellite owner perform this task. Fuel stored on the satellite is consumed making these small corrections. The life of the satellite is often based on the amount of fuel it can carry at liftoff. When the fuel is used up, orbital correction becomes impossible, and the satellite gradually drifts off station, becoming useless for reception using fixed antennas.

In this section, we discuss the issues that go into installation and operation of a receive-only earth station for video. Design equations are presented that will estimate the requirements for, and performance of, an earth station. This is not intended as in-depth coverage of the material.

The two popular bands for satellite service today are C band (downlink frequencies 3,700 to 4,300 MHz) and Ku band (downlink 11.7 to 12.2 GHz). The satellites themselves consist of a number of channels called *transponders*. Each transponder is a complete subsystem so that if one fails the others are not affected. A transponder consists of an input amplifier, bandpass filter, frequency translator, and output amplifier. It is usually a single conversion device. Most satellites today have 24 transponders on board though the number may vary.

Normally, one transponder will handle one analog television signal, which is frequency modulated onto a carrier. On occasion, two signals have shared a single transponder, in what is known as a *half transponder* mode. In the more common full transponder mode, the transponder is normally operated somewhat into saturation, meaning that it is slightly clipping the signal. This doesn't bother a frequency modulated signal and ensures maximum efficiency from the transponder. Each transponder is normally 40 MHz wide, including guard band between channels, with the actual signal occupying about 36 MHz.

In order to make the best use of the frequency spectrum, frequency reuse is practiced, in which signals of two polarities are transmitted. For example, some signals can be transmitted using vertical polarization, and others can be transmitted using horizontal polarization. The polarities are said to be *orthogonal*. The dictionary definition of orthogonal is that it has to do with a right angle, or something rectangular. An antenna that is horizontally polarized is theoretically insensitive to vertically polarized signals, and vice versa.

In U.S. satellite practice, the polarizations are linear, nominally vertical and horizontal, though the axes may be rotated by the time the signal travels from satellite to earth. In many international satellite applications, the orthogonal polarizations are right- and left-hand circular polarization. Either set may be used successfully.

Satellites are assigned orbital slots by the FCC for U.S. use and by the appropriate authority in other countries. Slots are assigned to each country by common consent under the auspices of the International Telecommunications Union. The orbital slots assigned to the United States are rather crowded, and the FCC has assigned orbital slots in two-degree increments. This places burdens on antennas to discriminate against another orbital slot two degrees away from the one to which it is aimed.

7.3.1 Antennas for Reception of Satellite Signals

Many different antenna types are used for reception of satellite signals. This discussion is restricted to those commonly used in cable television work today. Variations are common between different manufacturers. Almost all antennas

are of the parabolic ("dish") type, in which a relatively large parabolic surface is used to collect signal power. That signal is focused on the actual receive antenna, known as the *feed*. At the feed, the signal is amplified by a low-noise amplifier (LNA) and usually block converted to a lower-frequency band. The device used to amplify and convert to a lower band has come to be known as a low-noise block converter (LNB). The most common band to which signals are converted is 950–1,450 MHz though other frequency plans have been used. The electronic hardware is described in more detail in Chapter 8.

The Parabola

Almost all satellite antennas are based on the parabola, whose geometry is shown in Figure 7.9. (Of course, practical antennas are realized in three dimensions rather than the two shown here for clarity. You can envision the third dimension by rotating the figure around the x axis.) The interesting characteristic of the parabola is that parallel rays approaching the parabola from the right, as illustrated, are all reflected to one point, called the *focus* or *focal point*, of the parabola. If the feed is placed at the focus, it will capture rays striking any point on the parabola, so long as those rays are coming from the direction shown. That direction is known as the *boresight* of the antenna.

The parabola as defined by the equation shown in Figure 7.9 extends to infinity along the x axis, with the slope of the parabola asymptotically approaching the horizontal. Real antennas are limited to some useful dimensions. The larger the area of the dish, the more power the antenna can collect. Antennas are characterized by their size, which is defined by the length of a line from one side of the dish to the other.

An important quality of a real dish is the degree to which it conforms to the parabolic shape. Real dishes are constructed out of some sort of metallic surface, often of pieces bolted together in the field. A figure of merit for the dish is the RMS error in the conformity to the parabolic shape. The locations of many positions on the parabola are measured and stated as an offset, in the x axis for reflector orientation as shown in Figure 7.9, from the ideal location as defined by the parabola formula. The RMS error is then computed as follows:

$$\text{error(RMS)} = \sqrt{\frac{(x_1' - x_1)^2 + (x_2' - x_2)^2 + \ldots + (x_n' - x_n)^2}{n}} \tag{7.6}$$

where

x_i' = actual position
x_i = theoretical position
n = number of points measured

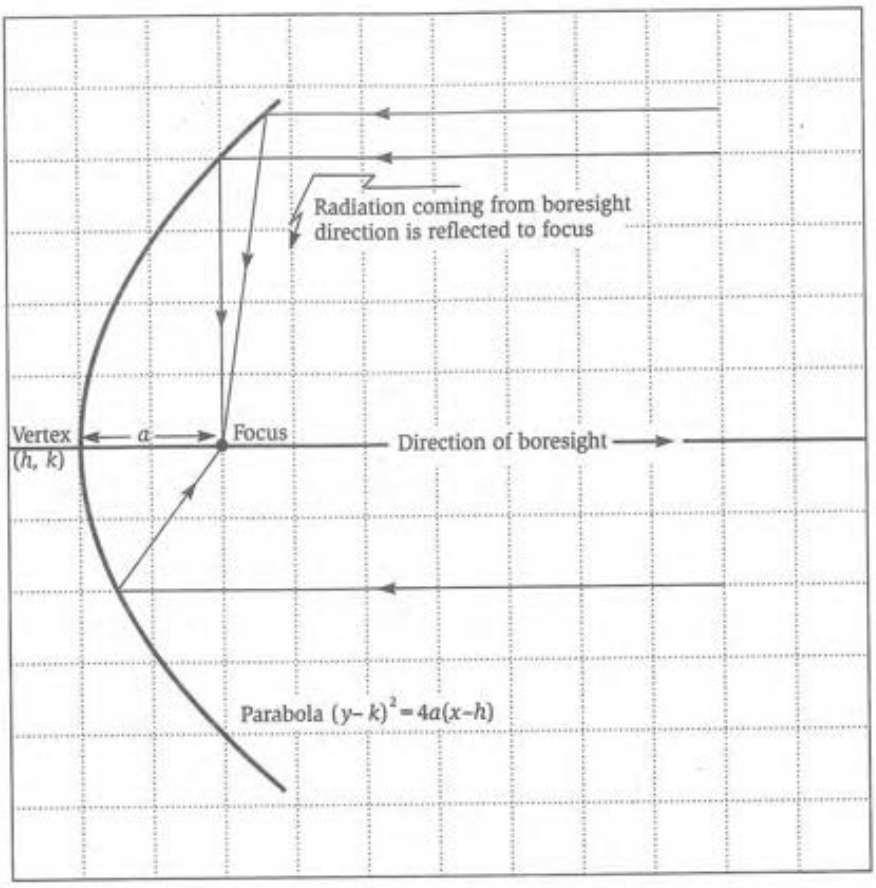


Figure 7.9 Parabolic surface.

The error may also be measured as offset positions normal to the dish surface.[10]

As the operating frequency of the dish increases, the surface errors become more important. Thus, a dish that has low enough surface tolerances to be useful at C band (3.7–4.2 GHz) may not provide satisfactory performance at Ku band (11.7–12.2 GHz). Surface errors reduce the gain of the antenna and spread out the pattern: power radiated from the feed and striking an element of the reflector not positioned correctly will be reflected in the wrong direction. The effect is distortion of the pattern. Since some of the power is going in the wrong direction, less is available in the correct direction, so the gain is lower.

Prime Focus Feed

The most basic satellite antenna type is the prime focus feed, which places the feed at the focus of the parabola, as in Figure 7.10. Support spars hold the feed in position. The feed itself and the support spars offer blockage to some of the power coming from the satellite and also reflect power coming from other directions, somewhat compromising the side lobe performance of the antenna. Nonetheless, the prime focus feed is popular until the dish size becomes quite small, where the blockage and reflections are more serious.

Note that the pattern of the feed is critical for maximizing the gain and the directivity of the antenna. The feed pattern should ideally illuminate the entire surface area of the dish evenly, as shown in the dotted pattern. If the feed pattern is too narrow, the outer portions of the dish are not illuminated, so they cannot

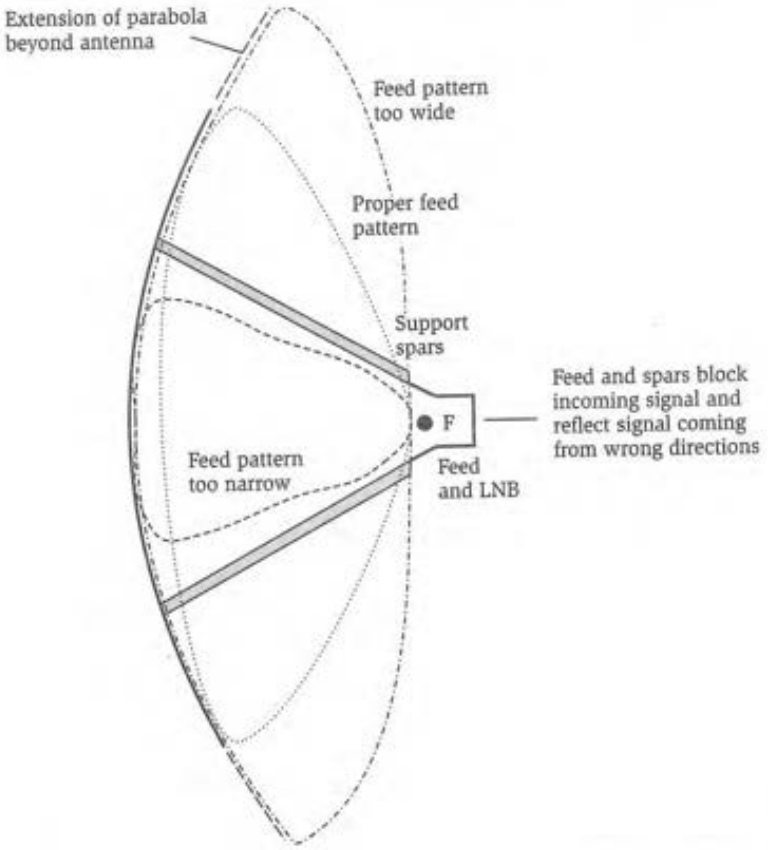


Figure 7.10 Prime focus feed.

aid in focusing a transmitted signal. Similarly, a received signal will not be captured if it strikes the dish outside the feed pattern because the power will be rejected by the feed when it reaches the focal point. If the feed pattern is too wide, some of the power radiated from a transmitting antenna will miss the dish surface and go off in the wrong direction, spoiling the pattern of the antenna. Received power will not be collected efficiently because the feed antenna gain is lower than optimum in the direction of the dish surface.

Matching the feed pattern to the dish is an art in itself. Of course, no feed is perfect, but the closer you can come to the ideal, the better the antenna gain and pattern. Feed antennas include dipoles above ground planes, Yagi or log periodic structures, and microwave feed horns.

Prime focus feeds are relatively simple to build and align. A disadvantage is that the feed, which contains the active electronics of the LNB, is relatively inaccessible, being held several feet in front of a parabolic surface that is facing up at some angle. Some would argue that the feed in this location is subject to relatively more damage from electromagnetic pulses that occur when lightning strikes.

Cassegrain Reflector

Another very common type of antenna is the Cassegrain reflector derived from the telescope design by William Cassegrain. It is the same basic antenna as the prime focus feed. The difference is that a second reflector is placed between the vertex and the focus of the parabolic reflector. The most common type of Cassegrain system is illustrated in Figure 7.11. A hyperbolic subreflector is used to reflect signals from the main reflector to the feed. The feed is located either at or in front of the vertex, depending on the design of the system. The hyperbolic subreflector is located such that its first focus coincides with F1, the focus of the parabolic main reflector. The second focus of the hyperbola is the location of the feed. (The second focus is from the solution of the hyperbolic equation defining the subreflector.)

The Cassegrain reflector places the feed electronics (usually a low-noise block converter) in a position where it is considered to be more protected and easier to access. With a Cassegrain configuration, it is not necessary to route the signals along a spar, leading to somewhat lower cable loss. On the other hand, this is not especially critical with the block conversion systems commonly used in cable television. The subreflector must be supported by spars, and the subreflector and spars create blockage and pattern degradation, as in the case of the prime focus feed. The Cassegrain system is somewhat more difficult to align since the installer must align not only the feed but the subreflector as well.

In general, you would choose a prime focus feed where low side lobes are important and a Cassegrain where gain for a given size antenna is most important. This latter statement is true for antennas whose diameters are large compared with the wavelength of the signal being received (on the order of 40 wavelengths); smaller antennas usually suffer too much diffraction on the edges

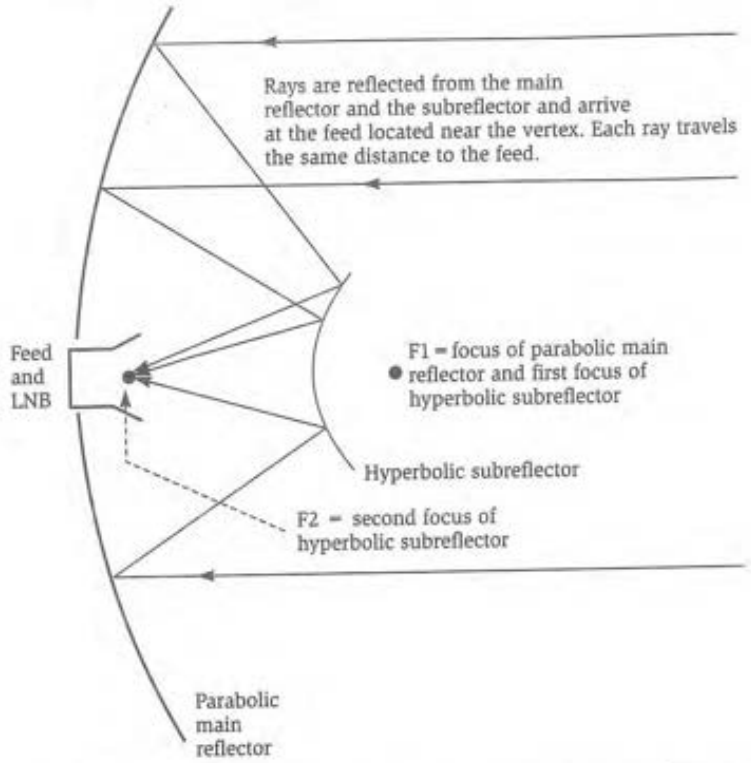


Figure 7.11 Cassegrain reflector.

of the subreflector to make the Cassegrain antenna useful. Gain is enhanced in Cassegrain antennas because the phase of a plane wave impinging on the antenna is preserved: each ray of the signal travels the same distance to the feed. This is not true of prime focus systems, but their generally smaller blockage of the primary reflector results in a better pattern.[11]

Offset Feeds

As the size of the dish gets smaller, the blockage and reflection of the spars and feed or subreflector become more of a problem. The solution can be an offset feed. The offset feed is simply the top portion of the prime focus antenna, with the bottom portion cut off. Figure 7.12 illustrates an offset feed. The illustrated system uses exactly the top half of the parabola though some offset feeds make use of less of the parabola.

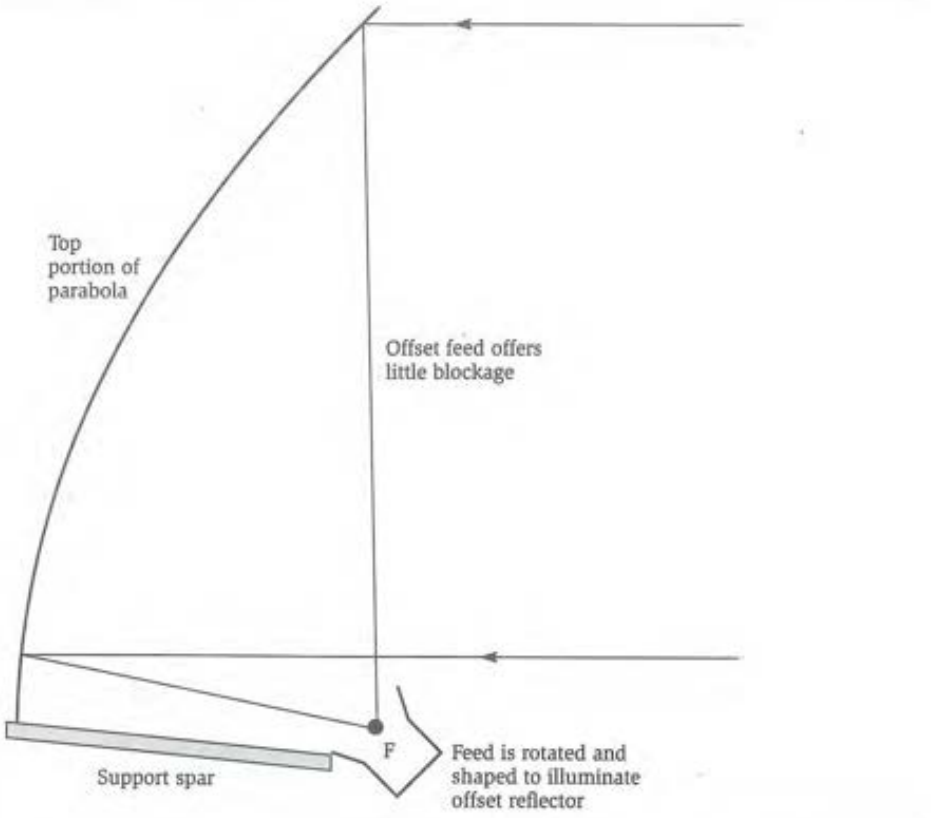


Figure 7.12 Offset feed.

The offset feed tends to be somewhat difficult to construct and align but in smaller dish sizes is a good choice because of the lack of blockage. There is some reflective pattern distortion from the spar and feed structure, but not as much as with symmetrical antennas. Because of the lack of blockage, this feed type is popular in very small aperture antennas, or very small aperture terminals (VSATs), and direct broadcast satellite (DBS) antennas.

Other Antenna Types

Several other variations of satellite antennas are used. Multiple feeds may be mounted on one reflector, which can be either a prime focus parabola or one of several shape variations. Each feed is able to look at a different satellite, so long as the satellites are close to each other in orbit. Compared with a single feed system of similar size, gain and side lobe performance are sacrificed, but the

technique can be used to reduce costs and the amount of land needed for multiple dishes. A particularly interesting multiple feed antenna is a torus antenna. The reflector has a circular contour in the scan plane and a parabolic contour in the orthogonal plane. The scan plane must be carefully aligned with the orbital arc plane. However, once aligned, that antenna is capable of quite good performance over a wide section of the geosynchronous arc.

Shaped reflectors may be used rather than parabolas or hyperbolas to optimize performance of the antenna in some respect. Of course, the feed must be designed to properly illuminate whatever shape is used.

Shrouds are sometimes placed around the circumference of an antenna to reduce the magnitude of side lobes. These antennas are sometimes called ultra directive antennas (UDAs).

Horn antennas are sometimes used. They tend to be big and difficult to align, and require awkward supporting structures. However, their performance in terms of gain and side lobe suppression can be impressive. Horn antennas are used in certain high-density terrestrial microwave applications. They look something like gigantic cones with a dome on the wide end. A feed in the apex of the cone launches a signal, which is reflected from a shaped reflector at the wide end.

7.3.2 Earth Station Geometry

In this section, we deal with the computation of pointing angles from an earth station to a satellite, and the signal loss in the downlink. Following this, we present computations to illustrate the performance you can expect from an earth station.

Many different antenna mounts are in use. Each has advantages and disadvantages, proponents and opponents. The description here concentrates on the elevation over azimuth coordinate set, a traditional coordinate set that is relatively easy to understand and work with. Antennas designed for other than communications with synchronous satellites often employ this set of axes. Antennas designed for communications with synchronous satellites may use different axes to permit the antenna to be moved between satellites by changing only one adjustment. Such systems tend to have proprietary elements, so you should consult the manufacturer for instructions on aiming the antenna.

The elevation-azimuth (or "el over az") coordinate system described here has its lower plane of rotation (azimuth) parallel to the local horizontal plane. A weather radar antenna, for example, rotates on the azimuth plane. The second axis, the elevation axis, is at right angles to the azimuth axis so that the antenna tilts up and down on the elevation axis.

Three pieces of information are of interest in defining how to point to the satellite: the azimuth (A) of the antenna, the elevation (E) of the antenna, and the path distance (D) to the satellite (also known as the *slant range*). These may be computed from knowledge of the satellite longitude (remember that the satellite latitude is zero, that being the only latitude for geosynchronous orbits). You must also know the earth station longitude and latitude.

Let the following definitions apply:

X = site latitude (degrees)
Y = site longitude (degrees)
Z = satellite longitude (degrees)
A = earth station azimuth angle (degrees)
E = earth station elevation angle (degrees)
D = slant range from site to satellite
C = difference between satellite and site longitudes

First compute the difference in longitude between the satellite and the earth station (the "site"):

$$C = Z - Y \tag{7.7}$$

Then the earth station azimuth angle is given by

$$A = 180^\circ + \tan^{-1}\left(\frac{\tan C}{\sin X}\right) \tag{7.8}$$

The elevation, E, of the antenna is given by

$$E = \tan^{-1}\left(\frac{\cos C \cos X - 0.15126}{\sqrt{\sin^2 C + \cos^2 C \sin^2 X}}\right) \tag{7.9}$$

By convention, longitude is positive west of the prime meridian and negative east. Latitude is positive north of the equator and negative south. The earth station azimuth angle is measured from true north, rotating clockwise as viewed from above the earth station. Use true north, not magnetic north, as the reference. The elevation angle is zero if the antenna is pointing horizontally and 90° if it is pointing vertically.[12] The constant 0.15126 is the ratio of the earth's radius to the satellite's orbital radius.

The slant range, or distance from the site to the satellite, is given by[13]

$$D = 26{,}199\frac{\cos C \cos X - 0.15126}{\sin E} \text{ statute miles}$$

or

$$D = 42{,}164\frac{\cos C \cos X - 0.15126}{\sin E} \text{ kilometers} \tag{7.10}$$

7.3.3 Earth Station Performance Computations

From the slant range and certain other information, we can compute the expected performance of an earth station. The signal level to be expected at the antenna may be computed from the published effective isotropic radiated power (EIRP) of the satellite and the path loss, for which the slant range is needed.

Concept of EIRP

EIRP is the power that would have to be supplied to the antenna *on the satellite* if that antenna radiated equally in all directions. If the antenna on the satellite radiated equally in all directions, a typical transmitter output power on the satellite would be somewhat above 3,000 watts. This is hardly practical, but then it is not necessary for the satellite to radiate signal in all directions. It is necessary only to radiate signal in the direction of that portion of the earth that is to receive the signal. By concentrating power in that direction, the power required on the satellite is reduced drastically.

For each satellite, the owner publishes a contour map showing the EIRP at locations within the satellite's *footprint*, or intended area of coverage. This map is consulted to determine the EIRP at the location of the receive station.

Path Loss

The next quantity required is the path loss, sometimes called spreading loss because it is caused by the signal from the satellite spreading as it travels through space. Path loss may be computed from

$$L = 20\log\left(\frac{4\pi Df}{c}\right)\text{dB} \tag{7.11}$$

where

D = slant range (from Equation (7.10))
f = frequency (Hz)
c = speed of light

Slant range and speed of light must be in the same distance units. From the equation, path loss increases with increasing frequency. However, gain of the receiving antenna increases as frequency increases, keeping the required antenna size the same. If the slant range is expressed in miles and the frequency in GHz, Equation (7.11) reduces to the form shown in Equation (14.12):

$$\text{path loss(dB)} = 96.6 + 20\log(f) + 20\log(D)$$

where

f = frequency (GHz)
D = distance (statute miles)

Station G/T

The final quantity that must be known is the G/T of the earth station. G/T (pronounced "G over T") expresses the gain of the station, divided by its noise temperature. Noise temperature is a particularly useful expression of the noise contributed by the earth station from all sources. It is not a physical temperature, but rather is the temperature that will produce the same equivalent noise in a resistor. If a resistor is connected to a power meter, the power meter will read a certain power, given by

$$P_n = kTB \tag{7.12}$$

where

P_n = noise (watts)
k = Boltzmann's constant = $1.38 \times 10^{-23} \frac{\text{Joules}}{^\circ\text{K}}$
T = temperature (°K)
B = bandwidth (Hz)

This noise power, P_n, is known as thermal noise, and is a basic law of physics: a resistor at a certain temperature will produce a certain amount of noise according to this equation. This doesn't have to be a physical resistor: it is often the terminals of an antenna, which looks electrically like a resistor. Real circuits, such as LNBs, add additional noise above and beyond the noise that would be predicted solely from the thermal noise. Noise temperature is a measure of the amount of excess noise, over and above thermal noise.

Other sources contribute to the noise temperature of a receive station. Space noise, which is produced by a number of objects in space, is a factor, but with small earth stations, it is usually a small factor. An exception is the sun, which is a very significant noise-producing object in space: when the receiving station, satellite, and sun are lined up, communication from the satellite is lost, overpowered by noise from the sun. Also, the earth has a rather high noise temperature. When the elevation of the antenna is less than perhaps 20° or 30°, the earth's noise temperature can be significant. Of course, the contribution from the earth depends on the antenna pattern in the direction of the earth.

In addition, the LNB produces excess noise over and above what you would expect from thermal noise. The excess noise can be translated to an equivalent noise temperature by solving Equation (7.12) for T and substituting the equivalent excess noise at the input of the LNB for P_n. An equivalent concept some what more familiar to cable television engineers is that of noise figure. Noise figure expresses the same thing that noise temperature expresses, but in satellite work, we often work with noise temperature because certain computations are simplified.

The G/T of a station is a function of the gain of the dish, the pattern of the dish if the elevation angle is low, the noise figure of the LNB, and any losses in the structure. Technically, losses after the LNB must also be taken into account. However, they tend to have only secondary effects on the G/T of the station unless the loss between the LNB and the receiver begins to approach the gain of the LNB. The manufacturer of the station can supply appropriate G/T values.

Higher G/T will provide better carrier-to-noise ratio, and hence better baseband signal-to-noise ratio on the video. G/T can be increased by going to a larger dish size, improving the efficiency of the dish, or using an LNB with a lower noise temperature. Some slight improvement may be obtained by reducing losses following the LNB.

Carrier-to-Noise Ratio of Received Signal

Combining the preceding concepts, the carrier-to-noise ratio of the receive station can be expressed as

$$\frac{C}{N_0} = EIRP - L + \frac{G}{T} + 228.6 \tag{7.13}$$

where

$\frac{C}{N_0}$ = carrier-to-noise ratio (dB) per hertz

$EIRP$ = satellite effective isotropic radiated power (dBW)

L = path loss (dB) from Equation (7.11)

$\frac{G}{T}$ = figure of merit for earth station (dB/°K)

All quantities are expressed in decibels. The carrier-to-noise ratio expressed is per hertz. As shown in Equation (7.12) for noise power, the total noise is proportional to the bandwidth over which the measurement is made. The measurement bandwidth is the *effective noise bandwidth* of the receiver. Noise bandwidth is that bandwidth a filter would have if it were perfectly rectangular and admitted the same noise as does the real filter.

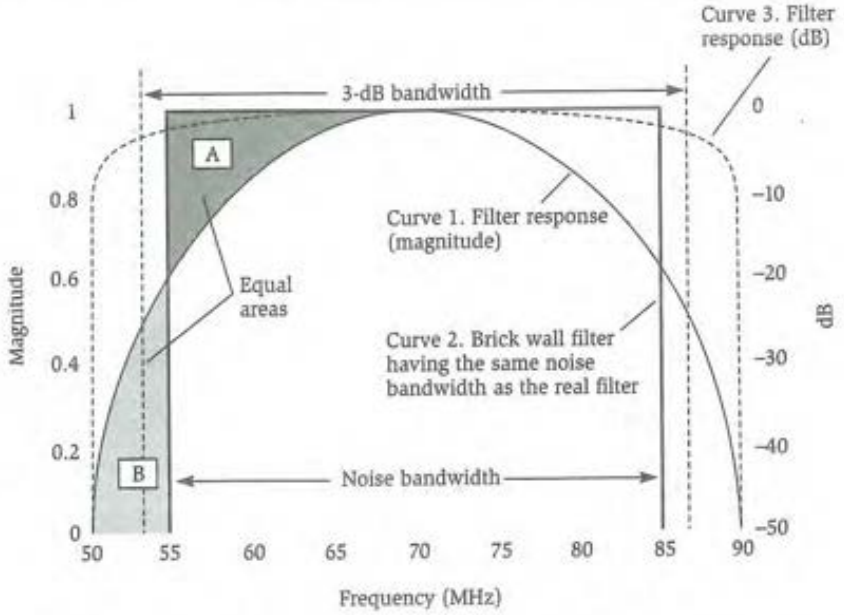


Figure 7.13 Filter response and noise bandwidth.

Figure 7.13 illustrates the concept of noise bandwidth and how it relates to the actual bandwidth of a filter. The majority of the filtering comes from the IF filter in the earth station receiver. A filter response is shown expressed in decibels in dashed curve 3. The same filter response expressed as a magnitude ratio, [(power out)/(power in)], is shown as a thin solid line in curve 1. In determining the noise bandwidth of the filter, we work with the ratio quantity, not the decibel quantity. Filter bandwidth is usually expressed at the −3-dB points, corresponding to the half-power (10 log 0.5) response. The −3-dB bandwidth of this filter is 33.6 MHz. On the other hand, the noise bandwidth is a little narrower, 30.5 MHz. The noise bandwidth is computed by calculating the area under the magnitude response curve, then determining the bandwidth of an ideal "brick wall" filter that has the same area. Curve 2 is that of the ideal filter enclosing the same area as curve 1.

In order for curves 1 and 2 to enclose the same area, area A must equal area B. Area A defines the region where the actual response, curve 1, is below the brick wall response, curve 2. Area B defines the opposite condition, where the actual response exceeds that of the brick wall response.

Note that the noise bandwidth is an artificial quantity not representative of the actual filter response. It is merely a mathematical construct used to permit computation of the noise in the bandwidth of the receiver. This quantity is needed to compute carrier-to-noise ratio from the result of Equation (7.13).

Equation (7.13) is expressed in terms of carrier-to-noise ratio in a 1-Hz bandwidth. From Equation (7.12), the total noise is proportional to the bandwidth of the system, which is expressed by the effective noise bandwidth of the filter as shown. In order to correct Equation (7.13) for the noise bandwidth, a term must be added for the noise bandwidth. The manufacturer of the receiver will supply the noise bandwidth of the receiver. Equation (7.13) then becomes

$$\frac{C}{N} = EIRP - L + \frac{G}{T} + 228.6 - 10\log BW_n \tag{7.14}$$

where

$\frac{C}{N}$ = carrier-to-noise ratio seen by receiver

$EIRP$ = satellite effective isotropic radiated power (dBW)

L = path loss (dB) from Equation (7.11)

$\frac{G}{T}$ = figure of merit for earth station (dB/°K)

BW_n = noise bandwidth (Hz)

FM Receiver Characteristics

Now consider the satellite receiver. The characteristic of interest is the baseband signal-to-noise ratio versus the RF carrier-to-noise ratio. By convention, baseband signals are characterized by a *signal*-to-noise ratio (S/N), whereas RF (and IF) signals are characterized by a *carrier*-to-noise ratio (C/N). At this point, we need to determine the receiver output S/N given the C/N computed in Equation (7.14).

Figure 7.14 illustrates a typical curve that relates baseband signal-to-noise ratio to IF carrier-to-noise ratio.[14] The measuring conditions must be stated and usually follow closely the transmission characteristics nominally used in satellite transmission (recommended by the CCIR):

Reference deviation: ΔF peak = 10.75 MHz

S/N per CCIR 421–3

525-line preemphasis

IF noise bandwidth: according to receiver (32.5 MHz in this example)

The solid line in Figure 7.14 represents the actual signal-to-noise ratio measured on the test receiver. At higher C/N, it is a straight line. As the C/N decreases, the S/N departs from the straight line, falling off more rapidly. If we

were to extrapolate the curve back to a C/N of 0 dB (dotted line), we would find that the S/N was strangely high for a C/N of 0. The point at which the actual curve diverges from the straight line is called the *threshold* of the receiver. In order to provide a repeatable measuring point, the threshold is defined as the point at which the two curves depart by 1 dB.

In determining the needed characteristics of an earth station, it is necessary to establish a nominal operating point sufficiently above threshold so that we can tolerate real-world reductions in performance. Pointing errors can cause signal loss. Polarization discrimination can suffer because of Faraday rotation and rain, causing loss of discrimination against the cross-polarized signals. Signal can be attenuated because of rainfall: the more rain the signal travels through and the harder the rainfall, the greater the attenuation.

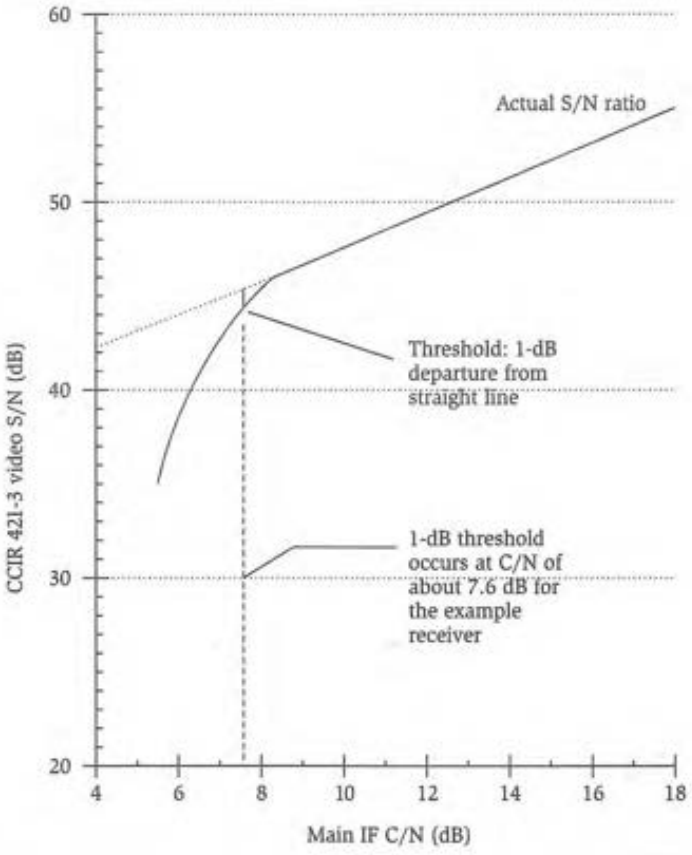


Figure 7.14 Typical receiver S/N versus C/N.

Terrestrial interference is a way of life at C band though arguably it may not be as severe as it was before many applications began migrating to fiber-optic cable. The C band is shared with terrestrial communications. Terrestrial interference has been known to make some receiving sites unusable on some transponders. However, in most cases, it can be controlled by erection of shielding fences or by putting the dish in a depression in the earth. If the interference is not too close to the center of the channel (the carrier), then sometimes a trap inserted in the IF of the receiver may prove effective at mitigating the interference. Many receivers include an IF loop-out feature to allow insertion of such traps.

Threshold Extension

In order to make use of smaller and smaller earth stations, which have lower and lower G/T, it is useful to find techniques that will enable the threshold to be reduced, enabling operation at lower C/N. Several proprietary techniques exist to provide a receiver with *threshold extension*. Regenerative techniques have been used, but a common threshold extension technique involves narrowing the noise bandwidth of the IF filter in the receiver. Some manufacturers have used a circuit that steers the filter, based on the luminance content of the scene. You can narrow only so far before the picture is adversely affected.

A common symptom of too low an IF bandwidth is streaking of colors in brightly colored areas of a scene. Red areas of the picture are particularly susceptible. This is a symptom of picture power that falls outside the passband of the IF filter of the receiver. If the filter was effectively widened, this effect would be reduced, but a higher carrier-to-noise ratio (higher station G/T) would be required.

Similarly, black dots ("sparklies") in the picture can indicate too narrow a filter bandwidth: as the carrier deviates with picture content, some power falls outside the filter passband, or the filter response is severely reduced at the frequency of the power. The effect is very similar to that observed when a station is operated too close to the threshold.

Pre- and Deemphasis

Frequency modulated signals have the characteristic that, when demodulated, the noise spectrum is triangular, increasing at frequencies farther removed from the carrier. In order to overcome this, many FM systems include preemphasis of the signal before modulation and deemphasis after demodulation. In audio work in North America, the traditional preemphasis network is a single break having a time constant of 75 μs. In satellite video work, the International Telecommunications Union prescribes a preemphasis curve that breaks upward at about 150 kHz, and breaks toward a flat response at about 860 kHz.

Figure 7.15 illustrates the recommended preemphasis curve. This shows that the color subcarrier is amplified nearly 15 dB with respect to low-frequency luminance information. Recall that this emphasis applies only between the

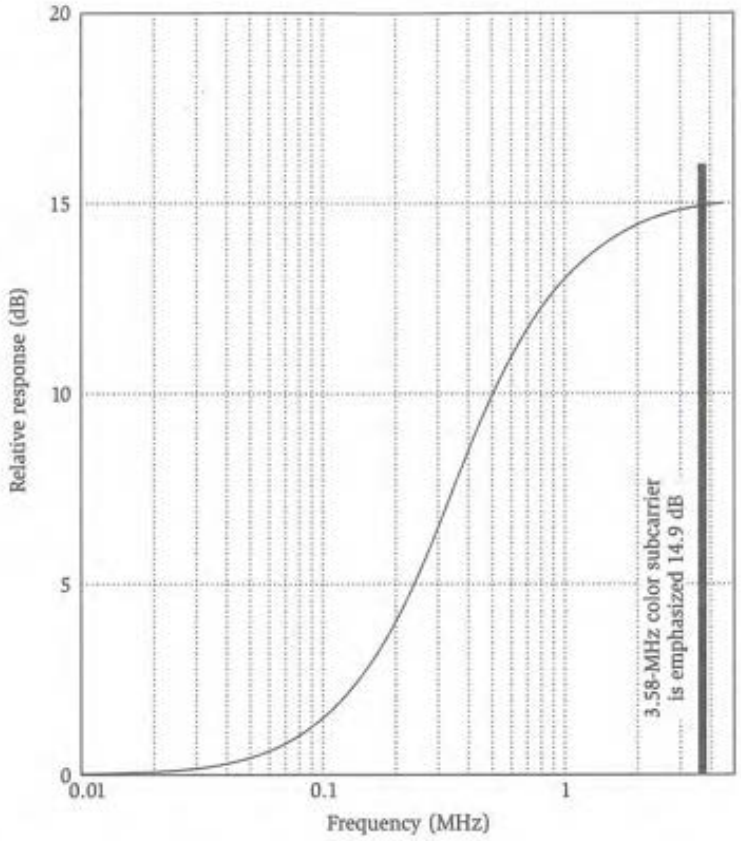


Figure 7.15 CCIR recommendation 405-1 preemphasis curve for 525-line systems.

uplink exciter and the receiver at the headend. It does not apply to the signal out of the receiver. A deemphasis circuit in the receiver restores the signal to its flat frequency response.

However, this does illustrate that the majority of the frequency deviation is taken up by the color subcarrier in the satellite portion of the transmission path. In computing the relationship between RF carrier-to-noise ratio and baseband signal-to-noise ratio, the deemphasis must be taken into account. This is done in Figure 7.14.

Audio Transmission

In the early days of satellite transmission, signals were sent in the clear. To prevent theft of signals, it became necessary to scramble the signal. The scrambling system currently employed replaces the horizontal sync with digitized,

encrypted audio. Enough bandwidth is available to provide for stereophonic transmission. If additional audio channels are required, as for cue tone and second language transmission, subcarriers are used. The subcarriers were used for transmission of the primary audio prior to the advent of scrambling.

A number of frequencies have been used for sound subcarriers, but the most common in North America are 6.8 and 6.2 MHz. The audio source is modulated onto the subcarrier (normally using the standard 75-μs preemphasis), then the subcarrier is combined with the video to form the complete signal transmitted through the satellite.

Digital Satellite Reception

As shown in Chapter 5, digital satellite transmission is primarily QPSK, with some movement toward 8-PSK transmission. The satellite transponder is simply a frequency translator and power amplifier. As such, it can handle analog or digital signals. However, in order to achieve maximum efficiency when digital signals are transmitted, the transponder is operated in or close to compression, the input signal level above which an increase in input signal does not result in a similar increase in output level. This means that we have to use modulation formats that don't depend on amplitude changes to communicate information. 8-PSK and QPSK meet this requirement.

Because the satellite transmission is digital, there is no color carrier and no audio carrier, this information being carried in the digital datastream with the video and supplemental information. Threshold extension is not practiced, and there is no preemphasis and deemphasis.

You can plot graphs similar to that in Figure 7.14 for digital transmission, where the operational parameter on the vertical axis is E_b/N_0, explained in Chapter 4. E_b/N_0 plays the same role as does carrier-to-noise ratio, normalized for the data rate transmitted. The E_b/N_0 required depends on the degree of error correction being used. Table 7.1[15] shows the expected performance of a

Table 7.1 Expected E_b/N_0 for Various Error-correction Rates

Inner code rate	*Required E_b/N_0 (dB), BER = 2×10^{-4} after Viterbi, QEF, after Reed–Solomon*
1/2	4.5
2/3	5.0
3/4	5.5
5/6	6.0
7/8	6.4

demodulator for different error-correction levels. A safety margin of about 2 dB on top of these E_b/N_0 values is recommended.

7.3.4 Satellite Frequency Plans and Interference Issues

The more established band for satellite reception is the C band, 3.7 to 4.2 GHz. This band is shared with terrestrial microwave links, and so interference becomes a significant issue. The issues are probably not as severe as they were a few years ago, however, since more terrestrial services have been migrating to fiber-optic cable.

Polarization Diversity and Frequency Plans

Figure 7.16(a) illustrates the frequency plan used at C band, effectively mandated by FCC rules.[16,17] In order to use the spectrum more efficiently, two polarizations are used, as explained in the introduction to satellite communications. In North America, linear polarizations are used. In many other places,

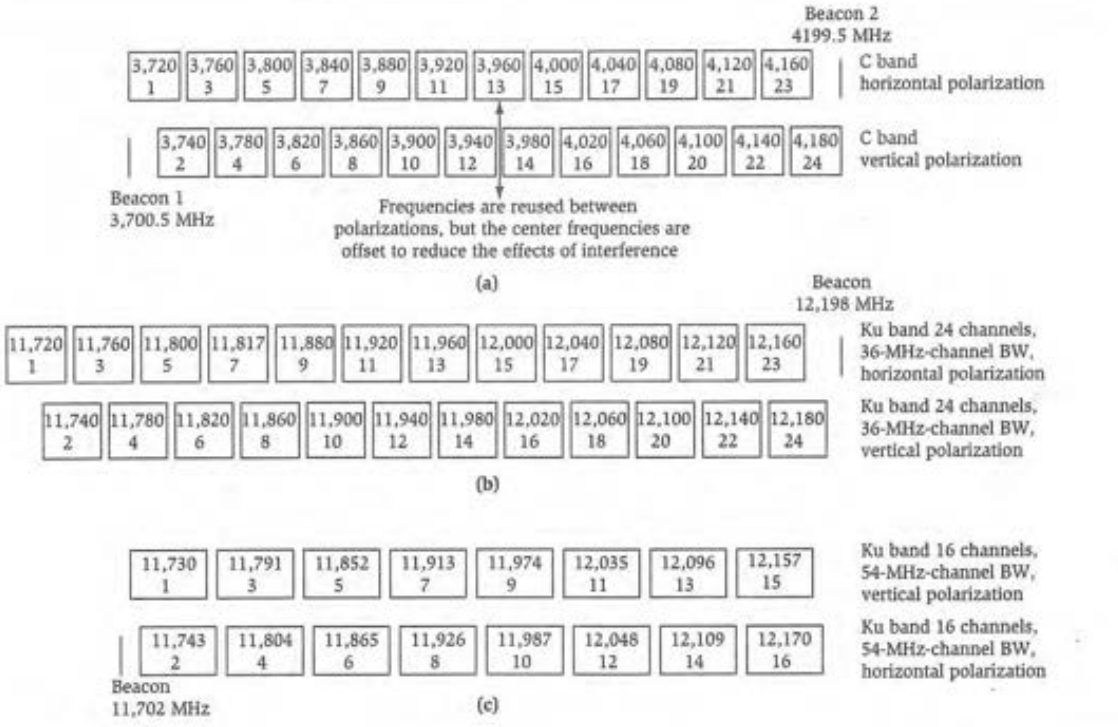


Figure 7.16 Satellite downlink frequency plans. (a) C band. (b) Plan for Ku band of AMC-1 dual-band satellite. (c) Ku-band plan for AMC-5 Ku band only.

right- and left-hand circular polarization is employed. Polarization diversity allows transmission of two signals in the same frequency band without interference. However, as a practical matter, the discrimination that we can achieve is not perfect. In order to improve the performance of the system, signals using the two polarizations are also offset in frequency from each other.

A channel is 40 MHz wide, with the occupied bandwidth of the signal approximately 36 MHz, leaving a 4-MHz guard band between channels on the same polarization. The opposite polarization is offset in frequency such that its picture carriers are in the guard band, as shown in Figure 7.16(a).

Ku-band plans are not as standardized as is the C-band plan. Figure 7.16(b) illustrates the Ku band downlink plan of the AMC-1 satellite. This is a combined C/Ku-band satellite with Ku-band channels 36 MHz wide having a 4-MHz guard band, just as in the C-band plan. Figure 7.16(c) shows another Ku-band plan, using 56-MHz-wide channels having a 7-MHz guard band. This plan, used in the AMC-5 satellite, allows one transponder to be used for two analog TV programs or a higher-data-rate digital signal.[18] Note in all plans that on top of the channel we show the center frequency, in megahertz, for that transponder, and we show the channel number under it. The satellite bands are 500 MHz wide in North America, whereas in Europe they are 800 MHz wide.

Each channel corresponds to one transponder on the satellite. A transponder works much like a cable TV signal processor, explained in Chapter 8. It is simply a frequency translator and power amplifier. Signals are uplinked to the satellite at a frequency roughly 2 GHz higher than the downlink band. On the satellite, the transponder receives and amplifies the signal, converts it to the output frequency, amplifies it, and feeds it to the transmit antenna. The *beacons* are narrow-bandwidth signals that transmit telemetry information related to the condition of the satellite. These beacons, sometimes called *telemetry* (T/M) *beacons*, are used by satellite controllers.

It is common for the downlink frequency band to be converted to a lower-frequency band at the antenna, before the signal enters the headend. This is done in the preamplifier located at the feed point of the antenna. The device that provides preamplification and down-conversion is called a low-noise block converter, or LNB (see Figure 8.16). In North America the common band to which signals are down-converted is 950–1450 MHz. In Europe, several bands are used at frequencies up to 2300 MHz. Since the signal is down-converted at the LNB, the receiver is the same for either C or Ku band.

Polarization Rotation

Even if the signals leave the satellite polarized exactly vertically and horizontally, they don't necessarily arrive at the earth so polarized. Faraday rotation will cause both signals to rotate. Faraday rotation in the northern hemisphere is clockwise looking from the satellite toward the earth station for a downlink signal. Rain will also cause rotation of the polarization. The

net effect is that the earth station antenna's ability to reject the undesired polarity is limited.

Faraday rotation is generated in the ionosphere. The ionosphere is a neutral gas containing charged particles called a plasma. In the presence of a magnetic field (from the earth), a plasma becomes an anisotropic medium, which will cause rotation of an electromagnetic signal propagating through it. Michael Faraday observed the same thing in light traveling through gyrotropic crystal with an applied magnetic field, so the phenomenon came to be named for him.

The amount of Faraday rotation depends on the electron density in the ionosphere, the magnetic field intensity, and the component of the wave traveling along the magnetic field vector. It will thus vary, as the concentration of electrons in the ionosphere varies with the season and with the time of day. The amount of Faraday rotation is inversely proportional to the square of the frequency, so it is more severe at C band than at Ku band. A typical peak value at 4 GHz is 4°. The loss of polarization discrimination at the earth station is given by the sine of the angle of rotation, so 4° results in polarization discrimination of about 23 dB. Of course, the feed may be adjusted for an average Faraday rotation, and if this is done correctly, the loss of polarization discrimination may be much less. Depolarization due to rain may be more severe in a number of locations.

Rain can cause depolarization because the raindrops are elliptical and cause amplitude and phase distortions in the two orthogonal linear polarization components of any incident polarization. Rain depolarization tends to be more severe at higher frequencies. Depending on the rain rate and the elevation angle, rain depolarization can cause the polarization discrimination to drop to between 30 and 10 dB (very high rain rates).[19]

Thus, though polarization can normally yield some 20 dB of discrimination against the undesired polarity, there are times that we wouldn't be able to expect much more than 10 dB of discrimination, thus the need for frequency offset as well as polarization discrimination.

Numerical Observations Regarding Satellite Receiving Stations

The FCC uses a 2° separation standard for satellites serving the United States. Many satellites are so spaced, mostly near odd numbers of degrees. The physical separation is referenced to the center of the earth. Since real earth stations are located at the surface of the earth, about 15% of the distance from the earth's center to the synchronous orbit, the apparent spacing is going to be larger than is the specified orbital spacing.

Cook has presented an equation showing the apparent separation between two satellites from any place on the surface of the earth:[20]

$$\cos\alpha = \frac{2C_2 - C_3\cos X(\cos\Delta L_A + \cos\Delta L_B) - 2C_1(1-\cos\beta)}{2\sqrt{C_2 - C_3\cos X\cos\Delta L_A}\sqrt{C_2 - C_3\cos X\cos\Delta L_B}} \qquad (7.15)$$

where

$C_1 = 43.70532$
$C_2 = 44.70532$
$C_3 = 13.222$
X = station latitude
ΔL_A = difference in longitude between first satellite and station (degrees)
ΔL_B = difference in longitude between second satellite and station (degrees)
β = satellite separation (degrees)

From this and other equations shown earlier, we can illustrate the geometrical issues for two satellites separated by 2° for several locations in the United States. Choose the eastmost positions in the orbital arc (84° and 86°) for which the satellites would be visible at all locations illustrated. (For illustration, we're using even-numbered degrees though most satellites over the Americas are at odd-numbered degrees.) The orbital slots shown are not occupied by actual satellites, but are intended to give a feel for the range of values we might encounter at moderate latitudes. Table 7.2 shows the azimuth and elevation for earth stations located in several places, for the satellite at 84°. The table also shows the apparent separation between that satellite and one at 86° and the slant range and space loss for the satellite at 84°.

The satellite locations are toward the east end of the orbit. They are roughly south of the Florida panhandle, one south of the eastern part and one south of the western part of the panhandle. The slant range from the farthest earth station (Alaska) is about 14% greater than the closest (Florida). However, the space loss, caused by spreading of the beam from the satellite, varies only by about

Table 7.2 Earth Station Geometry and Space Loss for Representative Locations in the United States for Desired Satellite at 84° and Interfering Satellite at 86°

	Lat. (deg)	*Long. (deg)*	*Az. (deg)*	*El. (deg)*	*App. Sep. (deg)*	*Slant range (statute miles)*	*Space Loss (dB)*
Equation:			(6.8)	(6.9)	(6.15)	(6.10)	(6.11)
Location							
Alaska	67	150	112	0.4	2.02	25,867	197
Hawaii	20	155	96.7	9	2.06	25,270	197
Washington State	47	120	135	26	2.16	24,239	196
Illinois	45	90	172	38	2.22	23,581	196
Florida	27	82	184	58	2.30	22,742	196

1 dB. A more dominant factor in determining station performance is the pattern of the antenna on the spacecraft. Typically, the EIRP in a particular direction will vary by a few decibels compared with that in another direction.

The apparent separation between the satellites is always at least as great as is the actual separation (as seen from the center of the earth). This may not be intuitive: you might expect that an earth station far from the subsatellite point (the point on the surface of the earth directly under the satellite), and near the equatorial plane, might see less separation than 2°. In this example, Hawaii is the station you might expect to experience the small difference in apparent separation. However, because the earth is round, Hawaii is "below" the normal plane at the subsatellite point, leading to more apparent separation.

7.3.5 Earth Station Design Example

A brief example of the parameters of an earth station design can be obtained from the preceding numerical examples. Assume a satellite at a longitude of 84°, as in Table 7.2 (no real satellite is assigned to this location). Compute the azimuth and elevation, which tell how to point the antenna. The space loss is computed once the slant range is determined. In order to compute the performance of the earth station, it remains to determine the carrier-to-noise ratio at the receiver. The first step is to obtain the EIRP of the satellite at the chosen location. This is obtained from the satellite operator. A typical C band number might be 34 dBW. Assume the Washington State location, with a path loss of 196 dB, and an earth station with an effective G/T of 20 dB/°K. Finally, assume that the receiver manufacturer specifies an IF noise bandwidth of 32.5 MHz. Equation (7.14) now becomes

$$\begin{aligned}\frac{C}{N} &= EIRP - L_s + \frac{G}{T} + 228.6 - 10\log(BW_n) \\ &= 34 - 196 + 20 + 228.6 - 10\log(32 \times 10^6) \\ &= 11.55\,\text{dB}\end{aligned}$$

The receiver will receive a signal having a carrier-to-noise ratio of 11.55 dB. If the S/N versus C/N curve is as shown in Figure 7.14, the expected video-signal-to-noise ratio will be on the order of 46 dB. The receiver is operating some 4 dB above threshold, before allowance for fade. Rain fade, antenna misalignment, polarization errors, foliage blockage, and other factors will reduce the actual carrier-to-noise ratio. Also, nothing has been allowed for terrestrial interference, interference from adjacent satellites, or the uplink (usually much better than the downlink). A prudent operator will allow several decibels for such.

7.4 Other Methods of Receiving Programming at Headends

By far the most common methods of receiving signals at headends is either by off-air antennas or by satellite. However, other methods are in use, and may even increase in importance in the future. These are covered briefly.

7.4.1 FM Terrestrial Microwave

In earlier days of the industry, it was very common to use terrestrial FM microwave to bring signals to the headend from distant television stations. Many of these links have now been replaced with fiber-optic cable or with closer television stations, but some links remain in operation. The equipment was the same as that used at the time by the telephone industry for microwave transmission, with the exciter changed to handle video. The standards used for modulation were very similar to those used for satellite work, so the concepts discussed earlier are valid for terrestrial microwave also. Propagation issues relating to microwave transmission are similar to those discussed in Chapter 14.

7.4.2 Direct Feed from Television Stations

Because of the perils of the radio path, it is sometimes not possible to deliver high-quality signals to a headend using pickup off the air. A moderate number of television stations now are using direct feeds from their studios to the cable headend. Typically, the baseband output of the master switcher, which supplies the broadcast transmitter, is taken as the input to the cable system.

A related activity today is cooperation between a local television station and one or more cable systems to produce local programs. Typically, the news department of a television station prepares a special cable edition of their newscast, or the off-air newscast is repeated on the cable. Other programming may also be involved, produced by either the television station or the cable operator.

Several methods are used to transport signals to the headend. Today, the most popular methods are analog using either coaxial cable or fiber optics. The trend is toward digital transmission using fiber optics. Some of the more common options are listed next.

Baseband Transmission

It is possible to transmit baseband signals several miles on coax, but the signal must be carefully conditioned to compensate for the considerable rolloff in response of a coax cable in the baseband spectrum. Also, hum pickup is a possibility that has to be dealt with. For stereophonic sound, the audio will have to be picked up off the air, brought down to the 4.5-MHz intercarrier frequency (or perhaps the 41.25-MHz sound IF), and combined with the video after modulation. Alternatively, the

sound signal could be modulated onto a 4.5-MHz subcarrier at the TV station, and the sound subcarrier combined with the video for transmission. Most higher-end modulators can handle signals delivered in this fashion.

AM Supertrunk

It is possible to locate a cable television modulator at the TV studio and transport signals on a dedicated coax or fiber-optic path to the headend. At the headend, the signal is supplied to a processor of the same type used for off-air signals, which converts its frequency to the final cable channel and connects the signal to the outgoing cable plant. This is probably the oldest technique in use. If stereo transmission is to be employed, then a stereo encoder must be located at the studio so that BTSC stereo is transmitted from the studio. At the headend, most cable processors will handle BTSC stereo signals without modification. However, some operators have picked up the off-air signal using a second processor, and combined the sound IF from that processor with the picture IF from the processor handling the video coming from the studio. Engineers must be careful in doing this so that they don't exceed the frequency tolerance of the intercarrier frequency (see Chapter 8). The FCC requires that the intercarrier frequency error not exceed ±5 kHz on the cable (the broadcast tolerance, respected by high-end modulators, is ±1 kHz). If one processor is used to take an incoming picture carrier, convert it to a second channel, and combine with a sound carrier coming from an independent source, it is not too likely that the frequency tolerance will be met. Engineers must do custom-tuning of the oscillators involved in order to make this happen.

AM Sublow Band

Often if a cable operator wants to get programming originating in the community back to the headend, he or she will *backhaul* it on the sublow portion of a trunk cable. Many cable systems carry local sports events, government meetings, and government-sponsored information. The backhaul is frequently done by installing two-way amplifiers in the required portion of the plant. A sublow modulator is used at the point of origination, and at the headend the signal is converted to the downstream channel, again using a processor. This is essentially the same as the technique just discussed, except for the use of the sublow spectrum on the necessary portion of the plant.

By common convention, the industry has divided the return band into 6-MHz "channels," which are designated T channels. Table 7.3 shows the traditional designations used. Generally, the higher-frequency channels provide the higher-quality signal, due to noise on the return path. As shown in Chapter 16, frequencies generally below 15 or 20 MHz tend to experience a lot of interference on the return path. Thus, higher frequencies are preferred, especially for analog video, which tends to be more susceptible to interference than is digital transmission.

Table 7.3 Sublow Channel Designations

Channel	*Picture carrier (MHz)*
T-7	7
T-8	13
T-9	19
T-10	25
T-11	31
T-12	37
T-13	43

One factor, though, forces us to not put the return video too high in the spectrum. As shown elsewhere, the diplex filters used in two-way distribution amplifiers produce group delay near the crossover frequency. If the signal must be transported through too many amplifiers, the group delay may add to an unacceptable number. The usual countermeasure for this is to reduce the frequency of the analog transmission.

FM Supertrunk

A similar technique is to use FM transmission on a dedicated cable. This was popular a few years ago since the quality signal delivered to the headend could be quite good. The same conditions apply to audio as described before.

Digital Transmission with No Data Compression

Equipment is available today to allow noncompressed digital transmission to take place over a fiber-optic cable. The systems are low in cost but require a relatively high data rate for each channel. Data rates of 120–240 Mb/s per video channel are not uncommon in these applications. Several video signals can be multiplexed, with the total data rate being 2.4 Gb/s or higher. The technique is also used to interconnect headends, where it is desired to transport several high-quality signals between headends. A dedicated fiber is usually used, though wave division multiplexing (WDM) is beginning to be practiced today.

Digital Transmission with Moderate Data Compression

Equipment is available to allow moderate compression in a transmission path. Generally, this technique will be more costly than use of noncompressed transmission, but it may be preferred if there is a scarcity of bandwidth.

Compression is based on vendor proprietary compression schemes, most of which use *linear predictive coding* (LPC). In an LPC system, an algorithm is used to predict the content of the next picture (usually either a field or frame) based on past data. The encoder includes a decoder, which works identically to the one at the receiving end. The predicted output of the decoder is compared with the picture before encoding. The difference is an error signal, which is transmitted to the receiver. Because TV pictures tend to have a lot of commonality between frames, the amount of information transmitted is much less than if no compression were used. Data rates of around 45 Mb/s can yield excellent-quality pictures and are routinely used for backhaul of programs to broadcasters. Data rates in the neighborhood of one-half of this have met with limited acceptance.

Digital Transmission with Extreme Data Compression

This usually means some form of MPEG or MPEG-derived compression algorithm based on the *discrete cosine transform* (DCT) and possibly other aspects of MPEG compression. Because of the cost, no known applications exist today that utilize full MPEG compression for point-to-point transmission of video though this may happen in the future. However, it is not that expensive to implement just the DCT and maybe some other modest data reduction techniques. Data rates of under 20 Mb/s for broadcast quality have been reported. Chapter 3 includes more information on compression.

7.4.3 Local Origination

Finally, local origination (LO) is practiced in many systems. LO is done from studios operated by the cable company or sometimes by other parties. Typically, the studios will be similar to those at broadcast stations, except that the amount and quality of equipment is less.

Much programming originates on tape. Common tape formats in use include consumer VHS (very low end), super VHS, and 3/4 inch, in increasing order of quality and cost. Better operations employ time base correctors to compensate for the unavoidable timing errors that develop because of inconsistent tape transport speed. Computers may be used to control the tape machines, or they may be controlled manually.

The leading edge in local origination storage is the use of video file servers. Programs are digitized and compressed. The compressed program is stored on a large file server, not unlike those used in business applications to store massive amounts of data. At the programmed time, the file server "plays" the file to an MPEG-to-NTSC converter (for analog channels), which converts the signal back to analog. This type of system saw its earliest successes in the commercial insertion (ad insertion) market, is now used for playing out long-form programs.

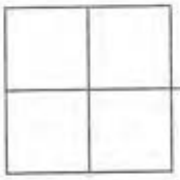
7.5 Summary

This chapter has discussed the underlying technology used to bring signals into the headend. The vast majority of signals are received today from satellites, with the next largest number being picked up off the air. Other techniques are used and are outlined briefly here. In the next two chapters, we will discuss headend-related concepts beyond getting signals into the headend. Chapter 8 will cover handling signals in individual pieces of equipment in the headend, and Chapter 9 will cover handling the signals outside those pieces of equipment.

Endnotes

1. Richard C. Johnson et al., *Antenna Engineering Handbook*. New York: McGraw-Hill, 1984, pp. 4–11.
2. Mark J. Wilson, ed., *The ARRL Handbook for the Radio Amateur*. Newington, CT: American Radio Relay League, 1988, p. 17.2. Regrettably, this figure seems to have been removed from recent editions.
3. Computed from Equations (12) and (14) of International Telephone and Radio Corporation, New York, *Reference Data for Radio Engineers*, 4th ed., 1967, pp. 676–677.
4. Scientific-Atlanta, Inc., Atlanta, GA, *CATV Data Book*, available at www.sciatl.com.
5. Richard C. Johnson et al., op. cit., pp. 29–17ff.
6. Thomas Milligan, *Modern Antenna Design*. New York: McGraw-Hill, 1985, p. 383ff.
7. Scientific-Atlanta, Inc., Atlanta, GA, *CATV Data Book*, rev. 6, August 1990, p. 77.
8. Marty Barringer, Radio Wave Propagation, *Engineering Handbook*, 7th ed. Washington, DC: National Association of Broadcasters.
9. Ibid.
10. Rick Barker and John Friesz, Earth Station Antennas, RF Considerations. Scientific-Atlanta Communications Symposium, 1983.
11. Ibid.
12. F. M. Fonda, *Earth Station Geometry*. Scientific-Atlanta Communications Symposium, 1983.
13. Jack Sanders, Satellites: The Birth of an Industry. *Installer/Technician*, May 1988, p. 15.
14. J. Searcy Hollis, *Satellite Link Analysis*. Scientific-Atlanta Communications Symposium, 1983, p. 12.

15. European Telecommunications Standards Institute, *Digital Broadcasting Systems for Television, Sound and Data Services: Framing Structure, Channel Coding and Modulation for 11/12-GHz satellite services*, ETS. 300 421, December 1994. Available at http://www.etsi.org/.
16. 47 CFR sec. 26.211. U.S. Government publication, available at Government Printing Office bookstores in major cities.
17. J. Searcy Hollis, Principles of Satellite Communications. Scientific-Atlanta Communications Symposium, 1983, p. 25.
18. www.ses-americom.com
19. Rick Barker, *Re-Use Earth Station Antennas*. Scientific-Atlanta Communications Symposium, 1983.
20. James H. Cook Jr., *Interference Analysis*. Scientific-Atlanta Communications Symposium, 1983. (The equation is in Appendix A.)



Chapter 8
Headend Signal Processing

8.1 Introduction

This chapter deals with the electronic devices most commonly used in headends to format signals for transmission on a cable plant. These devices include signal processors, RF devices used to take television signals off the air or from an incoming cable and put them on the cable. Also included are modulators, miniature transmitters that accept baseband audio and video signals and convert them to a channel. Demodulators do the opposite, converting an RF signal back to baseband video and audio. Stereophonic encoders convert left- and right-channel audio into the composite stereo format used in television transmission. Earth station receivers convert signals coming from communications satellites to baseband audio and video. Often these are used with descramblers that deal with the proprietary scrambling format currently used with most analog satellite transmissions.

Block diagrams and descriptions of the operation of the equipment are presented, along with information regarding the proper application of the hardware. The block diagrams are representative of commonly available equipment. They are intended to illustrate how the functions are accomplished; they are not intended to represent particular pieces of equipment. Variations will be encountered in the field. Some equipment may have controls that don't work as do the controls shown here, as a consequence of differences in block diagrams.

Chapter 7 covered inputs to the headend from various RF sources. Chapter 9 will cover methods used to couple signals from the equipment covered in this chapter to the cable. You may wish to review the fundamental composition of baseband and RF signals in Chapters 2, 3 and 4.

8.2 Signal Processors

RF signal processors are normally used to transfer incoming VSB-AM signals from off-air antennas or incoming cables to the cable plant. The processor can change the channel (frequency) of the signal and the relative sound carrier level. The signal is bandpass filtered to remove any undesired adjacent channel power.

Automatic gain control is employed to stabilize the output level. Optionally, a standby carrier can be substituted when the incoming signal goes away. Switching of IF sources can be done to maximize use of a channel.

8.2.1 Input Section

Figure 8.1 illustrates a typical block diagram of a processor. A number of variations are possible. Many modern processors are frequency agile. Differences between agile and nonagile processors are pointed out as appropriate. The incoming RF signal is supplied to a bandpass filter, BPF1, which is tuned to the incoming channel. This bandpass filter is typically of fairly low order and has only a little adjacent channel rejection. Its primary purpose is to reduce off-channel power to minimize distortion in the input circuitry. It, along with BPF2, provides image rejection, as described in Chapter 9.

The signal passes through a variable attenuator, AT1, which is part of the delayed AGC circuit. "Delayed AGC" is a somewhat unfortunate term because it has nothing to do with a time delay. Rather, the term *delayed* refers to the fact

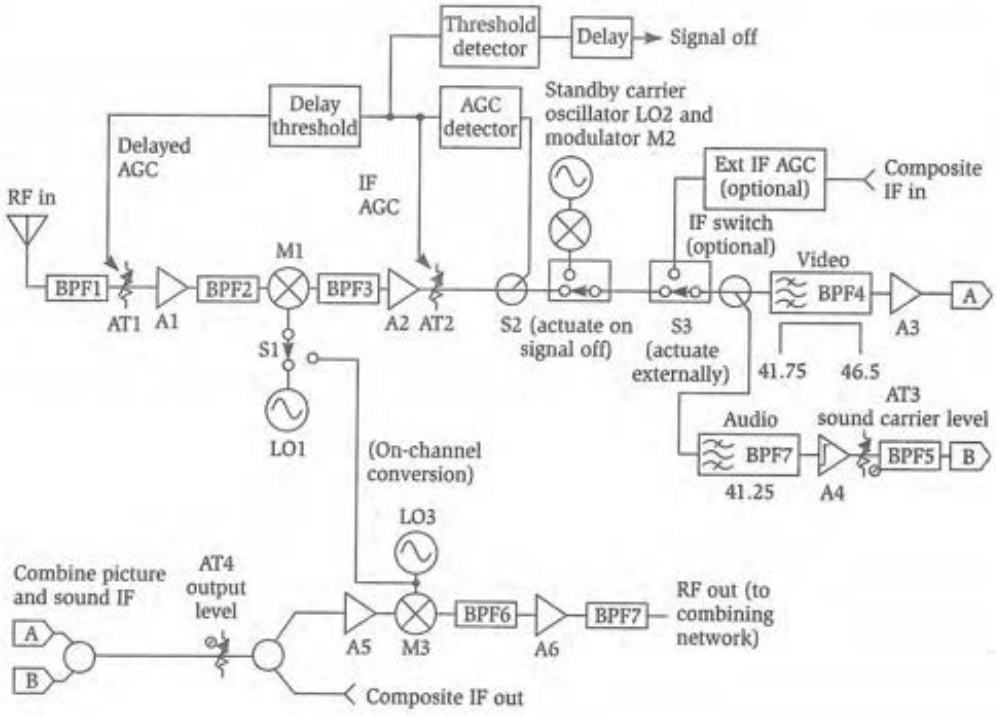


Figure 8.1 Signal processor block diagram.

that AT1 is at minimum attenuation (the front end is at maximum gain) at low signal levels. As the signal level increases (above a predetermined threshold), it becomes necessary to reduce its amplitude before it reaches RF amplifier A1 and mixer M1. Failure to reduce the amplitude would result in unacceptable distortion. At lower signal levels, the gain of A1 is needed in order to overcome the noise introduced in the mixer and following circuitry.

8.2.2 IF Section

The signal is converted to the intermediate frequency (IF) in mixer M1. The commonly used North American IF is 45.75 MHz for the picture carrier and 41.25 MHz for the sound carrier.* Note that the spectrum is inverted at IF compared with the spectrum at RF, where the sound carrier is at a higher frequency than is the picture carrier. This is a consequence of using a high-side local oscillator (LO); that is, the frequency of the local oscillator is higher than is the RF frequency of the signal being converted. Use of high-side oscillators helps control spurious responses from the mixing process.

Following mixer M1 is another bandpass filter, BPF3, tuned to the IF band, normally 41 to 47 MHz for NTSC signals. Amplifier A2 provides gain to help overcome the insertion loss of later circuits. The IF AGC attenuator is AT2, which is adjusted such that the input to the AGC detector is constant regardless of the incoming signal level. Additional amplification may be provided, and at some point in the signal chain, a sample of the signal will be routed to the AGC detector. The output voltage of the AGC circuitry changes as the signal into the AGC detector increases. This change causes the attenuation of AT2 to increase, compensating for the increasing level of the incoming signal.

If the signal level increases sufficiently, the IF AGC voltage reaches the delay threshold, above which the gain of the input attenuator, AT1, is reduced to effect stabilization of the IF signal level. Depending on the design of the AGC system, the IF attenuator, AT2, may continue to drop in gain, or it may not. Usually, the best carrier-to-noise performance of the processor is achieved at input signal levels just below the delayed AGC threshold. If sufficient level is available, this is the preferred operating point for the processor.

Returning to the main IF signal path, as illustrated, the signal next is supplied to switch S2, which selects the standby carrier upon loss of input signal. The standby carrier was originally employed to prevent TVs from showing snow when a signal goes off the air. It is also employed if the output channel is used as a pilot for the distribution system. A pilot signal is used to set the gain of distribution systems, and if it disappears when a TV station goes off the air, the operating point of the distribution system will be upset.

---

* PAL countries use 38.9 MHz for the picture carrier IF and 33.4 (PAL-B/G) or 32.9 (PAL-I) MHz for the audio IF. In China, 38 MHz is used for the picture carrier IF.

Some processors include the ability to modulate the standby carrier with pseudosync pulses, a low-duty cycle 15.734-kHz square wave that simulates normal TV modulation. This is necessary to support distribution AGC systems that need sync tips to operate, rather than a CW signal.

Switch S2 is normally activated by a drop in the IF AGC voltage sufficient to indicate that the input signal level is lower than a minimum needed for useful performance. Following switch S2 is switch S3. This switch is optional and is used when it is desired to switch to an alternative IF signal. This is done, for example, when the primary signal leaves the air and it is desired to substitute another signal on the same channel. For example, the cable operator may carry a local television station that goes off the air at night. It is desired to substitute another signal on the channel when the local station goes off the air. The alternative signal, processed in another modulator or processor, is supplied at the "Composite IF in" connector. Optionally, a separate AGC circuit may be used to normalize the level of the alternative IF signal. In this example, switch S3 would be connected to the IF AGC signal such that when the input signal drops, S3 is activated, selecting the alternative IF signal.

The picture and sound carriers are separated in the example processor. Aural IF, at a frequency of 41.25 MHz, is routed to a bandpass filter, BPF7, which eliminates any power associated with the picture signal or other channels. Following BPF7 is a limiter, A4, which functions identically to the limiter in an FM radio. Its output is constant regardless of the input signal level, within a reasonable level range. Following A4 is a manually adjusted attenuator used to adjust the sound carrier level. The sound carrier is normally adjusted to be −15 dB from the picture carrier level. A bandpass filter following AT3 removes harmonics of the sound carrier, which are introduced in the limiting process. The sound IF is then combined with the picture IF.

An alternative sound IF circuit uses a notch of variable depth to adjust the sound carrier level. This circuit works but cannot remove any residual AM on the sound carrier. Further, if not adjusted properly, it can add FM/AM noise, which hurts operation of many scrambling systems.

The visual signal at IF is supplied to BPF4, which provides the bulk of the selectivity of the processor. We describe the response of this filter later. After amplification, the visual IF is combined with the sound carrier.

8.2.3 Output Section

After combining, the composite IF signal is applied to output level control AT4. The location of this control in the signal chain allows it to simultaneously adjust the picture and sound carrier levels. After level control, the composite IF signal is split. One side of the splitter is supplied as the "Composite IF out" of the processor, to be used if needed as the alternative IF input to another processor or modulator. The other output of the splitter is amplified in A5 before being supplied to mixer M3, the output mixer. It is mixed with the output of local

oscillator LO3. The signal is now on the assigned output channel. Bandpass filters BPF6 and BPF7 eliminate the image of the second conversion process and any spurious signals coming from mixer M3.

Amplifier A6 increases the signal level to the desired output level. Its performance is critical: any third-order distortion it introduces will generate frequencies too close to the output channel to be rejected by BPF7. The most common distortion produced is a third-order beat between the picture and sound carriers. This beat, appearing 4.5 MHz below the picture carrier, is derived from twice the picture carrier less the sound carrier. Another beat appears 4.5 MHz above the sound carrier but is lower in amplitude.

If the output is *frequency agile*, meaning it can be set to any of a number of channels, the output section shown is replaced with a dual conversion output section, with a first IF above the maximum frequency the processor can output. The IF is filtered to remove the image, and then it is down-converted to the output channel. Sometimes up-conversion is done externally to the modulator; this technique, along with more information regarding up-conversion, is covered in Section 8.4.1. The mechanism that creates images is covered in Chapter 9.

8.2.4 Local Oscillators

Normally, LO1 and LO3 are fixed, crystal-controlled oscillators in a fixed channel processor. If the processor is agile on the output channel, LO3 and M3 will be replaced with a dual conversion output converter, which converts the IF signal to a high frequency (typically in the high hundreds of MHz), where it is filtered and down-converted to the output channel using yet another mixer and a phase locked local oscillator.

A processor with an agile input will have the input section replaced with an agile oscillator and tracking filter. In higher-quality processors, the front end will be double conversion, with the second IF at the normal TV band of 41–47 MHz.

If the processor is used as an on-channel processor, with the input and output on the same channel, then it is desirable that the output signal be phase coherent with the input signal to avoid beat products. To effect this phase coherency, local oscillator LO1 is not used for on-channel conversion. Rather, a sample of the output local oscillator, LO3, is supplied to mixer M1. Thus, the same local oscillator is used for both down- and upconversion. This may be shown to ensure phase coherency. Switch S1 is thrown to the alternative position for on-channel conversion.

If it is desired that the output be phase locked to an external reference signal at the frequency of the output channel, then LO3 will be replaced with a phase locked circuit.

8.2.5 Processor Frequency Response

Figure 8.2 illustrates the visual channel IF response of the processor, overlaid with the spectrum of signals that may be present. An important function of a

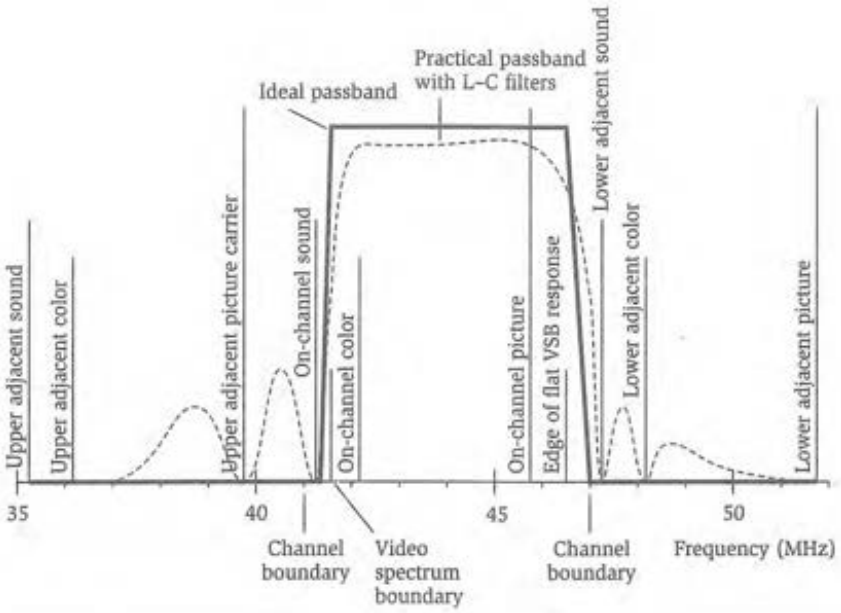


Figure 8.2 Processor visual channel IF response.

processor is to provide off-channel rejection. This is important in cases where the processor is used to process signals coming in from another cable. For example, it is sometimes practiced to route signals by AM microwave or cable to a hub, where they are reprocessed and supplied to subscribers. In this case, adjacent channel power coming into the processor must be rejected.

Also, if a station is being received off the air along with an adjacent channel from another city, the adjacent channel must be rejected before the signal is supplied to the cable. It is accepted that, in order to be invisible, an undesired signal should be 60 dB below the level of a desired signal. For specification purposes, this is the desired adjacent channel rejection of a processor. In the case of incoming cable signals, adjacent signals will be at about the same level, so 60-dB adjacent channel rejection is adequate. If an adjacent channel has a greater signal level than the tuned channel, as could be the case in off-air processing, then more rejection may be required. This is considered to be handled best with external selectivity rather than to burden all processors with more rejection capability.

Figure 8.2 illustrates the ideal processor passband as a heavy line and a typical passband as a dashed line. The location of the relevant carriers is shown at IF, where the picture carrier is at a higher frequency than the sound carrier. Chapter 8 shows why the spectrum is inverted from that broadcast. Ideally, the

visual IF passband would pass signals from the edge of the visual spectrum, 4.2 MHz from the picture carrier, to the edge of the flat portion of the vestigial sideband region, 0.75 MHz above the picture carrier at IF.

This ideal spectral response may be approximated if a surface acoustic wave (SAW) filter is used as the IF filter (BPF4 in Figure 8.1). This filter sets the overall visual IF response of the processor for all practical purposes. Some processors employ L–C (inductor–capacitor) filters for this application. Practical L–C filters are unable to achieve the ideal passband shape. Bandpass sections are combined with traps, which provide high rejection over a very narrow band of frequencies. These traps are placed on the adjacent carrier frequencies as shown. Most of the power in an NTSC signal is concentrated at the picture carrier, the sound carrier, and the color carrier. The power density at other frequencies in a channel is relatively low. Because of this characteristic, it is possible to specify 60-dB rejection even if that much rejection is not available at all frequencies. The specification of 60-dB rejection is interpreted as meaning that power in an equal-level TV signal appearing in an adjacent channel is more than 60 dB below the on-channel picture carrier.

Even when a SAW filter is employed, it may be necessary to supplement its response with some traps. Practical SAW filters exhibit ultimate rejection of 40 to 50 dB, against a requirement of up to 60 dB depending on the frequency. Traps may be needed to achieve this last little attenuation.

L–C filters exhibit group delay. This means that signal components passing through the filter near either edge are delayed with respect to signal components passing through near the middle of the passband. The delay is a natural consequence of the attenuation of the stopband and is predicted from the mathematics describing the response of the filter. If the transition from passband to stopband occurs in a smaller frequency band, the group delay is greater.

Group delay causes components of the visual signal that occupy frequencies near the edge of the passband to propagate through the filter more slowly than do components near the center of the passband. This causes undesirable distortion of the video signal. To remedy this, another filter, called a *delay equalizer*, is incorporated with the bandpass filter. The delay equalizer must be designed for the bandpass filter with which it is to be used. It adds more delay to signals at the center of the passband, complementing the delay response of the bandpass filter.

A delay equalizer is necessary when an L–C filter is used as the bandpass filter, but not when a SAW element is used. Because of the way a bandpass filter is realized using a SAW device, group delay is inherently low.

8.2.6 Digital Off-Air Processing

To meet some early needs to carry off-air digital signals on cable plant, a simple system was used. An analog signal processor was retuned to widen the IF passband shown in Figure 8.2 such that it would pass the 8-VSB signal picked up off the air. The AGC was bypassed by placing the processor in the manual mode. Signals were up-converted to an off-air UHF frequency (which differs

from the cable UHF frequency plan by 2 MHz, as explained in Chapter 9). This is not particularly efficient method of carrying off-air signals, but it sufficed in the early days, when adequate spectrum was available on the cable plant.

As shown in Chapter 4, off-air digital signals use 8-VSB modulation with a transport stream payload data rate of 19.393 Mb/s. The data rate of a 64-QAM video signal on cable (after removing transmission overhead) is 26.97 Mb/s; for 256-QAM it is 38.811 Mb/s. You can see that a 256-QAM channel can accept two complete 8-VSB modulated channels. Furthermore, in many cases a cable operator may want to carry some but not all of the data in the off-air channel, freeing up more spectrum. When cable operators carry off-air digital channels it is important to minimize the spectrum required, so there is a strong need to combine two off-air digital channels into one cable channel. This can be done with a *digital television processor*.

Figure 8.3 illustrates digital television processing, the digital equivalent to analog signal processing. A digital signal is recovered off-air in tuner #1 and supplied to the PSIP processor and remultiplexer (remux). As explained in Chapter 3, digital video signals include the *program and system Information Protocol* (PSIP), a series of tables that, among other functions, tell the receiver what programs are included in the multiplex (i.e., the 6-MHz channel carrying several programs) and how to find them. These tables must be modified if some programs are removed from the multiplex and others are added from another off-air channel or other source.

In the configuration shown, a second program stream may be added from either another off-air channel via tuner #2 or from another signal source. This program transport stream is supplied to the PSIP processor and remux. Its PSIP tables are merged with those from the first program stream and multiplexed into a higher-bandwidth stream that is supplied to a 256-QAM modulator. The modulator may include up-conversion to the frequency the combined transport stream occupies on the cable plant, or the up-conversion may be separate. Since each program stream included in each of the two input streams shown may be transmitted in

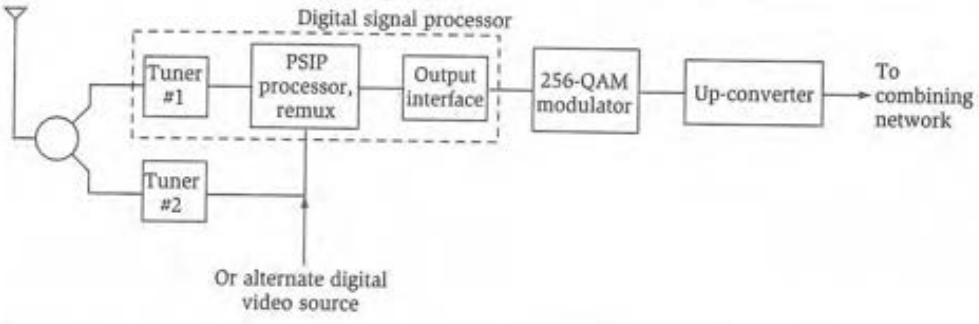


Figure 8.3 Digital television processing.

variable-bit-rate format (VBR — explained later), there are some issues that must be addressed relating to the amount of data that is carried in the combined transport stream. In the case of combining two off-air channels, there is enough bandwidth to carry two complete channels. In many cases, however, only the HD elements are desired, and the remultiplexing process is capable of dropping not only the other television services but other data services as well. This would leave two off-air HD signals, with perhaps enough remaining bandwidth for another SD service of the operator's choosing. The remultiplexing process must also derive and recombine all of the PSIP information from the desired incoming sources.[1]

8.3 Modulation

Figure 8.4 illustrates the video baseband, IF processing, and RF processing found in a typical modulator. (Upcoming Figure 8.5 shows the audio circuitry.)

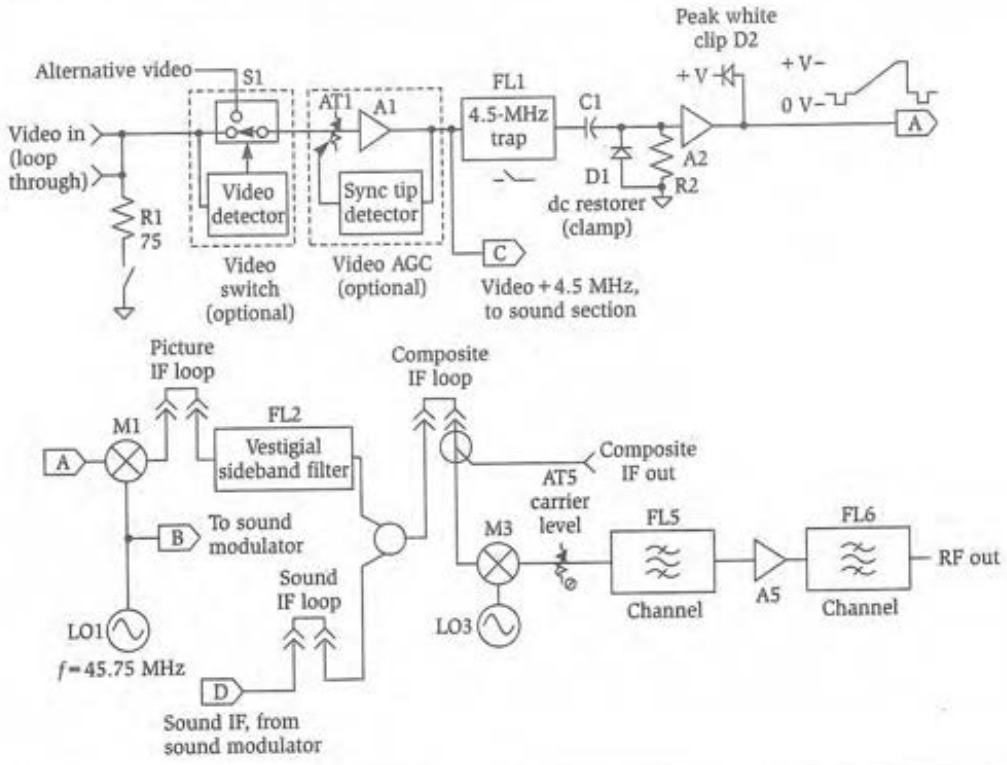


Figure 8.4 Modulator block diagram, baseband and RF.

Modulators are used with many more channels than are processors since the majority of channels are programmed from satellite-distributed programming or local sources. The output circuitry of the modulator is identical to that of a processor, but the front end is significantly different.

8.3.1 Video Switch and AGC

Video is introduced into the modulator in the top left-hand corner of Figure 8.4. An optional loop-through facility is shown on the input. Where a video source must be supplied to several places, it is possible to loop the video through from one piece of equipment to another by using a high input impedance in each. A termination is used only at the last piece of equipment in the signal chain. Section 8.9.1 covers this topic in more detail.

As shown, a video switch is optionally available to allow selection of an alternative video source. The switch may be activated manually, from an automation system, or automatically upon loss of the primary video.

Another option available with some modulators is a video AGC circuit. Typically, the amplitude of the sync tip is measured. If it is in error, attenuator AT1 is adjusted until the sync amplitude is normalized. It is assumed that the ratio of active video signal to the sync tip is correct. The proper level of video depends on the video content, which cannot be known at the modulator. However, the sync tip amplitude is well known. An error in sync tip amplitude is indicative of a gain error in video-processing circuitry. By correcting the sync tip amplitude, the active video amplitude is corrected.

Although the ideal correction strategy for the video AGC is to correct sync tip amplitude, a few modern modulators feature additional correction strategies. It is a fact of life that certain video signals handled in headends do not conform to the expected characteristics of amplitude and sync-to-video ratio. In order to deal with these signals, some video AGC circuits include a mode in which the amplitude of the sync tip to peak video signal is normalized rather than the sync amplitude. This mode should not be used except under certain conditions. If the video level drops because the picture intentionally goes dark, this mode would cause the sync level to increase and the video to be stretched to white unnaturally. However, the mode might be useful for video sources that are always expected to include peak white, but that are not always expected to be at the correct amplitude.

8.3.2 Sound Subcarrier Trap

Following the video AGC is a trap tuned to 4.5 MHz. In some cases, video is supplied combined with sound on a 4.5-MHz subcarrier. This is done, for example, when video is supplied via a terrestrial FM microwave link. The base-

band video and subcarrier sound must be separated before modulation. The aural signal is coupled out through flag C (to Figure 8.5). The aural subcarrier is removed in FL1, a 4.5-MHz trap. Frequently, the trap is switchable because it does introduce some group delay. It may not be needed, and should not be used, if the incoming video doesn't contain significant power at 4.5 MHz.

Under some conditions other than the existence of a 4.5-MHz subcarrier, use of the trap may be desirable. If the video contains significant power at 4.5 MHz owing to distortion or other processing, that power would appear as noise to the sound subcarrier. This can adversely affect the sound–signal-to-noise ratio, and in certain circumstances, could affect descrambling (see Chapter 21).

8.3.3 DC Restoration and Peak White Clip

From the trap, video is supplied to a dc restoration circuit. Video signals may be interfaced between pieces of equipment with dc restored or ac coupled. As video content changes, the average signal level, and hence the dc value of the signal, changes. However, since it is required to transmit modulated signals with a fixed sync tip amplitude, the dc level must be reestablished before the signal is modulated onto the RF carrier.

A number of circuits have been used for dc restoration. One of the simplest is shown in Figure 8.4. If any portion of the video signal tries to drop below ground, diode D1 conducts, charging C1 positive at the right end. Since the video signal voltage is all positive with respect to the sync tip, when the charging of C1 is complete, the sync tip will be at ground level (ignoring the diode drop in D1), with all other elements of the video signal being positive with respect to it. This dc level is maintained to the actual point of modulation. Resistor R2 is used to discharge C1 so that if the sync tip rises too high the capacitor will discharge until D1 again conducts on sync tips.

After the video signal is clamped, it is applied to a peak white clipper. Visual carrier modulation is strictly limited to 87.5% in the NTSC system. If the carrier modulation is too deep, the carrier will at some point be cut off, or very nearly so. Because of the way sound is demodulated (see Sections 8.5.5 and 9.5), carrier cutoff will cause intolerable sound buzz. In order to prevent this, many modulators include peak white clip circuits. If the video voltage exceeds voltage $+V$ in Figure 8.4, then diode D2 conducts, holding the voltage on the video line constant.

8.3.4 Modulation Process

The clamped and clipped video signal is coupled to modulator M1. It is no mistake that the modulator is shown modeled as a mixer with the multiplication sign. Amplitude modulation is a multiplicative process, as is frequency mixing. The circuits used for both amplitude modulation and mixing are essentially

identical and perform a multiplication process. Chapter 2 provides background on amplitude modulation theory.

Conventional amplitude modulation of a carrier by a sine wave is expressed by the equation

$$\begin{aligned} V_0 &= \cos \omega_c t(1 + M \cos \omega_m t) \\ &= \cos \omega_c t + \frac{M}{2}\cos(\omega_c - \omega_m)t + \frac{M}{2}\cos(\omega_c + \omega_m)t \end{aligned} \tag{8.1}$$

where

ω_c = carrier frequency

ω_m = modulating frequency

M = percent modulation of the signal (though TV signals use a different expression for modulation)

From the first form, we can see that the envelope (represented by the term inside parenthesis) never goes negative. (We can show that if the 1+ term is omitted, the carrier is suppressed, which is done when we modulate the color subcarrier.) The second form is obtained by trigonometric expansion of the first. It shows that the frequency domain includes the carrier and two other frequencies, spaced above and below the carrier by the modulation frequency. These upper and lower modulation products are called *sidebands*. The information contained within the modulated carrier is carried redundantly by the upper and lower sidebands.

A video signal is much more complex than a sine wave. The single sinusoida modulation term $\omega_m t$ is replaced by a sum of sine waves of different frequencies, which taken together, define the video waveform (see Chapter 2). The resulting modulated signal has a series of sidebands above and below the carrier. One set may be removed without losing the information contained in the signal.

8.3.5 IF Processing

From modulator M1, the signal is often looped out for possible external signal processing. (The term *loop out* is applied to the routing of a signal from one piece of equipment to another and then back into the first piece for further processing.) Many common scrambling systems (described in Chapter 21) need access to the modulated signal at this point. They attenuate the picture carrier during the horizontal and, usually, the vertical blanking intervals by attenuating the IF signal. Though some older systems attenuated the sync by applying a sine wave attenuation, more modern systems use so-called gated sync suppression, achieved by switching an attenuator into the signal path. The switching operation introduces high-frequency modulation components, which will interfere with adjacent

channel signals if not eliminated. By arranging FL2 so that it follows the addition of the scrambling signal, the undesired modulation components are removed.

After scrambling (when used), the signal returns to the modulator and is supplied to vestigial sideband filter FL2. Vestigial sideband modulation and filtering were described in Chapter 2, and the precise shape of a VSB filter for NTSC is described later. It is used to eliminate one set of modulation sidebands. Modulation sidebands more than 0.75 MHz from the picture carrier are removed from one side of the picture carrier. If the signal has been scrambled using gated sync suppression scrambling, then FL2 will also filter those modulation sidebands.

After filtering, the IF signal is combined with the sound carrier signal. The combined signal may be looped out, at the composite IF loop, to provide for processing by an older sine wave sync suppression scrambling system. It is likely that the signal will be split, with a portion supplied to the composite IF out on the rear panel of the modulator. This signal is used with other processors or modulators in IF switching applications, as described earlier.

The composite (visual and aural) IF signal is up-converted to the RF channel in mixer M3, driven by local oscillator LO3. Attenuator AT5 controls the output level of the composite visual and aural signals. Remaining processing is identical to that described in the signal processor section of this chapter. The signal is filtered, amplified, and filtered again. A phase locking system can be used to phase lock the carrier to another source, and is described later.

8.3.6 Aural Modulation

Figure 8.5 illustrates the modulation process. Baseband audio is supplied to an optional switch S2. If video switching (S1 of Figure 8.4) is used, then audio switching is likely required, too. Normally, the audio switch will be driven by the video switch: when the video switch activates, it activates the audio switch. Sometimes a third audio input, intended for emergency override, is provided though it is not shown in the Figure. Frequency response of the override input may be limited to voice band only. The reason it would be limited is to reduce any noise that may appear on an audio source that is outside the roughly 300–3,000-Hz band needed for understanding voice communication.

Preemphasis

Preemphasis is employed, as in FM radio broadcasting, to reduce the noise introduced by transmission. An FM demodulated signal can be shown to have rising noise density with frequency. In order to compensate, preemphasis is used. Above a certain frequency, the gain in the modulator increases 6 dB per octave (doubling of frequency). The demodulator employs deemphasis, which attenuates the signal by the same amount. In the process, the noise spectrum is flattened, and the total noise heard by the listener is reduced.

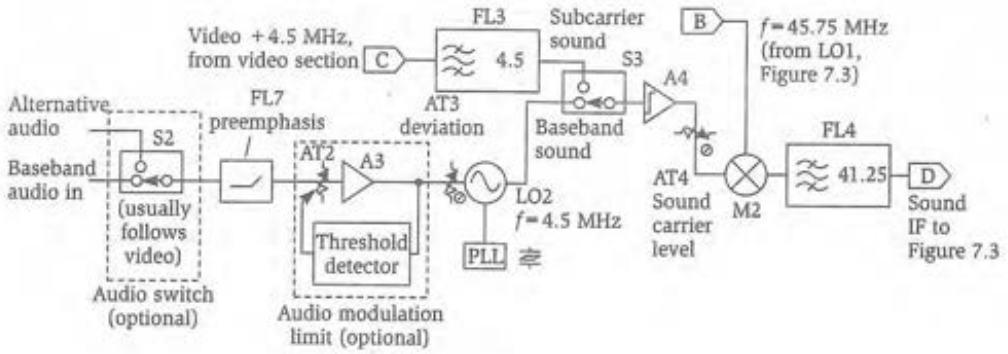


Figure 8.5 Modulator block diagram, aural.

In North America, the standard emphasis curve followed is called the *75-microsecond* emphasis curve because the break frequency, at which emphasis is 3 dB, is given by a 75-μs RC break:

$$\begin{aligned} e_0 &= (1 + j\omega RC)e_i \\ &= [1 + j\omega(75\mu s)]e_i \end{aligned} \tag{8.2}$$

where

$$j = \text{imaginary operator}$$
$$\omega = \text{angular frequency } (\omega = 2\pi f)$$

The corresponding 3-dB frequency, at which the imaginary component is equal to 1, is 2.122 kHz. In Europe and elsewhere, it is common practice to use an emphasis time constant of 50μs, corresponding to a 3-dB break frequency of 3.183 kHz. This affords slightly less reduction in noise, all else being equal, while providing slightly more protection against overmodulation.

The issue of overmodulation with constant amplitude signals arises in the following manner. Consider that you are normalizing the audio level of a program using any of several accepted criteria. You should not set program level based on a preemphasized signal because that is not the way a listener hears the program. On the other hand, the program is transmitted with preemphasis. If the dominant program material occurs at a frequency significantly above the preemphasis break frequency, then overmodulation is possible with normal program level audio. Fortunately, most program material tends to be dominated by frequencies below either the 75-μs or the 50-μs break frequencies. However, sometimes this is not the case, and problems can develop. The 50-μs break frequency provides somewhat more immunity. In order to compensate for the

reduction in signal-to-noise ratio caused by the higher emphasis break frequency, countries using 50-μs preemphasis normally employ wider deviation of the sound carrier (50 kHz compared to 25 kHz for 75-μs countries).

When stereo audio is used, the preemphasis is more complex. The preemphasis circuit in the audio baseband circuits shown is disabled so that the input is flat. Stereo audio is covered in Chapters 2 and Section 8.6.

Aural Modulation Limit, or AGC

An optional modulation limiter may be provided. This circuit monitors the audio signal level, and if the level increases such that the deviation would be excessive, then the modulation limiter temporarily turns down the gain of the signal path by adjusting AT2 of Figure 8.5. After the overmodulation ceases to exist, the gain is returned to its previous value over some length of time, which reduces the audible impact of the change.

Some modern modulators include a second mode of operation, in which the audio level is measured and AT2 is adjusted to provide for a consistent sound level regardless of the incoming modulation. This mode would not be appropriate for a number of audio sources since it will make low-volume sounds the same level as high-volume sounds. (A whisper and a shout would be the same loudness.) It might be appropriate, though, if the program audio consists of normal speech only.

Some modern modulators also include audio clipping, in which audio that would result in significant overdeviation has its peaks clipped, eliminating the overdeviation. In the United States, the FCC requires that a television aural transmitter be able to deviate to at least ±40 kHz, so the appropriate clip point would be above this. The clipping process produces considerable distortion, so operation into the clipping region should never be done intentionally. However, it is a useful way to avoid gross overdeviation, which could cause interference into video or, worse, problems with scrambling recovery (see Chapter 21). It may be applied at all times, so long as the audio level is carefully set to avoid hitting the clipping level during normal programming.

Modulation Process

Audio is frequency modulated onto an aural carrier. Frequency modulation is usually accomplished by applying the audio directly to an oscillator, such that the frequency of oscillation depends in part on the instantaneous modulating voltage. As shown in Figure 8.5, the modulation is applied to LO2, a 4.5-MHz oscillator (other frequencies may be used with conversion).

Some method must be employed to stabilize the center frequency of the oscillator. Though not the only means to accomplish this, a commonly used stabilization method is a phase locked loop (PLL). It generates a frequency correction to the oscillator if the center frequency wanders. The bandwidth of

the PLL is very low to avoid having the loop "Fight" the modulation, which is intentionally forcing the frequency of LO2 to deviate instantaneously.

Audio from Other Sources

If audio is supplied already modulated on a 4.5-MHz subcarrier, then it is introduced into the audio section via flag C (see also Figure 8.4). At this point, the aural subcarrier may be combined with the baseband video, so bandpass filter FL3 is used to separate out only the sound subcarrier. It is passed to one input of switch S3, which selects either the audio from the modulated oscillator, LO2, or the external signal. Switch S3 may be actuated permanently by manual selection of the audio source, or it may be activated by some other event.

A common situation is to activate the switch upon loss of main video. Suppose a modulator ordinarily handles a signal delivered at baseband from a television station. The baseband link may be backed up by an RF link that picks up the off-air signal and demodulates it. The primary audio is delivered in baseband format, and so is applied to the main input to switch S2. The backup demodulator supplies sound on a 4.5-MHz subcarrier, combined with the video. If the video switch (S1 of Figure 8.4) detects loss of video, it activates to couple in alternative video from the backup demodulator. At the same time, it drives the 4.5-MHz switch (S3 of Figure 8.5) to select the external aural subcarrier of flag C.

Following the switch is a limiter whose purpose is to ensure that the sound carrier level is constant regardless of the source. After the limiter is a manual sound carrier level adjustment, AT4, which sets the level of the sound carrier relative to the visual carrier level. Finally, the aural carrier is up-converted from 4.5 MHz to the sound IF at 41.25 MHz using the same local oscillator as was used to generate the picture IF (LO1 of Figure 8.4). The aural IF is coupled at flag D back to Figure 8.4, where it is combined with the visual IF signal.

8.3.7 Vestigial Sideband Filtering

As described earlier, the majority of one sideband is removed by a device known as a vestigial sideband (VSB) filter (FL2 of Figure 8.4). Figure 8.6 illustrates the operation of the vestigial sideband filter in an NTSC modulator. Recall that the spectrum is shown at IF, intermediate frequency, where the picture carrier is higher in frequency than the sound carrier. This is reversed at RF, by the use of a high-side local oscillator.

The VSB filter has a response that is flat from 0.75 MHz above the picture carrier to at least 4.2 MHz below. The reason for the cutoff at 4.2 (really 4.18) MHz is that this is the maximum baseband video frequency that is handled in the NTSC transmitted signal. The color carrier and all its sidebands are contained within the passband of the VSB filter.

The high (vestigial) side cutoff is rather carefully defined. The response should be flat to 0.75 MHz above the picture carrier, then drop to zero at the

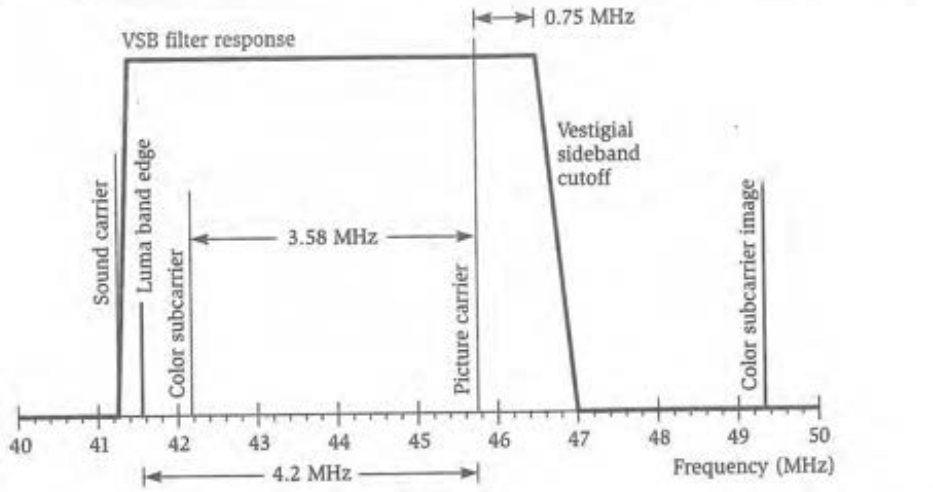


Figure 8.6 Modulator vestigial sideband filter response.

band edge, 0.5 MHz higher. This frequency, 1.25 MHz above the picture carrier, is at the channel edge, 47 MHz at IF.

No filter has truly zero stopband response, so it must be decided how much attenuation is "enough" at the band edge and frequencies farther out. In the section on processors, we argued that the power spilling into an adjacent channel should be at least 60 dB down from the picture carrier. This number is applied to the VSB filter also. The requirement does not mean that the filter must exhibit a stopband response of −60 dB because the power of the modulated TV waveform is not constant throughout the channel. Most of the power is concentrated at the picture carrier. Typically, other carriers are −15 dB or lower with respect to the picture carrier. The only frequency removed by the VSB filter that is likely to be that high is the color subcarrier image at about 49.3 MHz. Thus, if the VSB filter has 45-dB rejection, that is good enough. Even if the filter has only 40-dB rejection, that may be adequate given a small attenuation by other filters.

In some modulators, the sound signal may be supplied through the VSB filter, in which case a slightly wider response than shown is needed. However, with the modulator architecture of Figures 8.4 and 8.5, this would not be necessary.

The critical requirements for the VSB filter made it an early candidate for use of a surface acoustic wave (SAW) device, and in fact, this is the first known application of SAW devices in the cable television industry. SAW devices exhibit sharp cutoff regions, good temperature stability, and low group delay. However, they do exhibit higher propagation delay than do conventional L–C filters. Propagation delay is the absolute delay of a signal as it propagates through the

device. The propagation delay is related to the construction of the SAW device and has decreased somewhat with improvements in the state of the art. All else being equal, the sharper the skirts (the narrower the rolloff region), the greater will be the propagation delay. Delay times of 3 to 5 microseconds tend to be typical of SAW filters built for this application. The propagation delay can be an issue when events must be synchronized between those transmitted on the picture and sound carriers. Such a situation develops in the commonly used scrambling systems where synchronization data is transmitted on the sound carrier (see Chapter 21).

A SAW filter utilizes the propagation of an RF signal on the surface of a suitable material, such as a quartz or lithium niobate crystal. A pair of transducers is built onto the surface of the crystal, using the same techniques as are used to deposit metalization on the surface of integrated circuits. One transducer launches the signal into the crystal, where it propagates as a mechanical wave on the surface. The other transducer converts the wave back to an electrical signal. The shape of the transducers determines the passband of the filter. The loss of a SAW filter tends to be rather high, but the skirts can be made very steep without incurring the group delay that accompanies a sharp L–C filter.

Digital Modulators

Digital modulation is covered in Chapter 4. In general, digital modulators are simpler than their analog cousins. Interfaces with digital modulators are shown later. One issue of concern is when scrambling is added in a digital transmission stream. We use *scrambling* in the cable TV sense of the word, "to render the signal unintelligible to those who are not authorized to receive the signal." In some headend systems, there is an external scrambler that adds scrambling prior to the modulator. In other headend systems, the scrambling is an integral part of the modulator. In some cases, the modulator includes the up-converter to the final channel assignment; in other cases, the up-converter will be external to the modulator.

8.4 Phase Locking of Carriers

The output carriers of both processors and modulators may be phase locked to external sources. Two primary reasons exist for phase locking the output. If the frequency is used for local off-air broadcasting, then there is a high probability that off-air signals will leak into some television receivers or, much less likely, into the cable plant itself. This leakage, or *ingress*, will produce a visible moving pattern on the screen if the on-cable picture carrier is not exactly equal to the ingress frequency. The industry commonly calls this pattern a *beat*. The processor or modulator output can be phase locked to the off-air signal to reduce the visibility of the ingress. This works because the phase locking process forces the frequency of the signal on the cable to be exactly equal to that of the ingress.

The other application for phase locking is to lock all picture carriers on the cable to a master reference frequency. Such techniques can be shown to reduce the interference caused by distortion in the distribution plant. Two techniques have seen fairly widespread use in North America. They are the harmonically related carrier (HRC) technique, in which all picture carriers are locked to harmonics of a master reference oscillator, and the incrementally related carrier (IRC), where all frequencies supplied from the master generator are offset 1.25 MHz. Neither HRC nor IRC techniques tend to be used as much today as in the past, but they will still be found in some locations. Chapter 9 describes the standard HRC and IRC frequency allocation plans.

Yet a third possible requirement for phase locking outputs is described in Chapter 9, in the discussion of node-specific programming. More than one program may exist in a headend on the same channel, and if leakage develops from one to the other, phase locking the two picture carriers will lessen the visibility of the beat. This is true if the subject program is analog. If digital, phase locking is probably not necessary.

The term *phase lock* refers to a technique of forcing one RF carrier to maintain a constant phase relationship with a second carrier. The second carrier is called the *reference carrier* because it is a reference to which the first is locked. If the two carriers have a constant phase relationship to each other, it follows that they are identically at the same frequency. Indeed, a common technique for verifying a phase locked condition is to display one carrier on an oscilloscope that is triggered from the other. The displayed waveform should appear stable on the screen. Note that it is not necessary that the two carriers have a particular phase relationship to each other (though some applications require this), and in general, the phase relation might drift over time.

8.4.1 Phase Locked Circuits and Agile Outputs for Processors and Modulators

Figure 8.7 illustrates one possible block diagram of a phase locked circuit. This can be used with either Figure 8.1 (processor) or Figure 8.4 (modulator). In either case, the output local oscillator, LO3, is replaced by the circuit of Figure 8.7. Mixer M3 is the mixer M3 of Figures 8.1 and 8.4.

The reference frequency, a sample either of the off-air signal or from the comb generator (for HRC and IRC applications) is selected in bandpass filter FL10. It is mixed to the common intermediate frequency of 45.75 MHz in mixer M5. The local oscillator used in this mixing process must be LO3, the same local oscillator used to mix the output channel.

After conversion to IF, the signal is filtered in FL12 before being supplied to limiter A11. This limiter must be used because, if the reference source is an off-air signal, it will have modulation on it, and the modulation must not interfere with the mixing process. From the limiter, the signal becomes one input to phase detector M4. The other input to M4 is a sample of the IF signal developed in the

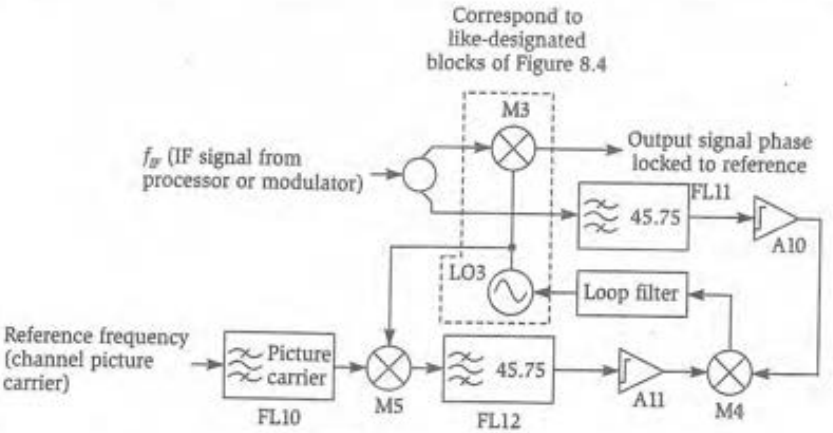


Figure 8.7 Phase lock option for processor or modulator.

modulator or processor. This signal has been filtered (if needed) and limited in FL11 and A10.

Again, a multiplier is shown as the phase detector. A multiplier makes a good phase detector, just as it makes an ideal modulator or mixer. If the reference signal (at frequency ω_r) and the variable signal (at ω_v) are multiplied, the result is expressed as follows:

$$\begin{aligned} E_0 &= \sin \omega_r t \sin(\omega_v t + \theta) \\ &= \frac{1}{2}\cos 2\omega t - \frac{1}{2}\cos \theta \end{aligned} \tag{8.3}$$

where

$$\omega = \omega_r = \omega_v$$

This expansion uses the relation of Equation (2.5). The first term in the second line represents the second harmonic of the carrier frequency and is removed by the loop filter. The second term represents the phase difference between the two carriers, θ.

The output of the phase detector is supplied to the loop filter, which rejects the second harmonic and performs required filtering on the error voltage representing θ. These functions often include providing extremely high gain at dc, which forces $\cos\theta$ to be near zero. The phase locked loop will do whatever is necessary to the frequency of LO3 to keep θ constant. Note that, in order for the cosine of the phase angle to go to zero, the two signals at the mixer will be held

in quadrature, that is, with a 90° phase difference. This phase relationship is typical of multiplicative phase detectors.

If it is not intuitive that the output frequency is phase locked to (that is, the same frequency as) the reference, you are encouraged to write equations for the output frequency in terms of the frequency of the IF, of LO3, and of the reference frequency. Assume that the frequencies of the two inputs to M4 are equal (a condition forced by the phase locked loop), and solve for the output frequency in terms of the reference frequency.

Some modern agile modulators offer an alternative way to achieve phase lock between channels. If all relevant frequencies in all modulators are synthesized from the same reference frequency, then phase lock is accomplished.

Comb Generation Technique

Though not used extensively today, it is instructive to consider briefly how a comb of frequencies might be generated. The comb is used for phase locking processors and modulators in HRC and IRC frequency plans. Figure 8.8 illustrates an HRC comb generator in part a and the modification to make it into an IRC generator in part b.

HRC Comb Generation

The HRC comb is simply all the useful harmonics of a 6-MHz (for NTSC) generator. A simple way to produce the comb is illustrated in Figure 8.8(a). A precision 6.0003-MHz oscillator is used to generate the basic frequency. (The nominal frequency is 6.000 MHz. See Chapter 9 for an explanation of why an oscillator at 6.0003 MHz is used.) After amplification, the signal is supplied to a fast recovery diode and to an inductor. When the signal from A10 goes positive, D10 conducts, setting up current in L10. When the signal from A10 begins to go negative, D10 becomes back biased and cannot conduct. The input impedance to A11 should be high so that no significant current exists in its input. When D10 is back biased, a current path to complete the circuit through L10 does not exist. If a closed circuit does not exist, then the current through L10 must stop. However, if you try to instantaneously interrupt the current in an inductor, the voltage across the inductor must go to infinity (less in a practical case, due to losses in the circuit). This produces a voltage spike that approximates an impulse. An impulse may be shown to contain all harmonics of the fundamental frequency at equal amplitudes. After amplification in A11, the resultant may be applied to filters that remove unneeded harmonics. The small spectrum diagram near the output illustrates the comb.

IRC Comb

Figure 8.8(b) illustrates how the IRC comb is developed from the HRC comb of 6.000 MHz. The IRC comb is the HRC comb except offset up in frequency by

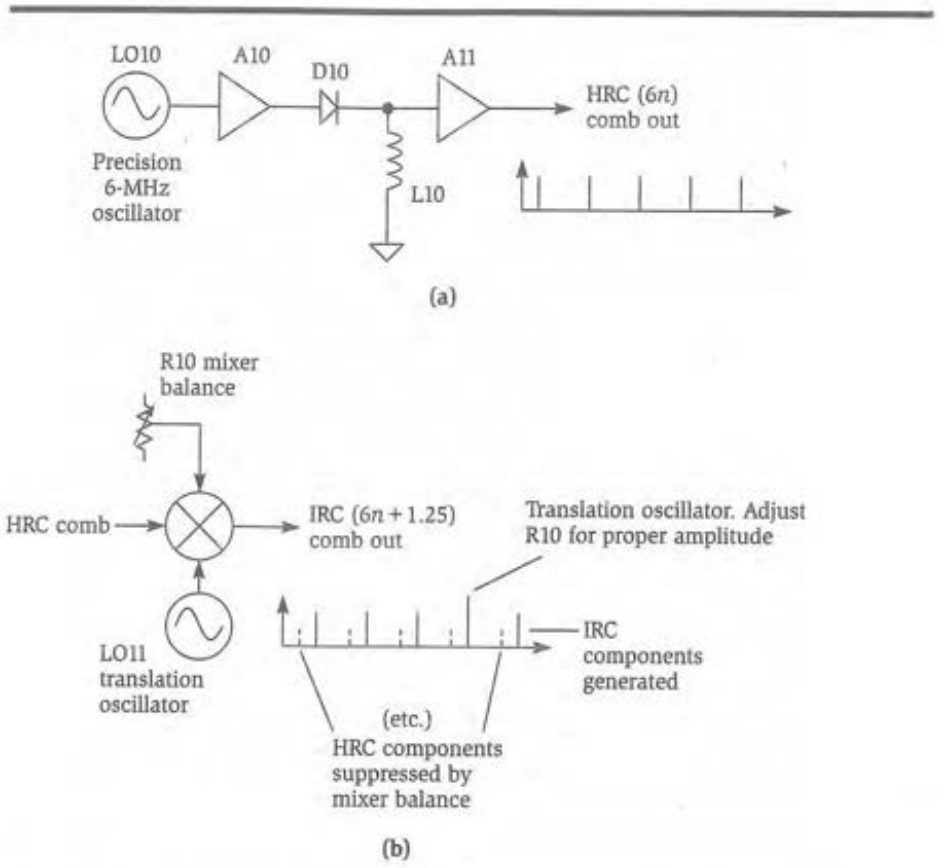


Figure 8.8 Comb generator block diagrams. (a) HRC comb generator. (b) IRC comb generator.

1.25 MHz. This is done by band limiting the HRC comb and using it as modulation on a carrier positioned at a picture carrier frequency of about one-half the maximum frequency. The effect is to generate a carrier at the center of the comb spectrum at one of the comb frequencies. The modulation sidebands are at the other frequencies required of the HRC comb. A double balanced mixer is used for the translation. Since it will try to suppress the frequency of LO11, a mixer balance adjustment is used, and the mixer is intentionally unbalanced to the translation oscillator, enough to cause the translation carrier signal to bleed through at the same level as the sidebands. The HRC comb elements should be completely suppressed, but if they are suppressed by only a few decibels, the phase locked loop in the processor or modulator should not have any trouble recovering the correct element.

Dual Conversion

Figure 8.9 illustrates two features of modern headend processing. One is the use of an agile up-converter to allow the output of a modulator (analog or digital) or CMTS to be placed on any frequency. This generally requires a dual-conversion up-converter to provide low spurious outputs. The genesis of the spurious-output problem is with the output filter after the final conversion. A simple example will show why this is so. Refer to Figure 8.4 for a moment; then we shall return to Figure 8.9.

In Figure 8.4, suppose it was desired to make the modulator frequency agile so that the output could be set to any channel, maybe within a range of several hundred megahertz. We can certainly operate the output local oscillator, LO3, on any frequency (using a phase locked loop to control its frequency), and this will allow us to set the output frequency to any frequency we want. The problem is that we have filters FL5 and FL6, which are tuned to a single output frequency. We will have to either eliminate these filters or make them electrically tunable if we are to allow the output to be on any frequency over a wide range.

But there were some very good reasons as to why we have these filters and why they are only one channel wide. Consider that we want to put the output on channel 13, with the picture carrier at 211.25 MHz (see Appendix A). With a picture IF of 45.75 MHz, this means that the frequency of LO3 will be $45.75 + 211.25 = 257$ MHz. Mixer M3 produces the difference frequency between LO3 and the IF, 211.25 MHz. But the mixer is quite happy also to produce the sum frequency, $257 + 45.75 = 302.75$ MHz, the image of the conversion. We can't let this frequency out of the modulator because it will interfere with channel 37 (and the sound carrier with channel 36). That is one function of filters FL5 and FL6 in Figure 8.4.

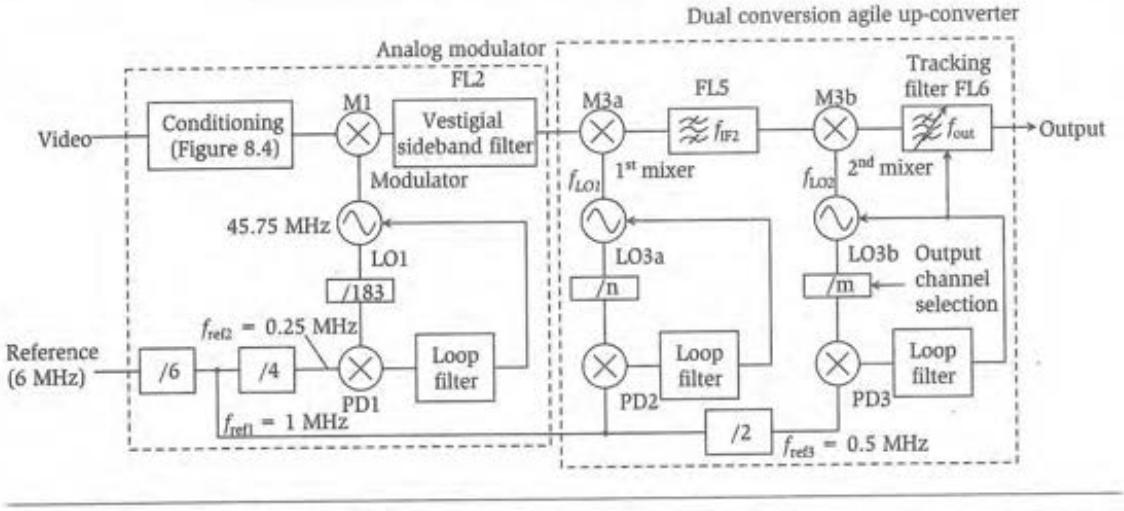


Figure 8.9 Dual-conversion output with optional phase lock to a 6-MHz reference.

So if we want an agile output, we have to modify the block diagram of Figure 8.4. We can use a dual-conversion system such as in Figure 8.9. Like components in Figures 8.4 and 8.9 have the same reference designations, so you can see how they perform similar functions. Video is supplied to all the circuitry shown at the top of Figure 8.9, which we have referred to here simply as *Conditioning*. The actual modulation is done in M1, just as in Figure 8.4. M1 is driven by the IF signal from local oscillator LO1. For phase lock purposes in Figure 8.9, LO1 will need to be phase locked; but it can be simply crystal controlled if phase locking is not to be done (or if it is to be done using the method shown earlier).

The IF signal passes through vestigial sideband filter FL2, just as in Figure 8.4. (Note that this diagram might apply to a digital modulator or to a CMTS, in which case the filter function is still there but is not a vestigial sideband shape and would not be called that.) For simplicity, we'll omit the sound carrier and the IF loop, though they are present. Now what we are going to do that is different from Figure 8.4 is to pass the signal to a dual-conversion agile up-converter. This one is phase locked to an external 6-MHz source, although that is not necessary. The two local oscillators involved in the dual-conversion process will probably be phase locked to a reference, but the reference may be internal to the up-converter rather than external as shown.

The first local oscillator will up-convert the IF to some high frequency. For illustration, assume it converts to a second IF of 927.25 MHz, a somewhat arbitrary frequency. The first local oscillator, LO3a, thus operates at $f_{LO1} = 45.75 + 927.25 = 973$ MHz. The second IF filter, FL5, is set to this frequency and has a width of at least one channel. Its main job is to remove the image of the first IF, $973 + 45.75 = 1018.75$ MHz. We don't show amplification, but there may be some amplification at this second IF frequency before the signal is converted to the final output frequency in mixer M3b, driven by local oscillator LO3b. If we want the output frequency to be 211.25 MHz for channel 13, the frequency of LO3b will be $927.25 + 211.25 = 1138.5$ MHz. Notice that there is still an image of the second IF, but the image frequency is $1138.5 + 927.25 = 2065.75$ MHz, well above any frequency we are interested in on the cable plant, and so far from the output frequency that it is easy to filter.

There is a need for a final filter, however. We show a tracking filter at FL6, which is located after any amplification we provide at the output (amplification not shown). The reason is that there is some noise generated in the mixing process and in amplifiers. If we don't filter this noise, it will be combined with all the other signals and could harm the output C/N of the headend. Thus, a tracking filter can be provided to eliminate this noise. As a practical matter, no tracking filter can simultaneously provide excellent adjacent channel attenuation and flat response in-band while being tuned over a wide range. Thus, some compromises must be made in its ability to attenuate noise in close-in channels versus the flatness and tunability of the filter.

We have shown an alternative way to provide phase lock in Figure 8.9. This uses a reference, in this case 6 MHz, to phase lock all oscillators. The 6-MHz

reference is divided, in frequency dividers, to frequencies that allow it to serve as a reference for all three oscillators. If phase locking is not needed, then there is no need to phase lock the first IF oscillator, LO1. Local oscillator LO3a can either be phase locked or crystal controlled, at the manufacturer's discretion. Local oscillator LO3b will typically be phase locked in any case, since it must be changed in frequency as you set the up-converter to different channels.

Yet another consideration with phase locked oscillators is phase noise produced by the oscillator. For several reasons, the phase noise is typically much greater from a phase locked oscillator than it is from a crystal-controlled oscillator. Phase noise can hurt the bit error rate of a digital signal, as explained in Chapter 4. For analog signals, phase noise can be converted to amplitude noise in the receiver's Nyquist slope filter and can damage the signal-to-noise ratio of the demodulated signal. Considerable effort and expense go into minimizing the phase noise of oscillators in agile up-converters.

8.5 Demodulation

A third type of RF product used in the headend is a demodulator. It is useful to understand the demodulation process in order to make sense of a lot of phenomena observed in cable systems. Also, demodulation is used for quality checks, for recovery of off-air signals for backup purposes, to put signals on a digital transportation route, and as interfaces for FM microwave systems. A few headends employ demodulation and remodulation instead of processing for off-air signals. This is advantageous if considerable signal switching is done. The use of a quality demodulator as a test tool is, alone, justification for studying the technique.

8.5.1 RF Input Section

Figure 8.10 shows the RF, IF, and video demodulation portions of a demodulator. The sound section is shown in upcoming Figure 8.11. The demodulator input is identical to that of the processor of Figure 8.1. Both processor and demodulator are used to recover signals from off the air or from an incoming cable. Thus, the input circuitry through attenuator AT2 is identical.

8.5.2 IF Processing Section

IF processing in a demodulator is significantly different from that of a processor. Power for demodulation of the sound carrier is taken at flag A. Sound demodulation will be covered later in conjunction with Figure 8.11. The AGC operates to hold the visual signal level after AT2 at a constant level regardless of the input level.

Zero chop switch S1 is optional but is often provided on professional demodulators as a way to measure visual signal depth of modulation. When

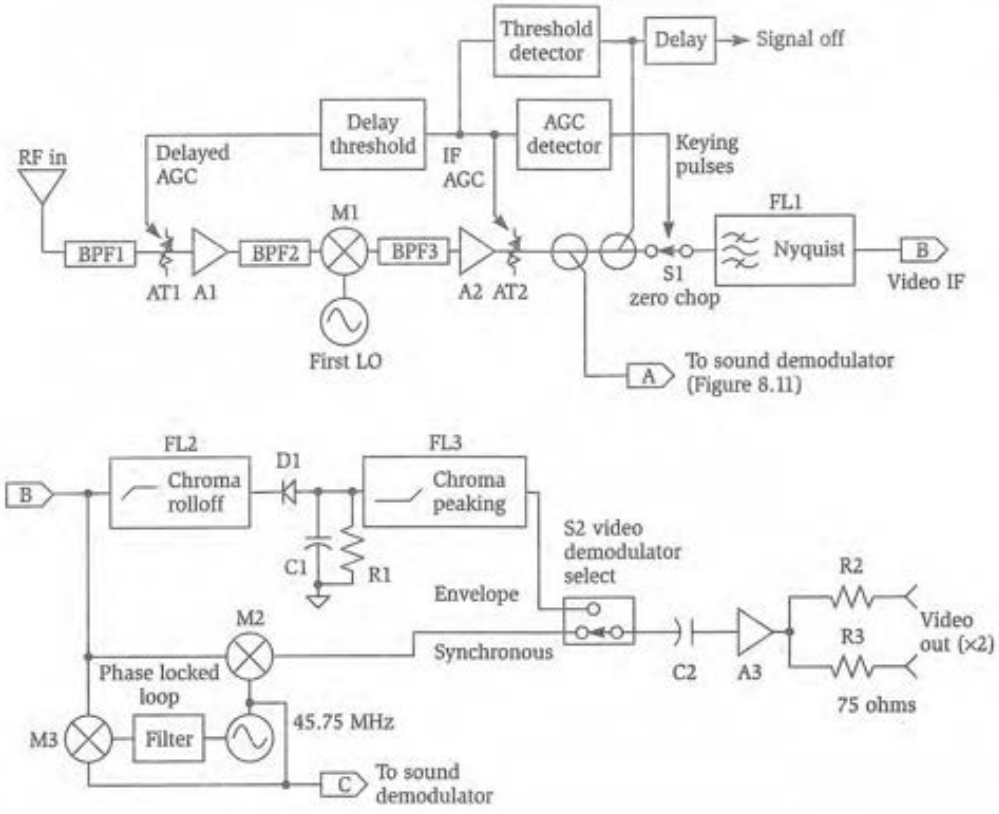


Figure 8.10 Demodulator block diagram, input and video.

enabled, the chopper momentarily interrupts the signal at IF, usually during the vertical blanking interval. The interruption allows a waveform monitor or oscilloscope to be calibrated to measure visual depth of modulation very accurately. Chapter 9 describes measuring depth of modulation using a zero chopper. The chopper is coupled to the AGC detector in the illustrated case so that the switch may be restored to the IF path during sync pulses. By doing this, the zero chop pulse may extend over several lines without interrupting sync.

From the zero chop circuit, the signal is supplied to the Nyquist slope filter, FL1, which provides the main filtering function for the visual signal. The filter is named after Harry Nyquist, a Bell Laboratories engineer who developed this technique for recovering vestigial sideband signals. The filter linearly reduces the amplitude response in the double sideband region (picture carrier ±750 kHz) such that the picture carrier is reduced 6 dB. The reason for doing this is to equalize the frequency response in the double sideband and single sideband regions.

8.5.3 Visual Carrier Demodulation

After the Nyquist slope filter, the signal is supplied to one of two demodulation circuits, an envelope detector or a synchronous detector. Home television receivers use one detector or the other, but it is advantageous to have both in a professional demodulator because the two will yield different information about a signal. Some of the differences are discussed in Section 8.5.7.

Envelope Detector

In the illustrated system, the envelope detector consists of a chroma rolloff filter, FL2, and following circuits. The chroma rolloff filter reduces the amplitude of the IF signal at the color subcarrier frequency. The purpose of the chroma rolloff filter is explained later. Diode D1 performs the actual detection process, and capacitor C1 filters the signal. The detector, except for component values, looks just like the detector for an AM broadcast receiver.

The direction of the diode is not arbitrary. Recall from Chapter 2 that the sync tip is the most negative portion of the video signal as the signal is normally interfaced. The sync tip also corresponds to the highest amplitude of the signal. With the diode in the direction shown, the voltage across C1 will be such that the highest amplitude of the RF signal (the sync tip) will be the most negative voltage across C1. Other voltages will be positive with respect to that. Of course, in a real demodulator, it is possible to detect the signal with the opposite polarity and then invert the video, but this is often not done for practical reasons.

Following the detector is the chroma peaking filter FL3, which restores flat chroma response, complementary to the operation of the chroma rolloff filter FL2. The signal is then applied to switch S2, which selects either the envelope detector or the synchronous detector.

Synchronous Detector

The synchronous detector is shown below the envelope detector in Figure 8.10. It consists of a phase locked loop used to generate a carrier phase coherent with the incoming picture carrier (at IF). In contrast to other applications described earlier, the phase of the locked carrier with respect to the input IF signal is of paramount importance: the two carrier phases must be the same. Because the phase of the carrier varies with modulation and, in addition, can vary owing to defects in transmission, the phase locked loop is normally keyed (closed) on either sync tips or the back porch portion of the modulated signal.

When a modulated carrier is multiplied with an unmodulated carrier at the same frequency, the result is recovery of the baseband signal. The advantage of the synchronous detector over the envelope detector is primarily associated with demodulation of signals having asymmetrical sidebands, which is the case for all

analog television signals. An envelope detector introduces distortion that is avoided with a synchronous detector.

From the synchronous demodulator M2, the signal is low-pass filtered to remove the carrier and harmonics, and supplied to S2, the switch that selects either the envelope or synchronous detectors.

8.5.4 Video Output

Switch S2 is normally actuated manually to select either the envelope or synchronous detectors if both are provided. By comparing the results of demodulation using the synchronous and envelope detectors, we can deduce a lot about the performance of the RF components handling the signal. Under normal conditions, the synchronous detector will yield a more accurately demodulated signal than will the envelope detector.

A coupling capacitor C2 removes dc components of the demodulation process, as well as any bias voltages introduced in the demodulation process. It has been customary to ac couple the video output since this minimizes the power supplied or dissipated in the load and often simplifies circuitry. Today, both dc and ac coupling of the output are employed. If dc coupling is employed, capacitor C2 would likely be used anyway, with a dc restoration circuit in the output amplifier.

It is essential with video that the coaxial cable impedance be matched both on the source end and the load end. Failure to do so can cause unacceptable frequency response errors, due to reflections, if the interconnecting cable is long enough. A typical way to achieve multiple well-matched outputs is to provide a very low output impedance in amplifier A3 and multiple matching resistors R1 and R2 at the outputs.

8.5.5 Aural Carrier Demodulation

The aural demodulation process is shown in Figure 8.11. The sound IF is supplied from Figure 8.10 on flag A. The other input to the sound demodulation

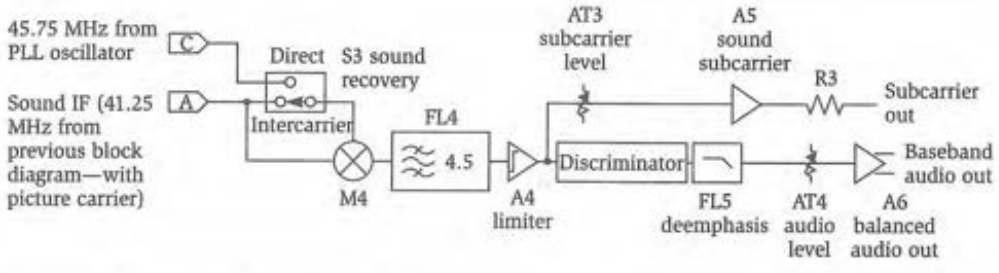


Figure 8.11 Demodulator block diagram, audio.

process is a sample of the signal from the 45.75-MHz phase locked oscillator, used in the synchronous demodulator (flag C, which is omitted if a synchronous detector is not supplied).

The intercarrier detection mode is the one used almost exclusively in consumer applications. The sound final IF frequency, normally 4.5 MHz, is derived from mixing the sound first IF, 41.25 MHz, with the picture carrier, 45.75 MHz. This final sound IF frequency is often called the *intercarrier frequency*. Use of intercarrier detection solves several problems related to television transmission. If any process in the transmission of the television signal introduces phase noise, then the phase noise will seriously damage the signal-to-noise ratio of the audio: phase modulation is a form of frequency modulation, so once phase noise is introduced, it cannot be removed. However, since phase noise will affect the picture and sound carriers identically, when the intercarrier signal is developed by taking the difference in the picture and sound carriers, phase noise is eliminated from the aural signal. The IF signal on flag A contains both the aural and picture IF carriers, so by applying this signal to both inputs of a mixer, intercarrier mixing results. (Additional filtering may be used.)

On the other hand, if phase noise is not a problem, then there are advantages to a direct detection process. The picture carrier may have some incidental phase modulation on it, and it certainly has visual modulation sidebands. These will cause interference with the aural signal if used in the intercarrier detection process. In direct detection, the CW signal locked to the picture carrier is mixed with the sound carrier to derive the intercarrier frequency. (In this mode, the sound carrier path may include filtering to remove the picture carrier though this filtering is not shown in Figure 8.11.)

After the intercarrier detector, M4, the signal is filtered in a 4.5-MHz bandpass filter FL4. Following the bandpass filter is a limiter, A4, which removes any residual amplitude modulation from the aural carrier. At this point, a sample of the aural carrier is usually taken. It is passed through a manually controlled attenuator AT3, amplifier A5, and a current source resistor R3. The purpose is to provide a sound subcarrier for external use. Often this signal is combined with the picture carrier to provide a signal to be transmitted by FM microwave link. The resistance of R3 is high compared with 75 ohms. If one of the video outputs on Figure 8.10 is paralleled with the subcarrier out connector, then R3 will act as a current source, adding the 4.5-MHz subcarrier to the video without significantly upsetting the impedance of the video circuit.

The other path from the limiter is through a discriminator, which recovers baseband audio. Note that the signal recovered at this point is either monaural (mono) or a stereo composite signal, as defined in upcoming Section 8.6. If the signal is mono, then a deemphasis filter FL5 is included. If it is desired to recover a composite stereo signal for subsequent demodulation or processing, the deemphasis is omitted in favor of more complex processing in the stereo decoder.

Finally, the signal level is adjusted at attenuator AT4, normally a manual control. It is amplified in output amplifier A6 and provided on balanced outputs.

Balanced outputs are used in professional equipment to provide for reduction of pickup of 60-Hz ac components ("hum"). "Balance" refers to a transmission technique in which the two sides of the output have equal impedance to ground, as opposed to unbalanced transmission, where one side of the circuit is at ground potential. Section 8.9.2 describes use of balanced interconnection for audio.

From a long-standing telephone specification, the impedance of the audio output is normally 600 ohms between the two balanced lines. In much equipment today, the source impedance is less than 600 ohms, but the performance of the amplifier is specified when the amplifier is working into a 600-ohm load. The lack of a defined 600-ohm output impedance is not a problem in most cases since the length of audio cable is very small compared with the wavelength of any audio frequency. If the output is to drive a line several miles long, then the impedance should be controlled more carefully.

8.5.6 Nyquist Slope Filter

The Nyquist slope filter shown as FL1 in Figure 8.10 has a shape shown by the solid line in Figure 8.12. At the low-frequency side (aural carrier at IF), the shape is fairly straightforward. The bandpass characteristic should pass the color carrier and its sidebands while rejecting the sound carrier.

The high-frequency side (near the picture carrier) is rather unusual. The purpose of this shape is to complement the shape of the vestigial sideband (VSB) filter used in the modulator, and shown in Figure 8.12 as a dotted line. The

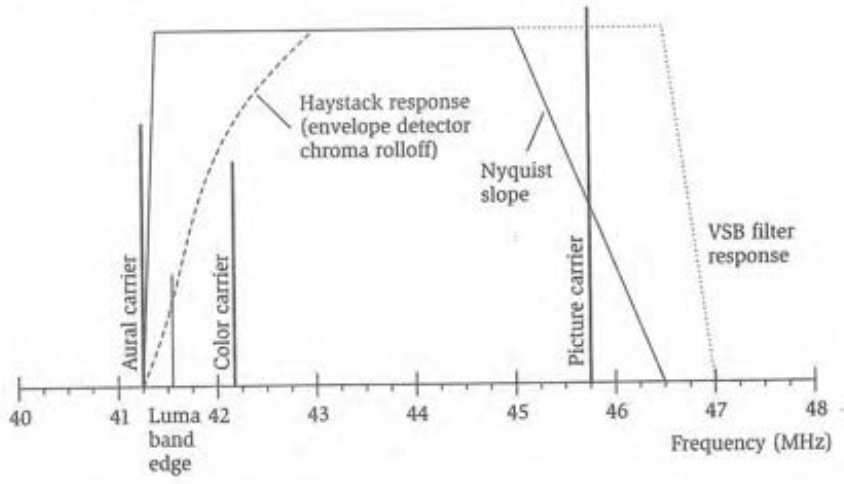


Figure 8.12 Nyquist slope filter response, with haystack.

Nyquist shape is flat to 0.75 MHz below the picture carrier; then it begins falling off at a prescribed rate above that frequency. The rolloff must be linear with frequency on a voltage plot, not a logarithmic plot. The response of the filter must be 0.5 at the picture carrier, continuing to drop until the response is zero at 0.75 MHz above the picture carrier.

Other features of the Nyquist slope filter are emphasized in the next section, in which we describe the differences between synchronous and envelope detection. One of those features is not really a part of the Nyquist slope filter response. It is shown in Figure 8.12 as a dashed line and was mentioned in conjunction with the chroma rolloff filter (FL2 of Figure 8.10). This is the reduction of response near the color carrier applied in the case of an envelope detector. It is called a *haystack* response after traditional television usage: taken with the Nyquist slope around the picture carrier, the filter shape looks like a haystack. The haystack response is used to reduce the effects of luminance-to-chrominance crosstalk.

In consumer television receivers using envelope detectors, the haystack response is built into the Nyquist slope filter (FL1 of Figure 8.10). In professional receivers, the chroma rolloff may be realized independently of the Nyquist slope filter so that it is not in the signal path in the synchronous detector mode. Not that the haystack response would not hurt the response of the synchronous detector if it were perfectly compensated at baseband. Rather, it is difficult to compensate perfectly, and for measurement purposes, it is better not to introduce a response error that must later be compensated.

In modern practice, the Nyquist slope filter is normally realized as a SAW filter in consumer receivers though the shape is compromised to allow the SAW filter to be built with good cost effectivity. In professional demodulators, the filter may be realized either with a SAW device or as a conventional L–C filter with delay equalization.

8.5.7 Synchronous Versus Envelope Detection

Both synchronous and envelope detectors may be used for operational demodulation. Synchronous detection yields the higher-quality results absent a few distortions that prevent its proper operation. The most common problem that interferes with operation is phase noise, or phase modulation, on the picture carrier. A synchronous detector will not work as well as does an envelope detector under this circumstance.

Envelope detectors are susceptible to the practical limitations of diodes though there are ways to mitigate this problem. More seriously, envelope detection is susceptible to quadrature distortion errors that arise as a result of asymmetrical sidebands in the television signal. Quadrature distortion occurs when the upper and lower sidebands are not symmetrical. As shown earlier and in Chapter 2, the television signal presented to the detector does not have symmetrical sidebands at any frequency: at modulation frequencies above 1.25 MHz,

only one sideband is transmitted. Below 0.75 MHz, symmetrical sidebands are transmitted but are turned into asymmetrical sidebands by the Nyquist slope filter. The lowest modulating frequencies approach sideband symmetry, but as the frequency increases, the symmetry is quickly lost.

Phasor Representation of the Modulated Signal

The mathematics of modulation were shown in Chapter 2. A useful graphical representation that shows the function of the two demodulators is to invoke the mechanism of phasors.[2] A phasor represents any steady-state sinusoid as a vector whose length is a function of the magnitude of the signal and whose angle is a function of the angle between a reference and the phasor. From Equation (8.1), an amplitude modulated signal can be represented as

$$\begin{aligned} V_0 &= \cos \omega_c t(1 + M \cos \omega_m t) \\ &= \cos \omega_c t + \frac{M}{2}\cos(\omega_c - \omega_m)t + \frac{M}{2}\cos(\omega_c + \omega_m)t \end{aligned} \tag{8.4}$$

Figure 8.13(a) represents this equation in phasor form. The first term, representing the carrier, is defined as having a phase such that its phasor is straight up (this is arbitrary). The upper sideband has a phasor that rotates clockwise (with respect to the stationary carrier) at an angular frequency of ω_m. The lower sideband term, $\cos(\omega_c - \omega_m)$, is represented by a counterrotating phasor. The resultant phasor is shown to the right in Figure 8.13(a). The vector sum of the three phasors is always in phase with the carrier. An envelope detector responds to this resultant vector. Analysis of the vector at all different phase angles θ would show that the resultant traces out a sinusoid, which is the waveform used to modulate the carrier (the $M \cos \omega_m t$ term of Equation (8.1)). All is well.

Now consider the situation shown in Figure 8.13(b), a single sideband signal. The modulating waveform is the same sine wave as in Figure 8.13(a), but one sideband has been removed by the VSB filter (FL2 of Figure 8.4). Most notably, one sideband of the color subcarrier is removed, so the color subcarrier is transmitted as a single sideband signal. The vector sum of the carrier and single sideband phasor is shown to the right of Figure 8.13(b). Here the resultant is *not* in phase with the carrier. The resultant can be resolved into an in-phase component and a "quadrature" component at right angles to the carrier phase.

Response of an Envelope Detector

The envelope detector responds to the magnitude of the resultant carrier. By applying trigonometric relations to the resultant vector of Figure 8.13(b), we can compute the magnitude of the vector for various conditions. Figure 8.14 illustrates the result obtained using an envelope detector for two different relative levels of

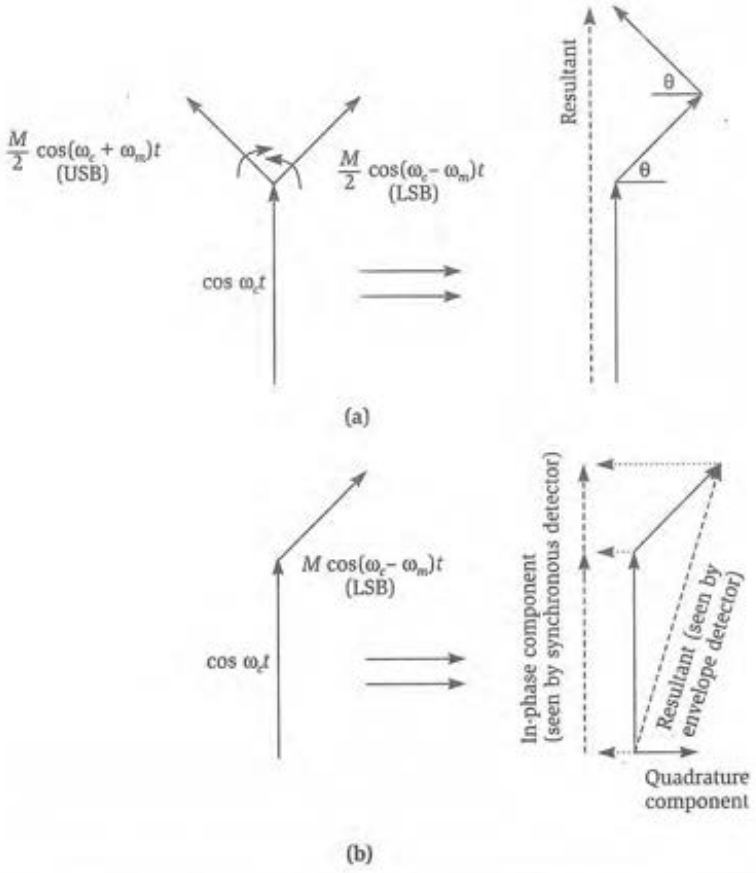


Figure 8.13 Phasor diagram, double and single sideband carriers. (a) Symmetrical sidebands. (b) Single sideband.

carrier and sideband.* The dashed line is the sinusoid with which the carrier was modulated. One of the two cases represented (the middle trace) results when the sideband is −6 dB (one-half the voltage) of the carrier. The third waveform represents the waveform produced by equal amplitude carrier and sideband. This is a realistic condition for a television signal: a scene with medium to high brightness has a relatively low carrier level during the lighter portions of the

* The Figure is displayed as would be a television signal, with the lowest resultant amplitude at the top. The gain has been normalized in the three cases to produce the same peak-to-peak signal.

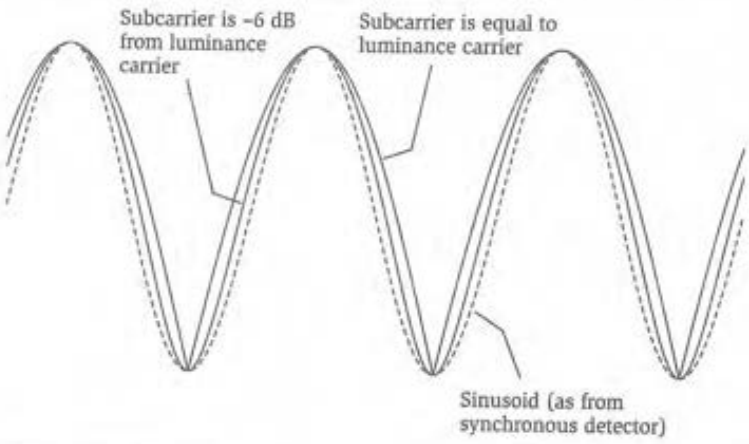


Figure 8.14 Demodulated results, envelope detector.

scene. If the object is brightly colored, the chroma carrier amplitude will be high, so when the camera is scanning the object, the chroma can be as high or higher in amplitude than the picture carrier. The resultant tends to be "scalloped," with the negative peak too wide and the positive peak too narrow.

From this illustration, it is evident why the chroma response of the Nyquist slope filter is reduced (haystacked) when an envelope detector is used (see Figure 8.12). The chroma carrier is the highest amplitude sideband of the luminance carrier when brightly colored objects are being scanned. The higher the chroma signal is with respect to the picture carrier, the greater the distortion. By reducing the chroma amplitude before detection, the effect is reduced. Of course, the chroma amplitude must be corrected before the signal is presented for measurement purposes; hence the chroma peaking circuit FL3 of Figure 8.10. (In a consumer TV set, it is not necessary to use this peaking circuit because the chroma is separated from the luminance, and the correction can be approximated by adding more gain to the chroma channel.)

If the Fourier transform of the chroma waveform is taken, it will be seen that the amplitude of the 3.58-MHz fundamental component is less than what it would have been without quadrature distortion. This causes the chroma saturation (proportional to chroma amplitude) to be lower than it should be — the color appears washed out compared with the way it should look. An envelope detector that doesn't reduce the amplitude of the color carrier prior to demodulation will exhibit excessive differential gain and luminance-to-chrominance and chrominance-to-luminance crosstalk.

The average level of the chroma signal is no longer in the center of the waveform of Figure 8.14. Rather, it is lower than it should be. This means that

the luminance level is lower than it should be, making the object appear darker than it should. Thus, a brightly colored object of relatively high luminance level (bright yellow is an example) will appear too dark and somewhat washed out owing to the quadrature distortion introduced by the envelope detector. Whether or not the distortion is objectionable depends on a number of factors.

One useful way to show the difference between envelope and synchronous detectors is to study the various sinusoids involved in the multiburst waveform. Switching back and forth between the two detectors will readily show the difference in waveform. Some bursts will show more distortion with an envelope detector than will others, depending on the sideband amplitude seen by the detector.

Another convenient test is to study the difference between a sine squared (2T) pulse when demodulated with synchronous and envelope detectors. The demodulated shape will be correct with the synchronous detector. With the envelope detector, the pulse will appear too narrow, and possibly too tall. This would have the effect on a picture of narrowing a vertical white line on the screen.

Response of a Synchronous Detector

The synchronous detector can be shown to respond only to the in-phase component of the phasor of Figure 8.13(b). In turn, it can be shown that the in-phase component, the projection of the sideband on the phase of the carrier, is an accurate representation of the modulation.

Application of Test Demodulators Having Both Detectors

Some demodulators have a second synchronous detector, in which the phase of the reconstructed carrier is 90° with respect to the incoming picture carrier (at IF). This quadrature detector is sensitive to the quadrature component of Figure 8.13(b). By displaying the quadrature component on the y axis of an oscilloscope and the in-phase component on the x axis, we can easily spot incidental carrier phase modulation (ICPM) in the modulator. Normally, a luminance ramp is supplied as video for this test. ICPM can be caused by improperly operated power amplifiers, and is a common problem with older broadcast transmitters, particularly UHF transmitters. It tends to be less of a problem with cable modulators because of lack of high power stages, but it can be introduced in an amplifier stage that is distorting. ICPM can cause excessive intercarrier buzz if it affects only the luminance carrier.

A problem with some older and lower-cost cable modulators is poor balance in the video modulator stage. This will distort the luminance level of bright portions of a picture. The presence of this may be deduced by measuring differential gain with the envelope and synchronous detectors. A good differential gain reading with the synchronous detector and a bad reading with the envelope detector indicate possible problems with the modulator's modulation stage.

Subtle amplitude and group delay problems in a piece of equipment can be masked when measured with an envelope detector. Amplitude and group delay problems can be evaluated by looking at the sine squared (2T) and modulated 12.5T pulses. These pulses are particularly convenient for use in evaluating problems because the results are presented in terms that directly relate to the visible distortion of a video signal. However, some problems produce waveform distortions that are similar to the distortions produced by an envelope detector. Thus, a synchronous detector is more useful when studying response problems.

8.6 TV Stereo

At least three TV stereo systems are in use worldwide, with several variations to each. North America uses the BTSC system, named after the committee that adopted it. The Broadcast Television System Committee was formed under the auspices of the Electronic Industries Association (EIA). It was active from the late 1970s to the mid-1980s. The Japanese use a variation of BTSC stereo. In Great Britain, the BBC and IBC developed a digital system known as near instantaneous amplitude companding and modulation (NICAM). With variations, it is in use in a number of European countries and elsewhere. The German IRT earlier developed a two-carrier analog system, which is used there, in Korea, and elsewhere.

8.6.1 BTSC Stereo Generation and Application

The BTSC stereo system is described in Chapter 2. Figure 8.15 shows one possible block diagram of a BTSC stereo encoder. Left- and right-channel audio are supplied to identical low-pass filters, whose cutoff frequency should be as close to 15 kHz as possible, while providing a lot of attenuation at the horizontal sweep frequency of 15.734 kHz. Because of the sharpness demanded of these filters, their complexity is high. The filters are often implemented using digital signal processing (DSP) technology today. Failure to provide adequate reduction of power at 15.734 kHz will result in compromised stereo performance due to interference with the pilot carrier. Further, the difference channel approaches this frequency, and poor rejection could result in artifacts arising from spectral crossover between the sum and difference channels.

Note that the left- and right-channel program audio levels must be metered as shown to ensure the correct deviation. The monaural meter on the modulator is not usually suitable for this purpose.* In fact, it can be very misleading. Similarly, SAP level is monitored at its input.

* If the monaural meter drive signal is limited to a 15-kHz frequency response it may be marginally useful as a deviation indicator, but this is not recommended.

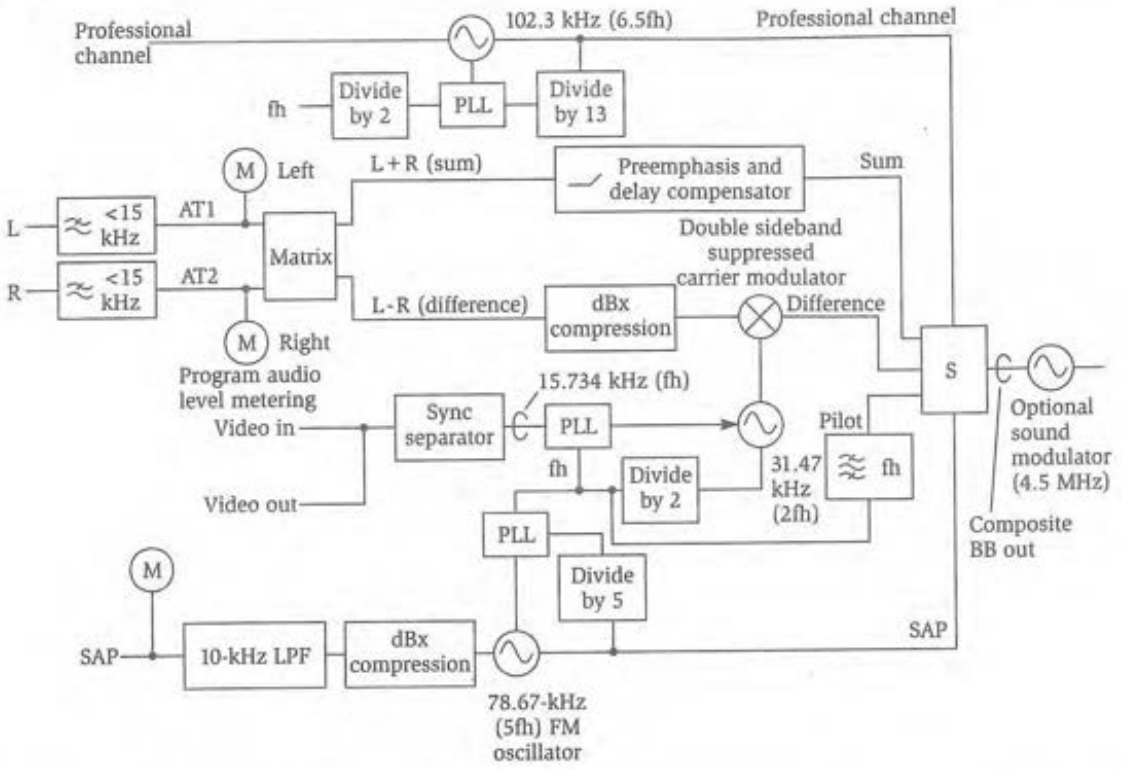


Figure 8.15 Block diagram, BTSC stereo encoder.

From the filters, the signals are supplied to the matrix, which is simply a pair of networks that produce the sum (L + R) and difference (L – R) signals. The sum channel is supplied to a preemphasis circuit and a delay compensator, which compensates for the delay in the difference channel. See the discussion of emphasis in Section 8.3.6. It is necessary to use preemphasis in the sum channel in order to maintain backward compatibility with the mono system. From the preemphasis circuit, the sum signal is supplied to an adder that combines the different parts of the stereo signal.

The difference signal is supplied to the dBx encoder, which compresses it in order to reduce the effects of poor carrier-to-noise ratio in the transmission path. The compression is frequency selective and follows an algorithm proposed originally by the dBx Corporation. (For test purposes only, the dBx compression may be replaced with a 75-μs preemphasis network.)

From the compression circuit, the difference signal is applied to the modulator. It is double sideband suppressed carrier amplitude modulated onto a

carrier at 2fh (twice the horizontal scan frequency), 31.47 kHz. The carrier oscillator is phase locked to the horizontal sync rate, fh, derived from the video signal. This need for phase lock is the reason that video must be looped through the stereo encoder. Failure to lock to video could result in beats and separation "pumping" from oscillators that are free running on almost but not quite the same frequency. A portion of the fh signal, obtained by dividing the output of the 2fh oscillator by 2, is filtered and supplied to the summing circuit as the pilot carrier. This fh signal is also supplied to a phase locked loop, whose output adjusts the 2fh oscillator as needed to maintain the correct phase relationship between its two inputs, one from the divider and the other from the sync separator. These three components — the sum, difference, and pilot signals — make up the basic stereo signal.

If the SAP is not used on a particular channel, that circuitry is not likely to be included in the stereo encoder. If it is, the circuitry at the bottom of Figure 8.15 is supplied. The SAP audio is compressed then frequency modulated onto an oscillator whose center frequency is locked to the fifth harmonic of the horizontal frequency, 5fh. This SAP signal is then summed with the stereo components. If the professional channel is used, then the blocks at the top of the page are supplied. It is rarely used in cable television.

Interfacing and Setting Levels in BTSC Service

A number of methods may be used to connect the BTSC encoder to the modulator. Selection of the interface method to use is based on inputs available on the modulator and outputs available on the stereo encoder. Of particular concern is the ability to set the deviation of the sound carrier correctly. Failure to do so will result in decreased stereo separation. For example, separation between the left and right channels will be degraded to less than 20 dB if there is a 1.9-dB error in setting the deviation.[3]

Several methods may be used to interface the encoder with the modulator. The most basic, but not necessarily the best, is to defeat the preemphasis in the baseband input of the modulator (because preemphasis is handled in the encoder) and supply the broadband signal to the baseband audio input of the modulator (Figure 8.5). This works, but you must be careful to set the deviation of the modulator correctly. Failure to do so will result in poor stereo separation, as shown earlier.

In order to use this interface, some modulator designers have arranged metering circuitry such that a modulation meter measures only the sum signal. The object would be to set the level using the left and right meters on the encoder, then set the deviation on the modulator using its meter. Setting deviation only with reference to the sum deviation is not at all accurate and will almost surely result in poor stereo separation. However, metering the sum channel at least makes a meter on the modulator make sense when handling a stereo signal.

In addition to metering the program audio, the encoder manufacturer may provide a test tone. When turned on, the test tone replaces the composite stereo signal supplied to the modulator. The modulator *deviation* adjustment (AT3 of Figure 8.5) is adjusted for the correct deviation of the test tone, usually 25 kHz. If this is done and the program audio levels are set correctly on the encoder, then the deviation will be correct. This method is far superior to metering only the sum channel program audio on the modulator.

A preferred interface method, if available, is to have a 4.5-MHz FMed oscillator built into the stereo encoder. An optional sound carrier modulator (oscillator) is shown in Figure 8.15. This way, the encoder manufacturer can ensure that the deviation is set correctly. The encoder output at 4.5 MHz is supplied to flag C of the modulator block diagram, Figure 8.5. It is best to *not* combine the sound subcarrier with the video since doing so will reduce the carrier-to-noise ratio in the aural channel and will create more group delay problems in the video channel. The modulator must be capable of accepting a separate input to FL3.

Another possible but somewhat less desirable interface method is to locate a 41.25-MHz FM oscillator in the encoder and supply sound at this frequency to flag D of the sound IF loop at the modulator, Figure 8.4. This provides the same level calibration advantage as does the 4.5-MHz interface, but exacerbates the problem of maintaining accuracy in the frequency spacing between the picture and sound carriers. In any of the other methods described, sound carrier spacing depends only on a single 4.5-MHz oscillator. The FCC-required accuracy in spacing for cable television is ±5 kHz. Applied to the 4.5-MHz spacing, this represents an accuracy requirement of 1,100 parts per million, a trivial number for a crystal-controlled oscillator, and barely possible with an L–C oscillator (used in set top terminals but definitely not appropriate for professional use). If the interface is at 41.25 MHz, then two independent oscillators are involved, one at 41.25 MHz in the encoder and another at 45.75 MHz in the modulator. This means that each must be accurate to about 57 parts per million. This is not excessive for a crystal oscillator but is getting to the point that care must be exercised.

In addition, there is some possibility with a 41.25-MHz interface that ICPM could exist on the picture carrier and not on the sound carrier. This would cause a degraded signal-to-noise ratio in the intercarrier demodulator in the TV receiver. (See the discussions of ICPM and intercarrier detection in Section 8.5.)

Regardless of the interfacing method used, it is necessary to ensure that the deviation of the main sound carrier by the composite stereo signal is correct, then to set the level of the left and right channels using only attenuators AT1 and AT2 of Figure 8.15. The stereo encoder should provide metering to allow accurate setting of the program audio levels.

8.6.2 The NICAM Stereo System[4]

NICAM's formal designation is NICAM 728 because the bit rate is 728 kb/s. The system is a digital transmission system developed in the mid-1980s in the United

Kingdom. It has been adopted for use in other countries as well. It is capable of carrying one stereo pair, or two independent monaural programs, or 352-kb/s data and one monaural program, or 704-kb/s data only.

Data is initially sampled at 14 bits per sample and down-converted to 10 bits (plus parity). The data is divided into blocks of 32 samples, and the scaling of the block is based on the highest amplitude sample in that block. A scaling factor is transmitted with the data. Transmission is divided into 1,000 frames per second. Each frame consists of 704 bits of data and parity, with the remaining 24 bits being overhead and synchronization.

Modulation is differential QPSK (see Chapter 4). As practiced with the CCIR-I television system (Great Britain), the modulator and demodulator both use a root raised cosine filter with an alpha of 100%. When NICAM is used in countries employing the CCIR-B/G television format, the filter alpha is reduced to 40% to accommodate the narrower available bandwidth.

In the CCIR-I system (used in Great Britain), the NICAM carrier is located 6.552 MHz (nine times the 728-kHz bit rate) above the picture carrier. The normal analog sound carrier, 6 MHz above the picture carrier, is transmitted −7 dB with respect to the picture carrier, and the NICAM signal is set to $-20\ dB_{PC}$. The power in the NICAM signal does spill out of the channel, slightly invading the upper adjacent channel. This has not been a problem.

In the -B/G system (used in much of continental Europe except France and Russia), the NICAM center frequency is slightly outside the channel, with energy spillage even farther out. At UHF (system -G), the occupied bandwidth falls in the 1-MHz guard band between the channels. At VHF (system -B), the power spills into the upper adjacent channel. By narrowing slightly the vestigial sideband filter for the upper adjacent channel, it is claimed that a cable system can successfully add the NICAM signal while employing adjacent channel transmission.[5]

8.6.3 The IRD Two-Carrier Stereo System

An earlier stereo system developed in Germany employed two analog carriers. It is used there with the CCIR-B/G television system. It has also been used, with modification, in Korea, where NTSC is the television system in use. The normal 5.5-MHz (352fh) sound carrier is modulated with a left plus right signal. A second carrier is added at 5.7421875 MHz (367.5fh). It is modulated with the right channel only. This second signal is demodulated and supplied directly to the right loudspeaker. The left signal is obtained by subtracting the right signal from the L + R signal on the main aural carrier. The reason for doing this rather than placing a difference (L − R) signal on the second carrier is that noise from the second, low-amplitude, carrier theoretically gets canceled by virtue of its phase being opposite on the two channels.

The main sound carrier at 5.5 MHz above the picture carrier is carried 13 dB below the picture carrier. The second sound carrier is carried −20 dB from the picture carrier and is also frequency modulated with a deviation of ±50 kHz.

A pilot carrier of 64.6875 kHz (3.5fh) contains modulation that identifies the broadcast as mono, stereo, or dual sound (two mono channels, as for two languages). In addition, mode identification is carried in the vertical blanking interval of the video.

8.7 Satellite Earth Station Receiving Equipment

The modern cable television industry was enabled by the availability of communications satellites that could distribute many television programs simultaneously to all cable headends in North America. The watershed event was the September 30, 1975 (U.S. time), Ali–Frazier prizeFight, the "Thriller from Manila," seen at the two or three headends in the United States that had received 10-meter earth station antennas by then.

Figure 8.16 illustrates the major electronic components of the earth station used at cable headends. Signal is coupled from the "dish" antenna (see Chapter 7) to a feed located at the focal point of the antenna. The signal is amplified and, typically, down-converted to a lower frequency, by a low-noise block converter

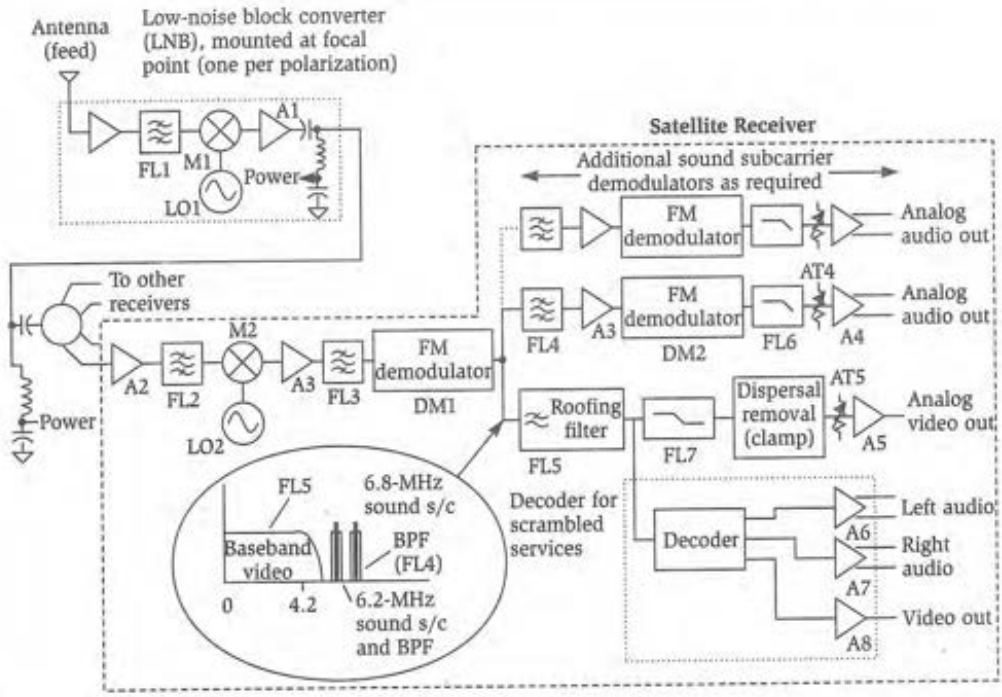


Figure 8.16 Analog earth station electronics, block diagram.

(LNB) located at the feedpoint. The LNB is located at the feedpoint to minimize signal loss. Communication satellite downlinks today operate on two bands, C band (3.7 to 4.2 GHz) and Ku band (11.7 to 12.2 GHz). Coaxial cable loss is significant at these frequencies though older earth stations used large, high-quality cable to route C band signals into the headend.

The most common band to which signals are converted is 950–1450 MHz (roughly L band). Some older equipment used 270–770 MHz. At these frequencies, the signal can be routed into the headend using fairly small coaxial cable. The cable may also carry power to the LNB as shown. Fiber optics may also be used to bring signals to the headend.

Once in the headend, the signal is split to serve a number of receivers: today, up to 12 receivers may be connected to a single LNB. A second LNB located at the same feedpoint can handle up to 12 additional channels coming from the same satellite in the same band, with different polarization.

At each receiver, the signal is again amplified by amplifier A2, filtered, and down-converted to another IF (dual conversion downconversion will be used if the receiver is frequency agile). The downconversion is done in mixer M2, using local oscillator LO2. After amplification and filtering, the signal is demodulated in FM demodulator DM1. As shown in the insert of Figure 8.16, the spectrum coming out of DM1 consists of baseband video (probably scrambled) and possibly one or more sound subcarriers. A number of subcarrier frequencies have been used for sound, but the two most common are 6.2 and 6.8 MHz. Today, these subcarriers are generally not used for primary program audio, that being carried digitally in the video. The subcarriers may be used for cue tones and supplementary audio. Cue tones are used to initiate local advertising inserts.

If an aural subcarrier is used, it is filtered in bandpass filter FL4, amplified, and detected in FM demodulator DM2. The audio is preemphasized using the standard 75-μs preemphasis, and so must be deemphasized in FL6. Attenuator AT4 allows the output level to be set, and differential output amplifier A4 supplies the output gain.

In the video path, a roofing filter FL5 may be used to remove power above the video spectrum. If not scrambled (rare today), the video passes from the roofing filter to deemphasis filter FL7. The video is emphasized to reduce the effects of noise on the signal, as is audio. A particular emphasis and deemphasis curve is used, as prescribed by the CCIR. After deemphasis, a dispersal removal clamp removes any dispersal signal that may be imposed on the video. Dispersal is used in C band satellite circuits because the same 3.7–4.2 GHz band used for satellite downlinks is also used for terrestrial microwave communications, primarily by the telephone industry. The fear is that if modulation is removed from the satellite signal, enough power could be concentrated at one frequency to interfere with terrestrial transmissions. A dispersal signal was required in the early days to prevent that from happening. The dispersal signal is a 30-Hz triangular waveform synchronized to the vertical blanking interval (VBI) and changing direction during the VBI. It is removed using a clamp or dc restorer,

somewhat as illustrated by C1 and D1 of Figure 8.4. From the deemphasis filter, the video level is set using attenuator AT5. After buffering in A5, nonscrambled video is output from the receiver.

Dispersal waveforms are not always used in modern practice. If a transponder is used full-time for video, then the downlink power will not be concentrated at one frequency. For transponders having occasional video service, dispersal waveforms are still used, but they are frequently used only to replace video when it is removed but the uplink remains on the air (possibly for test purposes).

Almost all signals delivered to cable headends today are scrambled. Programmers went to scrambling during the 1980s to prevent theft of service by privately owned earth stations that had not paid for service. The analog scrambling system used almost universally in the cable industry today is a sync replacement system (similar to sync suppression — see Chapter 21), with the horizontal blanking interval (HBI) replaced by encrypted audio. The idea is to make the video scrambling hard to break. Then the audio is "hard" encrypted, such that it could not be decoded without authorization. Descrambled audio and video outputs from the receiver are amplified in A6 through A8, and supplied to the outside, normally to modulators and stereo encoders as described earlier.

Signals to be descrambled are passed to the decoder, which accepts authorization and then descrambles the signals. The system in use has the capability to deliver two audio signals so that a stereo pair can be transmitted. If a second language program is transmitted, it uses one of the FM subcarriers described earlier.

8.7.1 Digital Satellite Receivers

Digital satellite receivers look similar to analog receivers at the block diagram level, with the exception that the demodulator is QPSK, OQPSK, or sometimes 8-PSK. All of these modulation formats are defined in Chapter 4. Typical receivers include an MPEG decoder, and they supply both analog video (from one channel in the received multiplex) with stereo sound and the entire digital transport stream in digital form. Audio outputs may include analog and sometimes a professional digital audio interface.

Figure 8.17 illustrates a digital earth station electronics block diagram; compare it with Figure 8.16 for the analog equivalent. Note that the antenna and LNB are identical, as is the front end of the receiver. In fact, the same antenna and LNB can be used to receive analog on some transponders and digital on others. The only difference is in the receiver.

As shown, the front end of the receiver is the same as in the analog receiver, but rather than an FM demodulator, a QPSK demodulator (most common today) is used. Its output supplies a descrambler that removes the signal scrambling done at the uplink to protect the programs for being pirated. Commonly, an analog decoder is supplied through a program selector, to recover an analog signal from one of the signals being transmitted. Its output is normal analog audio and video. An ASI driver formats the digital video signal for output, as described shortly.

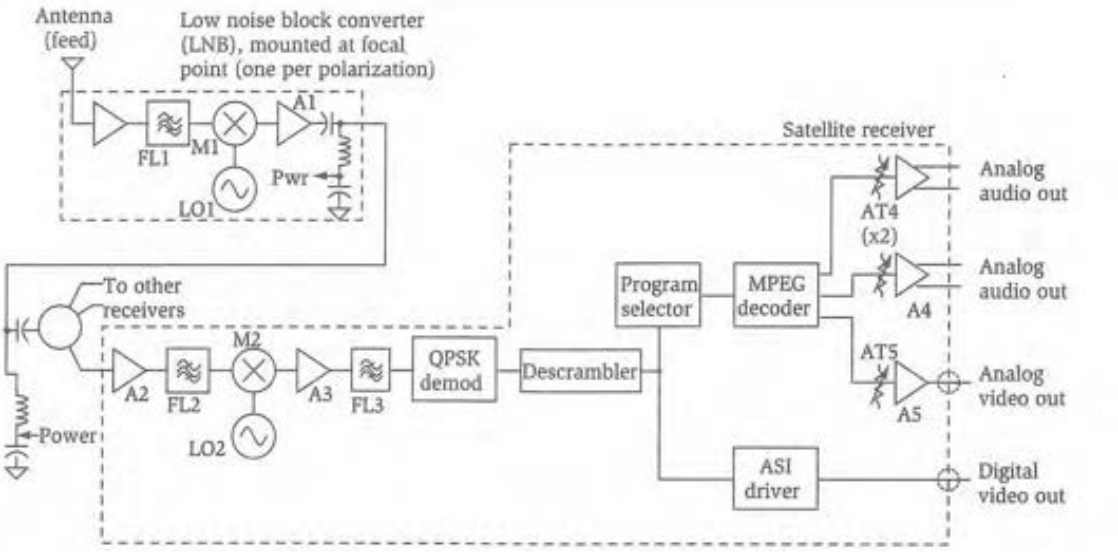


Figure 8.17 Digital earth station electronics, block diagram.

8.8 Digital Video Interfaces

Digital video interfaces connect one or (usually) more MPEG-2 transport streams (TS) from source equipment, such as a satellite receiver, an encoder, or a video-on-demand (VOD) server, to information sink equipment, most commonly QAM modulators. The MPEG-2 transport stream is described in Chapter 3. It includes video and audio as well as ancillary data for one or more programs. Note that multiple audio streams may be associated with one video stream. These audio streams can include surround sound and one or more languages. Since the audio, video, and ancillary data are transmitted in the same transport stream, there is only one connection for a program; there are no separate audio and video connections as you have in analog transmission. Also, it is most common to have more than one program in the same transport stream.

8.8.1 Asynchronous Serial Interface (ASI)

A common interface for digital video is the *asynchronous serial interface* (ASI), specified by the DVB Project Office. This is a serial interface, as the name implies, transmitted over a single 75-ohm coaxial cable. Besides coaxial cable, there exists a standard for an ASI using fiber-optic cable.

ASI is a one-way data transmission standard, with no acknowledgment of receipt of information and no built-in error correction. Being derived from the fiber channel standard, the signal is 8B/10B encoded, as explained in Chapter 19. For the coaxial cable version of the ASI, the transmitter supplies a baseband digital signal at an amplitude of 0.8 volts p–p into 75 ohms. The receiver must successfully detect a signal amplitude of 0.2 volts p–p. The return loss of the 75-ohm system is ideally 17 dB minimum.

The wire speed (actual data speed) of the data is 270 Mb/s with 8B/10B coding. Removing the 8B/10B coding results in a data speed of 80% of this, or 216 Mb/s. Included is a layer of ASI sync. One ASI link is able to contain enough data to fill several channels with 64- or 256-QAM data.

Figure 8.18 illustrates the logical ASI. The input at the transmitter (internal to a piece of equipment, so you normally won't see it) is a packet-synchronous MPEG-2 transport stream. As it enters the ASI transmitter, the signal is 8B/10B encoded. Chapter 19 explains 8B/10B encoding. Since the MPEG signal does not operate at the ASI data rate, there will be times when the ASI transmitter is ready to send more data and yet no data is ready to be sent. At such times, *filler code* words are inserted. This must be done to keep a constant bit rate on the ASI link. The signal is converted from parallel to serial for transmission, amplified and impedance matched, and then coupled to the connector. Transformer coupling is anticipated in the specification.

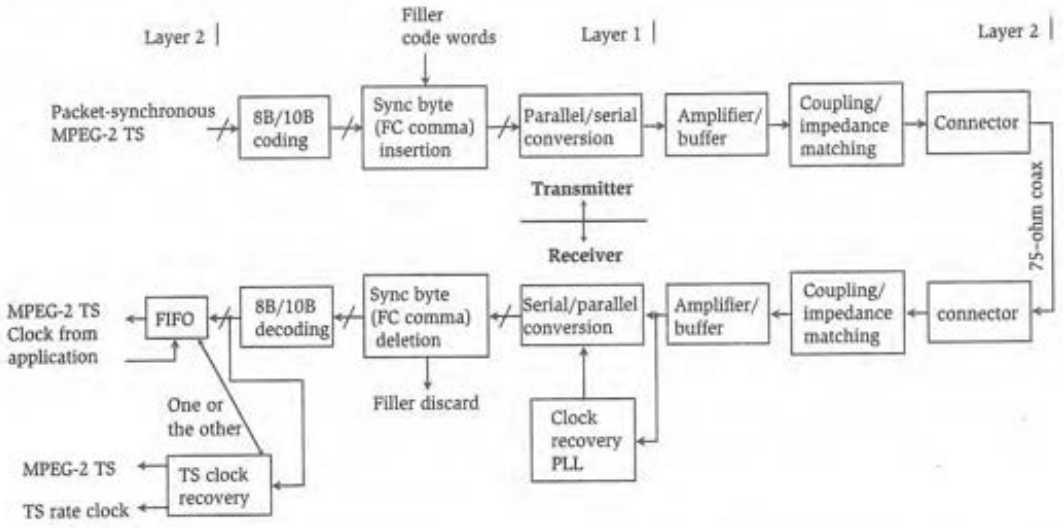


Figure 8.18 Digital video ASI.

After transmission, the signal is coupled into an amplifier through the input-coupling and impedance-matching mechanism. The bit clock can be recovered from the transmitted signal, and controls conversion of the signal from serial back to parallel for further processing. The ASI sync is detected, and the filler code words added in the transmitter are removed. 8B/10B coding is removed. The receiving circuit has a choice of how to recover the transmission stream data sync. Either the signal can be reclocked from the receiving application's clock or the transmission stream clock from the incoming signal may be recovered and used in the receiving application. In the former case, the signal is supplied to a first-in-first-out (FIFO) register, which is used for retiming a signal. It is clocked in from the transmission stream's clock and clocked out from the application's clock. In the latter case, the transmission stream clock is recovered and supplied to the application, along with the transmission stream itself.[6,7]

8.8.2 Ethernet for Video Transmission

Gigabit Ethernet is also popular as a method for transferring digital video in headends. The primary advantage is speed. ASI has a practical transfer capacity of about 160 Mb/s after all inefficiencies and overhead are removed. On the other hand, the practical transfer capacity of a dedicated gigabit Ethernet link can be as high as 900 Mb/s.[8] This means that when coupling signals between high-density video sources and modulators, such as in a video-on-demand system, fewer interfaces are required with gigabit Ethernet interfaces than with ASI. Recall that there does not have to be a 1:1 correspondence between headend interface cables and channels; one headend interface can serve several 6-MHz channels' worth of video, so long as the video is coming from one place (e.g., a VOD server) and going to one place (e.g., a cluster of QAM modulators).

When gigabit Ethernet is used for video transmission, it is common to embed the MPEG-2 transmission stream into RTP packets (explained in Chapter 6—this same packetization is done for VoIP). The RTP packets allow for removal of jitter introduced in the transmission process; they also keep track of packet numbers to allow reordering of packets received out of order. (Out-of-order packets should not be a problem for local point-to-point transmission of video.) The RTP packets are embedded in UDP packets, which enable one-way or broadcast transmission of packets. In turn, the UDP packets are embedded in IP packets, as explained in Chapter 5, and the IP packets are embedded in Ethernet packets.

8.9 Signal Handling in Headends

This section includes information that should be useful to practitioners setting up and maintaining headends. Included are notes about video and audio signal handling. RF signal handling is covered in more detail in Chapter 9.

8.9.1 Video Loop-Through

Figure 8.4 included the illustration of video loop-through. This is a technique whereby you can supply video to several using devices such as modulators, video monitors, or switchers, without using amplifiers. It is normally recommended that no more than three devices be cascaded with a loop-through because there is adverse effect on the signal from each device. This section explains what loop-through is and the most common alternative to it.

Figure 8.19 illustrates video loop-through. Figure 8.19(a) illustrates video being looped through three modulators. The devices could also be a video switcher, a picture monitor, and a modulator. A common application in headends is to loop the video through a BTSC stereo encoder, a scrambler, and then terminate it at a modulator. Each modulator in the example has the ability to either terminate the video in a 75-ohm resistor or not. It is necessary to terminate video in the characteristic impedance of the coaxial cable carrying video so as to prevent reflections that would damage the frequency response of the picture.

If a signal is not terminated, it will be reflected toward the source. This reflected signal will combine with the direct signal, resulting in a frequency

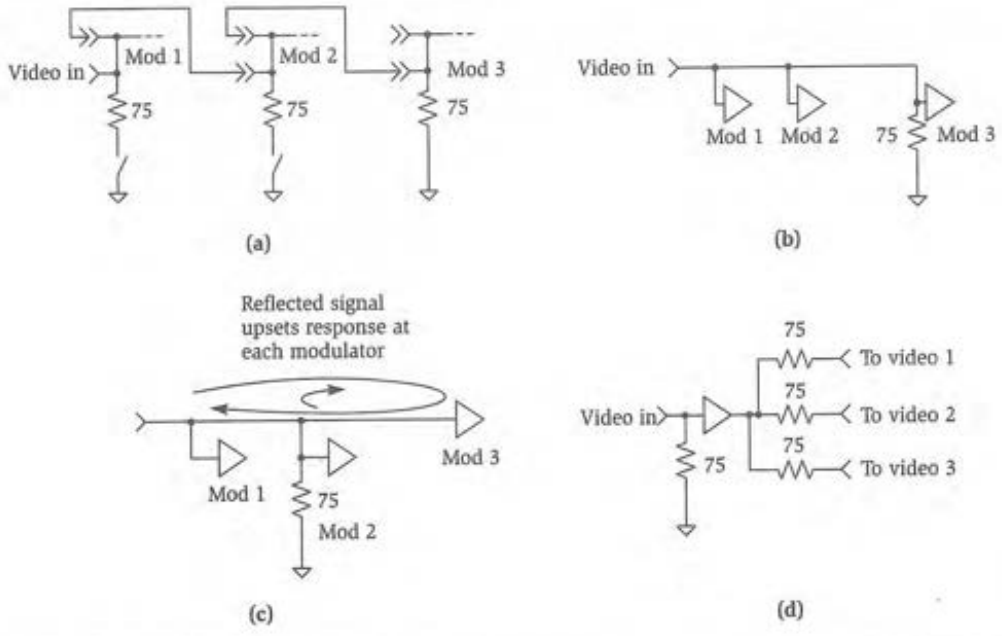


Figure 8.19 Video loop-through. (a) Looping through three devices. (b) Equivalent circuit for (a). (c) Incorrect application of loop-through. (d) Video distribution amplifier alternative.

response that peaks at frequencies where the two signals are in phase and nulls where the two cancel.

So long as the coax is terminated at the last device in a chain of devices, there is not a problem. This is illustrated in Figure 8.19(a), where three modulators or other devices are in a cascade, with the video being terminated by the resistor in the last device. The modulators shown all have the ability to terminate or not. This is usually selected by a switch on either the back or front of the modulator.

Figure 8.19(b) illustrates the equivalent circuit of the arrangement of Figure 8.19(a). High input impedance amplifiers are used to sample the signal from the coax transmission line without affecting significantly the impedance of that line. At the last modulator, a 75-ohm resistor terminates the line. As illustrated in Figure 8.19(c), if the termination is placed other than at the last device, reflections will result. Signal power will be reflected from the last device in line, modulator 3. When it arrives back at modulator 2, several bad things happen. First, depending on the frequency and the length of cable between the two modulators, signal components at some frequencies will add, increasing the amplitude, whereas others will cancel. The frequency response at modulator 2 is thus corrupted.

Furthermore, reflected energy arriving at modulator 2 will see a double termination: the 75-ohm terminating resistor in modulator 2 and the 75-ohm impedance of the transmission line leading back toward the source. This impedance mismatch creates a 6-dB return loss, so one-half the signal propagates toward the source (to the left), and one-half propagates again toward modulator 3, upsetting the response there. Thus, it is essential that a looped-through signal be terminated at the last, and only at the last, device in the signal chain.

Similarly, it is unacceptable to parallel more than one signal path because reflections will cause frequency response problems. This would happen if you branched off from device 1 and routed signals to devices 2 and 3 on different cables.

Each device the signal is looped through must use a very high impedance amplifier to bridge the video signal. As a practical matter, it is not possible to prevent some signal degradation as you loop video from one device to another. Because of this, the number of devices a signal is looped through should be limited. A rule of thumb is that three devices can be looped through without excessive deterioration in signal quality. Another limitation is that if a device the signal is looped through must be removed, then service will be interrupted to other devices in the loop-through chain. Finally, looping through may result in awkward headend wiring. An alternative, when video is to be distributed to many locations, is to use a video distribution amplifier (VDA).

Figure 8.19(d) illustrates a common form of VDA though most have more than three outputs. An amplifier provides a voltage gain of 2 to the video signal so that a 1-V p-p signal at the input becomes a 2-V p-p signal at the output. A number of outputs are provided, each with a series 75-ohm resistor to establish

the source impedance of the path. This resistor and the 75-ohm terminating resistor in the following device form a voltage divider, reducing the voltage back to the video standard interface of 1-V p-p.

8.9.2 Audio Wiring

Many of the same factors that apply to video also apply to audio, except that audio is usually less critical in terms of signal wiring than is video because of the lower bandwidth. It is possible, and many times recommended, that audio distribution amplifiers (ADAs) be used for much the same reasons that VDAs are used. However, it is sometimes acceptable to wire a number of audio-using devices in parallel without regard for reflections since typical headend wiring distances are insignificant compared with the wavelength of audio. It is important, though, to provide one and only one 600-ohm termination in an audio chain.

Professional audio sources are usually specified to drive 600-ohm loads. Some sources may actually exhibit an output resistance of 600 ohms though it is not at all uncommon for the source resistance to be less. Since audio lines are usually of insignificant length compared with a wavelength, this lower driving resistance is not a problem unless it is desired to drive a signal several miles. In this case, more care should be taken to match resistance on both ends.

A major headache in audio wiring is hum elimination. Most hum is due to ground loops and other sources of potential difference. Ground loop mitigation is the most magic of all magic arts: no matter how many words are written about it, the problems encountered in the field will not be covered. However, some pieces of advice are sufficiently general to be worth mentioning.

Figure 8.20 illustrates options in audio interconnection between an audio source and load. Figure 8.20(a) represents the interconnection method strongly recommended and practiced almost exclusively in professional audio applications. Balanced transmission is used, in which the source consists of two signals having equal resistance to ground (not shown) and exactly opposite phase. One way to produce such signals is to use two amplifiers, one for the + phase and another inverting that signal to generate the – phase, as shown. Another way is to use an audio transformer.

The signal is connected to the load on the right of Figure 8.20(a), using a balanced, shielded cable. Several cables made for this purpose are available. The two wires of the cable are terminated in a differential amplifier, whose output is an amplified version of the difference between the two input signals on the + and – inputs. By using balanced transmission, any hum pickup will be the same on the two wires of the balanced pair, so the hum will be subtracted out at the receiving differential amplifier (or transformer).

Notice that the shield is connected to ground at only one end. It is not correct to ground the shield at both ends since this can create a ground loop.[9] Ground loops exist any time there is more than one path for current between pieces of equipment. As a practical matter, this is hard to prevent: racks must be bonded

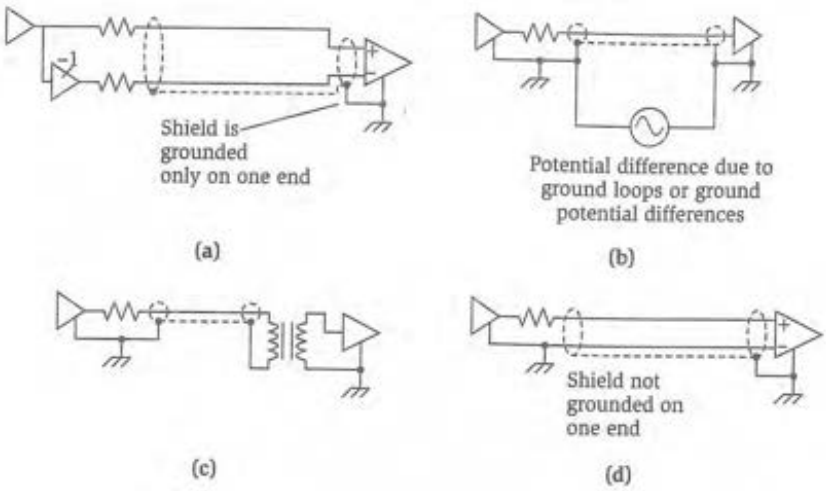


Figure 8.20 Options in audio interconnection. (a) Balanced audio. (b) Unbalanced audio with hum source. (c) Unbalanced source and load with transformer. (d) Unbalanced source and balanced load.

together and to ground, providing a path. Individual pieces of equipment must have third-wire safety grounds, providing a second path. After that, consider all the interconnecting wiring, much of which involves grounds. A loop can form the secondary of a giant transformer, the primary of which is all the ac wiring in the headend. The result is current induced in the ground paths, which sets up potential differences between equipment. Besides the transformer model, it is likely that many headends will experience potential differences between racks because of the various routes supply current can take to get back to the neutral of the primary power supply.

Figure 8.20(b) illustrates the unbalanced signal connection method used in most consumer equipment but not recommended for professional use. A single-ended output amplifier sends the signal to the receiving equipment, which receives it using a single-ended input. The shield of the cable serves as the return path. Also shown is a prime reason to not use this method in headends. If an ac voltage exists between the two ends, it will result in current on the shield. Because the shield has some resistance, a voltage drop will exist between the two ends. This voltage will be seen by the receiving amplifier, resulting in hum in the audio.

If the equipment at hand can accommodate only unbalanced audio, and if hum is a problem, then the solution of Figure 8.20(c) will have to be used. A high-quality audio transformer having a 1:1 turns ratio can be used at the receive side to isolate the ground on the two sides. The input side is connected to

the center conductor and shield, but not to ground. The output side is connected to the amplifier input and ground. Transformers *must* be operated at their specified impedance. Small, low-cost transformers will distort the audio, particularly at lower frequencies, though high-quality professional transformers are available that do a good job.

Figure 8.20(d) illustrates an option where the sending equipment is single ended, but the receiving equipment can accommodate a balanced input. Balanced audio cable is used, with the sending ground connected to one of the wires of the pair. Again, note that the shield is connected at one end only.

In the end, though, we can say only that ground loops are so different from each other, and so difficult to resolve, that whatever works is the right thing to do.

8.9.3 Controlling Hum in Video

Hum in video systems using NTSC and in 60-Hz power grids will show up as horizontal bars moving slowly up through the picture. The upward movement is due to the slightly slower than 60-Hz field rate of NTSC. Just as ground loops can cause great anguish with technologists attempting to interconnect audio circuits, it can be a problem with video. Because of the frequencies involved, balanced (differential) coupling of video is not as predominant as it is in audio applications though it is possible to interconnect video signals using balanced transmission. It is far more common to use single-ended equipment to interconnect video. This makes resolution of ground loop problems even more difficult for video than for audio.

Figure 8.21 illustrates two methods that can be used to mitigate hum in video circuits. Shown are video source and terminating equipment. Both the signal and chassis grounds are shown, emphasizing that they are usually one and the same in video equipment. They may be kept separate internally to the equipment, but for the most part, they are the same outside the equipment.

Figure 8.21(a) illustrates use of a hum bucking coil. These may be purchased commercially and consist of a number of turns of 75-ohm coax through a high-permeability ferrite core. Typically, BNC connectors are provided on the case of the hum bucking coil. The connectors must be insulated from each other. To a current on either the center conductor or shield, the arrangement presents a high inductive reactance, which opposes the current. However, to a current traveling in one direction on the center conductor and the opposite direction on the shield (correct differential mode signal transmission), the inductive reactance is not present: the inductance operates only on the net current in the coax, and in this case, the net current is zero.

So long as the inductive reactance at 60 Hz is high enough, it is not possible for significant ground loop current to develop through the coax. Thus, the hum bucking coil can offer significant reduction in hum, but only if it is not shorted between the input and output shields. The short would afford ground current a way to avoid the inductive reactance of the coil.

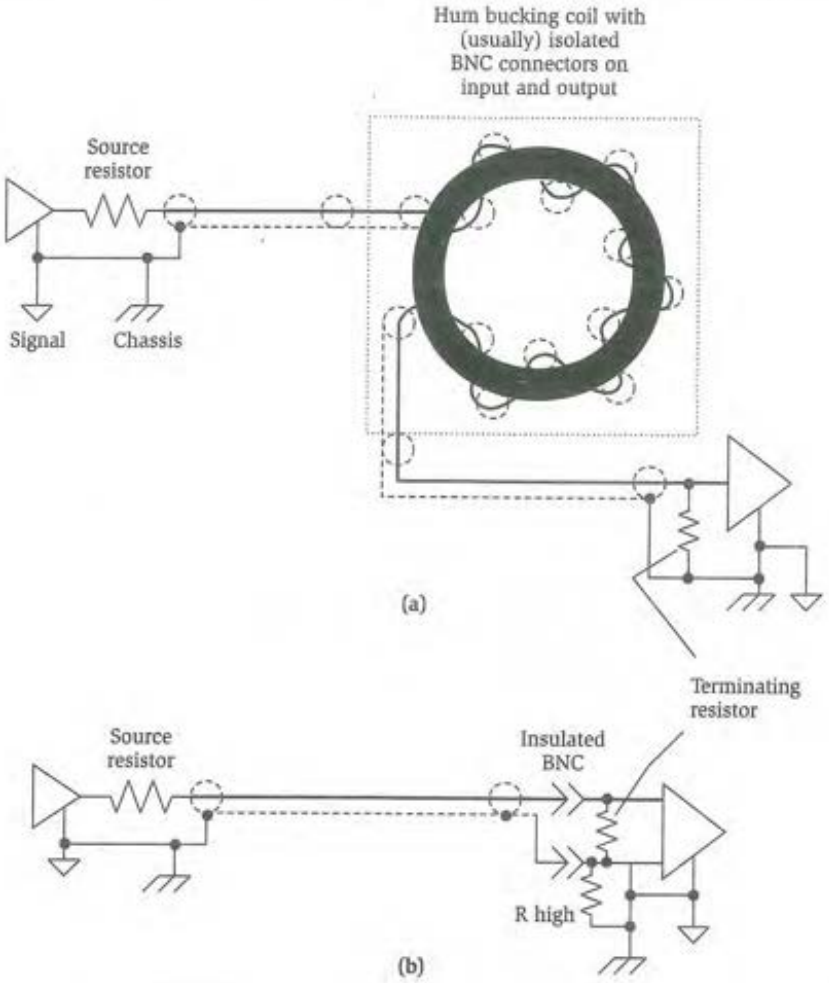


Figure 8.21 Methods used to mitigate hum in video circuits. (a) Hum bucking coil. (b) Floating input connector.

Some video equipment does offer a balanced input driven from an unbalanced source, as illustrated in Figure 8.21(b). This is analogous to the situation of Figure 8.20(d). Usually, a special BNC connector is used, which has its shield insulated from the chassis. A differential amplifier recovers the voltage difference between the center conductor and the shield. The resistor to ground is high and is used to prevent a static charge buildup on the shield if the other end is not terminated. A floating BNC connector as shown is used on some professional

equipment, but the cost prevents its use in lower-cost lines of videoprocessing equipment.

8.9.4 Handling Digital Video Signals in the Headend

Asynchronous serial interfaces should be routed point-to-point and not looped through from one piece of equipment to another, even if the same signals are bound for more than one place. Use a digital distribution amplifier to route the signals to more than one place.

ASI signals are handled on the same type of 75-ohm coaxial cable used for RF and analog video, so the same handling precautions apply. The digital signal will in general be more immune to interference pickup than is the analog signal. However, signal radiation from a digital interface to an analog interface is a possibility. The same measures taken to minimize interference in analog signal transmission will work here. Use the best-quality coaxial cable possible. Quad shielded cable is recommended for headend use. Pay particular attention to the shield connections, because poor shield connections can encourage signal radiation. If digital pickup in analog cabling is a problem, also try physically separating the cables. If this fails, try ferrite beads on both the digital and the analog cables. Make sure the ferrite beads are specified for the frequency range over which the interference is a problem. A ferrite optimized for use at power frequencies will perform very poorly at several hundred megahertz.

A lot of signal transmission is on Ethernet cables today. These cables consist of a number of twisted pairs handling balanced signal transmission. As shown in the telephony chapter, balanced transmission is a good way to minimize pickup or radiation of signals, but it is not perfect. If you are having problems with radiation from Ethernet cables, you can also try appropriate ferrite beads on the Ethernet cables. You can also try shielded Ethernet cables (STP). Many times, though, the best strategy is to separate the Ethernet cables from victim cables.

8.10 Ad Insertion

Insertion of advertising in the programs transmitted by various programmers is a significant source of income for many cable systems. Many basic cable networks charge cable systems for the privilege of carrying the network. In turn, the networks make available "avails," or time each hour that is available for sale of advertising by the local cable systems. Though practices vary, it is not uncommon that two blocks of time, each 2 minutes long, are made available at predetermined times each hour for insertion of advertising sold by the local cable system. Control of insertion is done by the network, which transmits cue tones to initiate the play of local commercials and cue tones to transfer back to the network.

In the past, ad insertion was done using analog tape players controlled by a computer. Today, file servers have replaced the tape players in new systems. Figure 8.22 illustrates a file server used for this purpose. Differences will exist from one system to another. The file server itself is usually a computer workstation optimized for transfer of large digital files from its internal hard disk drives to the outside world. It interfaces with a traffic system, which provides the schedule for playing commercials. Through a second interface, commercials are loaded onto the hard drives. Typically, the commercials are compressed using MPEG-2 encoding.

One file server is typically capable of supplying data fast enough to service a number of channels simultaneously. As shown in Figure 8.22, the file server "plays" the file out through a system interface, to an individual analog channel interface, which is shown in more detail in Figure 8.23. The channel interface receives analog video and audio from the IRD or other source, along with cue tones. Usually, the cue tones are transmitted on analog sound subcarriers since the audio is transmitted digitally with the video. FM demodulators, such as DM2

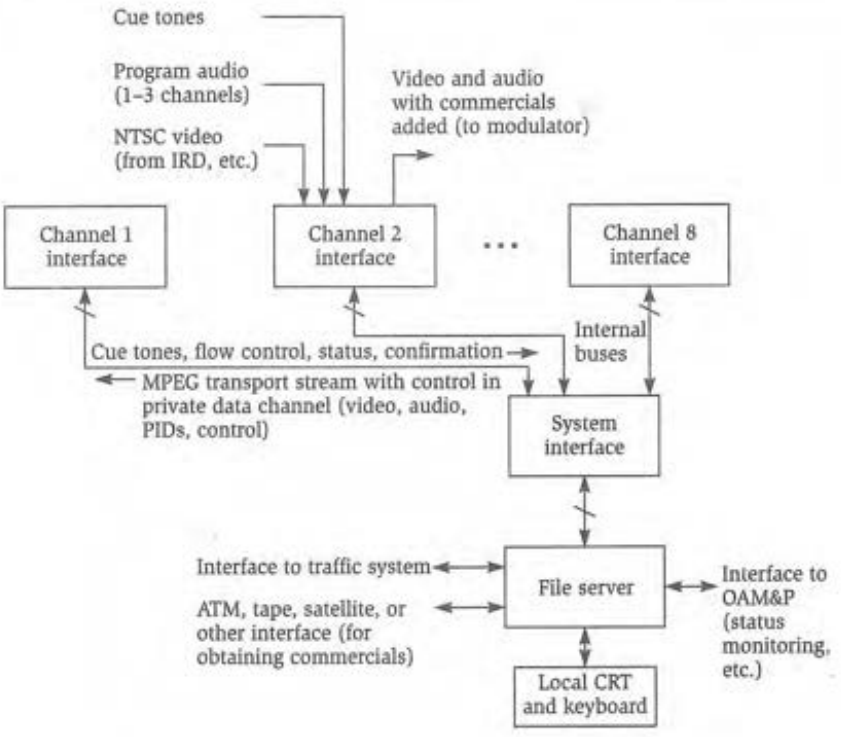


Figure 8.22 Video file server system used for ad insertion.

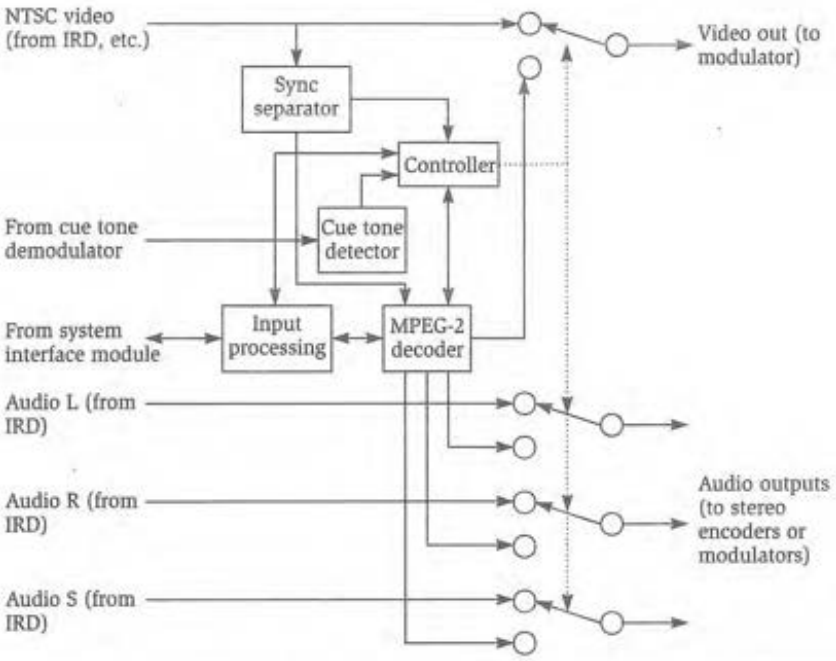


Figure 8.23 Individual channel interface in ad insertion system.

of Figure 8.16, recover the cue tones. The cue tones are passed to the file server, where they initiate playing of commercials.

Figure 8.23 illustrates an individual channel interface. NTSC video is supplied to a switch, which allows replacement of the incoming video with the commercial. Sync is extracted from the video to allow the inserted video to be synchronized with the incoming video, a requirement for smooth switching between sources. An MPEG decoder receives video from the file server and converts it to NTSC video. The cue tones are supplied to a detector, which passes them on, ultimately to the file server. One to three (if SAP is included) channels of audio are switched with the video.

The video and audio from the ad insertion system are supplied to a modulator and, usually, stereo encoder, for transmission to subscribers. If the inserted video is synchronized properly with the incoming video, and the actual switchover between incoming and advertising video is done in the vertical blanking interval, the subscriber does not see a change that indicates the video is being originated locally. Cue tones cannot be heard because they are transmitted on a channel other than the one used for program audio.

Ad insertion in digital programs is covered next.

8.11 Video On Demand and Digital Ad Insertion

Video on demand (VOD) is a popular service. VOD allows a subscriber to order his or her own custom video program from the headend. It is the cable equivalent of the video rental store, without late fees and rewind fees! Programs are stored on a server and listed in the electronic program guide (EPG) on the subscriber's STT. When a subscriber orders a VOD program, the request is relayed to the headend on the return data channel from the STT. The VOD server must determine a free time slot in a free channel on that subscriber's node. The STT is told what RF channel to tune to and what PID to select on that channel. The server then "plays out" the program, multiplexing it with several other signals bound for other subscribers on that node.

VOD is a personal service sent individually to one subscriber. However, in order to reach that one subscriber, the program must be sent to all subscribers on that fiber node. Only the one subscriber for whom the program is bound has the correct descrambling key to allow recovery of the program. Usually the subscriber has the ability to perform "VCR-like" functions, such as pause, rewind, and fast-forward. The subscriber is allotted so many hours to watch the program, an amount of time somewhat longer than the run time of the program. This will allow for meaningful pausing and rewinding.

Sometimes two or more levels of file server hierarchy are employed, with a file server in a hub close to a group of subscribers that has the more popular titles on it. A file server in the master headend or some other location keeps less popular programming and supplies it when requested. A significant issue in VOD is the need to accurately keep track of what programming is available and where it is located. The program owners may have strong feelings about controlling where the programming resides and how it is distributed from there.

A VOD file server can be constructed architecturally the same as an ad insertion file server, except more storage capacity is required. To improve reliability, VOD servers may employ *redundant array of independent (or inexpensive) disks* (RAID) technology, a category of disk drives that employ two or more drives in combination for fault tolerance and performance. In RAID systems, multiple disk drives are used to store different portions of the program data such that the failure of one disk drive does not cause the loss of any program material. There are several levels of RAID drives, of which levels 0, 3, and 5 are the most popular.

Level 0 provides *data striping* (spreading out blocks of each file across multiple disks) but no redundancy. This improves performance but does not deliver fault tolerance. Level 3 is the same as Level 0 but also reserves one dedicated disk for error-correction data. It provides good performance and some level of fault tolerance. Level 5 provides data striping at the byte level and also stripe error-correction information. This results in excellent performance and good fault tolerance.[10] In other words, RAID spreads out the bits of each byte in the file over several disk drives and puts the data back in the right sequence upon play out.

Error correction is added to allow the file to be recovered if any one disk drive fails. Typically, the disk drives are hot-swappable, meaning you can change a failed drive out without stopping operation of the file server. The server can operate with one disk down by doing on-the-fly error correction. As soon as the faulty disk drive is replaced, it can rebuild the file by error-correcting the surviving data.

A related service is a network *personal video recorder* (PVR). The idea is to use a video file server to capture programs as they are broadcast and then to allow subscribers to view them later as a VOD program.

8.11.1 Digital Video Rate Management

Cable systems have long inserted advertising on analog channels. But with the advent of digital television, ad insertion had to be extended to this medium. In principle, the block diagram of Figure 8.23 applies to insertion in digital video streams, but there are some new issues to consider. The first issue in digital ad insertion is that you have to exit the programming stream in favor of the advertising stream. You do this once you have transmitted enough information to recover all frames of the program picture that are to be displayed prior to starting the advertisement. Then you must start the advertisement with an MPEG I-frame, since it is the only frame that will contain enough information to allow the first frame of the advertisement to be displayed. It is not enough to do this just for video; you must make the audio match what the video is doing, and audio is compressed independent of the video. Typically, the audio will be muted for a short time during the switch-over. Finally, since the advertisement is being inserted into a multiplex of channels that have usually been statistically multiplexed, you have to ensure that introduction of the advertisement does not overrun the capacity of the RF channel.

The first job, digital program insertion, is covered in Section 3.5.4. Figure 3.26 shows the steps that must take place to leave the normal program stream and go to the ad insertion stream. The second issue, managing the data rate to stay within the available bandwidth, is covered here. This discussion applies any time you modify the content of a digital video stream in a VBR environment (see next subsection). This happens when you insert a local commercial and when you groom an incoming video feed to select some programs and combine with other programs on a different incoming stream.

Constant and Variable Bit Rate

Video programs may be transmitted using either a *constant bit rate* (CBR) or a *variable bit rate* (VBR). The terms are self-explanatory. If a program is sent using CBR, the data rate is set to whatever is necessary to provide a good picture at all times. If fewer bits are required, then filler bits are added ("stuffed") to round out the set bit rate. If by chance, however, a picture would benefit from a higher bit rate for a time, the picture can't have a higher bit rate; the encoder must make do with the bit rate allocated.

VBR transmission is more efficient because it transmits only whatever data rate is necessary to recover a good picture. VBR transmission frequently originates where video is encoded, using a *statistical multiplexer*, or stat mux. A stat mux monitors all the data coming from a number of encoders, which data is being multiplexed into one transmission stream. The transmission stream, of course, is at a constant bit rate, required to enable the receiver to lock to the transmission. But the transmission stream can be composed of a number of VBR streams managed by the stat mux. The stat mux buffers each encoder's output, multiplexing all outputs into the output transmission stream. If the buffer starts to empty, meaning that not enough bits are going to be available to make up the output transmission stream, then one of three things can happen: (1) The stat mux can call for more data from one or more of the encoders, even if the additional data is not needed for quality video; (2) the data stream is stuffed with filler bits; (3) opportunistic data is transmitted, that is, data held in queue until there is room for it.

On the other hand, there will be times when the stat mux buffer starts to get too full, indicating that there is not going to be room in the transmission stream for all of the bits. In that case, the stat mux must tell one or more encoders to back off and encode to a lower bit rate, even if the target video quality is compromised. Hopefully, the data peak is over in a fraction of a second, and viewers remain unaware that an encoder had to compromise its target quality momentarily.

Most VOD systems store material in CBR format, though there could be savings of maybe 30% on storage space by using VBR.[11] The difficulty with VBR is that when you multiplex several VOD streams together, you have a problem managing the data rate. This job is simplified if you work with CBR.

Many program multiplexes are encoded using VBR, however, in order to gain efficiency in transmission. The problem comes in when you pull some program streams out of a VBR transmission stream and combine them with other VBR streams, from an ad insertion system, from a broadcaster, or from somewhere else where you have no control over the data rate to which the material was encoded.

Figure 8.24 illustrates a VBR transmission stream with four programs. In order to provide better quality for programs that we want to look better, we have set a peak rate for program 2 that is three times that of programs 1 and 4, and we have assigned a peak rate for program 3 that is twice that of programs 1 and 4. You can see that in two instances in the time monitored, the total rate for the four programs exceeds the capacity of the channel. If the programs were being put together by a stat mux, it would have one or more encoders back off during the peaks. You can also see several valleys in the total rate, in which opportunistic data may be inserted if available.

The problem comes in where you have assembled this transmission stream from several independently encoded programs and you have no control over the incoming data rate of each. Suppose, for example, that program 4 is a commercial you have inserted into a transmission stream. You have no control of the three program streams you are passing through, and the commercial substituted for program 4 is preencoded, so you have no capability to influence its data rate

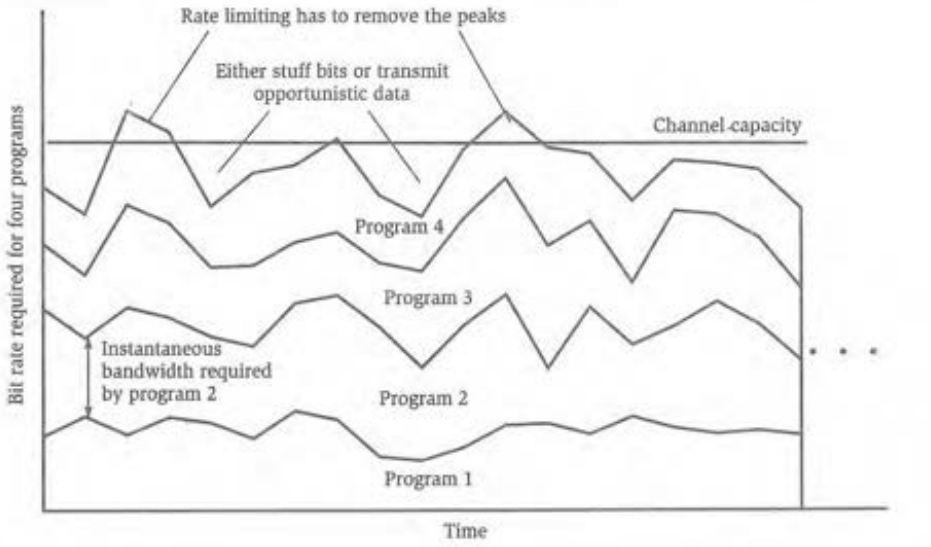


Figure 8.24 VBR transmission stream with four programs.

either. Specialized equipment called a *rate changer*, or a *groomer*, has been developed for this situation.

It is possible to decode each program back to the non-MPEG-2 encoded video, and to reencode using a stat mux system. There are a couple of difficulties with doing so. The first is that it is rather expensive, and it must be done with each program in the multiplex. The other problem is that, as pointed out in Chapter 3, MPEG-2 encoding is a lossy process. If you cascade multiple encoding, you are going to damage the picture. It is possible to find that the second encoding process needs a higher bit rate to handle artifacts introduced in the first encoding process.

In order to avoid these problems, a rate changer analyzes the picture and makes a judgment on what information can be dropped from the program stream without doing noticeable damage to the picture. It may be that the rate changer will decide that it doesn't need to encode the higher-order DCT coefficients with as many bits as was done originally, or it may decide that it can drop some of the higher-order coefficients altogether, for example. Figure 3.9 and the associated text shows what is going on here.

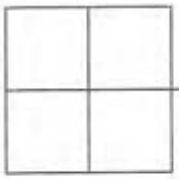
8.12 Summary

This chapter has provided information about the common types of equipment found today in headends. This equipment includes processors, modulators,

demodulators, stereo encoders, earth station receivers, and advertising insertion equipment. We have demonstrated typical circuit sets and features, recognizing that a number of ways to do each job exist, and that not every piece of hardware will have every feature shown. The next chapter will cover what happens to signals after they exit the individual pieces of equipment and will also discuss other techniques related to headend use.

Endnotes

1. www.wegener.com. See data sheets for DTV 700 Series.
2. Donald G. Fink, *Electronics Engineers' Handbook*. New York: McGraw-Hill, 1975, pp. 2–26.
3. Christopher Bowick, *The Importance of Setting and Maintaining Correct Signal and Modulation Levels in a CATV System Carrying BTSC Stereo Signals*. This has appeared in several places, but a useful source is the *BCT/E Certification Program Reference Bibliography Reprint Manual*, Society of Cable Telecommunications Engineers, 140 Philips Rd., Exton, PA 19341–1318. Telephone 800–542–5040. Many other useful reprints are also found in this volume.
4. IBA, BREMA, BBC, *NICAM 728: Specification for Two Additional Digital Sound Channels with System I Television*, 1988. The current status of this document cannot be determined. However, the Instituto de Communicagoes de Portugal has published substantially the same information, which can be found at http://www.icp.pt/legisuk/p316–93uk.htm.
5. Televerket Radio Laboratory (Sweden), *Digital Multisound in Television—Feasibility of Filtered 728-Kbit/s NICAM with 5.85-MHz Carrier Frequency in CCIR System B. Adjacent Channel Operation in Cable Networks*. Paper is dated January 16, 1987, and was privately furnished to the author. Publication information is unknown.
6. DVB Project Office, *Interfaces for CATV/SMATV Headends and Similar Professional Equipment*. DVB Document A010, rev. 1, 28 May 1997.
7. ETSI, *Digital Video Broadcasting (DVB); Professional Interfaces: Guidelines for the implementation and usage of the DVB Asynchronous Serial Interface (ASI)*. ETSI TR 101 891 b1.1.1 (2001–02).
8. Per private correspondence with Bob Gaydos, Concurrent Computer Corporation.
9. Linc Reed-Nickerson, Audio Levels Scream for Attention. *CED*, October 1997, p. 54ff.
10. www.webopedia.com
11. Per private correspondence with Bob Gaydos, Concurrent Computer Corporation.



Chapter 10
Coaxial RF Technology

10.1 Introduction

Cable television's technical roots are the distribution of analog television signals to the antenna terminals of customers' receivers. The least expensive way to accomplish that was to avoid the need for in-home equipment by carrying each video stream on a different, standard television channel so that subscribers could use their existing TV sets to select and view programs. Although cable systems have evolved since then into sophisticated multiservice networks using combinations of linear fiber optics and coaxial cable for transmission, the essential characteristics have remained unchanged. In particular, the last portion of the network is still a linear, broadband coaxial network that simultaneously carries many modulated RF signals, each occupying a different band within the spectrum — frequency division multiplexing (FDM) — as discussed in several previous chapters. In modern systems, many of the modulating signals may be digital rather than analog; however, the network must still be linear to avoid generation of unwanted additional signals due to mixing among the desired carriers. This is discussed in detail in Section 10.3.3.

This chapter will treat coaxial network technology in detail, including cable, amplifiers, passive components, and powering systems. Basic linear network concepts will be introduced that will also apply to the linear fiber-optic links whose unique characteristics are discussed in Chapter 12 and to the microwave links covered in Chapter 14. Chapter 11 will discuss coaxial design practices and cascaded performance.

10.2 Coaxial Cable

Coaxial cable is not the only option for transmitting broadband RF signals. Indeed, many early systems were built using open, parallel-wire balanced transmission lines, and a few even used an ingenious single-wire cable

known as G-line, which had only a center conductor and dielectric. Coaxial cable, however, offers the advantages of a high degree of shielding, coupled with relatively low cost and moderate signal loss. For that reason, most modern RF systems use coaxial cable as at least one of the physical media through which signals are transported.

10.2.1 Definition

Coaxial cable is constructed with a center conductor surrounded by a dielectric of circular cross section and by an outer conductor (shield), also of circular cross section. Signals within the normal operating bandwidth of coaxial cable have a field configuration known as transverse electric and magnetic (TEM). In the TEM mode, the electric field lines go radially between the center and outer conductor and are of uniform strength around a cross section of the cable, whereas the magnetic field lines are circular and perpendicular to the length of the cable (see Figure 10.1). In a cable with a continuous, perfectly conducting shield, no electric or magnetic fields extend beyond the outer conductor so that no signals can leak out of or into the cable.

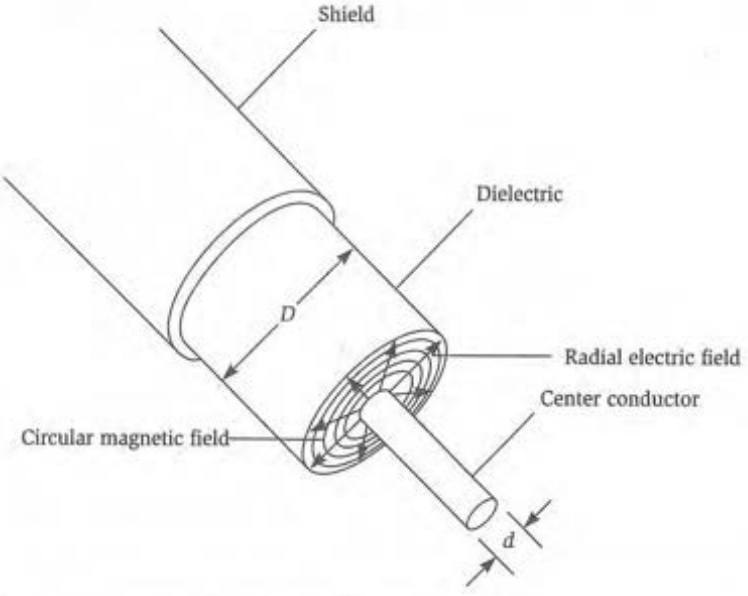


Figure 10.1 Coaxial cable basics.

10.2.2 Characteristic Impedance

Coaxial cables have a property known as surge impedance or, more commonly, characteristic impedance, which is related to the capacitance and inductance, per unit length, of the cable. The characteristic impedance is most easily thought of in terms of the effect on signals being transported: if a cable is connected to an ideal pure resistor whose value is equal to its characteristic impedance, a signal transmitted toward the resistor will be entirely absorbed by the resistor and converted to heat. In other words, no energy will be reflected up the cable.

The characteristic impedance, Z_0 (in ohms), is a function of the relative diameters of the center conductor and the inner surface of the outer conductor and of the dielectric constant of the dielectric:

$$Z_0 = \frac{138}{\sqrt{\varepsilon}} \log\left(\frac{D}{d}\right) \tag{10.1}$$

where

D = the inner diameter of the shield
d = the outer diameter of the center conductor
ε = the dielectric constant

10.2.3 Attenuation as a Function of Frequency

Signal loss (attenuation) through coaxial cable can occur through any of four principal means:

- Radiation out of the cable due to imperfect shielding
- Resistive losses in the cable conductors
- Signal absorption in the dielectric of the cable
- Signal reflection due to mismatches between the cable and terminations or along the cable due to nonuniform impedance

Even when cables have perfect shields, exact impedance matches, and uniform construction, imperfect dielectrics and resistive conductors will cause loss. The general equation for this residual cable loss is

$$\alpha = 4.344\left(\frac{R}{Z_0}\right) + 2.774F_p\sqrt{\varepsilon}\, f \tag{10.2}$$

where

α = the attenuation of the cable in dB/100 feet*
R = the effective ohmic resistance of the sum of the center and outer conductors per 100 feet of cable length at f
F_p = the power factor of the dielectric used
f = the frequency in MHz

The currents in the conductors are proportional to the strength of the magnetic fields at the conductor surface. If the conductors had no resistance, the RF currents would travel only on the surface. In real conductors, however, the current extends into the conductor, decreasing exponentially with depth. This property is known as *skin effect*, and the distance from the surface to where the current has decreased to $1/e$ (36.8%) of the surface amount is known as the *skin depth* at that frequency. It is related to frequency by the formula

$$\delta = \frac{1}{\sqrt{f}} 2.60 \sqrt{\frac{\rho}{\rho_C}} \tag{10.3}$$

where

δ = the skin depth in mils (thousandths of an inch)
ρ/ρ_C = the resistivity of the conductor relative to copper
f = the frequency in MHz[1]

The formula is valid only for nonferrous materials such as aluminum or copper.

Over a typical downstream frequency range of 54–750 MHz, the skin depth in copper will vary from 0.00035 to 0.00009 inch and will increase to 0.0012 inch at 5 MHz (the lower end of the upstream frequency range).

The effective resistance of the conductor is the same as if it were a tube of material whose thickness is equal to the skin depth with the current distributed equally throughout its volume. The resistance of this tube will be

$$R = 0.0996 \frac{\sqrt{f}}{d} \sqrt{\frac{\rho}{\rho_C}} \tag{10.4}$$

* In accordance with common North American practice, all attenuation values will be given in decibels per 100 feet of length. This has the advantage that signal loss becomes a linear function of length and of simple conversion to metric units of length.

where

R = the resistance per 100 feet of the conductor
ρ/ρ_C = the resistivity of the conductor relative to copper
f = the frequency in MHz
d = the conductor diameter in inches

Taking into account the resistance of both center and outer conductors, Equation (10.2) can be rewritten as

$$\alpha = \frac{0.433}{Z_0}\left(\frac{\sqrt{\rho_d/\rho_c}}{d} + \frac{\sqrt{\rho_D/\rho_c}}{D}\right)\sqrt{f} + 2.774(F_p\sqrt{\varepsilon})f \qquad (10.5)$$

where

ρ_d/ρ_c = the resistivity of the center conductor material relative to copper
ρ_D/ρ_c = the resistivity of the shield material relative to copper

This has one term that increases as the square root of frequency and one that increases linearly with frequency. For most cable constructions (discussed later), the conductor loss will dominate the dielectric loss so that the overall cable attenuation will increase approximately as the square root of frequency. Figure 10.2 shows the specified variation in loss for several cables in comparison to an ideal square-root relationship (with the curves matched at 100 MHz). Note that the agreement is excellent for cables ranging from drop sizes to the largest trunk cables, indicating the relatively small contribution from dielectric losses.

A common cable construction, known as P3, utilizes a solid aluminum outer conductor, a foamed dielectric ($\sqrt{\varepsilon} = 1.15$), and a copper-coated aluminum center conductor. It is available in a number of nominal sizes (referenced to the *outer* diameter of the aluminum shield in inches). For this family of cables, the attenuation can be approximated using

$$\alpha = \left[\frac{0.036}{D_{\text{nom}}}\right]\sqrt{f} + 0.0002f \qquad (10.6)$$

where

α = the attenuation in dB/100 feet
D_{nom} = the nominal shield outer diameter in inches
f = the frequency in MHz

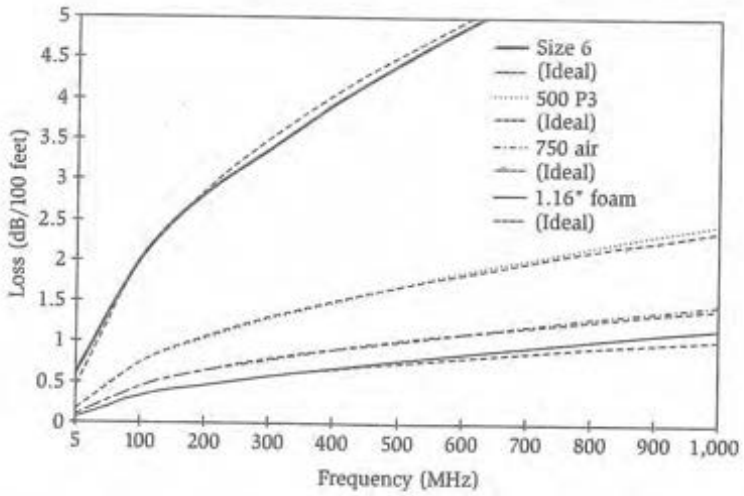


Figure 10.2 Cable loss: specified versus ideal.

Other cable types using thinner shields and/or air dielectrics will have slightly lower losses for the same outer diameter, whereas solid dielectric cables will have higher losses.

10.2.4 Attenuation as a Function of Temperature

As Equation (10.2) shows, the attenuation of a coaxial cable, in decibels, is a linear function of its conductor resistance, provided that conductor losses are much larger than dielectric losses. As we will discuss in Section 10.2.10, the most common cable construction uses an aluminum shield and copper-coated aluminum center conductor. Copper has a resistivity at 68°F of about 0.68×10^{-6} ohm-inches, whereas aluminum has a resistivity of about 1.03×10^{-6} ohm-inches, and each has a temperature coefficient of resistivity of 0.22%/°F. Since conductor resistance varies as the square root of resistivity, the resistance of both conductors, and therefore the attenuation, would be expected to change about 0.11%/°F. In fact, the attenuation of commercial coaxial cables with hybrid aluminum–copper construction is typically specified to increase by 0.1%/°F.

10.2.5 Attenuation as a Function of Characteristic Impedance

For a given outer diameter (the primary determinant of the amount of materials used and, therefore, cost and weight), different impedance cables will be optimized for different characteristics. If the dielectric loss is assumed to be of only secondary

importance and the usual construction is assumed (copper-plated center conductor with aluminum shield), then Equations (10.1) and (10.5) can be combined to find cable attenuation as a function of characteristic impedance:

$$\alpha = \frac{0.433\sqrt{f}}{D}\left[\frac{1.23 + 10^{\left(\frac{Z_0\sqrt{\varepsilon}}{138}\right)}}{Z_0}\right] \tag{10.7}$$

As Figure 10.3 shows, the loss minimum is at about 80 ohms, for air dielectric (dielectric constant = 1.0) and decreases as the dielectric constant increases. Cables with air dielectric also have lower overall loss since the center conductors are larger, and therefore have less resistance, for the same impedance. Although the choice of impedances near the minimum of the curve was not important, 75 ohms may have been chosen because it is also close to the feedpoint impedance of a half-wave dipole antenna. In order to minimize the need for repeaters, wide area distribution systems universally use 75-ohm cables. They are also used for local interconnection of baseband video signals within headends and broadcast facilities in order to minimize the differential loss across the video baseband spectrum.

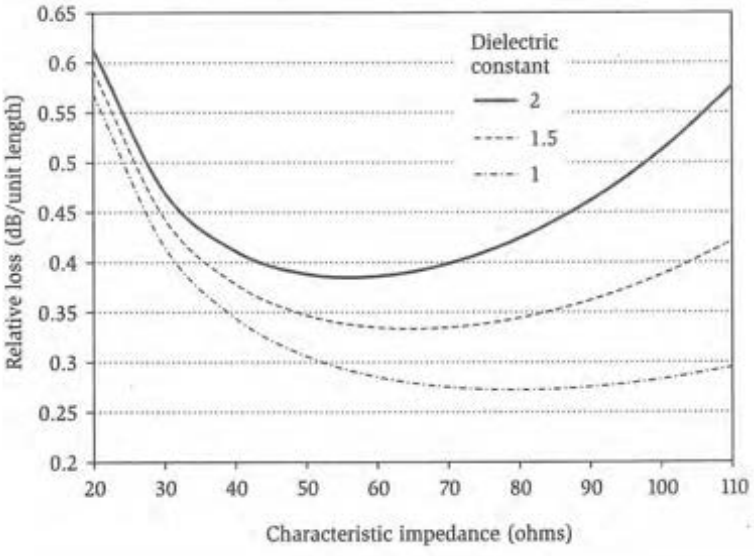


Figure 10.3 Coaxial cable loss versus characteristic impedance for various values of dielectric constant.

Although the losses are higher, 50-ohm cables are generally used by the broadcast and radio communications industries because the power-handling capability is higher, the cables are less fragile, and they are a close match to the feedpoint impedance of a vertical quarter-wave antenna.

10.2.6 Wavelength

Along the length of the cable, at any one instant in time, the electric fields vary sinusoidally in strength. At some point, the center conductor will be at its most positive with respect to the shield. Moving along the cable, the voltage will decrease to zero, then become negative, then back to zero, and finally be at its positive maximum again. The distance between the maximum points of one polarity is known as the wavelength at the frequency being transmitted and is equal to the distance the signal transverses during one period of its frequency.

In free space, signals travel at the rate of 3×10^8 meters per second or, in more convenient terms, about 984 feet per microsecond. Signals in cable, however, travel more slowly due to the higher dielectric constant of the dielectric material. The ratio between velocity in cable and the free-space velocity is the *relative propagation velocity* V_P and varies from about 0.85 to 0.95 for common cable types. V_P is related to the dielectric constant by

$$V_P = \frac{1}{\sqrt{\varepsilon}} \tag{10.8}$$

The wavelength of a signal in a coaxial cable is thus

$$\lambda(\text{feet}) = \frac{984 V_p}{f(\text{MHz})} = \frac{984}{f\sqrt{\varepsilon}} \tag{10.9}$$

10.2.7 Theoretical Size Limitation

If the size of a coaxial cable is sufficiently large, it will support modes other than TEM. In particular, if the mean diameter exceeds one wavelength in the cable, it will support a mode in which the electric and magnetic fields are not uniform around the cable. This *cutoff wavelength* can be expressed as[2]

$$\lambda_C = \pi(D + d) \tag{10.10}$$

where

D = the inner diameter of the shield
d = the outer diameter of the center conductor

Non-TEM mode generation is a problem since higher-order modes may not propagate at the same velocity and because coaxial components may not react the same to the higher-order mode. Standard practice is to limit cable sizes to those that do not support modes higher than TEM at the maximum operating frequency. The maximum size cable that will support a given frequency in TEM mode only can be found by substituting Equation (10.9) for λ in Equation (10.10), then solving Equation (10.1) for d, substituting in Equation (10.10), and solving the resultant expression for D:

$$D_{\max} = \frac{984}{\pi f \sqrt{\varepsilon}\left[1 + 10^{-\left(\frac{Z_0\sqrt{\varepsilon}}{138}\right)}\right]} \tag{10.11}$$

Alternatively, we can solve for the maximum frequency that can be transmitted in only TEM mode through a given cable:

$$f_{\max} = \frac{984}{\pi D \sqrt{\varepsilon}\left[1 + 10^{-\left(\frac{Z_0\sqrt{\varepsilon}}{138}\right)}\right]} \tag{10.12}$$

For example, 75-ohm cables with foamed dielectrics (ε about 1.3) will support 1-GHz signals at up to 0.222-foot shield inner diameter, or about 2.6 inches. Since this is more than twice the largest cable used by network operators today, higher-order mode suppression is not an issue with current operating bandwidths.

10.2.8 Precision of Match: Structural Return Loss

One measure of cable quality is how close it adheres to its nominal impedance. This has two aspects: the precision with which its *average* characteristic impedance matches the ideal value, and the *variations* of impedance along the length of a cable. As an alternative to attempting to measure the physical structure of the cable along its length, the most frequent measure of quality used is the percentage of incident power from a source that is reflected at the input of a cable being tested, when the cable's output is connected to a precision termination. If the source is calibrated using a precision, resistive terminator, then the result of the measurement is return loss. For cables, however, the variation of the reflection is often of more interest than the absolute match. In order to judge the variation, a variable bridge is used to match, as precisely as possible, the average surge impedance of the cable being tested. Then the reflection is measured over the full frequency range of desired operation, and the result is known as structural return loss (SRL). Its numeric value is the smallest measured ratio of incident to reflected power as the frequency is varied, expressed in decibels. Recommended worst-case SRL is 26–30 dB for trunk and feeder cable, whereas

20 dB is considered adequate for drop cable. The absolute impedance should be within ±2–3 ohms for trunk and distribution cables and ±5 ohms for drop cables.[3]

SRL is a compromise measurement, as the reflected power is a measure, at each frequency, of the composite effect of discontinuities at all distances. Because of the attenuation of the cable, signals traveling toward more distant discontinuities will be reduced in amplitude, and the reflections from those discontinuities will be further attenuated so that imperfections farther from the end of the cable being measured will have less effect than those closer. At higher frequencies, this effect will be increased. On the other hand, most cable imperfections cause a greater reflection at higher frequencies.

Where multiple reflections are spaced along a transmission line, the effects can add, cancel, or anything in between, depending on the length of line between them and the frequency. (Chapter 15 includes a more complete discussion of multiple reflections in coaxial transmission systems.) Absent line losses, small, identical discontinuities will approximately cancel when spaced odd multiples of a quarter wavelength, whereas they will reinforce when spaced multiples of one-half wavelength. This is especially important when considering multiple, equally spaced discontinuities arising, for instance, from an imperfection in a roller over which a cable passes during the manufacturing process. Each revolution of the roller produces an identical fault, spaced along the cable by a distance equal to the circumference of the roller. Testing such a cable over the full frequency range will show maximums and minimums of return loss as the effects of the discontinuities interact. This is why it is important to test cable over its full expected frequency range with sufficient frequency resolution to detect sharply resonant discontinuities.

Though structural return loss measurements accurately assess the composite reflection from a cable, it is a poor method of determining the location of major defects. If a cable includes only a single major defect, its location can be estimated from the change in test frequency between successive maxima or minima of return loss; however, in most cases, cable loss plus a multiplicity of reflections will prevent accurate determination.

10.2.9 Precision of Match: Time Domain Reflectometry (TDR)

An alternative measurement method, also a compromise, is used to more accurately determine discontinuity location but averages the magnitude of the discontinuity as a function of frequency. In this test, a pulse is transmitted down the cable, and the distance is determined from the time required for the energy reflected from a discontinuity to return to the transmitter. The magnitude of the discontinuity is related to the amplitude of the returned signal. Secondary information about the frequency "signature" of the fault can be determined by comparing the shape of the returning pulse with the original.

Although TDR testing of short, low-loss cables is an accurate process, there are many compromises in testing of the relatively long cables used in typical cable system construction. With very short cables, the classic pulse waveform is a very fast rise time step function. A Fourier transform of this voltage waveform reveals that its frequency content is inversely proportional to frequency, and therefore, the power spectrum falls inversely as the square of frequency (see Figure 10.4(a)). When such a pulse is used to test long cables, it suffers from two problems: (1) most of the energy is concentrated at low frequencies, whereas most discontinuities primarily affect higher frequencies, and (2) the relatively higher cable loss at high frequencies further masks the reflections.

For that reason, long cables are most frequently tested using a narrow "impulse" (see Figure 10.4(b)). This waveform has a spectrum that extends more uniformly to a frequency equal to approximately $1/P_W$, where P_W is the pulse width (the exact shape of the curve in the frequency domain depends on the shape of the voltage pulse). The trade-off is that the distance resolution is limited to approximately $cP_W V_P$, or about 1 foot per nanosecond of pulse width for typical foam dielectric cables. By increasing P_W, it is possible to increase the total energy (and therefore sensitivity) at the expense of frequency and distance

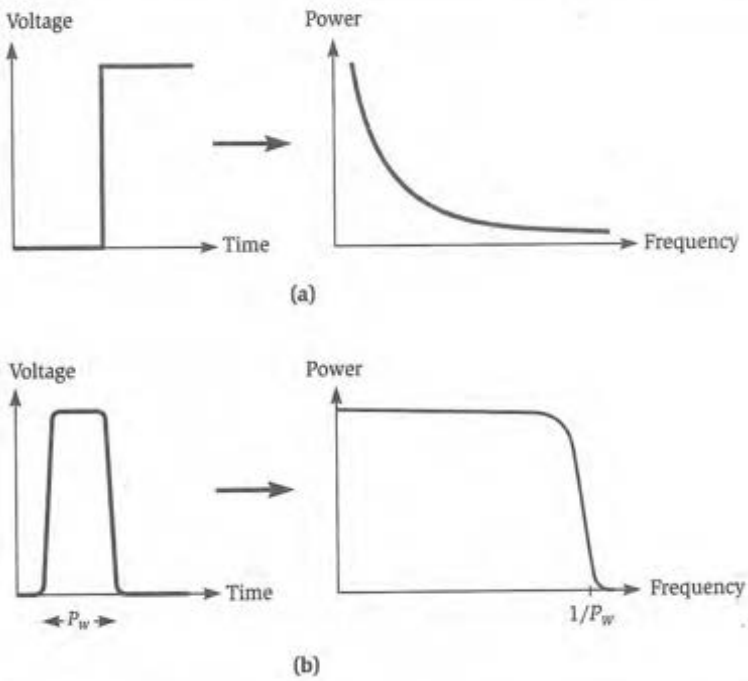


Figure 10.4 Frequency spectra of step and pulse TDR waveforms. (a) Step function. (b) Pulse.

resolution. As with step function pulses, impulse function testing suffers in accuracy because of the variation of cable loss with frequency. Because of the compromises, TDR testing of CATV-type cables is primarily limited to field operations such as measuring the cable remaining on a reel or locating an underground cut so that the cable can be repaired.

An instrument that is able to simultaneously resolve both frequency and time (distance) domain information is a network analyzer with gated TDR and Fourier transform capability. This instrument measures reflection coefficient, but "gates" the measurement so as to measure only reflections from a narrow range of distances at a time. A Fourier transform is then used to convert the waveform of the reflected signal into its frequency domain, resulting in a reasonably accurate measurement of the reflection coefficient of an individual discontinuity, even in the presence of other discontinuities at other distances. It is still, of course, affected by the cable loss and loss variation with frequency. As of the writing of this book, this technology has not been extensively used in field work in cable systems.

10.2.10 Practical Factors in Cable Selection

From an electrical standpoint, the ideal cable would have conductors with the highest possible conductivity, the largest diameter consistent with higher-order mode suppression, and a vacuum separating the conductors. Practical cables, however, are a compromise among several characteristics:

- Manufacturing cost (which will vary somewhere between being a linear function of diameter and as the square of the diameter)
- Weight (which affects the cost of the supporting structure for overhead lines)
- Diameter (which affects required conduit size for underground lines, visibility of overhead lines, and cost of connectors)
- Loss (which limits the distance between repeater amplifiers)
- Mechanical expansion coefficient (which affects the amount of sag that must be used with overhead lines)
- Connector cost (related to cable size)
- Shielding (although solid outer conductors provide nearly perfect isolation, braided shields may provide adequate shielding for some applications)
- Environmental protection (mechanical protection layers, compounds used to retard water ingress, integral support members, and so on)
- Handling characteristics (bending radius, reforming ability, crush resistance, pulling force resistance, and so on)
- Electrical resistance and current-carrying capacity at low frequencies

In oversimplified terms, the selection of optimum cable size for a distribution network design consists of comparing the cost of cable (and associated components) per decibel of loss with the cost per decibel of amplification to overcome that loss and choosing the least expensive combination. Of course, practical system design includes many additional factors that may affect the ultimate decision.

Trunk and Feeder Cable

Design optimization has resulted in several typical cable designs that are widely used. The most common distribution cable design in most of the world utilizes a solid aluminum shield, foamed polyethylene dielectric, and a copper-clad aluminum center conductor. The use of aluminum offers the best combination of cost, handling characteristics, weight, and strength, whereas the foamed dielectric offers a good compromise among loss, moisture ingress protection, and mechanical strength for the center conductor. Because of its rigidity, compared with cables that utilize braided wire shields, trunk and feeder cables using this basic structure are sometimes known as *hard-line* cable. The use of an aluminum core for the center conductor provides a match for the expansion coefficient of the shield, whereas the copper cladding provides a lower resistance to RF. Since the conducting area of the center conductor is much smaller than the shield, the total resistive losses are dominated by it so that copper cladding the center conductor is cost efficient though coating the inner side of the shield would not be.

The need to provide low resistance to the multiplexed power, as well as low RF loss, has led to the use of cables with outer shield diameters ranging from 0.412 to 1.125 inches, with larger cables being used where distance of coverage is paramount (trunk or express feeders), and smaller cables being used for shorter distances and cases where a substantial amount of the loss is due to tap losses rather than cable losses (tapped feeder cables). This family of cables is generically labeled by the outer shield diameter in thousands of an inch; for example, "500 cable" refers to a hard-line cable with an outer shield diameter of 0.500 inch. Common diameters include 0.412, 0.500, 0.625, 0.750, 0.825, and 1.000 inch though others are also used.

Versions of these same cables are available with solid copper center conductors, for use where 60-Hz power losses must be reduced. Protection options include combinations of none (bare aluminum), polyethylene jackets of various designs and degrees of sun and fire resistance, flooding compounds (a substance inserted between the shield and jacket to retard the ingress of water), and additional steel shields for crush resistance. Since the added environmental protection layers can considerably increase the cost and weight, builders generally utilize several designs in a single system, optimized by application.

A variation of this design utilizes air dielectric with periodic center conductor supports. This design has slightly smaller diameter for the same loss per foot

because of the lower dielectric loss and lower ratio of shield to center conductor diameter. Special design features are required to prevent water, which may enter the cable owing to a fault, from traveling down the center conductor/shield space. Other features are required to maintain the required concentricity and crush or bending resistance.

Though hybrid aluminum–copper cables are the most common, some network operators prefer solid copper for center conductors and, sometimes, shields.

Drop Cable

In a typical cable system, well over half the total cable footage is taken up by drop cables. For that reason, cost is an important parameter for that application, as well as weight and appearance. In North America, four common sizes of drop cable have evolved, as shown in Table 10.1. The approximate loss for each size, as a function of frequency, is shown in Figure 10.5.[4]

Table 10.1 Common drop cable designations and sizes

Size Designation	*59*	*6*	*7*	*11*
Approximate diameter over jacket in inches	0.24	0.27	0.34	0.40

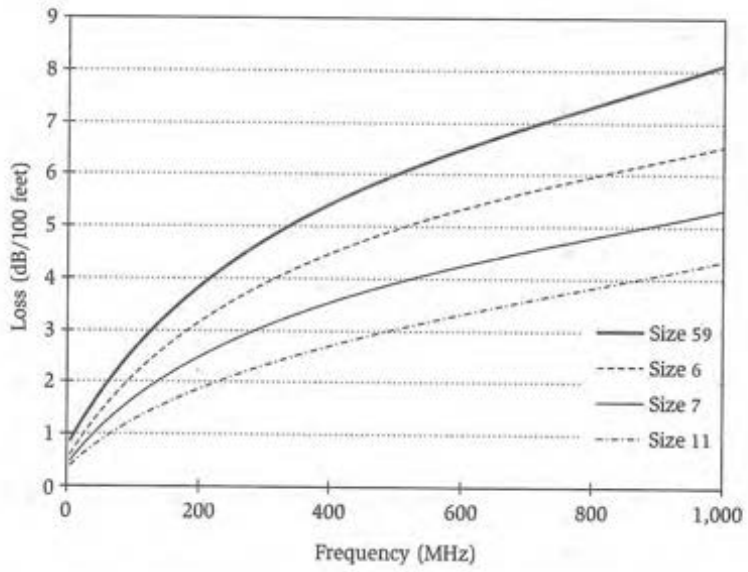


Figure 10.5 Drop cable loss versus frequency.

Drop cables are constructed using a copper-plated steel center conductor, foamed dielectric, a shield made up of at least one layer of an aluminum foil and one layer of an aluminum braid, and a plastic jacket. Additional layers of foil and braid are available as options for greater shielding effectiveness. Although size 59 cable was the historical choice of the cable television industry, size 6 cable is most frequently used in modern systems, due to the use of increasingly higher frequencies. The larger sizes are generally reserved for occasional longer drop cables.

The jacket material for general use is usually PVC or polyethylene. Cables used for overhead drops often also include a separate steel strength member (the *messenger*) enclosed in the same plastic sheath structure. See Figure 10.6 for a composite drop cable structure that includes a messenger.

An array of drop cable variations are available for special purposes, including those with *flooding compounds* to retard water ingress in case the plastic

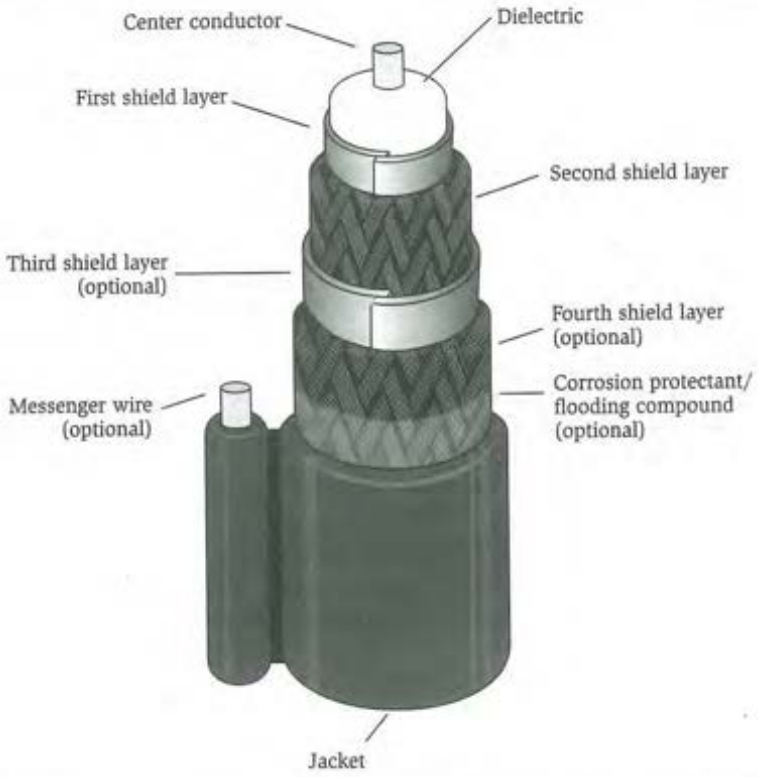


Figure 10.6 Typical drop cable construction.

coating is damaged, multiple cables within a common jacket, those with varying degrees of fire resistance for use inside buildings, silver-plated center conductors, copper braids, solid dielectric cables, and so on.

As with distribution cables, North American practice has been adopted by some countries, whereas others (for instance, Germany) have chosen to use semirigid drop cables with solid copper shields and center conductors for their underground drops.

10.2.11 Shielding Effectiveness in Drop Cables

Although distribution cables constructed with solid aluminum tube shields offer nearly perfect signal isolation from external fields, drop cables depend on the overlap in the aluminum wrap and the interaction between the aluminum braid and wrap for shielding integrity. It is a compromise that is justified on the basis of cost and cable-handling characteristics. The cable industry has long struggled to find nonambiguous testing methodology for evaluating shielding integrity. Although these are beyond the scope of this book, you may wish to refer to industry standard test procedures developed by the Society of Cable Telecommunications Engineers (SCTE) for this purpose.[5]

10.3 Amplifiers

In order to overcome the loss of the distribution cables and to provide power to drive end terminals, signal boosting amplifiers are utilized throughout the coaxial distribution system. These must be designed to add as little signal degradation as possible, particularly noise and distortion, consistent with providing the required gain and total power output.

10.3.1 Broadband Random Noise

Because of the random motion of electrons in conductors, all electronic systems generate an irreducible amount of noise power. This noise is a function of the absolute temperature and the bandwidth in which the noise is measured. As discussed in Chapter 7, the minimum thermal noise power (the *noise floor*) can be calculated using

$$n_p = kTB \tag{10.13}$$

where

n_p = the noise power in watts*

k = Boltzmann's constant (1.374×10^{-23} joules/°K)

T = the absolute temperature in °K

B = the bandwidth of the measurement in Hz

Signal power in cable television systems is commonly expressed in decibels relative to 1 mV† across 75 ohms (dBmV), so the thermal noise power at 62°F is approximately

$$N_p(\text{dBmV}) = -125.2 + 10 \log B \tag{10.14}$$

Finally, although the allocated channel bandwidth for analog NTSC signals is 6 MHz, the effective noise bandwidth of receivers is less. For a variety of reasons, a 4-MHz bandwidth is used as the standard for noise measurements of analog NTSC signals and has been codified in the FCC's rules.[6] At that bandwidth, Equation (10.14) becomes

$$N_p = -59.2 \text{ dBmV} \tag{10.15}$$

This value will vary by a few tenths of a decibel across the temperature range normally of interest to cable operators.

10.3.2 Amplifier Noise

Amplifiers generate added noise at various points in their circuitry. For convenience, however, the added noise is treated as if it were coming from an independent generator and summed into the input port (though it cannot be measured there, of course). The ratio of total effective input noise power to the thermal noise floor, expressed in decibels, is known as the *noise figure* of the amplifier. Thus, an amplifier with a noise figure of F_A dB will have a total equivalent input noise power of N_A, where

$$N_A \text{ (dBmV)} = N_p \text{ (dBmV)} + F_A \text{ (dB)} \tag{10.16}$$

---

* To avoid confusion, lowercase symbols will be used in this discussion to refer to scaler quantities (for example, watts), and uppercase symbols will refer to logarithmic quantities (for example, dBmV).

† In some countries, the reference is 1 μV rather than 1 mV. Levels in dBμV are therefore numerically greater by 60 than levels in dBmV. In either case, however, the reference is power, not voltage. Finally, for the convenience of those accustomed to working in systems referenced to 1 mW (dBm), 0 dBmV = −48.75 dBm. For consistency, this book uses dBmV as the units for RF power unless otherwise specified.

And the output noise power will be the input, increased by the gain, G, of the amplifier in decibels:

$$N_{out}\ (\text{dBmV}) = N_p + F_A + G \tag{10.17}$$

Similarly, a desired (noise-free) input signal, C_i, will be amplified by the same amount so that the desired signal output level $C_{out} = C_i\ (\text{dBmV}) + G\ (\text{dB})$, and the carrier-to-noise ratio will be*

$$\begin{aligned} \text{C/N}\ (\text{dB}) &= (C_i + G) - (N_p - F_A + G) \\ &= C_i - N_p - F_A \end{aligned} \tag{10.18}$$

When evaluated relative to the 4-MHz bandwidth commonly used for NTSC C/N measurements, this becomes

$$\text{C/N}\ (\text{dB}) = C_i\ (\text{dBmV}) + 59.2 - F(\text{dB}) \tag{10.19}$$

Note that, in an FDM system, C/N is evaluated on each signal independently, with the level of individual carriers compared with the noise in the bandwidth that affects that signal. As will be seen, the C/N may be considerably different for different carriers in the spectrum. (The design of systems with cascaded amplifiers will be treated in Chapter 11.)

10.3.3 Distortion

Amplifiers are not perfectly linear. That is, if the instantaneous output voltage is plotted against the instantaneous input voltage, the graph will not be a perfectly straight line. There are two causes of this nonlinearity: the inevitable minor small-signal nonlinearities of the semiconductor devices used and the compression that takes place as the amplifier nears its saturation voltage. Though the small variations define a distortion "floor," most of the distortion is due to saturation effects. They can take any of three related forms: even-order distortion, odd-order distortion, and cross modulation.

---

* As a reminder, the term C/N (carrier-to-noise ratio) is used in this book for RF modulated signals, whereas S/N (signal-to-noise ratio) is used for measurements of the noise level on baseband (unmodulated) signals. This is consistent with both regulatory terminology and common industry practice.

Even-Order Distortion (Composite Second Order—CSO)

A perfectly linear amplifier has a transfer function that is a constant; that is, the output voltage is a linear function of the input voltage. If the transfer function is not linear, then, in general, it can be expressed as a power series; that is,

$$e_o = Ae_i + Be_i^2 + Ce_i^3 + De_i^4 + \cdots \tag{10.20}$$

where

e_o = the output voltage
e_i = the input voltage
$A, B, C, D, \ldots$ = the gains at the various powers of the input voltage[7]

Those terms with even-numbered powers are called even-order distortion, and those with odd-numbered powers are called odd-order distortion. They act in very different ways in cable systems utilizing the most common channelization plans.

If a single sine wave signal $[e_i = E_i \sin(\omega t)]$ is transported through an amplifier with second- and fourth-order distortion, the output voltage will be

$$e_o = A[E_i \sin(\omega t)] + B[E_i \sin(\omega t)]^2 + D[E_i \sin(\omega t)]^4 \tag{10.21}$$

where

ω = the angular frequency
A = the linear component of gain
B = the amplitude of the squared nonlinear component
D = the amplitude of the fourth power component
E_i = the peak input voltage

Using standard trigonometric identities, this is equivalent to

$$e_o = \frac{BE_i^2}{2} + \frac{3DE_i^4}{8} + AE_i \sin(\omega t) - \frac{(BE_i^2 + DE_i^4)}{2}\cos(2\omega t) + \frac{DE_i^4}{8}\cos(4\omega t) \tag{10.22}$$

Analyzing this expression, we find that

- The first two terms of this expression represent a change in the average dc voltage that may correspond to a change in the power dissipation of the active device, but is not transmitted into the ac-coupled transmission system.

- The third term represents the desired output signal. Note that this amplitude does not change as the result of the even-order distortion.
- The fourth term represents an additional signal at twice the frequency of the input signal and in phase with it $\omega = \pi/2$ radians. In real amplifiers, B is substantially greater than D. Thus, for small values of E_i, the amplitude of this factor will increase as the square of the input voltage. At some input level, the fourth power term will begin to dominate, and the second harmonic level will increase more quickly. Since the level of the fundamental increases linearly with input signal level, the ratio of the fundamental to the second harmonic, at low levels, is a linear function of input level. Expressed in logarithmic terms, the "second-order ratio" degrades by 1 dB for every decibel increase in operating level so long as the amplifier is "well behaved" (meaning the fourth power terms are insignificant relative to the second power terms).
- The final term is the fourth harmonic of the fundamental. Like the fourth power term of the second harmonic, this product is generally small at normal operating levels and at all levels is at least 12 dB (one-fourth the voltage) below the second harmonic term. Note that this term is also in phase with the fundamental at $\omega = \pi/2$.

If we plot the fundamental and first two even harmonics on a time scale (Figure 10.7), we see an interesting relationship: the distortion-caused products are *antisymmetrical* with respect to the fundamental; that is, whatever voltage they have at the positive peak of the fundamental, they have exactly the opposite during the negative peak. The usual case of the combined composite signal is shown in the bold line in the figure, which shows one peak of the composite waveform flattened relative to the fundamental, whereas the other peak is sharpened.

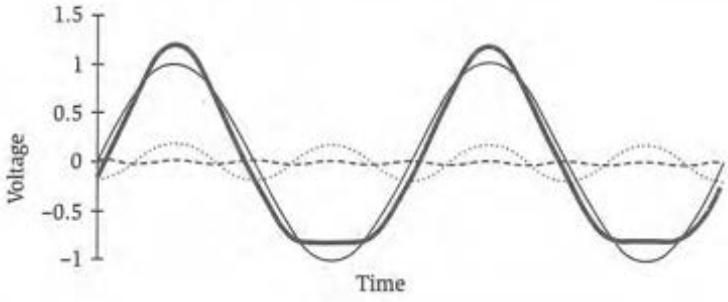


Figure 10.7 Even-order distortion.

As discussed in Chapter 9, the North American Standard channel plan (ANSI/EIA-542) places most analog television visual carriers approximately 6 MHz apart and offset approximately 1.25 MHz upwards from harmonics of 6 MHz.* Each such carrier can be expressed as

$$\sin[2\pi(6n + 1.25)t] \tag{10.23}$$

where the angle is expressed in radians and

n = the closest harmonic of 6 MHz
t = the time in microseconds

When two such signals (on different channels) are transported through an amplifier with second-order distortion (for simplicity, and because the effects are often small, the fourth-order term has been dropped), the output is

$$\begin{aligned} e_o = {} & AE_i(\sin[2\pi(6n + 1.25)t] + \sin[2\pi(6m + 1.25)t]) \\ & + BE_i^2(\sin[2\pi(6n + 1.25)t] + \sin[2\pi(6m + 1.25)t])^2 \end{aligned} \tag{10.24}$$

where the first term is the linear amplification of the input signals and

n and m = the multiples of 6 MHz closest to the two visual carriers
E_i = the peak voltage of either of the input signals (they are assumed to have equal amplitudes)

The last term has a dc component plus the following intermodulation products:

$$\begin{aligned} & -BE_i^2 \cos[2\pi(6[m + n] + 2.5)t] \\ & +BE_i^2 \cos[2\pi(6[m - n])t] \\ & -(BE_i^2/2) \cos[2\pi(12n + 2.5)t] \\ & -(BE_i^2/2) \cos[2\pi(12m + 2.5)t] \end{aligned}$$

The first product falls 1.25 MHz above the visual carrier nearest the $(m + n)$th harmonic of 6 MHz, whereas the second product falls 1.25 MHz below the visual carrier nearest the $(m - n)$th harmonic of 6 MHz.

* Variations from exact spacing are due to equipment frequency tolerances, broadcast station offsets of 10 or 20 kHz, and aeronautical band offsets of 12.5 or 25 kHz.

The third and fourth products will be half the voltage (−6 dB in power) and will each lie 1.25 MHz above another visual carrier (provided, in all cases, that the products fall within the occupied bandwidth of the distribution system). Note that these last products are identical in amplitude to those analyzed with a single carried signal and are, in fact, just the second harmonics of the two original signals.

The amplitude of each of these four distortion products, as with the single channel analysis, depends on the square of the input voltage so that the *ratio* of fundamental to second-order products varies linearly with input level. A similar analysis for fourth-order distortion, the next even power, will show low amplitude components at four times the original frequencies and additional components at twice the original frequencies.

When many similarly spaced carriers in the Standard channel plan are subjected to second-order distortion, each of the individual carriers and each of the carrier pairs will mix as analyzed above. This will result in clusters of products 1.25 MHz above and below each of the original carriers.* As a general rule, the number of beats falling closest to a specific carrier can be calculated using the expression

$$N_L = n - m - x + 1 \quad N_U = \text{INT}\left[\frac{x - 2m + 1}{2}\right] - \begin{cases} 0 & \text{if } x \leq n \\ x - n - 1 & \text{if } x > \text{n} \end{cases} \tag{10.25}$$

where INT indicates that the expression is to be rounded down to the nearest integer and

N_L = the number of lower $(A - B)$ beats
N_U = the number of upper $(A + B)$ beats
n = harmonic number of highest carrier
m = harmonic number of lowest carrier
x = harmonic number of carrier being evaluated

Since, in the ANSI/EIA-542 Standard channel plan, the carriers are not exactly spaced, the second-order products are not frequency coherent, but rather are clustered around the nominal product frequencies.

The usual measurement technique when evaluating this type of distortion is to use a filter bandwidth of 30 kHz, which is wide enough to contain all the

* Actually, there are also some products spaced ±750 kHz from the visual carriers because two of the channels (5 and 6) are offset from the comb in the Standard channel plan.

significant products in a group and to measure the total apparent power within the filter bandwidth. The power level of the upper composite beat cluster, relative to the sync peak power of the visual carrier, is known as composite second order (CSO) and is sometimes expressed in units of dBc (meaning decibels relative to the carrier power level, which is always a negative number). Some engineers prefer the notation C/CSO, the ratio of the carrier to beats in decibels (a positive number). This book will use this notation. The absolute values are the same in either case. Although the lower CSO cluster will generally have greater magnitude than the upper cluster on lower-frequency channels, it falls at the edge of the video channel and is often ignored since it has a minor effect on video quality compared with the in-band upper cluster.

Since analog video carrier modulation results in average signal levels that vary with program content, equipment is specified and evaluated when loaded with unmodulated carriers of a defined level in order to achieve repeatable measurements. When loaded with video carriers whose peak level is the same as the unmodulated levels used for the test, the average carrier power will be about 6 dB lower, resulting in an improvement of about 6 dB in C/CSO.

When the IRC channel plan is used, the visual carrier frequencies are derived from a common source so that all channels are phase coherent and offset from harmonics of 6 MHz by exactly the same amount. The result is that the second-order beats are also coherent.

When the HRC channel plan is used, the visual carrier frequencies are derived from a common source and *not* offset. Under those circumstances, the upper and lower second-order beats fall exactly on the visual carrier frequencies, and the difference frequency beats cannot be ignored.

The amplitudes of the individual products and the number in each cluster are a function of the nonlinearity of the amplifier, the levels of the carriers generating the beat, and the number of test signals used. Although the preceding analysis assumed a uniform carrier level, standard operating practice is to vary the carrier power in a controlled way as a function of frequency. This will result in varying individual beat levels.

When a comb of signals is subject to second-order distortion, the greatest number of products will fall on the highest and lowest channels. Figure 10.8, as an example, gives the total number of upper and lower 1.25-MHz second-order products in the Standard channel plan (in which channels 5 and 6 are offset by 2 MHz, and the 15th to 17th harmonics are not used).

The visual effect of the upper-side second-order products on analog NTSC television pictures is similar to that of a discrete interfering carrier — that is, closely spaced diagonal lines as discussed in Chapter 2. As discussed earlier, the lower CSO beats have relatively little effect on visual quality.

Many simple broadband RF amplifiers are single-ended class A devices, meaning that their output stages consist of a single active device that is biased to the approximate middle of its dynamic range. Such a device typically exhibits an asymmetrical compression behavior so that second-order distortion products

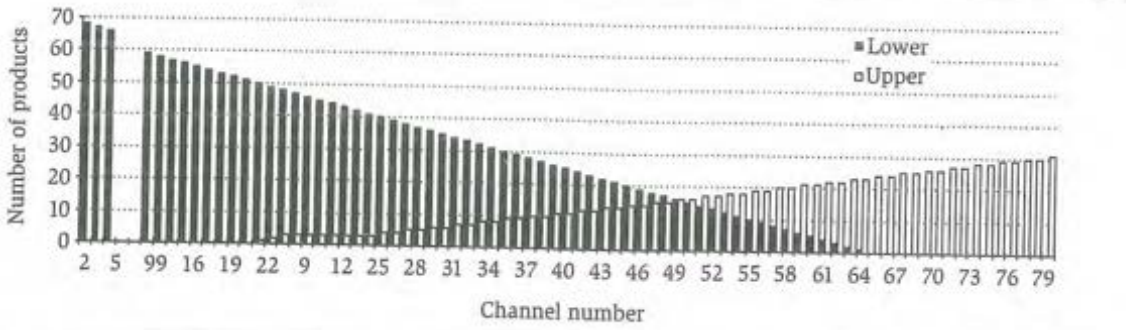


Figure 10.8 Number of 1.25-MHz CSO beats: 550-MHz Standard channel plan.

dominate. In recognition of this problem, the North American VHF off-air frequency plan designated channel boundaries such that neither second harmonic nor second-order beats from any combination of channels affect any other channels even if the drive levels to television receivers are high enough to result in compression. This was discussed in detail in previous chapters.

Early cable television amplifiers used single-ended designs for the same reasons. As the industry began using non–off-air channels, however, CSO quickly became the primary performance-limiting parameter.

Odd-Order Distortion (Composite Triple Beat—CTB)

In an attempt to reduce this problem, all modern cable amplifiers are "push–pull," meaning that their output stages contain two devices in a balanced circuit that assures symmetrical, or nearly symmetrical, compression. As will be seen, this will result in primarily odd-order distortion products.

If a single sine wave is transmitted through a circuit that has only third-order distortion, the output can be described by the following equation (derived from Equation 10.20):

$$e_o = Ae_i + Ce_i^3 \tag{10.26}$$

where A and C are the coefficients of the linear and cubed products, respectively, and

$$e_i = E_i \sin(\omega t)$$

Using a standard trigonometric identity, this reduces to

$$e_o = \left[AE_i + \frac{3CE_i^3}{4}\right][\sin(\omega t)] - \left[\frac{CE_i^3}{4}\right][\sin(3\omega t)] \tag{10.27}$$

C may be either positive or negative. In the case where the nonlinearity is due to compression, it is negative so that the third-order distortion results in a reduction in the amplitude of the fundamental (unlike second-order distortion that does not affect the amplitude of the fundamental) plus the addition of a product at three times the original frequency whose peaks are in opposition to those of the fundamental. The level of the third-order products rises as the cube of the input voltage (e.g., the rise is 3 dB for every 1-dB rise in input level) so that the *ratio* of fundamental to third harmonic at the output decreases by 2 dB for every 1-dB rise in the input level *so long as the ratio between them is high*. When the reduction in the amplitude of the fundamental becomes material, the ratio will decrease more quickly.

Figure 10.9 shows visually the fundamental, third harmonic, and the sum of the two. As can be seen, third-order distortion is equivalent to a reduced amplitude, symmetrically flattened waveform. If higher odd multiples of the original frequency are also plotted on the graph, it will be seen that all such harmonics will have exactly the same effect on both peaks of the original waveform.

If multiple signals are transmitted through a system exhibiting this type of distortion, the input can be described by the following general formula:

$$e_i = E_1 \sin[\omega_1 t] + E_2 \sin[\omega_2 t]) + \cdots + E_n \sin[\omega_n t] \tag{10.28}$$

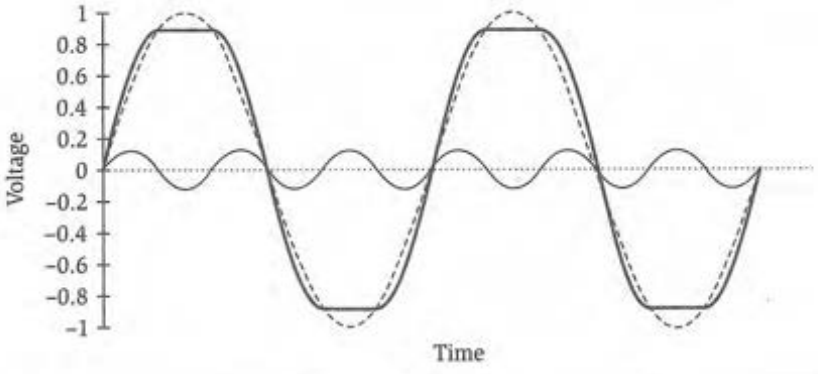


Figure 10.9 Effect of third-order distortion.

where

$\omega_{1,2,\ldots,n}$ = the angular frequencies of the input signals
$E_{1,2,\ldots,n}$ = the levels of each subcarrier (peak voltage)

If we insert this definition of e_i into Equation (10.26), it can be shown that the output spectrum will have signals at the fundamental frequencies of the input signals and all combinations of $3\omega_x$, $2\omega_x \pm \omega_y$, and $\omega_x \pm \omega_y \pm \omega_z$, where ω_x, ω_y, and ω_z are any of the input frequencies. In particular, single frequency terms, two frequency terms, and three frequency terms will be present:

- Single frequency terms:

$$\left[AE_x + \frac{3CE_x^3}{4}\right][\sin(\omega_x t)] - \left[\frac{CE_x^3}{4}\right][\sin(3\omega_x t)] \tag{10.29}$$

Just as with a single sine wave, this indicates that the output contains, in addition to the linearly amplified input signals, two terms proportional to the cube of the input voltage level of each of the carriers, one at the fundamental carrier frequency and one at three times the carrier frequency. When the input consists of carriers corresponding to the EIA Standard or IRC channel schemes (offset 1.25 MHz from harmonics of 6 MHz), the third harmonic term will fall 3.75 MHz above a harmonic of the root frequency (2.5 MHz above a carrier if one is present). With carriers that fall exactly on harmonics of a common root (HRC), the distortion products will also fall on harmonics of the root. The number of single frequency terms will be the same as the number of FDM carriers.

- Two frequency terms:

$$\left[\frac{3CE_x^2E_y}{2}\right]\sin(\omega_y t) - \left[\frac{3CE_x^2E_y}{4}\right]\sin([2\omega_x + \omega_y]t) + \left[\frac{3CE_x^2E_y}{4}\right]\sin([2\omega_x - \omega_y]t) \tag{10.30}$$

Each unique ordered pair of input signals mixes to produce a product at the fundamental of one of the input signals plus products at $2\omega_x \pm \omega_y$. The amplitude of the fundamental is twice that produced by the cubed term (assuming equal amplitude input signals). The voltage amplitudes of the other two terms are three times that of the cubed terms (+9.5 dB). When the input consists of carriers corresponding to the EIA Standard or IRC channel schemes, the $2\omega_x + \omega_y$ term will fall 2.5 MHz above a carrier assignment, while the $2\omega_x - \omega_y$ term will fall on a carrier assignment. With an HRC

channel assignment scheme, all products will fall directly on channel frequencies.

When the input signals consist of an FDM spectrum of q input carriers, the number of ordered pairs is $q(q-1)$. Since each ordered pair results in the generation of products at two new frequencies, the number of new products will be $2q(q-1)$. Thus, when the number of carriers is large, the total number of products will increase approximately as the square of the channel count.

- Three frequency terms:

$$\left[\frac{3E_xE_yE_z}{2}\right]\left[\sin([\omega_x+\omega_y-\omega_z]t)+\sin([\omega_x-\omega_y+\omega_z]t)\right.$$
$$\left.-\sin([\omega_x+\omega_y+\omega_z]t)-\sin([\omega_x-\omega_y-\omega_z]t)\right] \tag{10.31}$$

Each unique combination of three input signals mixes to produce terms at $\omega_x \pm \omega_y \pm \omega_z$. When the input signal amplitudes are equal, the voltage amplitudes of these terms are six times that of the cubed terms (+15.6 dB). When the input consists of carriers corresponding to the EIA Standard or IRC channel schemes, the $\omega_x+\omega_y-\omega_z$ and $\omega_x-\omega_y+\omega_z$ terms will fall on a carrier frequency, the $\omega_x+\omega_y+\omega_z$ term will fall 2.5 MHz above a carrier frequency, and the $\omega_x-\omega_y-\omega_z$ term will fall 2.5 MHz below a carrier frequency. Note that, though $\omega_x \pm \omega_y \pm \omega_z$ is sometimes negative, the only effect is the polarity of the magnitude since $\sin(-x) = -\sin(x)$.

When the input signals consist of an FDM spectrum of q input carriers, the number of unique three-signal combinations is $q(q-1)(q-2)/6$. Since each of these combinations results in the generation of four new products, the total number of new third-order products caused by three-signal mixing will be $4q(q-1)(q-2)/6$. Thus, when the number of carriers is large, the total number of products will increase approximately as the cube of the channel count.

In summary, when an FDM spectrum of q signals is transmitted through an amplifier exhibiting third-order distortion, the original signals will be amplified, and a number of new products will be generated:

- Some that affect the output amplitude of each original carrier.
- q new signals, one at the third harmonic of each of the original signals. With the ANSI/EIA-542 Standard or IRC frequency plan, these fall 2.5 MHz above visual carrier frequencies.
- $2q(q-1)$ new signals at frequencies $2\omega_x \pm \omega_y$, where ω_x and ω_y are any two input carrier frequencies. Each of these signals is about 9.5 dB higher in level than the third harmonic frequencies. With the Standard or IRC frequency

plan, half of these fall 2.5 MHz above, and half nominally on, visual carrier frequencies.

- $4q(q-1)(q-2)/6$ new signals at frequencies $\omega_x \pm \omega_y \pm \omega_z$, where ω_x, ω_y, and ω_z are any three different carrier frequencies. Each of these products will be about 15.6 dB higher in level than the third harmonic frequencies. With the Standard or IRC frequency plan, half of these will fall nominally on visual carrier frequencies, whereas one-quarter will be 2.5 MHz below and one-quarter will be 2.5 MHz above visual carrier frequencies.

The level of each of these products increases by 3 dB for every 1-dB increase in the levels of the input signals so that the ratio between the desired output signals and third-order products decreases by 2 dB.

Since the three-signal products ("triple beats") are the greatest in both number and magnitude, they quickly dominate other third-order effects. Since half of them fall nominally on visual carrier frequencies, those product groups are always the largest. Thus, composite triple beat (CTB) distortion is defined as the total power in a 30-kHz bandwidth centered on the visual carrier frequency without the visual signal present, relative to the peak level of the visual carrier in dBc (decibels relative to a reference carrier level). As with CSO, it can also be expressed as the ratio of visual carrier level to apparent beat power, or C/CTB, in decibels.

For a contiguous set of equally spaced carriers, the number of third-order beats falling in any given channel is approximately[8]

$$N_{CTB} = \frac{(a-1)^2}{4} + \frac{(a-b)(b-1)}{2} - \frac{a}{2} \tag{10.32}$$

where

N_{CTB} = the number of third-order beat products whose frequencies are nominally the same as the carrier being evaluated

a = the total number of contiguous, equally spaced carriers

b = the index number of the carrier being evaluated, i.e., the bth carrier from the bottom or top

For the center channel, which receives the greatest number of beats, this reduces to

$$N_{CTB} = \frac{3a^2 - 10a + 2}{8} \approx \frac{3a^2}{8} \qquad \text{for } n \gg 1 \tag{10.33}$$

At the end channels, which receive the fewest number of beats, it becomes

$$N_{CTB} = \frac{a^2 - 4a + 1}{4} \approx \frac{a^2}{4} \qquad \text{for } n \gg 1 \tag{10.34}$$

Actual cable systems using the EIA Standard channel plan, however, do not conform exactly to this model, with channels 4 and 5 being offset and some or all of the channels falling in the FM broadcast plan seldom used for video. Thus, "beat quantity tables" published for the industry show several percent fewer beats.

Figure 10.10 is a plot of the number of beats for some common, fully loaded, system bandwidths.[9] As can be seen, channels slightly above the middle of the spectrum are affected by the greatest number of beats. It does not follow, however, that the channel with the greatest number of beats is the channel most affected by third-order interference. Amplifiers are not typically operated with equal power carriers, but rather with a uniform upward "slope" of levels with frequency (for reasons explained later in this chapter), causing some beats to be stronger than others. More important, since signal quality is affected by the *ratio* between visual carrier level and interference level, the lower carrier levels at the low end of the spectrum will sometimes cause the worst interference to occur on a lower-frequency channel than the one with the largest number of beats.

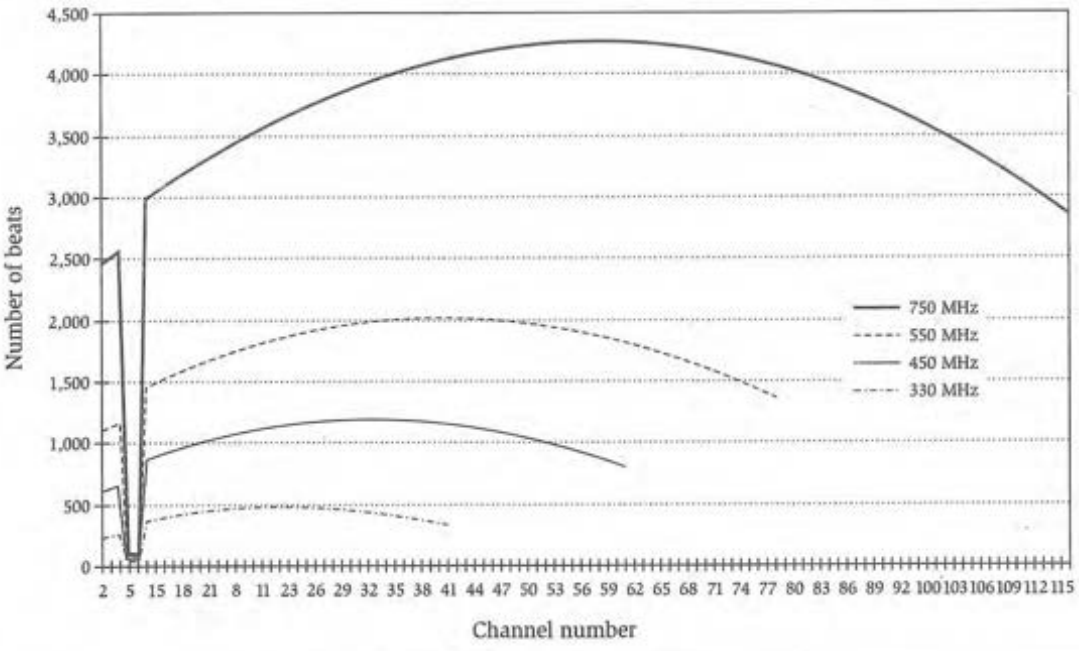


Figure 10.10 Number of triple beats per channel versus system bandwidth.

Since the third-order products are phase incoherent in most systems, their powers will add on a power, rather than a voltage, basis, and the measured CTB level would theoretically be proportional to 10 log(n), where n is the number of beats in a channel if the levels of all carriers are equal. In actuality, owing to the unequal carrier levels and beat amplitudes, and possibly differences in distortion across the spectrum, CTB may increase as slowly as 5 log(n) to 7 log(n).[10]

The visual effect of CTB on an analog television picture depends on whether the signals are harmonically related or not. With unlocked carriers, the differences in frequency between the desired carrier and the various beats produces low-frequency interference that has been described as "streakiness" in the picture. With locked carriers, the various products will be frequency coherent with the carrier and, like the fundamental frequency products, will alter the amplitude and phase of the original signals. The visual effects depend on many factors, including the amount of phase noise on each of the signals.

Since analog video carrier modulation results in an average signal level that varies with program content, equipment is specified and evaluated when loaded with unmodulated carriers of a defined level on each of the nominal visual carrier frequencies. This results in more repeatable measurements. When loaded with video carriers whose peak level is the same as the unmodulated levels used for the test, the average carrier power will be about 6 dB lower, resulting in an improvement of about 12 dB in CTB (though the improvement will vary over time depending on video modulation levels and timing and carrier phase relationships). Typical amplifiers provide a CTB of −70 to −90 dBc when loaded with unmodulated carriers at recommended operating levels, depending on design.

Cross Modulation (XMOD)

Another form of distortion occurs when amplifiers exhibiting third-order distortion are loaded with multiple amplitude modulated signals.

Although second-order distortion has no products that affect the gain of the primary FDM carriers, as Equations (10.27), (10.29), and (10.30) demonstrate, two classes of third-order products affecting every channel cause effective gain variations of the fundamental signal. In particular, the first term of Equation (10.30) demonstrates that the effective amplifier gain affecting each FDM subcarrier is modified by a series of terms that include the amplitude of every other signal in the spectrum carried. Thus, when the average levels of any other signals are varied because of modulation, the output level of the desired channel will also vary slightly. This effect is known as cross modulation.

If the modulation of the other signals were truly independent of each other and random, the effects of cross modulation from various channels could be expected to be noiselike. Unfortunately, video modulation is not random. As covered in Chapter 2, the change in amplitude from synchronizing peak to black level is identical for every channel and, furthermore, occurs at nearly

identical and quite precise frequencies. Thus, if the signals modulating multiple channels happen to be synchronized (which happens when they are locked to a common timing reference in a multichannel generation facility), or if they drift into alignment occasionally, the level of cross modulation can become significant.

The standard test for cross modulation is to load the amplifier with a comb of simultaneously modulated carriers plus one unmodulated carrier on the frequency to be tested. The level of modulation is measured on the nominally unmodulated carrier. Cross modulation is defined as the difference between the cross modulation sideband level and the sideband level that would correspond to 100% modulation expressed in decibels.

Since it is caused by third-order distortion, the level of cross modulation varies in the same way as CTB; that is, the relative level varies 2 dB for every 1-dB change in operating levels. Because of the difficulty in making the test and demonstrated inconsistency in the results, cross modulation has become less important than CTB as a means of characterizing third-order distortion. The inconsistency may be due to the many terms, whose effects may reinforce or cancel, affecting the cross modulation on any channel.

When the desired signal is an analog television program and the major cross-modulating signals are also analog television programs, the first visible indication of cross modulation takes the form of a single, faint broad horizontal line and another broad vertical line that may slowly move through the desired picture. These are caused by the horizontal and vertical synchronizing pulses (the highest carrier levels) of another television signal. In an improperly operated system, cross modulation can also be caused by a single very high amplitude channel that causes cross modulation onto other channels. In extreme cases, not only synchronizing bars but video modulation may be visible. The visual appearance of cross modulation is similar to that which results from direct pickup interference where an undesired signal "leaks" into the system and interferes with a cable signal on some channel.

Composite Intermodulation Noise (CIN)

The preceding analysis was based on the carriage of a spectrum of signals, each of which has a carrier that represents most of the energy in the channel. Though that is a valid model for analog television signals and many conventional AM and narrow band FM modulated signals, digital signals differ significantly. A QAM signal, for instance, has a suppressed carrier and, when viewed on a spectrum analyzer, looks like a flat block of noise occupying the entire communications channel (see Chapters 4, 5, and 6 for detailed discussions of QAM modulation).

When such signals are subject to second- and third-order distortion, they do not produce single frequency products, but rather noiselike bands of

energy whose amplitude frequency signature depends on the distortion mechanism. The products of this type of distortion are alternatively known as composite intermodulation noise (CIN), composite intermodulation distortion (CID), or intermodulation noise (IMN). This is an exception to our use of the term *noise* to refer to only thermal noise since CIN is actually a distortion product.

On the assumption that third order is the dominate distortion-producing mechanism and that, further, three frequency products ($\omega_a \pm \omega_b \pm \omega_c$) dominate, IM products involving digital carriers will be of three types:

- Products formed from the mixture of two analog carriers and one digital signal will appear as blocks of flat noise with a bandwidth equal to that of the digital signal.
- Products formed from the mixture of one analog signal and two digital signals will appear to have a symmetrical triangular-shaped amplitude-versus-frequency spectrum, where the triangle will have a spectral width equal to the sum of the widths of the two individual digital signals.
- Products formed from the mixture of three digital signals will appear to have a band-limited Gaussian shape with a spectral width equal to the sum of the widths of the three mixing signals.

If digital signals are carried lower in level than analog signals, then the preceding types of products are listed in decreasing order of amplitude.

Although standard operating practices have not yet been set for the cable industry, it has been reported that a 750-MHz plant, operating with a signal load of 77 modulated analog television signals (54–550 MHz) and 33 64-QAM digital signals (550–750 MHz) and typical operating conditions and levels, will suffer less than 1 dB of effective C/N degradation due to CIM with digital signal levels suppressed as little as 5 dB relative to analog signals.[11]

10.3.4 Group Delay Variation and Amplifier Stability

Group delay is a measure of the rate of change of phase shift through a device as a function of frequency. In equation form

$$\text{group delay (sec)} = \frac{1}{2\pi}\frac{d\gamma}{df} \tag{10.35}$$

where

γ = the phase shift through the device in radians
f = the frequency in Hz

In an ideal RF system, the transmission delay at all frequencies is a constant equal to the transit time through the length of line involved, and the group delay is zero.

If the group delay is not zero, signal components at one frequency will arrive shifted in time relative to other components. The degree of tolerance to group delay variation will vary according to the signal format. A particular case of interest to the analog NTSC format is chroma delay, a measure of the arrival time of the chroma signal relative to the luminance signal. As discussed in earlier chapters, the difference should not exceed 170 ns. Though modulators and demodulators are generally the primary contributors to chroma delay, channels close to forward or reverse transmission system band edges are also affected by amplifier design.

High-speed digital signals are also affected by group delay. For example, fast rise time waveforms will broaden if the higher-frequency components are delayed relative to the lower-frequency components. Occasionally, digital signals are carried below channel 2 in the downstream band and, thus, very close to passband edge.

The principal contributors to group delay variation in amplifiers are the diplex filters used at each of the signal ports to separate upstream and downstream spectra. Figure 10.11 illustrates the key components of a two-way amplifier. The input and output diplex filters each consist of a high-pass section connecting the forward gain block to the external ports and a low-pass section feeding the reverse gain block. Each of the filter sections, as well as the amplifier modules, will have a gain (negative in the case of the filters) that is a function of frequency.

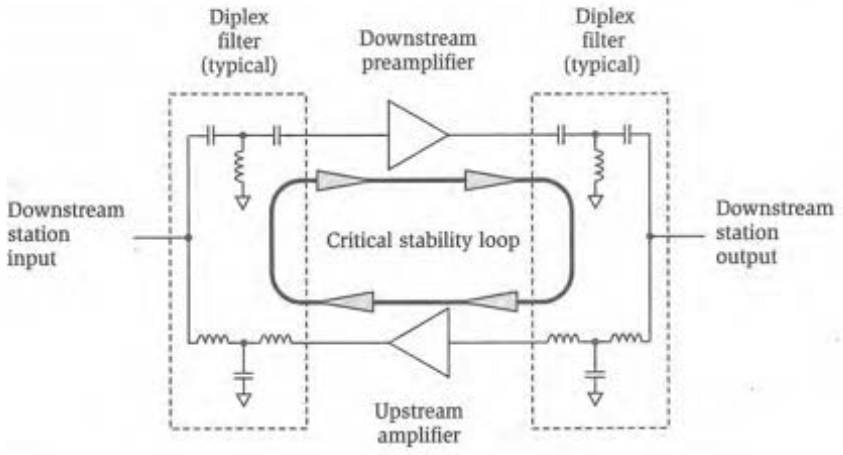


Figure 10.11 Bidirectional amplifier.

In order to assure absolute stability, the sum of the gains around the loop shown in the heavy arrow must be less than zero at all frequencies. This condition must be met not only for matched loads on the external ports but for any mismatch encountered in an actual field installation. With simple filter designs, this is most easily achieved by having the cutoff frequencies of the filters as far apart as possible and, thus, close to the edge of well-separated forward and reverse passband edges. Since group delay variation gets worse as the cutoff frequency is approached, there is an inherent conflict among (1) the separation between the top of the upstream spectrum and the bottom of the downstream spectrum, (2) the cost of the diplex filter, (3) the group delay variation at the extremes of the spectra, and (4) the stability margin of the amplifier.

Ensuring a nonoscillatory condition may not be adequate. Even if the loop gain is less than zero, frequency response variations may result unless there is adequate isolation. As a practical test, adequate performance may be deduced by observing the swept frequency response of an amplifier at all frequencies, including the diplexer crossover region. Peaking of the response indicates a possibility of a loop gain problem.

Figure 10.12 illustrates the downstream performance of a typical commercially available amplifier having a "42/52 MHz split" (upper edge of the upstream spectrum is at 42 MHz, whereas the lower edge of the downstream

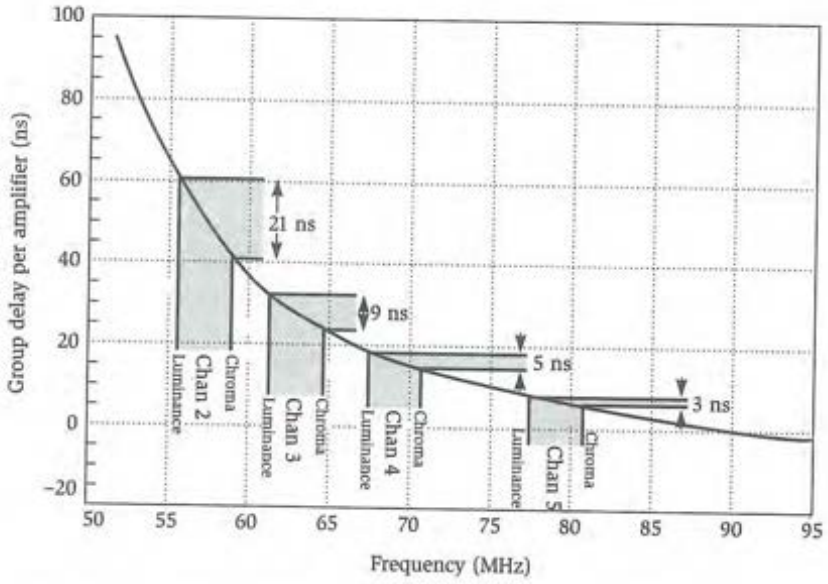


Figure 10.12 Downstream group delay due to diplex filters.

spectrum is at 52 MHz). The curve illustrates the group delay, with the shaded regions indicating the frequencies and chroma/luma delay difference for the four lowest NTSC television channels in the Standard ANSI/EIA-542 frequency plan. The direction of the chroma delay is to advance the chroma with respect to the luminance. Serendipitously, this is the opposite delay sense (luminance is delayed with respect to chrominance) to that which is commonly encountered in the modulation–demodulation process, but any delay is not good.

The required end-to-end group delay requirements and typical network performance are discussed in Chapter 15, and upstream group delay is discussed in Chapter 16.

10.3.5 Hum Modulation

Although amplifiers that plug into conventional wall outlets are available (and used in headend and apartment house applications where convenient), most distribution amplifiers are powered via ac voltage that is multiplexed with the signals on the common coaxial distribution cable. Along with their signal-processing circuits, amplifier stations include circuitry for separating the power and signal voltages and for connecting the power to local power circuits that convert it to direct current to power the various electronic modules used.

Amplitude modulation of the transmitted signals at the power line frequency (hum) can occur in two ways: excessive ac voltage at the output of an amplifier's local power pack and parametric modulation of magnetic component properties. Power pack "ripple" can modulate amplifier gain or be coupled into other circuits such as the automatic gain control (AGC). This ripple can result from faults in the power pack or loss of voltage regulation due to excessive voltage drop between the power supply and the amplifier's power pack.*

Parametric magnetic modulation is less obvious. Consider the circuit segment in Figure 10.13. The current is tapped off the common coaxial input and sent through L to the amplifier power pack (and possibly inserted into the output coaxial cable(s) as well). Series capacitor C serves to pass the signals into the amplifier while blocking the supply voltage. Finally, T is the input winding of a typical signal-processing element, such as a directional tap, splitter, or part of a band-splitting filter.

One possible source of hum modulation is L. This component is required to exhibit a very high impedance to frequencies covering the entire forward and return spectra. If the magnetic core becomes partially saturated at the peaks of the supply current, then the impedance match of the amplifier will vary

* In accordance with standard industry terminology, the term *power supply* is used to designate the device that generates the system power for multiplexing on the coaxial cable, whereas the term *power pack* is used to designate the device within each amplifier that converts the multiplexed power to the appropriate regulated dc levels for the circuitry.

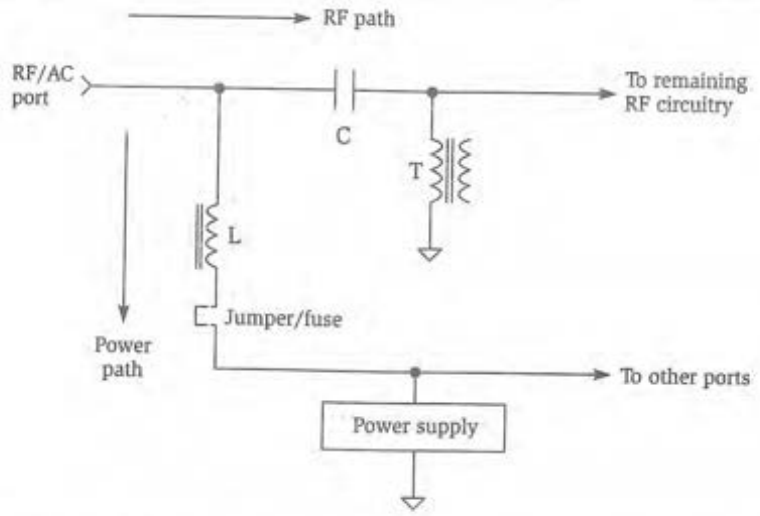


Figure 10.13 Power/RF separation in amplifiers.

accordingly, resulting in a level variation of the transmitted signal at the line frequency rate.

T is a much more sensitive component. Typically, these transformers are constructed of tiny cores that must have constant magnetic characteristics over an extremely wide frequency range. Though C nominally protects T from the supply current, there is still a displacement current to contend with. In particular, an ac current will flow through C and the primary winding of T whose magnitude is

$$i_C = C\frac{dV}{dt} \tag{10.36}$$

When the power supply is a sine wave with a frequency of f_p and a peak voltage of E_p, the current will be

$$i_C = C\frac{d}{dt}[E_p \sin(2\pi f_p t)] = CE_p 2\pi f_p \cos(2\pi f_p t) \tag{10.37}$$

which has a peak value of $CE_p 2\pi f_p$. If the voltage is a square wave with fast rise and fall times, then the displacement current can be much greater for the same peak supply voltage. To illustrate the trade-offs in design of these circuits, assume that it is desired that C have no more than 1 ohm of reactance at the low end of the upstream spectrum (generally 5 MHz) so as to affect the input match

as little as possible. This requires that C be at least 0.03 μF [$X_C = 1/(2\pi fC)$]. If the supply voltage is 90-volts rms (127 volts peak) at 60 Hz, then the displacement current will be 1.4 mA. If, under square wave powering conditions, the rise time is five times that under sine wave conditions, then the current could be as high as 7 mA. The designer's challenge is to ensure that this level of additional current in the winding of T does not change the transformer's characteristics.

In the North American NTSC analog television system, the frame repetition rate is 59.94 Hz. Since the commercial power frequency is 60 Hz, hum modulation generally appears as a horizontal bar or brightness variation whose pattern slowly moves upwards with time (about 16 seconds to move the entire height of the picture). Typical coaxial components (both amplifiers and passive elements) are specified to create hum modulation at no greater than −70 dB, relative to 100% modulation, at their full rated supply voltage and pass-through current.

10.3.6 Frequency Response Variation and Impedance Match

Practical amplifiers do not amplify signals at all frequencies equally. Part of this variation is intentional and is intended to compensate for variations in cable losses and to optimize distortion performance, as we discuss in the next section. Any deviation from the intentional variation, however, is known as frequency response variation (quantified as ±x dB) or, more commonly, peak-to-valley (quantified as y dB peak-to-peak). Response variation is defined by the SCTE as the maximum gain deviation in either direction from a line drawn from the gain at the low end of the defined bandwidth to the gain at the high end of the defined bandwidth.[12] Figure 10.14 illustrates this definition; the peak-to-valley (P/V) is the sum of the absolute values of +Δ max and −Δ max. This definition, however, is

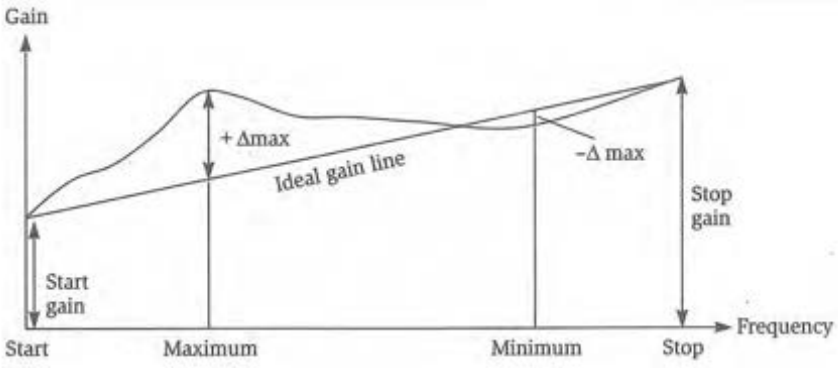


Figure 10.14 Definition of response variation.

somewhat controversial since it is highly affected by the endpoints. An alternative definition for frequency response is the variation above and below a line that is drawn so as to minimize P/V. For instance, if the lower end of the "ideal gain line" in the figure were raised slightly, the effect would be to lower $+\Delta$ max by more than the increase in $-\Delta$ max and thus reduce the total P/V.

The frequency response through a network may be affected by the imperfect impedance matching between the amplifier and attached transmission line. The effects of imperfect matches in coaxial components can best be understood in the context of their interaction. This is covered in detail in Chapter 15.

10.3.7 Practical Amplifier Design Choices

Complete amplifier stations may be simple or quite elaborate, depending on system needs. All share a common basic downstream structure consisting of several (generally two or three) cascaded gain stages, plus circuits that compensate for the loss variation of the cable before the amplifier (equalizers). Provisions are also made for adjusting the gain of the amplifier as needed to meet system needs. Amplifiers differ in how many stages are employed, internal splitting, redundancy provisions, where operating adjustments are in the chain, automatic gain control, and the performance of each gain block. Many have provisions for automatically adjusting amplifier gain to compensate for changing input signal levels (automatic gain control).

As will be seen, optimum operation of repeatered coaxial distribution systems results when the levels of the FDM subcarriers at the output of amplifiers are not equal but increase with frequency. The variation in level across the operating range is known as *tilt*. The relationship between operating levels, amplifier response, and interconnecting transmission losses is explored more fully in Section 10.3.8.

In the upstream direction, all the amplifier designs are similar. Since cable losses are much lower in that portion of the spectrum, a single amplifier module is sufficient. Diplexers are used at each port to separate the upstream and downstream signals. All the upstream signals are passively combined and fed to the amplifier module. Unlike downstream, where all level setting takes place at the input or midstage with the aim of setting a consistent *output* level, the common upstream practice is to align the system so that the *input* level to the module is constant and to set the output gain and slope after the gain stage so that the input to the next amplifier station upstream is correct. Chapter 16 covers the alignment and operation of networks in the upstream direction.

To allow the most flexibility in application and system design, jumpers or switches allow power to enter the amplifier via any desired port and to be routed to any combination of output ports.

Early power packs used conventional transformer/linear regulator designs. Transformer taps were available to optimize the winding ratio of the transformer to limit required regulator dissipation. Modern designs almost universally use switching regulators (whose design is beyond the scope of this book) that have the property of operating at a constant high efficiency regardless of input voltage. To do so, they draw a constant *wattage* from the coaxial cable, and thus the current is inversely proportional to the supply voltage. Though this results in the greatest overall powering efficiency, it has some unexpected consequences when a number of amplifiers are powered from a common source, as will be seen in Chapter 11.

Beyond their basic functional elements, amplifier stations may be equipped with a variety of enhancements, including redundant gain modules and power supplies, status monitors and transducers, and remotely controllable upstream path disconnect. Often fiber-optic receivers and transmitters are combined with amplifier station components, creating a fiber-optic node.

Although the labels attached to different configurations are somewhat arbitrary and subject to change, the following classifications are typical:

- Trunk amplifiers are the workhorses of traditional all-coaxial cable television systems. They are optimized for long-distance, repeatered signal distribution. As such, they are operated at output levels that cause only moderate distortion to allow long cascades. The gain is set as high as possible, consistent with low per-stage noise addition. Typically, the output levels of the highest channels are +35 to +40 dBmV, and the gain is about 22 dB. Trunk amplifiers are not used to supply signals to taps.
- Where a local distribution circuit is also desired, a portion of the trunk signal is fed to a separate output stage known as a bridger. This amplifier generally runs at a higher output level (and distortion) to allow feeding as many customer taps as possible. Though generally configured as an optional module within the trunk station, bridgers may be contained in separate housings along with directional couplers (midspan bridgers). Figure 10.15 is a block diagram of the RF portion of a typical trunk station, including bridger.
- Distribution amplifiers are basically stripped-down trunk/bridger stations that are available in a wide variety of configurations. Typically, they are available with multiple output amplifier modules whose level can be independently adjusted. Figure 10.16 shows a typical configuration with three output gain blocks.
- Line extenders are used to allow feeding more taps than possible with just a bridger or high-level distribution amplifier port. Since they are generally operated at similar output levels, the cascade of such devices must be limited to keep distortions at an acceptable level. Line extenders are the simplest of the classifications, as shown in Figure 10.17.

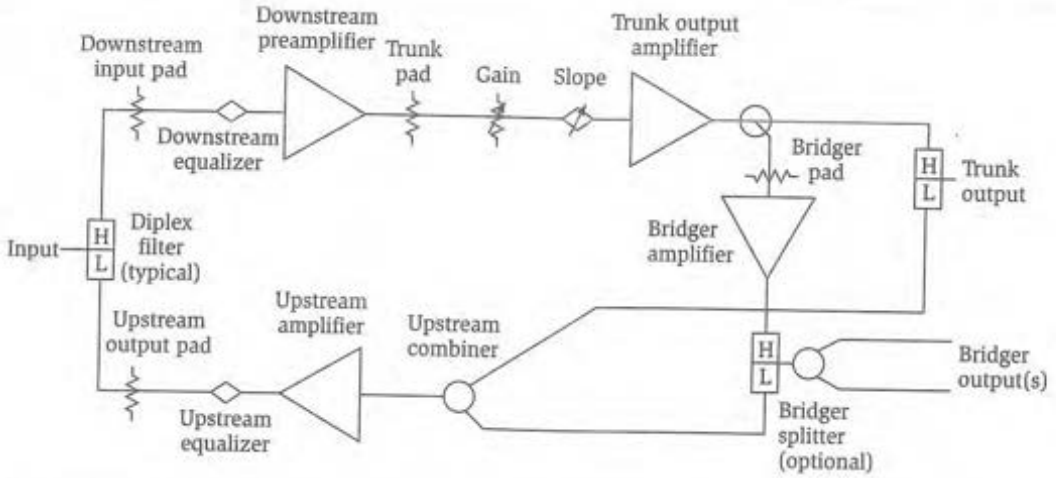


Figure 10.15 Typical trunk amplifier RF configuration.

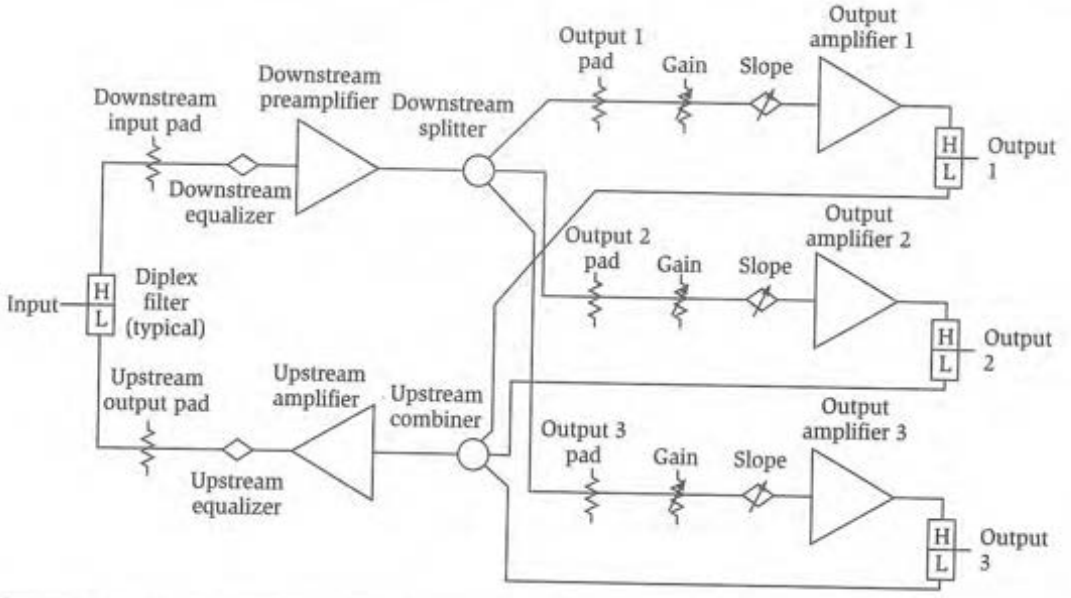


Figure 10.16 Typical distribution amplifier RF configuration.

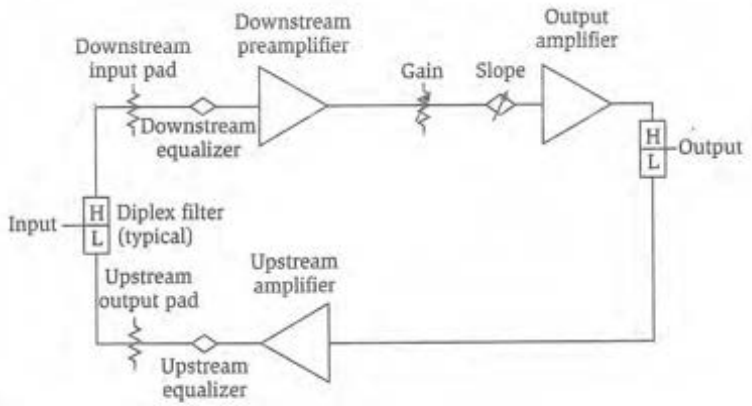


Figure 10.17 Typical line extender RF configuration.

10.3.8 Amplifier Operating Dynamics

The C/N degradation as a signal passes through an amplifier is related to the input signal level and the noise figure at any frequency (see Equation (10.18)). On the other hand, the loss of coaxial cable increases with frequency. Thus, the amplifier need not supply the same level for channels at the bottom of the spectrum as at the top to overcome cable loss. Reducing the levels of the lower-frequency channels, in turn, reduces the amplitudes of the intermodulation products to which those channels contribute, and thus CSO, CTB, CIN, and XMOD.

The relation between cable and amplifier characteristics and operating levels can best be illustrated by considering an amplifier and its driven cable as a system (see Figure 10.18). Only the downstream path will be discussed here; Chapter 16 will deal with the equivalent upstream issues.

The preamplifier will typically have a relatively uniform noise figure and gain across the spectrum. Between the preamplifier and output stage, a slope circuit is used to adjust the frequency response so that it uniformly increases across the downstream frequency range. Typical gain slopes are 6–10 dB for a 750-MHz amplifier. Additionally, pads are used to adjust the overall amplifier gain. The station noise figure and slope are illustrated in Figure 10.19(a).

Some combination of coaxial cable and passive RF devices is connected between this amplifier station and the input of the next station. Since both the cable and passive devices have a loss that varies across the spectrum (and, in the case of cable, increases as the square root of frequency), the variation in level reaching the next station will be reduced. Amplifier input circuits are designed to accommodate a variety of equalizers that are optimized to

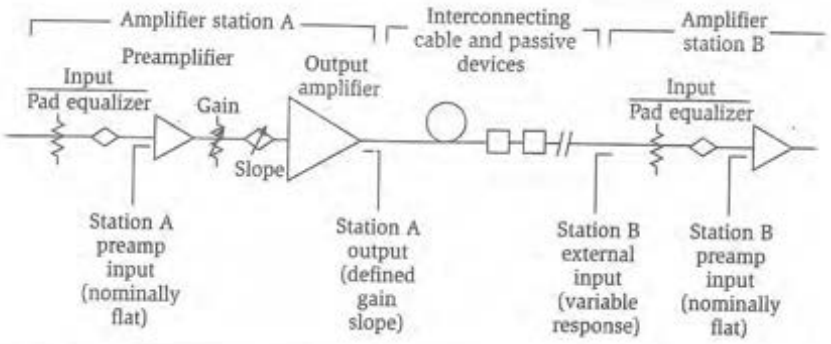


Figure 10.18 Downstream response and gain relationships and adjustments.

compensate for the residual response variation and pads that are used to set the correct drive level to the preamp input. Occasionally, external in-line equalizers may be used also, for example, as a tool to reduce the total variation of levels fed to customers.

In station alignment, the equalizers and pads are used to align the *output* of each amplifier station as closely as possible to the optimum levels and "gain slope." Thus, the system gain and frequency response, as measured from the output of one amplifier to the output of any other similarly aligned amplifier in a series-connected string, are ideally unity. Because of the variability in interstation losses, the levels and response at station inputs will vary; however, the response at the internal preamp input will be approximately flat. Amplifiers that do not feed customer taps (for example, trunk amplifiers) are typically operated at levels several decibels lower than those that do.

By contrast to the frequency response of the system (slope), the variation in levels of the signals carried is known as tilt. The terms are often confused, particularly since modern amplifier performance is often specified with a tilt at amplifier outputs that matches the gain slope. Historically (with narrower bandwidth amplifiers), it was more common to use "block tilt" with groups of channels operated at different levels. This was easier for technicians to understand since only a few levels needed to be memorized.

Modern 750-MHz amplifiers are generally specified under two signal-loading conditions. Figure 10.19(b) shows the nominal preamplifier input (after equalization) and output levels of an amplifier station loaded with a full spectrum of 110 video signals that are tilted to match the gain slope of the amplifier. This is the optimum condition for carrying signals of similar format and results in approximately equal C/N degradation for each channel (because of the uniform preamp input levels) and the lowest distortion (because the outputs at the lower channels are reduced in level to approximately match the interconnecting cable

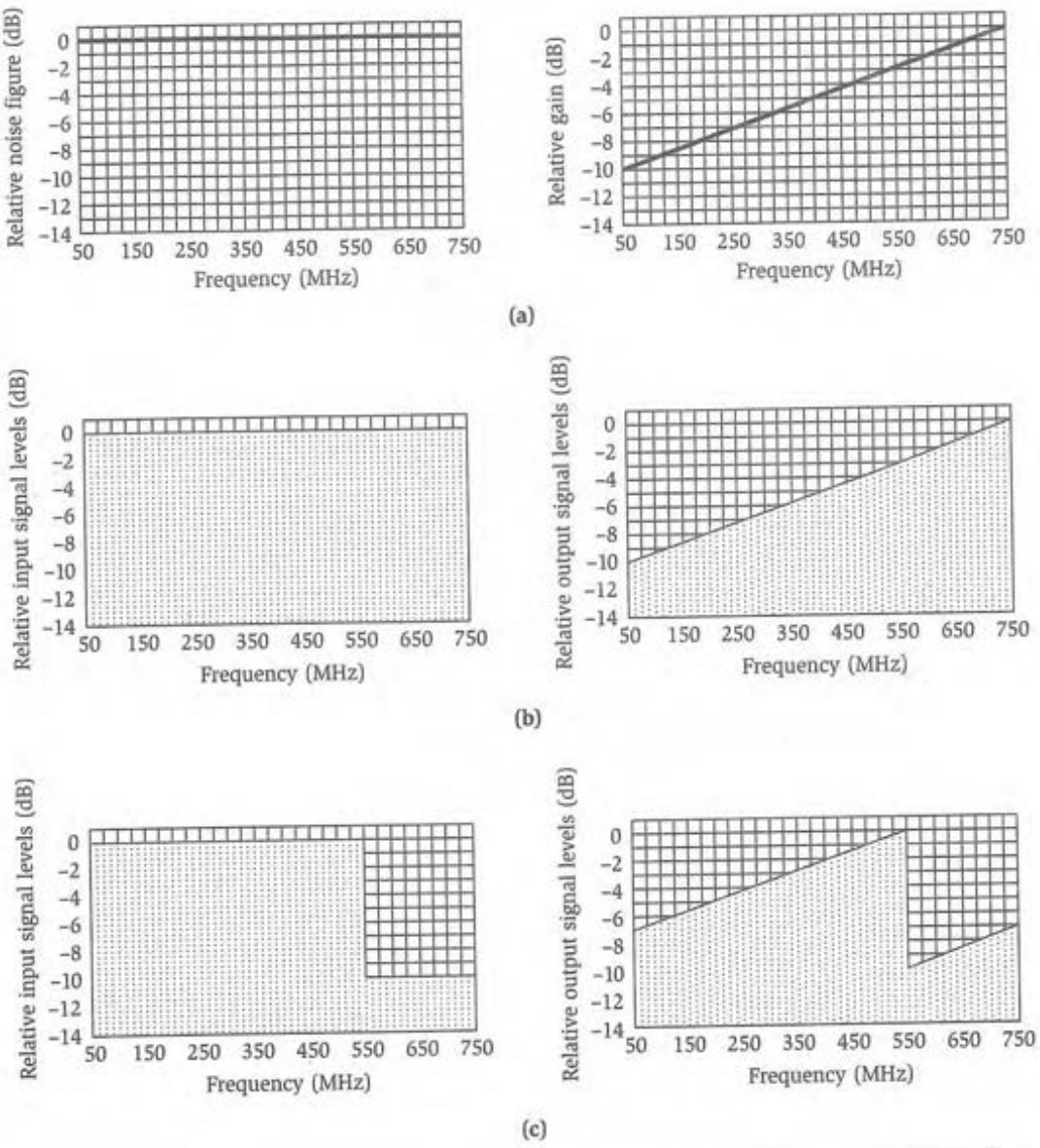


Figure 10.19 Amplifier gain and signal loading. (a) Amplifier characteristics. (b) Amplifier loaded with analog video signals. (c) Amplifier with hybrid analog–digital signal loading.

and component losses). Figure 10.19(c) shows an alternative loading condition with a superimposed −10-dB block tilt for signals above 550 MHz. The assumption is made that these signals are digitally modulated and do not require the same C/N and C/distortion as analog video channels and so are reduced in level to reduce overall amplifier loading.

Optimum amplifier gain is a function of noise figure and output capability (for a given distortion level). By way of illustration, assume that a given amplifier has an effective noise figure of 9 dB and a CTB of −78 dB at an output level of +45/35 dBmV (the dual numbers are the individual channel levels at the top and bottom of the downstream spectrum, respectively). If the gain of the unit is set at 35 dB (at the top channel frequency) and the input level is +10 dBmV on all channels, then, from Equation (10.19), the C/N of this stage alone will be 60 dB (10 + 59 − 9). The output level will be the input level plus the gain or +10 dBmV +35 dB = +45 dBmV, so the CTB will be the specified −78 dBmV.

If the input level is varied, the C/N will improve by 1 dB for every decibel increase in input level, whereas the CTB will degrade by 2 dB. In setting up a coaxial system, input padding is used to set the optimum operating point.

Now consider the situation if the amplifier had been designed with a gain of 40 dB. At first glance, this would seem like an improvement since such amplifiers could be spaced farther apart along the transmission line. When driven by a +10 dBmV input signal, however, the output level will be +10 dBmV+40 dB = +50 dBmV. At that level, the CTB will be only −68 dB, having degraded by 10 dB owing to the 5-dB increase in output level. If the input level is decreased by 5 dB to restore the distortions to the specified level, the station's C/N will decrease from 60 to 55 dB. In short, the excess gain forces the user to accept either higher distortion or higher noise.

Though line extender amplifiers often offer only an input padding option, some distribution and trunk amplifier designs also allow between-stage padding. This partially offsets potential excess gain problems by allowing the user to independently choose the operating level of each gain block though each gain block is still limited in performance by its individual noise figure, gain, and output capability.

The selection of optimum amplifier internal gains is very complex. It would seem from the preceding example that reducing gain would improve performance. Although that is true, more lower-gain amplifiers would have to be cascaded to reach the same distance, and that would lower reliability and increase power consumption. The industry's choice of operational gains ranging from 20 to 38 dB (depending on technology and application) is based on the best compromise among cost, reliability, and performance.

10.3.9 Amplifier Technology Choices

As previously discussed, modern broadband solid-state cable television amplifiers are all push–pull in nature, meaning that their output stages contain two

matched transistors connected in a symmetrical configuration. To the extent that any nonlinearities are also symmetrical, they will produce only odd-order distortion products. Amplifiers whose output stages each contain only a single push–pull circuit are known generally as *push–pull amplifiers*.

An improvement in distortion performance can be realized by splitting the input signal, feeding two push–pull output gain stages in parallel, and then combining the outputs into a single port. This type of amplifier, generically known as *parallel hybrid*, offers about twice the output level at the same distortion levels compared with push–pull units. Since two output stages are involved, the power consumption and dissipation are also higher than with push–pull units. Parallel hybrid amplifiers are generally used where high output powers are beneficial, such as bridgers and line extenders.

The third dominant amplifier technology is known as *feed forward*. In a feed-forward amplifier, the input signals are amplified via a conventional inverting amplifier. Samples of the input and output signals, adjusted for gain, are combined and amplified in a separate inverting *error amplifier*. The output of the error amplifier is summed with the output of the original amplifier so that the errors approximately cancel.[13] Figure 10.20 shows this configuration. The delay lines are required to match the phases of the signals. Modern feed-forward amplifiers offer about 12 dB of CTB improvement and 6 dB of CSO improvement over power doubling units at comparable operating levels. Furthermore, as discussed later, the remaining distortions are not necessarily phase coherent among cascaded amplifiers, which improves the performance of a repeatered system. Feed-forward amplifiers are generally used in long trunk lines where the lower cascaded distortion is beneficial.

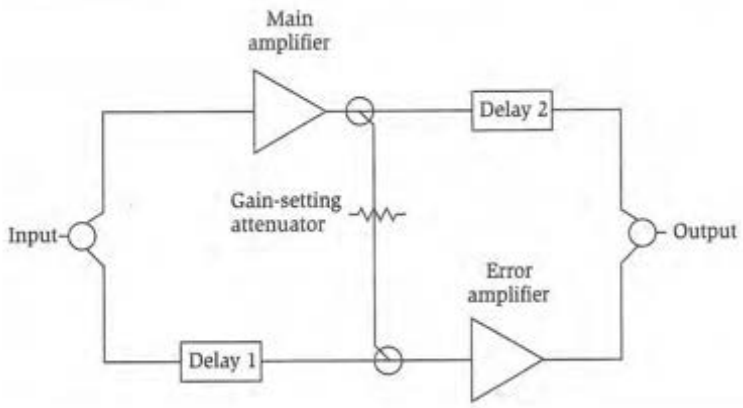


Figure 10.20 Feed-forward gain block.

10.4 Passive Coaxial Components

The nature of typical coaxial distribution systems is that the downstream path is split and resplit to create many endpoints from a common signal insertion point. In the upstream direction, the opposite is true, with signals from many potential insertion points combined at a common node. In the headend, signals from individual modulators and signal processors are combined to create the full downstream spectrum, whereas upstream signals are split to feed receivers for each service.

The signal splitting can occur either within amplifier housings or by use of stand-alone components. Those that split the signal equally are known as *splitters*, whereas those that divert a defined portion of the input signal to a side port are known as *directional couplers*. Finally, those that divert a portion of the input signal, then resplit the diverted signal to create individual customer drops are known as *taps*. All share certain characteristics and so will be treated in this section.

In general, all these are passive devices constructed using ferrite-loaded transmission lines and transmission-line transformers. They are bidirectional; that is, the same device can be used to split signals to feed multiple paths or to combine signals from multiple input ports. Important characteristics are signal loss, impedance match, and isolation between nominally isolated ports.

10.4.1 Directional Couplers

Figure 10.21 shows schematics of two directional coupler configurations. In both examples shown, the voltage transformer and current transformers have the same ratio so that, after combining, the ratio of voltage to current stays the same, meaning that the impedance of the side port is the same as the main line. In the main line, the relative polarity of voltage and current is reversed for signals traveling in one direction versus the other direction. The transformers sample voltage and current regardless of signal flow in the main line. The difference is that the sampled signal, after combining, will propagate either toward the termination resistor or toward the side port, depending on the relative polarity of voltage and current. This is what gives the coupler its directivity.

By varying the turns ratio of both transformers, various values of coupling are possible. Ellis[14] has published a tabulation of coupling values as a function of turns ratios, along with a good description of coupler operation. Table 10.2 lists some commonly used values and their theoretical signal loss values (practical units will exhibit 1–2 dB of excess loss due to imperfect components). The coupled loss is simply due to the voltage (or current) ratio defined by the transformers, whereas the input/output loss just reflects the loss diverted to the coupled port (or isolation resistor, depending on signal flow).

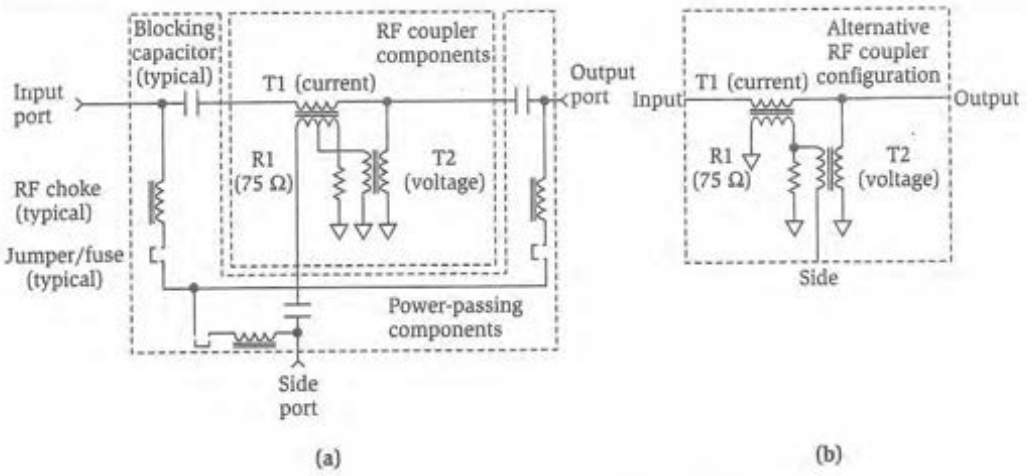


Figure 10.21 Directional coupler schematic. (a) Coupler plus power components. (b) Alternate RF configuration.

Table 10.2 Theoretical directional coupler losses

Turns Ratio (T1 and T2)	*Input to Output Insertion Loss (dB)*	*Input to Coupled Port Loss (dB)*
2.5:1	0.78	7.96
3:1	0.51	9.54
4:1	0.28	12.04
5:1	0.18	13.97

As is the case with amplifiers, added components in units intended for installation in trunk and distribution cables separate power from signal voltages and send them through different paths, as shown in Figure 10.21(a). Generally, means are provided to control transmission of power independently from RF signals so as to allow greatest flexibility in system design. Note that, if ac power is transmitted through all RF ports, the displacement current through the three blocking capacitors will add in phase through the primary of the voltage transformer. Thus, couplers generate parametric type hum modulation at levels as high as or higher than amplifiers.

The difference in sensitivity to downstream and upstream signals, as measured at the side port, is the *directivity* of the coupler. Sometimes, instead of directivity, the absolute upstream coupling factor, known as *isolation*, is specified. Numerically, directivity is simply the difference between coupling

factor and isolation. As we will discuss in Chapter 15, directivity and return loss interact to cause additional system group delay and response variation.

The coupling value may not be uniform across the entire frequency range of the device. Even though the overall variation in coupling factor may not be desirable or intentional, the slope and response variation are measured separately and are defined for passive devices the same as for amplifiers.

10.4.2 Splitters

Two-way splitters differ functionally from directional couplers only in that the signal splitting is equal. Other common configurations include four-way and eight-way equal splits. Three-way unequal splits are internally constructed by splitting two ways, then resplitting one of the legs from the first splitter to create a 50%:25%:25% output ratio. Equal ratio three-way splitters are less common.

Figure 10.22 illustrates a two-way symmetrical splitter, with expansion to four way. Figure 10.22(a) is the schematic diagram of a two-way unit, and Figure 10.22(b) is the standard symbol, with terminology. Transformer T1 is a 2:1 impedance step-down transformer that creates a 37.5-ohm source to feed the center tap of transformer T2. T2 is a transmission-line transformer derived from a Wilkinson coupler (which is used for symmetrical splitting of high-power, narrow band signals for directional antenna arrays, among other applications).[15] The two outputs at ports 2 and 3 are nominally in phase and at the same level. Resistor R1 is required to provide isolation between them.

Because power is equally divided between ports 2 and 3, the theoretical loss from port 1 to either output is $10\log(0.5) = -3.01$ dB. As with directional couplers, the practical loss will be higher, due to core losses in the ferrite core material in the transformers and resistive losses in the transformer wire.

The dimensions of transformers used in high-frequency applications are constrained by the necessity for the winding lengths to be a small portion of a wavelength at the highest frequency of operation. This, in turn, mandates very small cores and fine-wire gauges whose resistance is a factor in excess signal loss. The excess loss over and above that due to coupling is typically 0.5 to 1 dB to each of the output ports, making the total loss typically 3.5 to 4 dB.

Loss in a splitter is called *flat loss* because it tends not to be a strong function of frequency, as is the loss in coaxial cable. However, there is some increased loss as the frequency increases, partially due to skin effects, which increase the effective resistance of the wire as frequency increases. At the higher frequencies, the ferrite core on which the transformer is wound ceases to have significant magnetic properties, but rather acts as a coil form. Some loss may be associated with it, but the ferrite material is primarily included to improve low-frequency performance. At the lowest frequencies, loss also increases because the open-circuit reactance inductance of the transformer windings is no longer large relative to the circuit impedance.

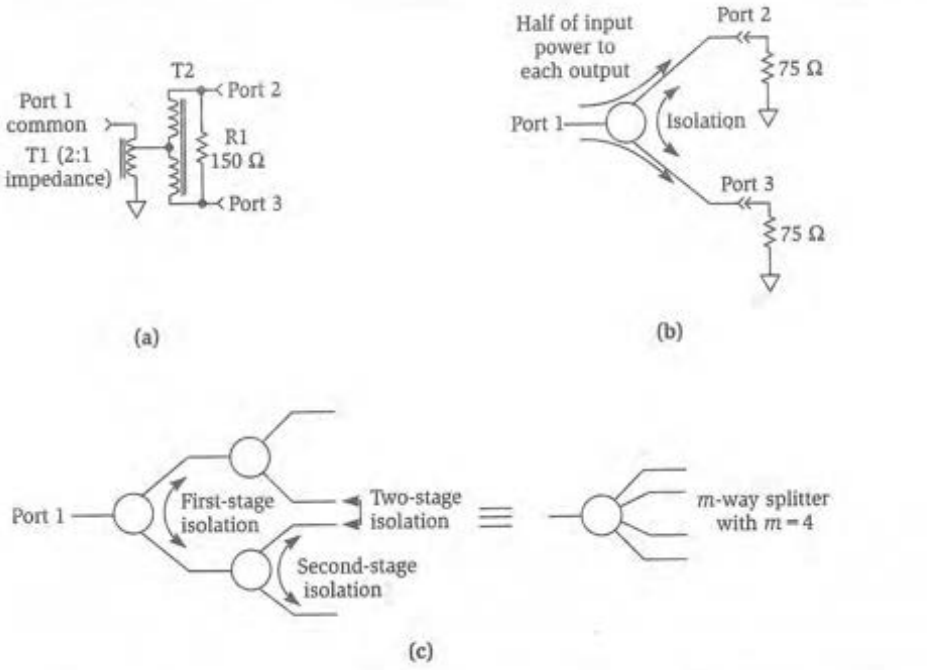


Figure 10.22 Splitters. (a) Schematic. (b) External power flow. (c) Multiple stages of splitting to create *m*-way splitter with enhanced isolation among some ports.

When the splitter is used as a combiner, incoming signals on ports 2 and 3 are combined and appear at port 1. The theoretical loss from port 2 (or port 3) to port 1 is also 3 dB, the other half of the power being dissipated in resistor R1. Theoretically, no power from port 2 appears at port 3 (if all ports are properly terminated) and vice versa. The signal loss between the output ports, with all ports terminated in 75 ohms, is defined as the isolation. Typical values of isolation for commercially available splitters range from 20 to 30 dB.

Other important specifications of splitters include insertion loss, normally specified as the total loss from port 1 to port 2, with port 3 terminated (or vice versa), and the return loss at each port, with the other ports terminated.

Lack of termination drastically reduces the effective isolation. For instance, power injected into port 3 is attenuated −3 dB at port 1. Should that port be unterminated, the power would be reflected into the splitter dividing equally between ports 2 and 3. Thus, the isolation from port 3 to port 2 would be only 6 dB. Similarly, the effective return loss at port 3 would be 6 dB as a result of power being reflected to that port.

Splitters having more output ports can be fabricated by combining two-way splitters in tree fashion, as shown in Figure 10.22(c). Each leg of the leftmost

splitter is connected to the input of another two-way splitter. At that level, four outputs exist. Each of those four can be connected to another two-way splitter, resulting in eight outputs, and so on. The practical loss at each stage will be 3.5–4 dB so that the four-way split loss is 7–8 dB. In an eight-way splitter, it is 10.5–12 dB and so on. In general, the loss in a multilevel splitter is given by

$$L_m = \left[\frac{L_1}{0.301}\right] \log(m) \tag{10.38}$$

where

L_m = the loss of the m-output splitter chain in decibels
L_1 = the loss in one two-way splitter (3.5–4 dB usually)

It is also possible to construct a splitter chain with a number of outputs not equal to 2^n by simply terminating the splits early on some of the chains. Equation (10.38) must be modified in this case.

Where output port isolation is important in multiport applications (such as headend combining) one alternative is to use ports in multilevel splitters that are not derived from common stages for critical applications. As shown in Figure 10.22(c), if two splitter outputs are connected to the same second-stage splitter, the isolation between them is merely the isolation of that stage. On the other hand, if two ports are not connected to the same second-stage splitter, then the isolation realized will include the effects of the first-stage splitter, as shown in the figure. Unfortunately, manufacturers may not indicate internal splitter configurations or specify nonadjacent port isolation.

Another higher-isolation alternative is to construct multiport splitters from cascaded directional couplers. Generally, side ports of adjacent, series-connected couplers will be better isolated than outputs from single-stage splitters.

10.4.3 Taps

The most common tap configuration consists of a directional coupler with the side arm signal split two, four, or eight ways to create customer drops. Although the input and through ports are typically provided with connectors appropriate for distribution cable, the side arms are equipped with connectors appropriate for drop cable. In North American systems, this means "5/8-24" ports for input and through ports and type F connectors for the drops.

A tap need not have a through port. At the end of the line, it is common to use a *terminating tap*, which is just a splitter whose output legs are equipped with drop connectors.

Until about 1995, taps supplied RF signals, but not power, to customer drop lines and were limited in main-line current-handling capacity to 7 amperes or

less since this was adequate for entertainment-based cable television systems. With the active consideration of HFC architectures for telephony service came increased demands on tap capabilities:

- Bandwidth expansion to 5–1,000 MHz
- The ability to supply power to side-of-house active *network interface devices* (NIDs) that could provide an interface between baseband twisted pair in-home telephone circuits and the RF carriers in the distribution system
- Provisions for changing tap values and configurations without interrupting service to customers farther along the distribution cable
- Increased voltage and current capability, typically 90 Vrms and 10 amperes, in order to allow transmission of sufficient power for both network amplifiers and NIDs over the same coaxial cables formerly used only for video service

The new generation of taps that have evolved are known as *power-passing taps* or, sometimes, *telephony taps*. In addition to normal tap capabilities, they tap into the distribution cable power and selectively provide current-limited power along with RF signals to a NID on each home. Depending on the design, this power may be multiplexed with the signals in the coaxial drop cable or carried along a separate twisted pair of copper wires in a hybrid cable configuration. In general, these taps also provide a make-before-break connection between the input and through ports when the tap "plate" (which contains all the circuitry) is removed from the housing so that both RF and power continuity are maintained to customers downstream.

A major concern for those using the drop center conductor to carry power to NIDs is that, if arcing occurs because of a faulty center conductor contact, the arc will transfer a very significant amount of RF power to the upstream plant, likely causing interference with all users of the reverse spectrum. On the other hand, an arc often causes healing of a bad contact and, thus, can be self-healing.

Figure 10.23 is a circuit diagram for a typical power-passing tap that uses the drop center conductor to pass power to the NID. The RF portion of the tap consists of a directional coupler and two-way splitter (the diagram is for a two-port tap). The directional coupler diverts a controlled portion of the downstream signal and passes it to the splitter, which divides the power to feed the subscriber ports.

Multiplexed ac power consideration in taps is similar to that on amplifiers and other network passives. Capacitors C1 and C2 isolate the multiplexed power supply voltage and current from the directional coupler RF components. Transformers T1 and T2 are built with very fine wire to permit operation at frequencies up to 1 GHz and would be unable to withstand the currents that would result if the primary of T1 were required to handle several amperes of ac current or if the primary of T2 had tens of volts of 60 Hz across it. The multiplexed ac power is shunted across the tap through the large RF choke L1.

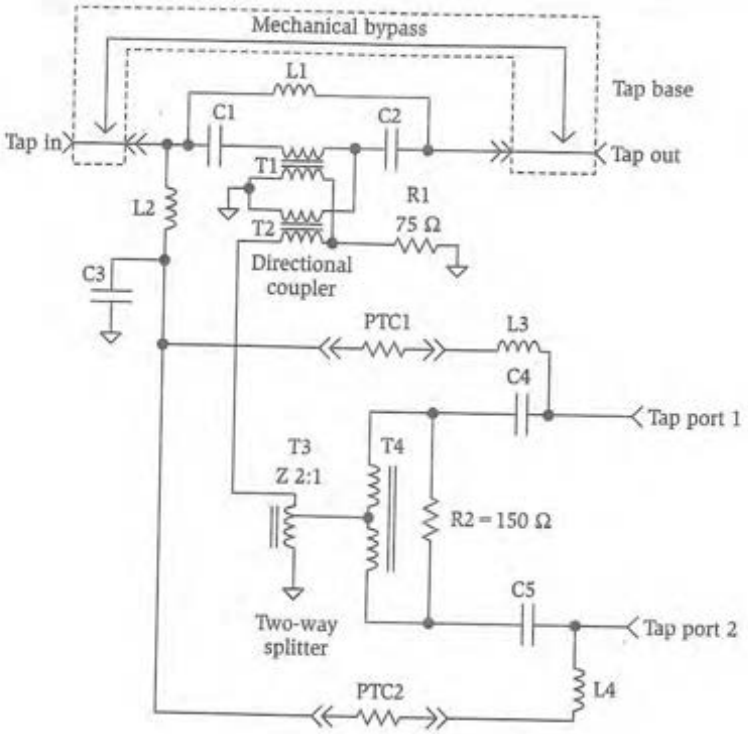


Figure 10.23 Power-passing tap using drop center conductor for power delivery to home.

A choke with somewhat less current capacity, L2, provides a path for the multiplexed power that will feed the customer ports. Since this choke is shunt-connected across the main RF path, it must have a higher RF impedance than L1. The unavoidable RF loss due to this choke is one reason power-passing taps have a slightly higher RF loss than normal taps.

In order to permit the tap plate (which contains most, if not all, the circuitry) to be removed without interrupting either power or RF continuity in the hard cable distribution plant, many modern taps incorporate a mechanical bypass switch that engages before the tap-to-baseplate connections are broken to provide continuity from input to through ports. During the removal or insertion of a tap plate, therefore, there will be a brief interval when both the bypass and main connections are made. This will cause a transient condition during which the primary of T1 will be shorted, but the primary of T2 will function normally. This momentary mismatch will cause some attenuation of the transmitted signals. Whether or not it causes a functional problem to the services offered will depend on the characteristics of the signals used.

Continuing with the description of the circuit, capacitor C3 is used to ensure a good RF ground on the "cold" end of L2. Positive temperature coefficient (PTC) resistors PTC1 and PTC2 limit the ac current that can be drawn through the customer ports. At normal currents, the PTC resistor (commonly called just a PTC) exhibits low resistance. As more current is drawn, its temperature increases, causing a resistance increase and effectively limiting the total current. Electronic current limiters have been proposed as an alternative to PTCs, but to date have not been widely deployed because of cost. New National Electric Code (NEC) regulations, effective in 1999, require a more sophisticated current-limiting system if the drop cable does not meet certain construction and installation criteria.

Finally, inductors L3 and L4 couple power to the two tap ports shown, again isolating the RF signals from the ac supply. As with L2, the reactance of these components must be sufficient to minimize RF losses, and their presence leads to slightly higher overall RF losses in commercially available power-passing taps.

The relationship between passive specifications and system performance will be explored in Chapter 15.

10.5 Power Supplies

Power supplies convert commercial power to lower voltages that are multiplexed with RF signals in the coaxial cables. Although not part of the RF circuit, certain characteristics of power supplies have an effect on the transmission of signals.

Direct current, very low frequency (as low as 1 Hz) ac, and 60-Hz ac have all been considered for system power, as have voltages between 30 and 90 volts. Direct current (dc) has been rejected primarily because of possible galvanic corrosion. Early cable television systems used 30 volts, but the increased power demands of wider-bandwidth, higher-output amplifiers forced the industry to convert to 60 volts for systems constructed after about 1970. Two factors are forcing consideration of even higher voltages: the desire to power larger plant sections from a single power supply to improve reliability (see Chapter 20 on HFC reliability for a detailed explanation) and the possibility of powering some terminal equipment through the distribution plant.

Along with the increase in voltage has come a requirement to increase currents as far as practical, with requirements to pass 10–15 amperes being common. Since current from supplies can be transmitted both ways along the cable and/or inserted into more than one cable, power supplies with 40 or more ampere capacity are available.

Field power supplies are roughly divided among nonstandby, standby, and UPS/standby types. The basic component of a nonstandby, line frequency power supply is the ferroresonant transformer. Though its magnetic design is beyond the scope of this book, ferroresonant transformers have a number of properties of interest to operators: (1) output voltage regulation (a typical specification is

±5% output voltage change when the input voltage changes ±15%), (2) inherent current-limiting and short-circuit protection, (3) transient protection, and (4) 90% or better efficiency at rated output current.

Standby power supplies add a battery charger, a set of rechargeable batteries, an inverter, and a changeover switch. The charger assures that the batteries are always held at full charge, whereas the inverter starts and supplies square wave power to replace the sine wave commercial power when it is interrupted (see Figure 10.24). Depending on the application, sufficient battery capacity may be provided for 2–8 hours or more of standby operation.

An important parameter for operation in the standby mode is the rise and fall time of the square wave voltage. As Equation (10.35) shows, the amount of ac power supply current that "leaks" into the low-level signal path in amplifiers and passives is directly proportional to the rise time of this waveform. The trade-off, for power supply designers, is that the efficiency of an inverter is an inverse function of the time the transistors spend in their active region (that is, not in a saturated or cutoff condition). Also, the power lost due to slow switching is dissipated in the transistors and must be conducted away by heat sinks. Unfortunately, the industry does not have a consensus on the best compromise on rise time.

When commercial power fails, there is an unavoidable but brief interruption while the relay switches between the ferroresonant transformer output and the inverter transformer output. Typically, the time to switch is 8–16 ms. Most supplies synchronize their inverters with commercial power before switching so that phase shift transients in the power supplies of the amplifiers can be avoided.

As will be discussed in Chapter 11, even with inverter phase synchronization, the short switching transient can be a problem when analyzed with respect to its effect on cascaded amplifiers. To avoid that, a third alternative is the uninterruptible power supply (UPS). Some UPS versions of field supplies avoid switching transients by utilizing the same ferroresonant transformer for both inverter and commercial power. During ac power outages, commercial power is disconnected, and the inverter continues to supply power via an additional primary winding.

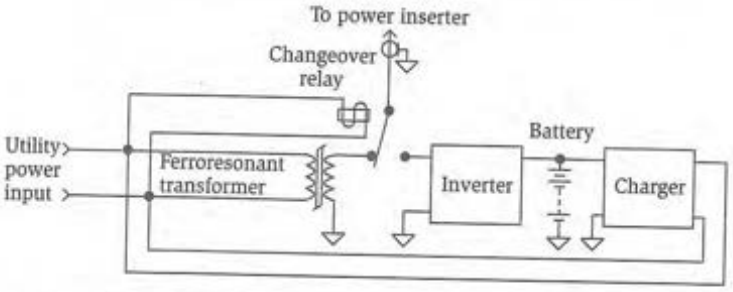


Figure 10.24 Basic standby power supply functional diagram.

10.6 Summary

This chapter has covered the technology behind coaxial distribution systems. The characteristics of the coaxial cable, amplifiers, passive components, and powering systems have been covered in detail, including both the theory behind operation and the performance of commercially available products.

Although other constructions are sometimes used, solid aluminum sheathed cable of 75-ohm impedance has been seen to offer a good combination of loss, cost-effectiveness, and handling characteristics and are in common use worldwide. Amplifier stations designed for cable systems are designed with unique characteristics that complement the characteristics of the interconnecting cable and passive devices. Integrated powering of active devices through the cable avoids the necessity of separately connecting each distribution system amplifier to the local utility, but requires special circuitry in every device to separate the power and signal paths.

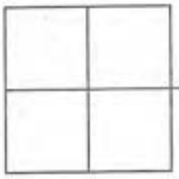
Before the mid-1980s, cable television systems used primarily coaxial technology (sometimes supplemented by broadband microwave links) to serve their customers. Chapter 11 will cover the design of coaxial distribution systems, including both the shared portions and individual in-home "drop" wiring. As will be seen, the performance of such networks is subject to certain unavoidable limitations that become more severe as bandwidth is increased. As a result, modern systems generally use fiber-optic "supertrunks" in place of the long, repeatered coaxial trunks to interconnect small, physically separate coaxial distribution networks with shared headends. Chapter 12 and 13 will cover the technology and performance of such links. Chapter 14 will then cover alternatives to fiber optics (primarily microwave). Chapter 15 will cover the end-to-end downstream performance of the broadband network, from headend to subscriber, and Chapter 16 will deal with upstream issues.

Endnotes

1. *Reference Data for Radio Engineers*, 6th ed. Indianapolis, IN: Howard W. Sams, 1977, p. 6–6.
2. Frederick Terman, *Electronic and Radio Engineering*. New York: McGraw-Hill, 1955, p. 121.
3. *NCTA Recommended Practices for Measurements on Cable Television Systems*, 2nd ed. Washington, DC: NCTA, October 1993, Part Six.
4. The size 59, 6, and 11 cables were formerly known as RG59, RG6, and RG11 because of their physical size similarity to standard military-grade 75-ohm cables carrying those designations. Since CATV cables do not attempt to

adhere to the military specifications, the RG notation has been dropped by the industry.

5. The SCTE is an ANSI-accredited standards-making body. It publishes a wide variety of standards covering both performance and test methods related to cable television products. In particular, SCTE-IPS-TP-011, *Test Method for Transfer Impedance*, and SCTE-IPS-TP-403, *Test Method for Shielding Effectiveness*, detail test methods for evaluating shielding integrity, whereas SCTE-IPS-SP-001, *Specification for Flexible RF Coaxial Drop Cable*, contains both mechanical and electrical minimum performance standards.
6. *Code of Federal Regulations*: C.F.R. 47 §76.5(w). Technically, the noise is defined as that occupying the band between 1.25 and 5.25 MHz above the lower band edge. Available from the Government Printing Office, Washington, DC.
7. An excellent detailed mathematical treatment of amplifier distortion based on a power-series expansion is given in the *Technical Handbook for CATV Systems*, authored by Ken Simons and published by Jerrold Electronics Corporation, Hatboro, PA (the third edition was printed in 1968 and was still in print in 1978). His results were also published in the *Proceedings of the IEEE*, in July 1970.
8. Schlomo Ovadia, *Broadband Cable TV Access Networks*. Upper Saddle River, NJ: Prentice-Hall, 2001, pp. 49–50.
9. Data taken from *Cable TV Reference Guide*, April 1992 edition, Philips Broadband Networks, 100 Fairgrounds Drive, Manlius, NY 13104.
10. Private correspondence from Archer Taylor based on his experiences with installed systems.
11. This treatment of CID is based on the paper by Jeff Hamilton and Dean Stoneback, The Effect of Digital Carriers on Analog CATV Distribution Systems, *1993 NCTA Technical Papers*, pp. 100–113. NCTA, Washington, DC.
12. SCTE IPS-TP-201 *Test Method for Frequency Response and Gain*. SCTE, Exton, PA.
13. A basic discussion of feed-forward gain block principles is contained in the paper by John Pavlic, Some Considerations for Applying Several Feedforward Gain Block Models to CATV Distribution Amplifiers, *Technical Papers, Cable '83*. NCTA, Washington, DC., June 1983.
14. Michael G. Ellis, RF Directional Couplers. *RF Design*. February 1997, p. 33ff.
15. Robert P. Gilmore, Applications of Power Combining in Communications Systems. *Application Note*. Qualcomm, Inc., San Diego, CA.



Chapter 22

Digital Set-Top Terminals and Consumer Interfaces

22.1 Digital Set-Top Terminals and Interfaces

With the widespread availability of digital video transmission on cable systems, digital set-top terminals (STTs) have taken over the marketplace. Extensive activities under the auspices of CableLabs and the SCTE, coupled with initiatives from the FCC and inputs from the consumer electronics industry, have resulted in a comprehensive set of standards for digital video transmission and reception. Many of the standards are recognized by ANSI, the American National Standards Institute. Where ANSI standard numbers were available at the time of writing, we refer to the ANSI number and also to the former standard numbers assigned by the SCTE.

A decision by the FCC to separate the STT from program denial created the need for a replaceable module to contain descrambling functions. The replaceable module is called a *POD* (point-of-deployment) security module also known as "cablecard". The idea was to allow consumers to purchase their own STTs and obtain a POD module from their cable company for reception of premium programs. In addition, TV manufacturers could choose to incorporate provision for POD modules in TV sets, eliminating the need for an STT altogether, even for premium programming. The POD module handles all vendor-specific in-band and out-of-band communications with the headend, supported by modulators and demodulators in the host device. *Host* is a generic term used to mean any device that accepts a POD module. This can include STTs, TVs, and any other device in which you might want to insert a POD module. The program input to the POD module is the demodulated 64- or 256-QAM datastream as it comes from the cable TV plant on the so-called *forward application transport* channel. (The program demodulator is located in the host.) The POD module descrambles the selected video stream from the incoming MPEG transport stream. It also handles out-of-band communications with the headend via the *out-of-band* (OOB) channel or the DOCSIS channel (advanced host). In addition, the POD module can handle additional services, such as emergency alert.

Concerns of the entertainment industry have resulted in the need for secure signal interfaces between STTs and display equipment. The ability to record signals is controlled by the copyright owner of the program. In this chapter we

outline the main issues involved in delivering digital TV to the subscriber. Of necessity, there is a tight coupling between this information and the information on digital transmission in Chapter 4 and on headend signal processing as shown in Chapters 8 and 9.

22.2 Relevant Standards

Several sets of standards contribute to the technologies described in this chapter. Most were generated under the auspices of CableLabs, in cooperation with the SCTE, and formalized into standards by the SCTE. The SCTE then submitted them to ANSI, which adopted them as U.S. standards. Additional standards have come from the Consumer Electronics Alliance (CEA), part of the Electronics Industries Alliance, and from European standards bodies.

Digital Video Subcommittee (DVS): A set of standards describing how TV signals are compressed and transmitted in cable television systems. This initiative by CableLabs has substantially completed the transition to ANSI standardization.

OpenCable: The OpenCable™ initiative was begun in 1997 with a goal of helping the cable industry deploy interactive services over cable. OpenCable provides a set of industry standards that help accomplish this goal via key objectives to define the next-generation digital consumer device, encourage supplier competition, and create a retail hardware platform.

OpenCable has two key components: a hardware specification and a software specification. The hardware specification allows a retail receiver to provide interoperability. The software specification of the OpenCable project, called the *OpenCable Applications Platform* (OCAP), is designed to solve the problem of proprietary operating system software by creating a common software platform upon which interactive services may be deployed.

VOD metadata: This is a newer initiative. It is a project to investigate the distribution of content assets (for example, movies) from multiple content providers and sent over diverse networks to cable operators. Metadata is descriptive data associated with a content asset package. It may vary in depth from merely identifying the content package title, to information for populating an EPG, to providing a complete index of different scenes in a movie or providing business rules detailing how the content package may be displayed, copied, or sold.

PacketCable: PacketCable is a CableLabs-led initiative aimed at developing interoperable interface specifications for delivering advanced, real-time multimedia services over two-way cable plant. Built on top of the cable modem infrastructure, PacketCable networks use Internet Protocol (IP) technology to enable a wide range of multimedia services, such as IP telephony, multi-

media conferencing, interactive gaming, and general multimedia applications. Much of PacketCable is covered in Chapter 6. Another early and continuing initiative, DOCSIS, is also covered in Chapter 5.

22.3 Cable TV Digital In-Home Processing

We could have described Figure 22.1 as a representative high-end digital STT, but it doesn't necessarily represent a real device.[1] Rather, it could be looked at as a superset of most real digital STTs or any device that accommodates POD modules. We'll refer to this as a *host*, an all-encompassing term used to describe any device that accommodates a POD module. You may not find all of the features shown on every host.

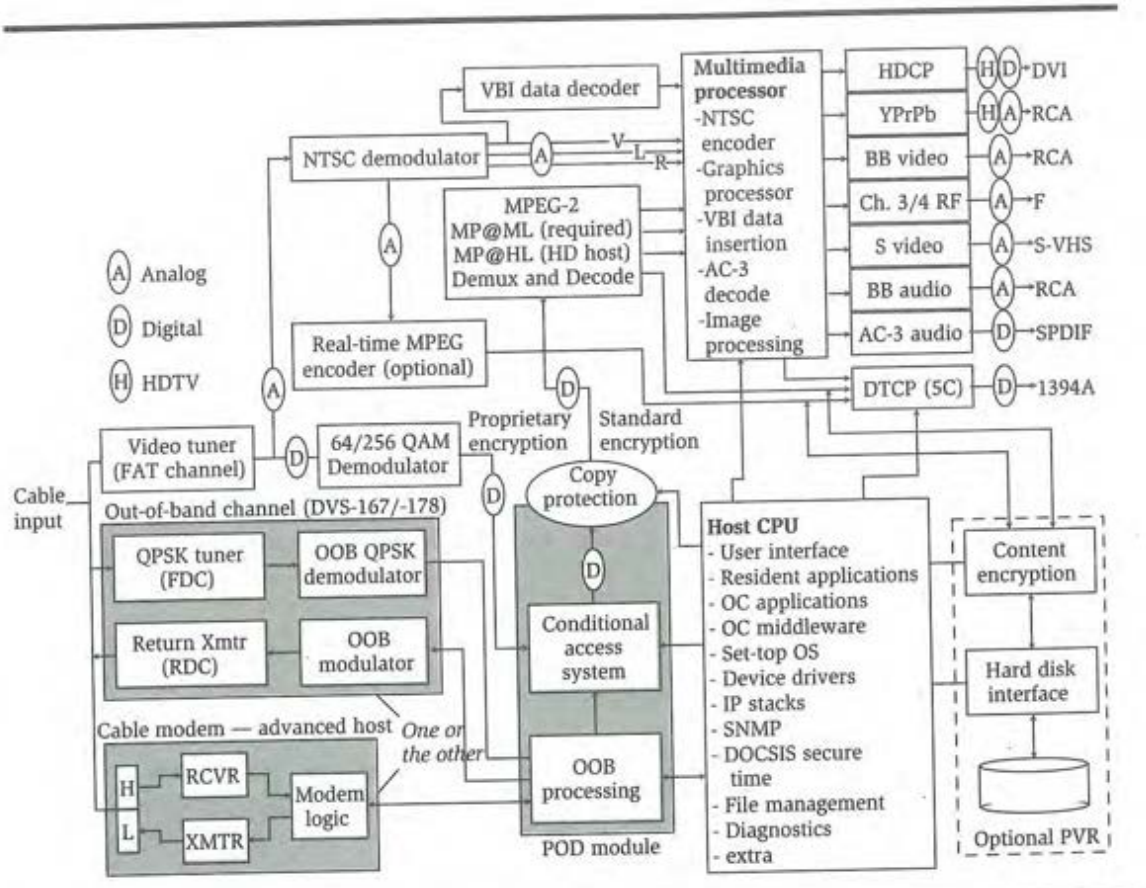


Figure 22.1 Representative digital host.

Many STTs are built with the POD module functions internal to the STT, so a separate POD module is not necessary. However, the intent of the FCC rules is to promote the availability of retail STTs that will use a POD module for customization of scrambling. The idea is to allow the consumer to purchase an STT, TV, or whatever with the features he or she wants and to get only the POD module from the cable operator. Also, it is possible to build TV sets, VCRs, and other consumer devices with POD module sockets, to allow them to operate as standalone devices, but to use the POD module to descramble digital signals. Note that the POD module is required only for scrambled programming; programs sent in the clear don't need a POD module to be installed.

22.3.1 Block Diagram Description

Video signals are tuned in the video tuner, which can be the same double conversion tuner used in analog STTs. Figure 21.9 illustrates a tuner in the blocks from the RF in through the second mixer M2. Dual conversion is used to achieve flat response over a very wide range of frequencies while not having to reject an in-band image. In Figure 22.1, the ovals with D or A in them indicate routes taken by digital or analog signals, respectively. An oval with an H indicates a high-definition signal path, which may also carry standard-definition signals.

Analog signals are routed to an NTSC demodulator of conventional design. A VBI decoder recovers VBI data to be inserted into the video signal in the multimedia processor. VBI data includes closed captioning text sent in the VBI of an analog signal, usually on line 21.

The multimedia processor accepts video in several forms as well as accepting VBI data from the VBI data decoder. Conceptually, the signals coming from the MPEG-2 decoder may or may not be in NTSC format (depending on the decoder used). Normally the multimedia processor will convert any format to RGB for insertion of graphics. The multimedia processor converts the signal to NTSC for conventional display and to whatever other formats are required, as described later. The graphics processor is the onscreen display of Figure 21.9, which adds VBI text data when called for and also displays text and graphics generated in the host. An example of host-generated graphics would be the grid of an electronic program guide.

The output from the multimedia processor may take a number of forms, depending on the capabilities intended for the host. We'll discuss the interfaces in more detail later, but can name them briefly here.

DVI is a digital interface used to transmit non-compressed HDTV pictures to a display. HDCP describes the copy protection specified for use with DVI. The YPrPb display is an analog component interface that is popular in early HDTV displays. It couples luminance (Y) and two color difference signals (Pr and Pb) to the display device.

Baseband, or BB, video is conventional analog NTSC (or PAL or SECAM) baseband video, as described in Appendix B and in Chapter 2. Ch. 3/4 RF output

is again the conventional RF interface used to couple signals to TVs that don't have any other interfaces available. Generally, the quality of the RF interface will be the lowest quality of any interfaces, though in many cases you will be hard pressed to see the difference between the RF interface and the baseband video interface. S-VHS is an analog interface that is the same as baseband video except that the luminance and chroma are transferred on separate wires. This has some distinct advantages in allowing for wider luminance bandwidth and in elimination of interference between the luminance and chroma information.

Baseband audio is conventional one- or two-channel analog audio for use by stereo systems. SPDIF is a digital audio interface originated by Sony and Philips (the Sony–Philips digital interface). It is a common digital interface for digital audio using the AC-3 format specified for digital transmission in North America.

Note that there is one digital interface that does not originate in the multimedia processor: the 1394 interface specified for standard-definition digital interfaces. DTCP is the copy protection system specified to be used with 1394 display interfaces. The reason it isn't shown originating from the multimedia processor is that it is able to transfer the MPEG-compressed video from the incoming video stream, plus information telling the display how to overlay any characters or graphics intended for display (the overlay information is provided from the multimedia processor). Thus, when using 1394, the display device, rather than the source device, is required to form the composite signal by overlaying graphics and/or characters on the picture. The multimedia processor does not have to assume this responsibility.

Returning now to a brief tour of Figure 22.1, an optional real-time MPEG encoder is shown. It is used if the host wants to be able to convert NTSC signals into digital signals for transmission to the display. This could be useful to allow consistent 1394 interfaces to the display. Another reason for including a MPEG encoder would be to accommodate a hard disk drive for storing video. The host would need such an encoder to accommodate analog signals. The storage function has come to be known as personal video recorder (PVR). Note the content encryption at the input to the hard disk interface. This is to ensure that content captured on the hard disk can be accessed only by the user who was authorized to capture the video and to limit its period of use if desired.

The POD module controls decrypting of digital video-encrypted signals. All of the program denial functions that allow the operator to sell premium services are embedded in the POD module, so the host without a POD module is useless for receiving anything except basic service.

Two options for out-of-band (OOB) communications are shown. The out-of-band channel is a QPSK tuner and demodulator plus a QPSK modulator and transmitter conforming to one of two OOB communications protocols recognized by the SCTE. The two are described later. The second OOB communications channel, anticipated for so-called advanced hosts, is a DOCSIS modem, as described in Chapter 5. The idea is to build in a DOCSIS modem, which handles host communications needs. The host could have a data interface to allow the subscriber to use this DOCSIS modem for all data communication needs.

Finally, the host CPU is the computer that controls all the functions of the host. These functions include user interfaces, such as the remote control and on-host controls, as well as any onscreen or on-host displays. The CPU will also house some resident applications that ship with the host. OCAP middleware running on the host CPU allows external applications to be written and loaded to the host. Other functions include external communications, the host role in network management, management of files, diagnostics, and whatever other functions the host manufacturer wants to include.

22.3.2 Cable Input

The cable input is standard cable TV signals, including return signals exiting the host. Input signals range from 54 to about 864 MHz, possibly increasing in the future. (The low frequency of 54 MHz is a North American standard. Other localities often use higher minimum downstream frequencies, allowing a wider upstream bandwidth.) The controlling document is SCTE 40 2001 (formerly DVS 313).[2] This document further references EIA-23 and EIA/CEA-542-B, specifications developed by the former NCTA/EIA Joint Engineering Committee. EIA-23 defined certain analog interfaces between cable systems and consumer electronics equipment. EIA/CEA-542-B defines channelization. The channel boundaries stated in this specification are reproduced as Appendix A. SCTE 40 2001 defines several types of signals that can be present on the cable plant:

NTSC: These are conventional analog TV channels on the cable.

FAT: This unfortunate acronym stands for *forward application transport* channels, which carry digital information via MPEG-2 transport streams. These are the digital video channels. Modulation is either 64- or 256-QAM.

NTSC and FAT channels may occupy the spectrum from 54 to 864 MHz and are channelized according to the table in Appendix A. (If history is any indication, the maximum frequency will increase in the future.) Signal level is measured at the end of the drop, except as noted. The maximum signal level is +15 dBmV for analog and FAT channels. For analog signals, the minimum signal level is 0 dBmV (governed by FCC rules) at the first consumer-owned device. Digital signal levels are carried at lower levels because they don't need as high a carrier-to-noise ratio (C/N) as do analog signals. 64-QAM modulated signals are carried nominally −10 dB with respect to the level of a hypothetical analog carrier on the same frequency. 256-QAM signals are carried nominally −5 dB from the hypothetical analog level. The specification allows a wider range of signal levels at the consumer device for digital signals, −15 dBmV to +15 dBmV for 64-QAM and −12 dBmV to +15 dBmV for 256-QAM.

C/N is an important quantity used to judge the quality of a signal delivered to subscribers. Today, C/N is sometimes replaced by the *carrier-to-composite-noise* ratio (CCN) or *carrier-to-noise-plus-interference* ratio (C/(N + I)). These

terms recognize that nonlinear distortions and leakage, which introduced essentially CW interference in analog systems, introduce interference that has the characteristics of noise in digital systems. C/N for an analog NTSC signal should be better than 43 dB (controlled by FCC rules), though good engineering practice dictates higher numbers. For 64-QAM the C/(N + I) should be no less than 27 dB, and for 256-QAM it should be no less than 33 dB. Chapter 4 explains why.

FDC: Out-of-band forward data channels carry control information to STTs and other devices that need it. This is also known as the downstream out-of-band (OOB) channel for addressable STTs. The information transported includes STT channel and program information (collectively known as *system information*, SI) and STT queries. The main information transmitted are *entitlement management messages* (EMMs), used to tell the STT or POD module what services it is entitled to receive. FDCs are usually modulated with QPSK (4-QAM) and set to −8 dB with respect to a hypothetical analog picture carrier at that frequency. FDCs are carried between 70 and 130 MHz.

RDC: Out-of-band reverse data channels carry return information from the STT and other devices to the headend. These channels may be carried anywhere between 5 and 42 MHz, depending on the cable system. Though reverse channel output levels may be up to +60 dBmV coming out of a device, the level allowed to appear at the input to another device is limited to +42 dBmV.

22.4 Out-of-Band Channels

The out-of-band FDC and RDC channels may take any of several forms. For advanced hosts, a DOCSIS modem may be included in the host to handle OOB communications. If a DOCSIS modem is included, the host may have an external interface for the DOCSIS modem to allow it to serve data needs for the entire home, not just for the host. The legacy OOB data channel shown in Figure 22.1 is based on one of two specifications: DVS-178 (SCTE 55-1 2002) or DVS-167 (SCTE 55-2 2002). Table 22.1 illustrates key features of the two standards: both the OOB FDC and the OOB RDC. SCTE 55-1 was originally developed by General Instrument; SCTE 55-2 was adapted by Scientific Atlanta from the DAVIC standard. In addition to two-way communications, it is important to provide for a one-way communications mode so that if the return channel fails for any reason, basic broadcast functionality, such as EPG, may be maintained.

22.4.1 Downstream OOB Channels

In the downstream direction, SCTE 55-1 operates at a wire rate (the actual data rate on the coax) of 2.048 Mb/s, with an information data rate of 2.005 Mb/s. SCTE 55-2 has two downstream grades, or modes, of which grade A is the only

Table 22.1 Comparison of out-of-band standards

	SCTE 55-1	*SCTE 55-2*
Formerly	DVS-178—mode A	DVS-167—mode B
Based on	MPEG, ATM	DAVIC
Downstream		
Data rate	2.048 Mb/s	1.544 Mb/s (grade A, required) 3.088 Mb/s (grade B, optional)
Nominal information rate	2.005 Mb/s	1.28 Mb/s
Frequency	70–130 MHz	70–130 MHz
Channel spacing	1.8 MHz	1 MHz (grade A) 2 MHz (grade B)
Modulation	D-QPSK, raised cosine $\alpha = 0.5$ (filter at receiver)	D-QPSK, root raised cosine $\alpha = 0.3$ (filter split)
Error correction	Randomization, R-S, interleaving	R-S, interleaving, randomization
Received carrier power	−10 to +5 dBmV	−18 to +15 dBmV
Segmentation and reassembly	Based on MPEG-2	ATM AAL5
Addressing	Broadcast Singlecast_unit Singlecast_network Multicast 40, −16, 24	Broadcast Singlecast
Upstream		
Data rate	256 kb/s	Grade A (optional): 256 kb/s Grade B (required): 1.544 Mb/s Grade C (optional): 3.088 Mb/s
Modulation and filtering	D-QPSK, $\alpha = 0.5$	D-QPSK, $\alpha = 0.3$
Access scheme	Polling, Aloha	Slotted TDMA synchronized to a common clock, reserved and contention slots
Packetization	ATM cell preceded by 28-bit unique word and 1-byte packet sequence, followed by 8 bytes R-S parity	ATM cell preceded by 32-bit unique word, followed by 6 bytes R-S parity and 1 byte guard band

Frequency	8.096–40.160 MHz	8–26.5 MHz
Channel spacing	192 kHz	Grade A: 200 kHz Grade B: 1 MHz Grade C: 2 MHz
RF output power	+24 to +60 dBmV	+25 to +53 dBmV
Received level at headend	+3 ± 10 dBmV	Four target settings, −5, +3, +11, +19 dBmV, all ±3 dB
Required C/N at receiver	20 dB @ PER < 10^{-7} after FEC	20 dB @ 10^{-6} packet loss rate after error correction
Acknowledgment protocol	STT waits for acknowledgment from headend or enters backoff algorithm	Positive acknowledgment sent for every successfully received ATM cell
Maximum upstream channels per downstream channel	6	8

one implemented as of this writing. It features a downstream wire rate of 1.544 Mb/s with a data throughput rate of 1.28 Mb/s. Both standards include the same downstream frequency range of SCTE 40, though in addition SCTE 55-1 shows a default frequency of 75.25 MHz.

Note a difference in the downstream filtering in the two standards. In Chapter 4 we showed that most digital transmission systems employ raised cosine bandpass filtering and that most split the filter evenly between the transmitter and the receiver. This leads to the description of a root raised cosine filter (the "root" terminology arises from the fact that the filter response at each end is the square root of the overall filter, so when the transmit and receive filters are cascaded, a raised cosine filter results). This is the technique used in SCTE 55-2, but SCTE 55-1 places all the filtering at the receiver. This results in a wider occupied spectrum on the cable but improves the C/N characteristics of the receiver.

The numbers in Table 22.1 were taken from the SCTE 55-1 and -2 specifications. There is a difference between them and the downstream received signal level specification of SCTE 40, but the difference is not of practical concern if you follow the recommended practice of placing the data carrier −8 dB from a hypothetical analog picture carrier at the same frequency. Note also that the downstream addressing for both standards allows addressing either individual hosts or all hosts at once.

22.4.2 Upstream SCTE 55-1 Channel

In the upstream direction, SCTE 55-1 uses a single data rate of 256 kb/s. The difference in upstream access schemes is of interest. SCTE 55-1 has a polling channel. A polling protocol may be thought of as a "speak only when spoken to"

protocol. Each host is polled and responds when polled and only when polled. This polling channel (which can be a separate return frequency) is used to allow the host to report its health and to report any pay-per-view programs that have been watched since the previous poll. Other non-time-critical data may be transmitted on the polling channel. Polling takes place on an infrequent basis — each Host may be polled perhaps once per day.

SCTE 55-1 also specifies one or more Aloha channels. These may be used by different applications running on the host. For example, a video-on-demand (VOD) system would like to use the host to request programs at any time and would also like to allow the viewer to command the program to pause and resume, rewind and fast-forward (so-called VCR functions). Aloha plays a central role in the vast majority of Ethernet networks in use today. The Aloha Protocol was developed at the University of Hawaii by Norman Abramson and colleagues in the late 1960s as a way to communicate by radio between the various islands and interisland shipping interests. The radio channels used in Hawaii had a lot in common with the return channel of cable systems today. The Aloha Protocol allowed any station to transmit when it needed to transmit, rather than having to wait until it was "invited" to transmit, using polling techniques. There was no guarantee that one station would be able to hear any other station that was transmitting at the same time. In the original application, this was because of the vagaries of radio paths. In cable plant, there is no way to route a signal from one host to another, so when one host is transmitting, no other hosts will know it.

In a pure Aloha system such as in SCTE 55-1, when a host has something to say, it transmits and then waits for an acknowledgment from the headend. If no one else is using the channel, the transmission gets through, the headend acknowledges it, and the host goes on about its business. If an acknowledgment is not received within a designated time, the host assumes its message didn't get through because another transmission at least partially overlapped with its transmission. This means there is at least one other host trying to communicate. So the host waits ("backs off") a random time and tries again. Chances are that the other host who was transmitting at the same time has backed off for a different random time and that on the next try both will get through and will be acknowledged.

Pure Aloha systems work well when there are a lot of potential "talkers," each needing the channel a small percentage of the time but needing it quickly when it is needed. A VOD application is a good example. Normally, VOD systems allow the subscriber to have VCR-like control, in that they can pause and resume the program, rewind, or fast-forward. A lot of subscribers can be watching VOD and sharing a return channel to send those commands to the headend. Rarely does any one subscriber want to use any of those VCR-like controls; but when he does, he needs a response fast from the headend.* Polling wouldn't work, because even

* There are some instances involving networked PVRs in which a number of people will want to simultaneously access the VCR-like controls, stressing Aloha.

fast polling of many hosts could take minutes if not hours. A VOD viewer would not wait that long to get a response to, say, a pause command. But since there are few commands issued by each subscriber, many can use the channel and not experience significant interference. A pure Aloha system can be utilized to about 18% before delays due to collisions become unacceptable.[3]

Figure 22.2 illustrates a typical STT communications channel based on SCTE 55-1 (formerly DVS-178). Some cable TV systems are intended to be locally controlled in terms of STT authorization, and these systems use a control computer. Other systems using SCTE 55-1 are designed to be controlled from a central point via satellite. These systems don't use a control computer but, rather, have a master integrated receiver-transcoder (IRT—a digital satellite receiver) that supplies control information. Either the master IRT or the control computer supply signals to an OOB modulator, which sends the downstream QPSK signal as described in Table 22.1. This signal is combined with all other downstream signals and supplied to the distribution network, commonly via a downstream optical transmitter.

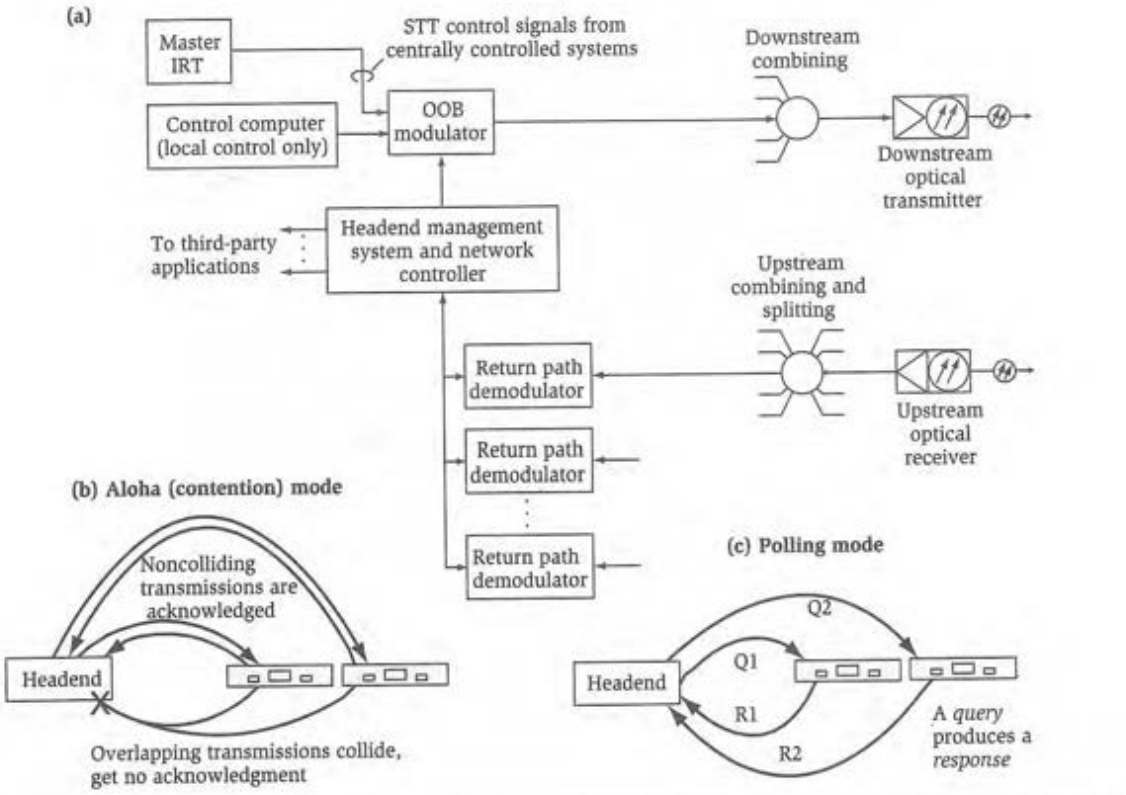


Figure 22.2 SCTE 55-1 communications system.

In the upstream direction it is common (though not universal) to make provision for several upstream return path demodulators for each downstream modulator. Upstream signals, either polling or Aloha Protocol, come to the headend, usually on an upstream optical receiver. Since there may be multiple combinations of upstream signals for different services, as described in Chapter 9, there will be some combining and splitting loss in the signal path before the upstream signals reach the assigned return path demodulator. The demodulators used may be assigned to different groups of STTs (by the way the upstream signal management is done), or one receiver may be assigned as the polling receiver and the others assigned as Aloha Protocol receivers, each supporting a different service (such as VOD or game playing). The receivers may be assigned by virtue of being tuned to different upstream frequencies. For example, the polling receiver might be assigned to 8.096 MHz (the lowest channel) and the VOD receiver (Aloha Protocol) to 8.288 MHz, the next channel up.

Each receiver demodulates the QPSK upstream signals and supplies it to the headend management system and the network controller. The headend management system receives the information on the polling channel and handles interfaces with the billing system and other systems needing access to the information. The network controller provides interface to third-party applications, such as a VOD system.

22.4.3 Upstream SCTE 55-2 Channel

The system used in SCTE 55-2 (formerly DVS-167) for return transmission is significantly different from the SCTE 55-1 system. It is called a time division multiple access (TDMA) system. SCTE 55-2 allows for three different upstream data rates, though only one is implemented as of this writing. Figure 22.3 shows the overall system. Systems using SCTE 55-2 were generally designed to operate with local host control. A digital network control system (DNCS) provides physical interface to applications needing access to the STTs. The control transmitter is the QPSK transmitter that generates the downstream signals shown in Table 22.1. As with SCTE 55-1 systems, several receivers may be used to improve the upstream throughput.

The upstream receivers demodulate the incoming signals and pass them to the control transmitter, which in turn passes signals to the DNCS. The DNCS manages some of the communications with hosts itself and relays other communications to and from other computers running various host tasks.

Upstream transmission is synchronized from a common clock at the headend. Downstream transmissions are divided into repeating signaling link extended superframes (SL-ESFs). Each SL-ESF consists of 24 frames, each carrying 193 bits (1 synchronizing bit and 192 data bits). Timing for the upstream begins with receipt of *SL-ESF START* from the headend. Upstream time slots are defined from this start of the downstream SL-ESF. Normally each Host that needs to talk reserves one or more of the upstream time slots in which to talk. Time slots may be assigned for one-time use or may be reserved for a longer duration

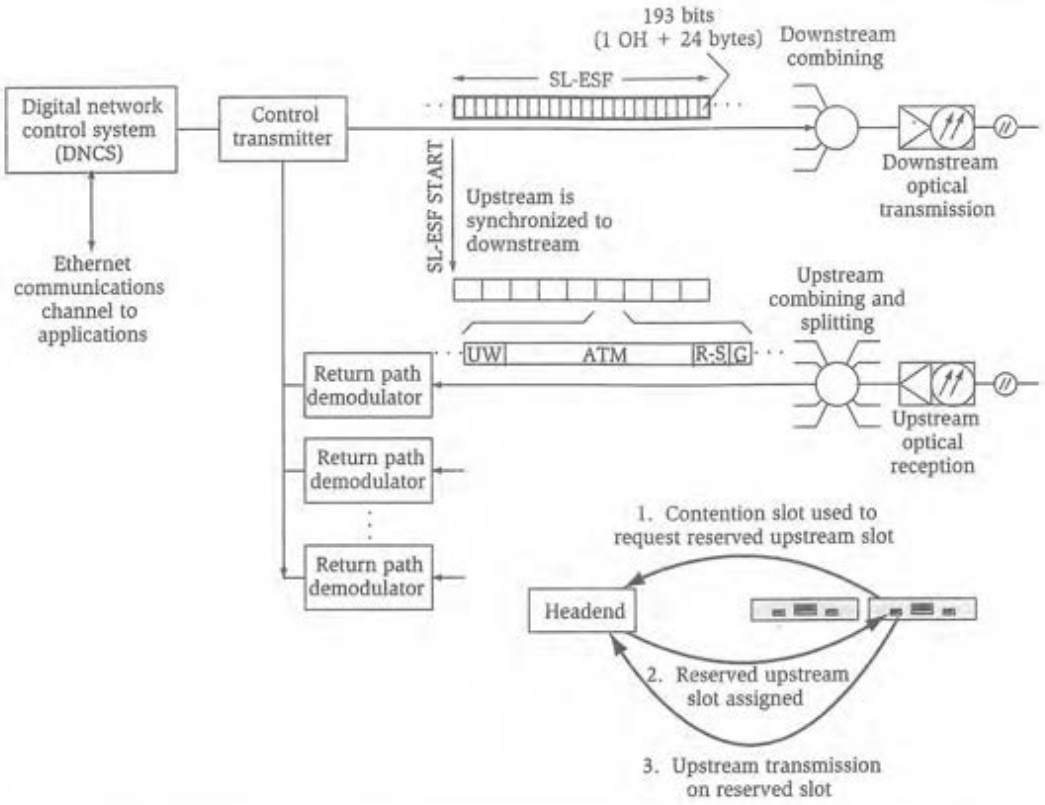


Figure 22.3 SCTE 55-2 communications system.

if an application needs it. An application may also reserve adjacent slots. For upstream transmission at 1.544 Mb/s, the most common speed, there are nine upstream time slots for each SL-ESF. These time slots consist of a unique word (UW) that serves as a run-in to synchronize the receiver clock. Following this is a 53-byte ATM cell. This is followed by a Reed–Soloman error correction word and finally by a guard time, which ensures that small timing errors don't cause transmission overlap from different hosts.

Propagation time between the headend and the host is important, since it is necessary to time all transmissions to arrive at the headend at the proper time. The propagation delay difference between a host that is close to the headend and one that is distant is significant, so a ranging function must be applied to each host when it is first added to the system. The headend measures the time delay between the "expected" response time from the host and the actual response time, and it tells the host how much to advance its timing so that the signal

arrives at the headend at the correct time. Other systems that depend on tight timing also have this ranging function.

In order to allow a host to gain access to a time slot, there must be provision for a host to inform the headend that it needs a time slot. Also, when a new host is added to the system, it is necessary to allow it to inform the headend of its presence. In order to meet these needs, a return time slot is occasionally set aside for contention. This time slot will work very similarly to the Aloha system described earlier, except that rather than the host's having the ability to begin "talking" at any arbitrary time, it must wait for a contention time slot, which is defined by the headend. This modification of the pure Aloha Protocol is called *slotted Aloha*. There is a time slot set aside periodically in which any host needing access to a reserved time slot so informs the headend in the contention time slot. The headend acknowledges the request and assigns an available time slot for that host. Whereas a pure Aloha system becomes unusable at about 18% utilization, slotted Aloha can operate with a utilization of up to about 37%, because there are no partial time overlaps.

The inset drawing at the lower right of Figure 22.3 illustrates the sequence of events that a host must go through in order to deliver an upstream message. Step 1 is to wait until a contention slot comes along. The host transmits its request for an upstream time slot to the control transmitter, which manages the assignment of time slots. In step 2, the control transmitter assigns a time slot to the host. In step 3, the host transmits its data. The upstream time slot is normally assigned in the second SL-ESF following the request. Some short messages may be sent in the contention slot rather than going through the process of getting an assigned time slot.

22.4.4 Advanced Hosts

As shown in Figure 22.1, a DOCSIS communications channel may be substituted for the SCTE 55 communications channel. This so-called DOCSIS set-top gateway, DSG, uses Ethernet tunneling to transmit data to the host, and so it doesn't need the QPSK channel described earlier. See Chapter 5 for a discussion of DOCSIS modems. If this is done, it is also possible that the host will include an interface to computers, so the modem may be used both for host communications needs and for computer communications. Possible interfaces include Ethernet, USB, and IEEE 802.11, among others.

22.5 Output Interfaces and Copy Protection

A number of output interfaces are shown for the model host of Figure 22.1. We shall first cover the various analog interfaces that are popular on early-generation digital STTs; then we shall cover the digital interfaces.

22.5.1 Analog Interfaces

BB Video

This is the common baseband analog video interface described in Chapter 2 and Appendix B. Sound is not included. The significant frequencies extend from nearly 0 to 4.2 MHz for NTSC video and from nearly 0 to 5 MHz for PAL video. Composite video (including sync, as described in Appendix B) is transmitted at a level of 1 volt p-p, with a source and load impedance of 75 ohms. Normally, the common "RCA" connectors are used, though occasionally other connectors, such as BNC, are employed. Impedance matching is important if the connecting cable is long, to prevent ringing or ghosting. The quality of the interface is better than that of the Ch. 3/4 interface described next but not as good as the other analog interfaces described, because the chrominance information is combined with the luminance, and the two must be separated before they can be processed. The filters used to separate the two are imperfect; even if perfect filters were available, there would be some inevitable interference.

Ch. 3/4 RF

This is baseband video modulated onto an RF carrier, identical to signals transmitted on cable systems and over the air. The one exception is the lower sideband, which is normally removed by the vestigial sideband filter (described in Section 8.3.7). Vestigial sideband filtering is done to allow adjacent 6-MHz channels to be used. Often the vestigial sideband filter is omitted in consumer products. This has no effect on the quality of the signal received, but it does preclude placing another signal in the lower adjacent channel. Also due to cost, the sound carrier frequency may not be as accurate, and the picture/sound carrier amplitude ratio may vary. Again, these differences should not pose a problem in consumer applications.

Although the quality of the Ch. 3/4 video interface can be quite good, of all the interfaces available, this one would be considered the lowest in quality, since it starts with composite video and modulates it onto a carrier, requiring demodulation in the receiver. The extra modulation and demodulation steps cannot improve the quality of the signal.

In North America it is common to use channel 3, except in areas having an off-air channel 3, in which case channel 4 is used. Channel 2 is not commonly used because it is the second harmonic of the old citizen's band radio service (CB), and harmonics of CB services have been known to impair reception.

The signal level coming out of the STT should be above 0 dBmV and less than approximately 15 dBmV. Sound is modulated onto a subcarrier at 4.5 MHz (NTSC), 5.5 MHz (PAL-B/G), or 6 MHz (PAL-I). F connectors are used almost universally in consumer equipment. It is important to ensure that the source and load impedances are 75 ohms and that 75-ohm cable is used, regardless of the length of the cable.

S-VHS

This standard originated in the mid 1980s as S(uper)-VHS, an improved version of VHS consumer videotape. The recording standard did not become popular for consumer use, though it has been used for semiprofessional purposes. Compared with normal VHS recording, there were some differences in the way the recording was made on the tape. The interface is now sometimes called S-video.

Luminance and chrominance (color) signals are transmitted to the display on different wires. The luminance information is transmitted as a baseband 1-V p-p signal including sync, just as in baseband video. The color subcarrier (3.58 MHz in NTSC and 4.43 MHz in PAL) is not superimposed on the luminance as it is in normal baseband video. Rather, it is transmitted on a different wire. The advantage is that no filtering is needed to separate the luminance and chrominance information. This permits wider bandwidth of the color signal, improving color resolution. Many TVs today have S-VHS inputs, and excellent quality can be achieved using this interface.

Besides allowing wider chrominance and luminance bandwidths, S-VHS eliminates cross-color effects that arise when the chrominance and luminance signals interfere with each other. Cross-color effects may be seen, for example, when fine diagonal stripes, such as in a striped jacket, are on the screen. You have seen spurious colors surrounding such images. Spurious color occurs when luminance signals with a lot of high-frequency content (as are generated by diagonal stripes) mix with the color spectrum, confusing the color demodulator. Since properly produced digital video is derived from a signal that has never existed in composite analog video format, use of S-VHS connection eliminates cross-color effects.

S-VHS uses a four-pin miniature connector. As shown in Figure 22.4, the luminance (Y) and chrominance (C) each have their own ground. The luminance signal is identical in every respect with the baseband luminance signal, including sync. The chrominance signal is a modulated 3.58-MHz (NTSC) or

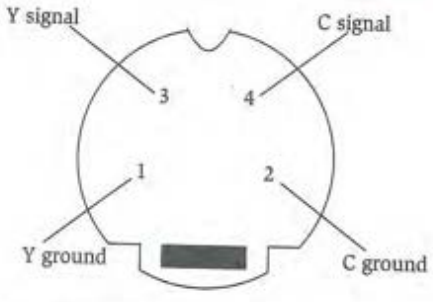


Figure 22.4 S-VHS pin function.

4.43-MHz (PAL) signal identical with the color subcarrier of the composite baseband signal. If you combine the Y and C signals, you will have a proper composite baseband signal. Again, the reason for the S-VHS connector is to keep the two signals separate so that they will not interact. A second improvement is that separating the chrominance from the luminance allows a wider chrominance bandwidth, resulting in sharper color edges.

Audio is not carried on the connector and is normally carried as left and right audio on a pair of cables terminated in RCA jacks.

YPrPb

This is an analog component interface using three cables to transfer the luminance (Y) and red and blue difference signals (Pr and Pb). It is used for standard definition video and may optionally be provided for HDTV display interface. Sync is carried on the luminance channel. Analog copy protection may be included, as described later. Three separate coaxial cables are used, often bound into a single cable. For consumer applications, RCA connectors are used, whereas BNC connectors are frequently used in professional applications. As with other video signals, 75-ohm cable should be used, though non-impedance-controlled cable is sometimes used in consumer applications.

As shown in Chapter 2, a color television signal is composed of red, green, and blue components. For transmission, these components are converted to a wideband luminance (Y) signal and two narrowerband color difference signals. When the signals are transmitted in analog form, the color signals are called the Pr and Pb signals, for the luminance minus red and luminance minus blue components. The same components, when in digital form, are called YCrCb.

The components are defined by CCIR Recommendation 601 for modern systems as follows:

$$
\begin{aligned}
Y &= 0.299R + 0.587G + 0.114B \\
P_B &= \frac{0.5}{1 - 0.299}(B - Y) \\
&= -0.169R - 0.331G + 0.5B \\
P_R &= \frac{0.5}{1 - 0.114}(R - Y) \\
&= 0.5R - 0.419G - 0.081B
\end{aligned}
$$

The physical interpretation is as follows. Assume the red, green, and blue (RGB) outputs from the camera range from 0 for black to 1 (the units can be volts or some arbitrary unit) for maximum of each primary color (the camera must be adjusted such that these maximums correspond to the "brightest" color to be reproduced). Then at each pixel (picture element, or the smallest resolvable element in the scene), R, G, and B will each have a value. The luminance (Y)

signal is computed from the first line of the equation, which approximates the response of the human visual system to each color. In order to make up the two color difference signals, the difference between Y and the R and B channels is taken, with the scaling factors shown. Notice that the color difference signals can take on negative values, even though there are no real negative colors. The two color difference signals usually have one-half the bandwidth of the luminance signal (less in analog transmission systems).[4]

The foregoing equations are usually written in matrix form,

$$\begin{bmatrix} Y \\ P_B \\ P_R \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Analog Copy Protection

An analog copy protection system was developed a few years ago to prevent consumers from recording premium programs on VCRs. It is used to prevent recording on analog video outputs from the host as well as on DVD players. The system takes advantage of the fact that VCRs use AGC on the video signal to set the peak-to-peak amplitude of the video signal before recording. Special signals are inserted in the vertical blanking interval that cause the AGC to set up to the incorrect level part of the time.

BB Audio

This interface is conventional analog transmission of baseband audio. In stereo systems, left and right channel information is transferred. For home theater use, 5.1 channel audio can be transmitted on six cables. The channels are left, center, and right front channels, left and right rear channels, and a low-frequency enhancement (LFE) channel. The LFE channel is limited to approximately 20–100 Hz and hence is considered "0.1 channel." Used to provide "feel" to movies in home theaters, it is an optional feature of digital TV. (There is a newer practice that uses a one or two rear channel for 6.1- or 7.1-channel audio, but this is not officially supported in most interfaces as of this writing.)

Common interface levels are about 1 V rms at full modulation. The terminating impedance is usually 2 kilohms or higher, and the source impedance is much lower. Because the frequencies of audio are so low, there is no need to maintain impedance match in consumer systems, and maintaining a much higher load impedance than the source impedance enables more than one load to be placed in parallel is desired. Compare this practice with professional equipment, where a load impedance of 600 ohms is used. The source impedance is often, but not always, either 150 or 600 ohms. Though the impedance of the cable is not

critical, it is important to use a high-quality shielded cable to minimize hum pickup and crosstalk.

Consumer equipment almost always uses unbalanced transmission, whereas professional equipment almost always uses balanced transmission, drastically improving the immunity to hum and spurious signal pickup. The issues related to balanced and unbalanced transmission are covered in Section 8.9.2. If long cables are used, there is some possibility of high frequency roll-off due to cable capacitance. This is normally not a serious problem in small home entertainment clusters, but if you find yourself running audio cables between rooms or connecting them with Y adapters, you should pay attention to the total cable capacitance. With a source impedance of 2 kilohms and a high load impedance, a total cable capacitance of some 800 pf could marginally impair frequency response (we are being conservative). Typical audio cables seem to exhibit capacitance of about 30 pf per foot, so as the total cable distance approaches 25 feet per channel, it could introduce some very minor response problems if the source has a 2 kilohm output impedance.

22.5.2 Digital Interfaces

DVI

This is the interface specified for high-definition digital displays. Uncompressed RGB (red-green-blue components, as opposed to luminance and color difference) digital video is transmitted, so the MPEG decoder is in the host. DVI has its roots in computer monitors. Manufacturers recognized that with high-resolution monitors the analog interfaces that had been used for many generations of product were now limiting performance. Furthermore, with LCD displays, the signal to be displayed was converted to analog in the computer and then sent to the display, where it was converted back to digital. DVI was created by the Digital Display Working Group[5] (DDWG) to convert analog signals into digital signals to accommodate both analog and digital monitors. The standard specifies a single plug and connector that encompass both the new digital and legacy analog interfaces as well as a digital-only plug connector. However, for television applications, the analog channel is not used. DVI handles bandwidths in excess of 160 Mb/s per channel and thus supports UXGA and HDTV with a single set of links. Higher resolutions can be supported with a dual set of links (not used for HDTV applications).[6,7]

The interface uses a physical transmission protocol known as transition minimized differential signaling (TMDS). In concept it is similar to 8B/10B encoding used with gigabit Ethernet and described in Chapter 19. Two additional bits are added to each word for the purpose of minimizing the number of transitions (in this case) while removing any DC component of the signal.

Figure 22.5 summarizes information concerning the physical interface used in DVI. Figure 22.5(a) shows the transmission of signals between a host such as

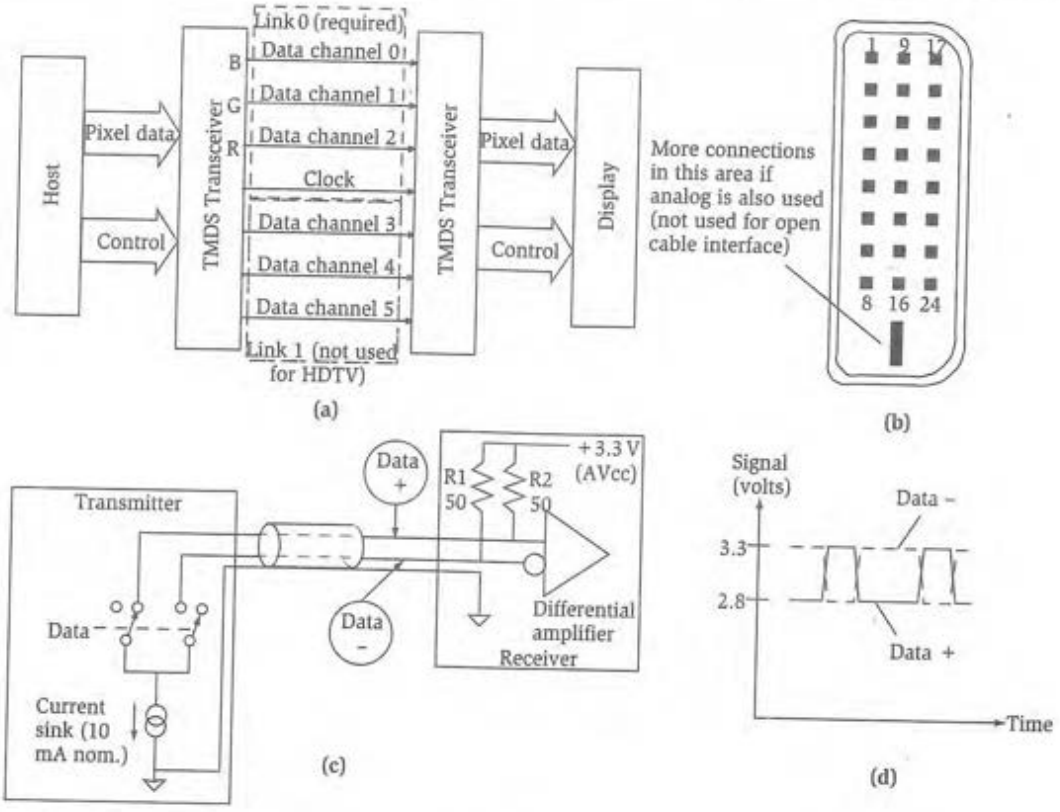


Figure 22.5 The DVI. (a) Communications channels. (b) DVI plug. (c) Equivalent electrical circuit. (d) Waveform.

an STT on the left and a display on the right. Pixel and control data are transferred to a TMDS transceiver and across a data channel (a DVI cable) to a TMDS transceiver on the display, which converts the signals back to pixel and control data for the display. Two complete digital links are included in the DVI specification, called link 0 and link 1. Each link consists of three data channels, for blue, green, and red data. A common clock line is used for all channels. The reason for two links is to double the effective data transfer speed. Only link 0 is needed for HDTV applications. Data channel 0 carries blue information, data channel 1 carries green, and data channel 2 carries red information. Horizontal and vertical sync information is carried in the blue channel. (Note that this is a different practice from that normally used in analog component video, where the green channel carries sync. Green is the logical choice in analog practice because the picture is made up predominantly of green. In digital, there is no particular reason for the sync to appear on one channel over any other.)

DVI can operate at data rates from 25 Mb/s to 165 Mb/s, depending on how much data is required to be transferred. At the same frame rate, there is up to 5.3 times as much information that must be transferred for HDTV as for SDTV, depending on the display format being used.

Figure 22.5(b) illustrates the connector used for DVI. Though not specified for HDTV use, there is an option for conventional analog signals in accordance with the VESA specification. If it is supplied, an additional four connections are used at the bottom of the connector. The digital signals are transferred on the 24 pins shown.

Table 22.2 shows the connections in the DVI connector. The two italicized rows in the center are used for link 1 and are not used for HDTV applications. Display and clock data are transmitted as balanced signals. Each data channel is individually shielded, with the shield ground being shared between link 0 and link 1. The DDC clock and data are used to identify the type of display, controlled by the VESA *data display channel* specification. Negotiations carried over this link allow the host and the display to agree on operating parameters. The hot plug detect on pin 16 places a ground on the host pin 16 when a display is plugged in. It allows a user to plug in a display after the host is powered up. The ground generates an action by the host to determine the type and capabilities of the monitor that has just been connected.

Figure 22.5(c) illustrates the method used for clock and data signal transmission between the host and the display. This is a balanced transmission technique using terminating resistors at the receiver and switched current sources at the transmitter. Note the similarities between this method of data transmission and balanced audio transmission as shown in Section 8.9.2. Balanced transmission is known for reliable transmission while generating low radiated emissions. Data actuates the switch in the transmitter, which alternately connects the data + and data– lines to the current sink, a circuit that pulls 10 mA to ground at all times.

Table 22.2 DVI connector pin functions

Pin	*Signal Assignment*	*Pin*	*Signal Assignment*	*Pin*	*Signal Assignment*
1	TMDS data 2–	9	TMDS data 1–	17	TMDS data 0–
2	TMDS data 2+	10	TMDS data 1+	18	TMDS data 0+
3	TMDS data 2/4 shield	11	TMDS data 1/3 shield	19	TMDS data 0/5 shield
4	*TMDS data* 4–	12	*TMDS data* 3–	20	*TMDS data* 5–
5	*TMDS data* 4+	13	*TMDS data* 3+	21	*TMDS data* 5+
6	DDC clock	14	+5 V power	22	TDMS clock shield
7	DDC data	15	+5 V ground	23	TDMS clock +
8	N.C.	16	Hot plug detect	24	TDMS clock –

When the switch is open for the data+ line, it is closed for the data– line, so current is pulled through R1, creating a nominal half-volt drop in R1. Since no current is being pulled through R2, the data+ line is at the supply voltage, nominally 3.3 volts. When the switch changes states with data, the opposite occurs. Figure 22.5(d) illustrates the waveform on data+ and data–. The voltage varies between +3.3 V and +2.8 V, with the two data lines at opposite polarity at all times. Thus, the total nominal swing between the two data lines is 1 V peak-to-peak. A differential amplifier in the receiver converts the differential signal to a single-ended signal for use within the receiver.

Since all signals are balanced with respect to ground, and assuming the two wires carrying each signal in the cable are twisted, then very little signal can radiate. There are rather wide tolerances on the voltages and currents shown, so you may measure a significantly different voltage swing in a real system. There is no absolute maximum specified length of the cable; indeed, it would depend on the clocking speed. For test purposes, the cable is to be 5 feet long. Nominal length to connect the cable between a STT and a monitor might be 6–12 feet; there is some thought that at least at lower speeds the cable could work to about 35 feet. In contrast to the IEEE 1394 interconnect, DVI is not intended for bus applications. It is not possible under the existing standard to transfer individual picture components and have the display put them together, so it is necessary for the multimedia processor to assemble the complete picture before sending it to the monitor.

Copy protection for DVI is described in Section 22.6.4.

1394a

This interface is specified for standard-definition digital displays and for recording. It is a serial bus designed, among other things, for interconnecting a cluster of entertainment devices, such as an STT, a display, a camera, a multichannel audio system, and whatever else the user wishes to connect. It is designed to transmit control information as well as program audio and video, and it can transmit various elements of the final picture separately, relying on the display to put them together. Whereas DVI is a point-to-point link intended to connect a video source to a video display, the IEEE 1394 link includes additional control features and is intended to be connected in bus form, interconnecting multiple devices. MPEG compressed video is transmitted on IEEE 1394a links, as opposed to the uncompressed video transmitted on DVI links.

IEEE 1394 is also known commercially as *FireWire*™ or *i.LINK*™. The Consumer Electronics Association (CEA) has specified compatibility standards for using IEEE 1394 in video interfaces, EIA-775-A, *DTV 1394 Interface Specification*.[8] Equipment conforming to this standard bears the logo for *DTV Link*™. Products bearing this logo also conform to the profiles of EIA/CEA-849-A, *Application Profiles for EIA-775-A Compliant DTVs*. The receiver or display has the capability of decoding an MPEG2 datastream, so the information passed over

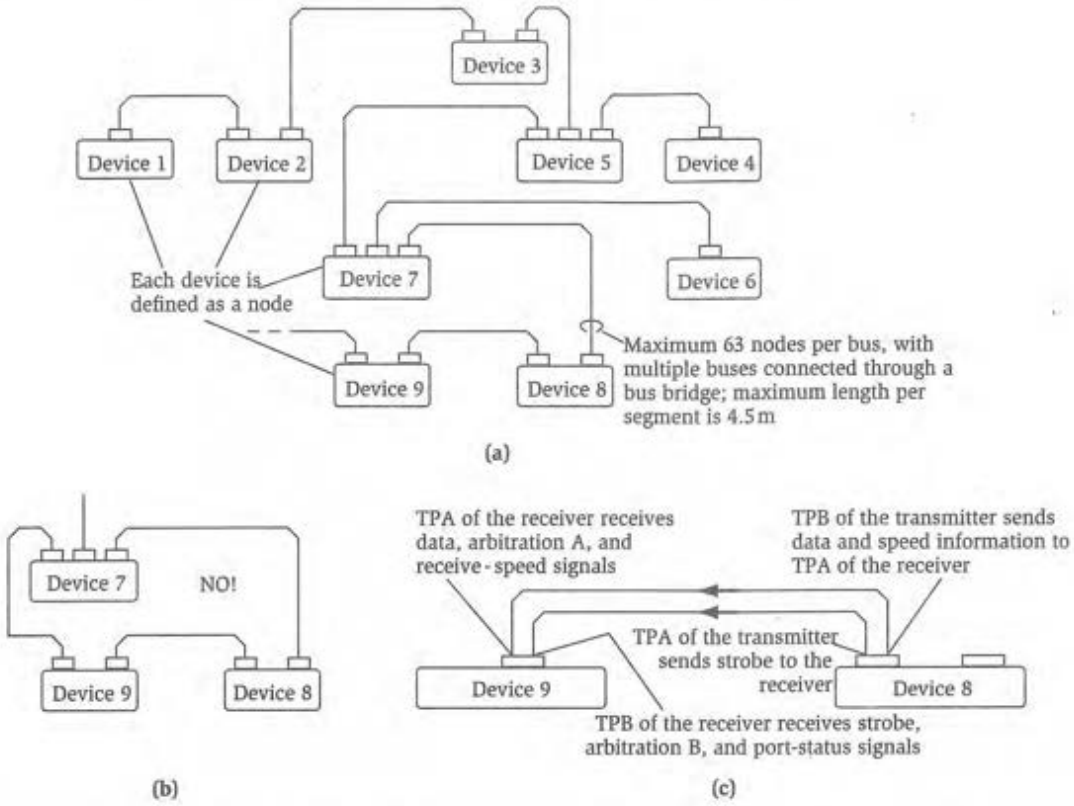


Figure 22.6 IEEE 1394 connection system. (a) Connecting multiple devices. (b) Incorrect connection. (c) Use of A and B twisted pair.

an IEEE 1394 link is not decompressed in the host. Furthermore, the host can pass a bit-mapped graphic overlay to the display, and the display will overlay it on the picture once it decompresses it. Thus, it is not necessary for the 1394-delivered picture to pass through the multimedia processor of Figure 22.1.[9] On the other hand, the multimedia processor does need to provide the graphics information to the 1394 interface so that it can be passed to the host.

Figure 22.6 illustrates several aspects of the IEEE 1394 bus as used in entertainment applications. A number of devices, called *nodes*, are connected together. The data rate is up to 400 Mb/s today, with higher data rates being developed. (Actual wire data rates are 98.304 Mb/s, 196.608 Mb/s, and 393.216 Mb/s. Conventionally the data rates are rounded to the nearest hundred Mb/s.) Termination of each connection is essential at these rates, so each device must regenerate the signal to pass it on to the next device, even if it is not otherwise involved in a

particular transaction. A maximum of 63 nodes may be connected together, as shown in Figure 22.6(a). If more nodes are desired, a bus bridge may be used to connect them. The maximum length of each cable segment is loosely specified to be 4.5 m (nearly 15 f), though this is not necessarily considered to be a limit. Advanced versions of the IEEE 1394 specification allow longer-distance operation on twisted-pair and fiber-optic cable.

Figure 22.6(b) illustrates an illegal connection of nodes. Here the three devices are interconnected by a loop, so there is more than one path for a data signal to take. This connection will cause the IEEE 1394 interconnection to fail. As long as a loop connection is not made, the system can automatically determine what devices are connected and can determine the capabilities of all other devices in the network. Connections may be daisy-chained and they may be branched, but they may not be looped.

As illustrated in Figure 22.6(c), the IEEE 1394 interface consists of two circuits that are crossed in the cable. Each circuit consists of a twisted-pair controlled-impedance cable, giving rise to the nomenclature of TPA and TPB. In the cable, TBA on one end is connected to TPB on the other, and vice versa. Both pair are used for data communication in both directions in a half-duplex communications protocol. (Half-duplex means that communication is carried in both directions but at different times.) Among other information, *data* is transmitted on TPB and received on TPA, and a *strobe* is transmitted on TPA and received on TPB. The transmission system is nonreturn-to-zero (NRZ) data in which each bit is represented by only one logic level. To aid in clocking, the *strobe* line makes a transition every time two consecutive bits are the same. This ensures one transition for every bit, which aids in clock recovery at the receiver.

Figure 22.7 illustrates the IEEE 1394 cable in more detail. The original 1394 specification, adopted in 1995, called out a six-conductor connector, which accommodated the two twisted pair (TPA and TPB), as well as a power source (8–40 Vdc at 1.5 A maximum). When 1394a was adopted in 2000, a four-pin connector was used, with the power requirement being dropped. This allowed a more compact connector, with each device responsible for its own power. Only the four-pin connector is specified in OpenCable.

Figure 22.7(a) illustrates a six-pin to four-pin cable. The shell is used for a shield connection, also called VG on the four-pin connector. This is the only ground connection between the two ends of the cable. Although balanced transmission is used, it is still necessary to maintain an approximate common ground between the two ends to prevent amplifier saturation and possibly static buildup problems. As shown in the pin function chart, TPA and TPA* (the opposite side of the balanced pair) on pins 5 and 6 of the six-pin connector are crossed to pins 1 and 2, the TPB and TPB* pins on the four-pin connector. Figure 22.7(b) illustrates the more common four-pin to four-pin connector. Note that each twisted pair is shielded, with the shields connected to the shell, VG, at each end of the cable. TPA/A* and TPB/B* are always crossed in the cable.

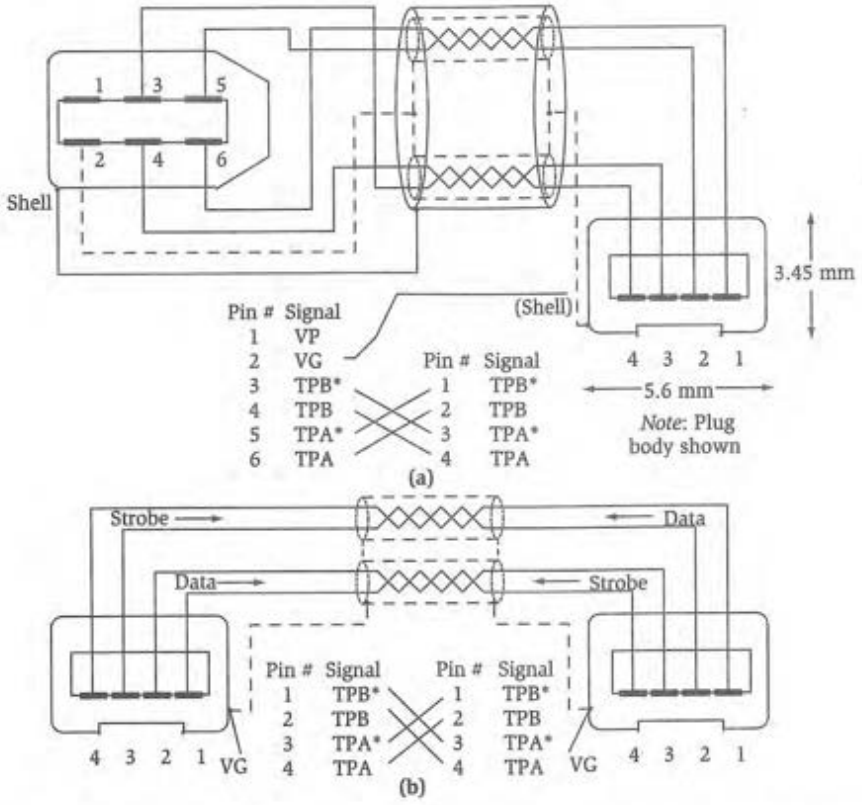


Figure 22.7 IEEE 1394 interconnections. (a) Six-wire to four-wire cable. (b) Four-wire to four-wire cable.

Figure 22.8 illustrates the electrical connections on the two ends of each twisted pair. TPA is illustrated on the left, and the mated TPB pair is on the right. Follow first the transmission from the strobe transmitter (on the left) to the strobe receiver (on the right). The differential impedance of the twisted pair is 110 ohms, or 55 ohms from either side to ground. The source impedance is set by resistors R1 and R2, the series combination of which is placed differentially across TPA and TPA*. The junction of the two is pulled to a bias voltage, T_PBias'. The strobe transmitter is a pair of switched current sources similar to those illustrated in the transmitter of Figure 22.5(c) for DVI (the current source in IEEE 1394 is closer to 2 mA than to 10 mA). By using switched current sources, the source impedance is set by R1 and R2 and is constant regardless of the state of the strobe transmitter.

The signal is transmitted from TPA/TPA* on the left to TPB/TPB* on the right over the IEEE 1394 cable. Resistors R8 and R9 form the differential

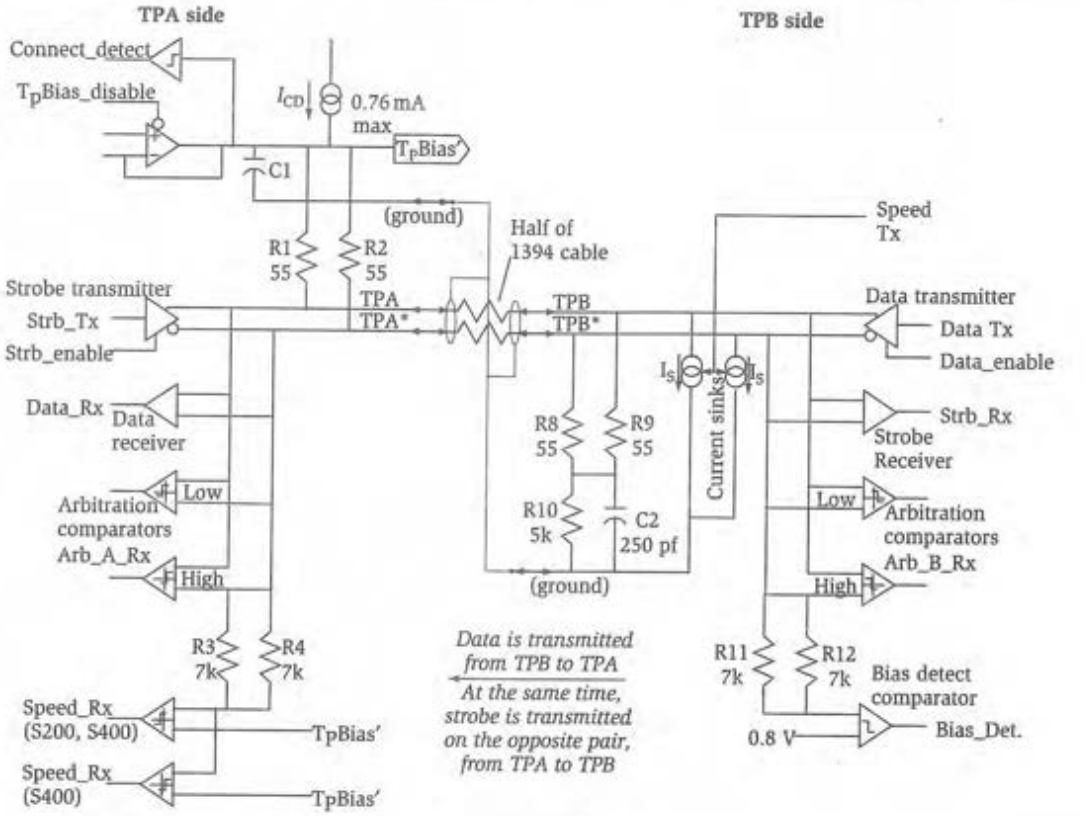


Figure 22.8 IEEE 1394 electrical circuits.

termination at the receive end. The junction of the two is pulled toward ground by R10 and C2. The strobe receiver recovers the logic-level strobe signal. It is a differential receiver, which puts out a logic 1 or a logic 0 depending on which of its two inputs is higher in voltage. The differential voltage amplitude on the signal and strobe lines is between 118 and 265 mV, depending on the speed of transmission.

Since the IEEE 1394 bus can operate at different speeds, it is necessary for two nodes to agree on the speed at which they will transmit and receive. This is done by adjusting the amount of current in the two current sinks I_S on the TPB side. If a node can operate only at 100 Mb/s, then it does not have the current sinks, and the value of I_S is 0. If the node wants to propose (or confirm) operation at 200 Mb/s, then it sets $I_S = 3.5$ mA. For 400 Mb/s operation, $I_S = 10$ mA. The current for these two current sinks comes from TPBias′ on the TPA side through R1 and R2. This current changes the common mode voltage on the twisted pair, with the common mode voltage being detected on the two speed

receivers in the lower left corner of the diagram. The common mode voltage is nominally 1.665 V for 100-Mb/s transmission and 1.438 and 1.03 V, respectively, for 200- and 400-Mb/s transmission.

The arbitration receivers detect the value of differential voltage on the pair during arbitration, the process by which nodes compete for the bus. Arbitration ensures that only one node at a time is transmitting. Upon completion of arbitration, the winning node is able to transmit a packet or initiate a short bus reset. Arbitration signaling is a protocol for the exchange of bidirectional, unclocked signals between nodes during arbitration. The differential voltage may be forced using a combination of the states of the strobe and data transmitters. Two arbitration receivers are needed to distinguish all the possible states of the connection.

The nodes must have some way to know when a new node is added; this is provided by the I_{CD} current source in the upper left. It provides a certain amount of common mode current on the bus when the T_PBias' is disabled. The current creates a voltage drop across R10, which is detected by the bias detect comparator in the lower right. This notifies the existing node that a new node has been connected to it, initiating the discovery process.[10,11]

Two classes of data communication between nodes are supported: asynchronous and isochronous. *Asynchronous* transmissions occur whenever they occur, not synchronized to any event and not at any particular interval. They are always acknowledged. *Isochronous* communications occur at regular intervals. An example of asynchronous traffic might be a control message telling a node to adjust its volume or to change a channel. An example of isochronous traffic would be a packet of video or audio information being delivered from an STT to a monitor. These packets must be delivered promptly and at a regular interval if good performance is to be maintained. IEEE 1394 provides for isochronous cycles 8,000 times a second. There is often little point in acknowledging an isochronous packet, because by the time a missed packet could be retransmitted, it is too late to use it. Isochronous cycles take priority over asynchronous traffic, but the protocol guarantees that a minimum portion of the bus time is reserved for asynchronous data transfers. An ongoing isochronous communication between nodes is referred to as a *channel*. Once a channel has been established, the requesting node is guaranteed to have the requested amount of bus time for that channel every isochronous cycle. Only one node may send data on a particular channel, but several nodes may receive the data. A single node may have multiple channels allocated, and additional channels may be added as long as isochronous capacity is available.[12]

Digital Audio—S/P DIF

Some people actually do try to make a word (spe-dif) out of this acronym. It stands for the *Sony/Philips digital interface*, a serial digital audio interface that can be used to carry either uncompressed stereo or compressed (that is, non-PCM data)

Dolby Digital (AC-3) with up to 5.1 discrete channels of surround sound. (AC-3 audio compression is described in Chapter 3.) The basic specification covering the S/P DIF connection is IEC 60958. It is a 75-ohm coaxial interface using RCA connections. Only one-way communication is supported.

The transmitter supplies 0.5 ± 20% V p–p into 75 ohms. The receiver must operate with signal levels as low as 0.2 V p–p. The signal is self-clocking at a data rate that is 64 times the *sampling rate* (the rate at which audio data is sampled). As explained in Chapter 6, in order to avoid aliasing, the audio data must be sampled at more than two times the maximum frequency of the signal. If you want to carry audio frequencies up to 20 kHz, you have to sample at greater than 40 ks/s (thousand samples per second). A common sampling rate used in CD audio is 44.1 ks/s, whereas for certain other high-quality applications, such as DVD and digital broadcast, the audio is commonly sampled at 48 ks/s, resulting in a data rate of 3.072 Mb/s on the S/P DIF interface. A sampling rate of 32 ks/s is also used for some applications. The sampling rate may be varied by up to 12.5% for varispeed applications. A self-clocking coding scheme called *biphase mark* is used in which there is a transition in level at the beginning of each bit cell. If the bit transmitted is a 1, there is an additional transition in the middle of the cell. If the bit is a 0, there is no midcell transition.*

When transmitting uncompressed two-channel audio, each frame consists of two subframes, one for left channel data and one for right channel data. Up to 24 audio bits are supported per subframe, with each subframe a total of 32 bits long. (Conventional CD audio data is limited to 16-bit word length and is sampled at 44.1 ks/s.) The additional bits are used for synchronization, auxiliary data, audio sample validity, a user bit, audio channel status, and subframe parity. Two subframes make up one frame, which thus includes one data sample of each of the left and right channels.

A two-channel Dolby Digital (AC-3) decoder will usually be available in the host in order to permit the recovery of digitally tiered audio programming. In some cases, programmers may offer discrete multichannel programming (that is, 5.1), and the host audio decoder must be capable of presenting this program either in stereo via the line level outputs or as a monophonic down-mix via the RF remodulator while the host is tuned to digital programming. This task is accomplished by down-mixing the multichannel audio program using a set of rules carried within the Dolby Digital bitstream using metadata. However, in many if not all cases, the host may default to placing Dolby Digital (AC-3) data on the digital audio interface (the data format is specified in IEC 61937) for use in home theater

* Biphase mark encoding is similar to, but not to be confused with, Manchester encoding, another self-clocking mechanism in which a transition always occurs in the center of the bit. The first half of the bit time contains the bit being transmitted; the second half of the bit time contains the complement of the bit being transmitted. Manchester encoding is described in Chapter 4.

systems. The receiving device will then convert the Dolby Digital (AC-3) encoded signal to discrete 5.1 channel audio. IEC 61937 defines the format for carrying nonlinear PCM-encoded audio bitstreams over the IEC 60958 interface. To summarize this format, the Dolby Digital (AC-3) data is formed into a sequence of data bursts, where each data burst is made up of a 64-bit burst preamble (Pa, Pb, Pc, and Pd) followed by the burst payload itself (that is, the encoded audio data). The burst preamble consists of a sync word (Pa and Pb), information about the burst payload (Pc), a bitstream number (Pc), and a burst length code (Pd). The complete data burst is then broken up into 16-bit chunks and placed into the audio payload area of the IEC 60958 subframe, in time slots 12 to 27. Both subframes are used to simultaneously carry 32 bits of data. This allows IEC 60958 to convey nonlinear PCM-encoded bitstreams such as Dolby Digital (AC-3) to home theater systems. For more complete information, refer to IEC 61937.[13]

22.5.3 Basic Concepts of Copy Protection

Before we discuss specific copy protection measures, we need to establish the several basic concepts that different copy protection methods have in common. These include the concept of encryption, public–private key pair, and a certificate. Hang in through this discussion, because after we cover this material, we'll show how these concepts are used in real host applications.

Encryption

Encryption (called *scrambling* in the context of cable TV program denial) is the process of rendering a packet of data unintelligible but capable of being returned to its intelligible, or *unencrypted*, original form by some process. Encryption uses a key or key pair known by the sender and the receiver. The keys may be identical (a *symmetrical* key system) or different (*asymmetrical* keys). A *key* in this context is a binary sequence used to encrypt another binary sequence. That is, I take so-called *plaintext*, which could, for example, be a packet of data representing a portion of a picture I am sending from a host to a display and/or a recording device. I need to encrypt the packet so that it cannot be read by another device, which might usurp the information for a purpose not intended by the owner of the content. In essence, I combine the plaintext with a key: some piece of data that renders the plaintext unable to be read (that is, encrypted). The intended receiver, maybe a video display, knows how to decrypt the packet, returning it to the original plaintext that can be displayed. That is, the intended receiver needs a *decryption key*, a piece of information used to recover the original information.

Figure 22.9 illustrates a basic encryption system. Suppose Bob wants to send a message to Alice. (Let "Bob" be the host and "Alice" be the display. For some unknown reason, the people who study encryption, cryptographers,

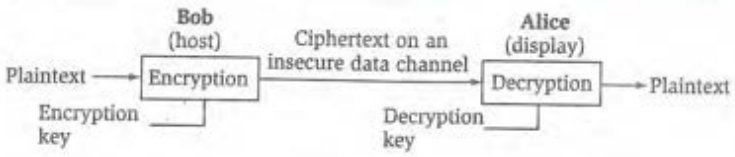


Figure 22.9 Basic data encryption.

traditionally refer to communication between Bob and Alice. We don't know what happened to Carol and Ted.) Bob has a piece of plaintext, maybe a packet of video display data, that he is ready to send to Alice, the display. But he must send it over an insecure channel, the cable between the two. This cable is insecure because it is possible for the subscriber to connect it to some pirate device, maybe to copy the program to a DVD that can be duplicated and given to people who have not paid for the program. Or the subscriber might do something else with the program that is not authorized by the content owner.

In order to allow communication over this insecure channel, the packet is encrypted before being sent to the receiver. A simple form of symmetric encryption is the use of an exclusive OR logic gate. The exclusive OR, XOR, operation is used to add two binary numbers; here it is used to encrypt data. Bit by bit, the data in the packet is presented to the XOR gate. An XOR gate puts out the input bit if the encryption key is a logic 0, and it puts out the opposite bit if the encryption key is 1. Table 22.3 shows the logical operation. When the signal reaches the receiver (Alice), the same encryption bit sequence is used to recover the original information. In order for this to happen, the sender and the receiver have to share a secret, the encryption key. Exchanging, or agreeing on, the key is a major issue in cryptography. (Note that in some of the protocols discussed later, the sender and the receiver don't have the same key, but rather they have related keys.) Somehow Bob and Alice must agree on a secret key or a pair of keys, but the only channel they have for communicating is an insecure one.

Table 22.3 XOR logic

Plaintext	*Key*	*Out*
0	0	0
0	1	1
1	0	1
1	1	0

Public–Private Keys

Exclusive OR encryption as shown in Table 22.3 is representative of simple symmetric key systems. There are also asymmetric key systems, some of which rely on an exponentiation operation rather than an exclusive OR operation. They tend to be much more computationally complex, but they yield some nice advantages, such as simplifying the key exchange problem.[14] Asymmetric keys are used in public–private key pairs.

Figure 22.10 illustrates a basic public–private key system. Suppose Bob needs to send that video packet to Alice. First Alice selects (or in some cases has stored at manufacture) a random number that only she knows. This random number is her private key. From that she can compute a public key (step 2) by applying certain prescribed mathematical operations to that private key. One method of computing a public key from a private key is given later in the description of Diffie–Hellman key exchange. The operations are such that the calculation she does is straightforward, but the inverse calculation is extremely difficult. We show a real example later in the chapter. Alice gives Bob her public key, which Bob uses to encrypt the message. Alice can decrypt it using her private key, which was not sent over the channel. It is not proved here that you can derive a public–private key pair that works this way, but since the mathematics is extremely complex, we must ask you to accept this as fact and refer you to texts that expand on the subject.[15,16]

An eavesdropper who learns Alice's public key still cannot decrypt the message because the keys being used are asymmetrical: Alice's public key encrypted the message, but it cannot decrypt it. Only Alice knows her private key, the only one that can decrypt the message. So the message is relatively safe being passed on the same insecure data channel used to send the encrypted data.

Either the private key can be used for encryption and the public key for decryption or the other way around: the public key can be used for encryption

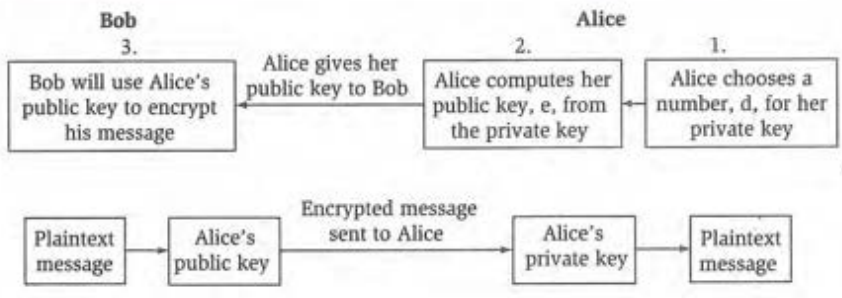


Figure 22.10 Public–private key exchange.

and the private for decryption. The same key cannot be used for both encryption and decryption.

Digital Signatures and Certificates

A problem remains. How does Bob know for sure that it is Alice who is sending her public key and not an interloper? For example, how does Bob, the host, know he is not sending video to a pirate box that invented a public–private key pair? After all, the method of calculating the public key from the private key is public knowledge. Maybe "Alice" is a pirate box that will decrypt the picture and record a DVD. This possibility is avoided by the use of digital certificates. There are many, but a common one is the X.509 certificate, promulgated by the International Telecommunications Union (ITU).

The first important concept in certification is not technical: There must be a trusted party to issue certificates. That party, called a *certificate authority* (CA) is a third party that both parties to the data exchange have decided to trust. There are companies set up to be this CA, and there are organizations that act as their own CA for specific purposes. If Bob and Alice wish to communicate, their manufacturers must get them certificates from a CA. Bob cannot send pictures to Alice unless he knows for sure that "she" is a legitimate display who will not usurp the video for a nefarious purpose. Alice cannot get a certificate until the CA has determined, using whatever steps it deems appropriate, that Alice is legitimate and will not violate the usage conditions established by the content provider.

So Alice's manufacturer must go to the appropriate CA and receive a digital certificate before he can build Alice. The digital certificate is then stored in Alice upon manufacture. Each and every copy of Alice, that is, each and every display that is built, will have its own certificate individually prepared by the CA. Alice's manufacturer sends data for the certificate to the CA, which encrypts portions with its private key and returns the finished certificate to Alice's manufacturer. Fortunately, once the credibility of Alice's manufacturer has been manually established, the certificates may be constructed automatically.

Figure 22.11 shows the current version, 3, of the X.509 certificate, the information stored in Alice, provided by the CA. Even Alice's manufacturer cannot construct the certificate, because it involves the CA's private key. The certificate must be individually prepared by the CA based on information supplied by Alice's manufacturer. The certificate is a public document. It does not include enough information for anyone to decrypt data. Rather, it is used only for the purpose of identifying the intended receiver of information and also to convey Alice's public key.

The *version* section of the certificate identifies the format of the certificate (version 3 is current as this is written). The *serial number* is a number unique within the CA by which it can identify the certificate. The *algorithm identifier* conveys the encryption algorithm used by the CA (often the RSA algorithm,

Version	Certificate version
Serial number	Unique number within the CA
Algorithm identifier: - Algorithm - Parameters	Identifies algorithm used to sign the certificate and any needed parameters
Issuer	Name of the CA
Period of validity: - Not before (date) - Not after (date)	Certificate is good only between these dates
Subject	Name of the user
Subject's public key: - Algorithm - Parameters - Public key	Algorithm name and necessary parameters, plus the public key
Issuer unique ID	
Subject unique ID	
Extension	
Signature	CA's signature

Figure 22.11 X.509 certificate.

named for its inventors, Rivest, Shamir, and Adleman, is used). There are some public parameters used in specifying the public key from the private key, and these are included with the algorithm identifier. (In early versions of the standard, this field is called the *signature*, but this is a misnomer.)

The next section of the certificate is the name of the CA. The *period of validity* specifies the beginning and ending dates that the certificate is good. Outside of these dates, the certificate is invalid and may not be used. *Subject* is the name of the user. The next section bears the subject's public key. It names the algorithm being used (often RSA) and any parameters associated with the public key. Finally, this section provides the public key that will be used by the sender to encrypt data.

The *issuer unique ID*, the *subject unique ID*, and the *extension* were recently added and may or may not be used. These are fields that may be used by the issuer or user as desired. The last section, the *signature*, is very important. The signature is the subject's (Alice in our example) public key plus other certificate data, encrypted with the CA's private key. This is the reason that the CA must provide individual certificates for each and every device made by a manufacturer. No one, and we mean no one, except the CA must know the CA's private key. If this private key is ever compromised, then all certificates issued by the CA are rendered invalid.[17]

During an initial discovery process, Alice sends her certificate to Bob. Bob must know the CA's public key. He can know it by virtue of its having been programmed into him during his manufacture or by obtaining it later. Bob decrypts Alice's public key and other information from the signature, using the CA's public key. He compares the public key and other information with the same data sent as plaintext in the signature. If the two sets of information match, then he must be talking to a legitimate Alice, because a pirate would not be able to get a certificate from the CA. Bob now has Alice's public key, so he can send data to her, encrypted with her public key.

The certificate is public knowledge. Can an interloper steal Alice's certificate? Yes. It wouldn't do any good, because it does not contain her private key, but the interloper could steal it. As an additional verification that there is a legitimate "Alice" out there, Bob can issue a challenge. He sends Alice a random number called a *nonce*, which she encrypts with her private key and sends back to Bob. Bob decrypts it with Alice's public key, which he received as part of her certificate. If the decrypted number matches the number Bob originally sent, he knows that the real Alice is out there and not someone who happens to have her certificate.

Can Bob be absolutely sure he is talking only to Alice and not also to an eavesdropper? No, but it doesn't matter. There could be a so-called *man in the middle*. For example, a man-in-the-middle attack could happen if the subscriber has connected a legitimate host (Bob) to an illegitimate device designed to capture the video on its way to Alice. The illegitimate device might then be connected to the real Alice, the legitimate display. The man in the middle passes data between Bob and Alice, but he eavesdrops, or hears all that is going on. When Alice sends her certificate to Bob, the man in the middle passes it through to Bob, who recognizes it as Alice's legitimate certificate. But it will do the man in the middle no good to have the certificate. He still doesn't have Alice's private key, which he would need in order to decrypt what Bob sends to Alice. Because Bob got Alice's public key in the certificate, he knows that he will be encrypting data for a legitimate display, regardless of who else is out there.

So Bob knows that the real Alice is out there because no one except the real Alice would have her certificate *and* be able to pass the challenge he sent. He knows the certificate is real because he was able to match information contained therein using the CA's public key to decrypt part of the certificate and to match it to the same information sent unencrypted. Only the real CA could have signed the certificate. And both sides have placed their trust in the CA. Bob can safely send data to Alice, knowing that the real Alice is receiving it. If anyone else is out there, the data will be useless to him.

In some instances the CA is designated by the originator of the standard utilizing the certificate. For example, for DVI, HDCP is the encryption used. It specifies that Digital Content Protection, LLC, an activity set up by the Digital Display Working Group that defined the specification, is the CA.[18] Anyone desiring to make a device that either sends video to or receives video from a

device conforming to the DVI specification must apply to them for licensing. For the IEEE 1394 interface, licensing of the DTCP copy protection algorithm and certification is provided by the Digital Transmission Licensing Administrator, DTLA.[19]

In the present application, a certain CA is usually specified. However, in the more general case it is possible to have two devices that were certified by different CAs. Certification may still take place through the use of a chain of certificates.

Certificate Revocation

From time to time it is possible that some device possessing a valid certificate may need to have its certificate revoked. This might happen if the CA's private key becomes known or, perhaps more likely, if it is discovered that the device has been modified into a pirate device. Many systems have a mechanism for disseminating to devices that they should not honor a certain certificate. One method of disseminating the information is that as soon as the breach of security is recognized, future content will have embedded in it the identity of the certificate that must be revoked. As that content is disseminated to different devices (hosts, legitimate digital VCRs, etc.), each remembers the revoked certificate and passes that information to devices to which it is, or becomes, attached. In this way, sooner or later most devices that may connect to the device with the revoked certificate will learn of the revocation.[20] In some cases, discussed later, the host must check with a database to confirm the validity of a certificate.

Key Length, Security, and Computational Complexity

One attack on an encryption key is called an *exhaustive search*. Encryption keys are simply long strings of 1s and 0s. In an exhaustive search, a would-be pirate simply tries all possible key combinations (sets of 1s and 0s) until the output makes sense. Often this can be automated, because incorrect keys yield results that can be distinguished, by computer, from logical results. The longer the key, the more difficult is an exhaustive search. If a key has a length (number of binary characters) of n, then there are 2^n combinations that must be tried. A short key might have only 24 bits, so there are a maximum of $2^{24} = 16.78$ million combinations that might work. If a computer can try 1 million keys per second, then the maximum time required to crack the encryption is 16.78 million divided by 1 million, or about 17 seconds. On the other hand, a 56-bit key could take 2300 years to crack at the same rate. Keys are used that have up to 256 bits, which could take 3.7×10^{63} years to break at the stated rate of trying keys. Average times, though, would only be half of these numbers.

Sometimes *churning keys* are used to improve security. This technique simply means that the key is changed fairly often. Each time a key is changed,

the pirate would have to begin again to break it. So a pirate may capture a scrambled program, but to unscramble it he would have to break a series of keys that might be changed every few seconds to every few minutes.

Symmetric Algorithms

Many asymmetric algorithms, such as the RSA algorithm used in a number of systems, involve taking a block of plaintext (treated as a binary number) to the power of the key, using some modulus. The computational power required as the key length increases becomes prohibitive for encrypting and decrypting video streams. Computational power may be traded off for time; but when you are transferring digital video, whether compressed or not, you don't have a lot of time to process the signal. Thus, many times a good asymmetric algorithm, such as RSA, will be used for certificate verification and/or key exchange, which are only done periodically. Then a symmetric algorithm will be used for actual data encryption. In the DTLA nomenclature, this symmetric key is called the *content channel key*. The key may be long, but the way it is used reduces the computational complexity to something manageable at the requisite speed.

Often the RSA algorithm is used for certificate validation. Then a symmetric key is agreed upon for the actual information scrambling and descrambling. The symmetric key is used in some manner to generate the actual encryption string, which in some encryption techniques is exclusive ORed with the plaintext to create the encrypted data, or ciphertext.

Figure 22.12 illustrates a simple four-bit shift register used to generate ciphertext.[21] Real shift registers are a lot longer, and multiple shift registers may be used. The shift register is clocked at the same rate as the plaintext data is clocked. Each time the shift register is clocked, the bits in each position shift one place to the right. (The clock is a signal that synchronizes operations dealing

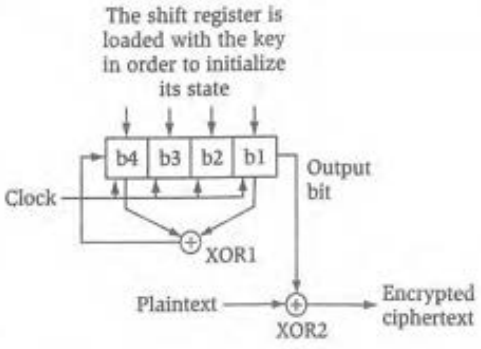


Figure 22.12 Linear feedback shift register used to generate ciphertext.

with delivery of the video.) The shift register is a series of flip-flops. On a clock cycle, the bit (1 or 0) in b4 is shifted to b3, the bit in b3 is shifted to b2, and so on. The bit in b1 is supplied to XOR2, which exclusive ORs it with the plaintext bit presented on the same clock cycle.

This type of shift register is called a *linear feedback shift register* (LFSR). The linear feedback is supplied by XOR1. It XORs the output of two bits, in this case, b4 and b1, feeding the result back as the input to the shift register, which is loaded into b4 on the following clock cycle. The bit sequence (a pseudo-random stream) obtained from a properly designed n-bit LFSR is $2^n - 1$ states (clock cycles) long before it repeats. It can be initialized with the secret symmetric key, which loads the starting state into the LFSR. A listener will not be able to duplicate the LFSR initialization without knowing the key used to initialize it.

Many other methods of generating the encrypting code are known. A different LFSR may be used for each bit of a parallel encryption system. The LFSRs may be initialized with the same key or different portions of the same key. Nonlinear generators may be used, in which AND functions combine the outputs of various stages to form the output. The outputs of several LFSRs may be combined to form the encryption signal.

Diffie-Hellman Key Exchange Protocol

Though by no means the only useful key exchange protocol, the Diffie–Hellman (D–H) protocol is used in quite a few systems. D–H permits both parties to an exchange (Bob and Alice) to supply part of the secret key. It is done in such a way that the entire key is not sent across the unsecured channel. Thus, a snooper is not going to gain the information needed to steal the secret key. The protocol proceeds as follows.

1. Alice chooses a random large integer x and sends Bob $X = g^x \bmod n$, where g and n are D–H parameters that must be passed between the two parties. For example, suppose D–H is used to agree on a key in an X.509 certificate. In the "subject's public key" portion of Figure 22.11, g and n would be passed as parameters, and D–H would be identified as the algorithm. X is the public key. The parameter n must be large and must meet certain criteria, but g may be a single-digit number.

   These parameters are not secret, and an eavesdropper will not know anything useful if he learns them. In some closed systems, g and n may be preprogrammed during manufacturing. The parameter that must be kept secret is x, and, as we shall see, Alice never divulges it in any form. The modulus, n, represents the maximum value the result can take (plus 1). It is the remainder after a division by the modulus. To calculate $X = g^x \bmod n$, take the remainder from the operation g^x/n.

2. Bob chooses a random large integer y and sends Alice $Y = g^y \bmod n$.
3. Alice computes $k = Y^x \bmod n$. Recall that Y was sent to her by Bob and that x is a secret parameter that only Alice knows.
4. Bob computes $k' = X^y \bmod n$. He got X from Alice and knows y. Now, from the foregoing (and not writing the modulus), $k' = X^y = g^{xy}$. Also, $k = Y^x = g^{xy}$, so k and k' are equal. Thus, Alice and Bob have the same secret key, and at no time did all the information required to know the secret key pass over the insecure link.

$k(= k')$ is the symmetrical key applied to the LFSR in foregoing. It is possible to derive a D–H key among three or more parties by expanding on the preceding process.[22]

In order for an eavesdropper to learn the secret key, he would have to derive x and y from X and Y, the two public keys that are passed over the link. Computing x and y is not feasible. And without them, knowing k is not possible. One might attempt to do an exhaustive search, but it will take many lifetimes to do an exhaustive search if x and y are large enough.

DES Encryption Algorithm

There is yet one more encryption algorithm we need to discuss, because it is used in handling digital video. This is the data encryption standard, or DES (usually pronounced "dez"). DES is based on work done at IBM in the early 1970s. The National Institute of Standards and Technology (NIST, formerly NBS, the National Bureau of Standards) established it as a national standard in 1976. The American National Standards Institute approved DES as a private-sector standard in 1981 (ANSI X3.92). It has been reaffirmed as a U.S. standard several times since then, though it is now no longer approved for federal information applications, having been replaced by AES in 2001. Just because DES is no longer the approved algorithm does not mean that it is broken, just that it has been replaced for federal information applications. Special-purpose DES chips are available.

DES is a 56-bit-key symmetrical algorithm, meaning the same key is used for encryption and decryption. Though complex, it uses only simple operations, such as exclusive OR, permutation (changing the order of bits), and shifting of bits. There are four modes of operation: electronic code book (ECB), cipher block chaining (CBC), output feedback (OFB), and cipher feedback (CFB). We shall describe the complete algorithm, though only a portion of it is used in some scrambling systems.

Figure 22.13 illustrates the operations performed on the data. DES is a block cipher; it operates on a block of 64 bits of data, sends that data on, and starts working on a new block of 64 bits. If a block has fewer than 64 bits, additional bits are added (*bit stuffing*) to bring the total to 64 bits. An object of DES is to

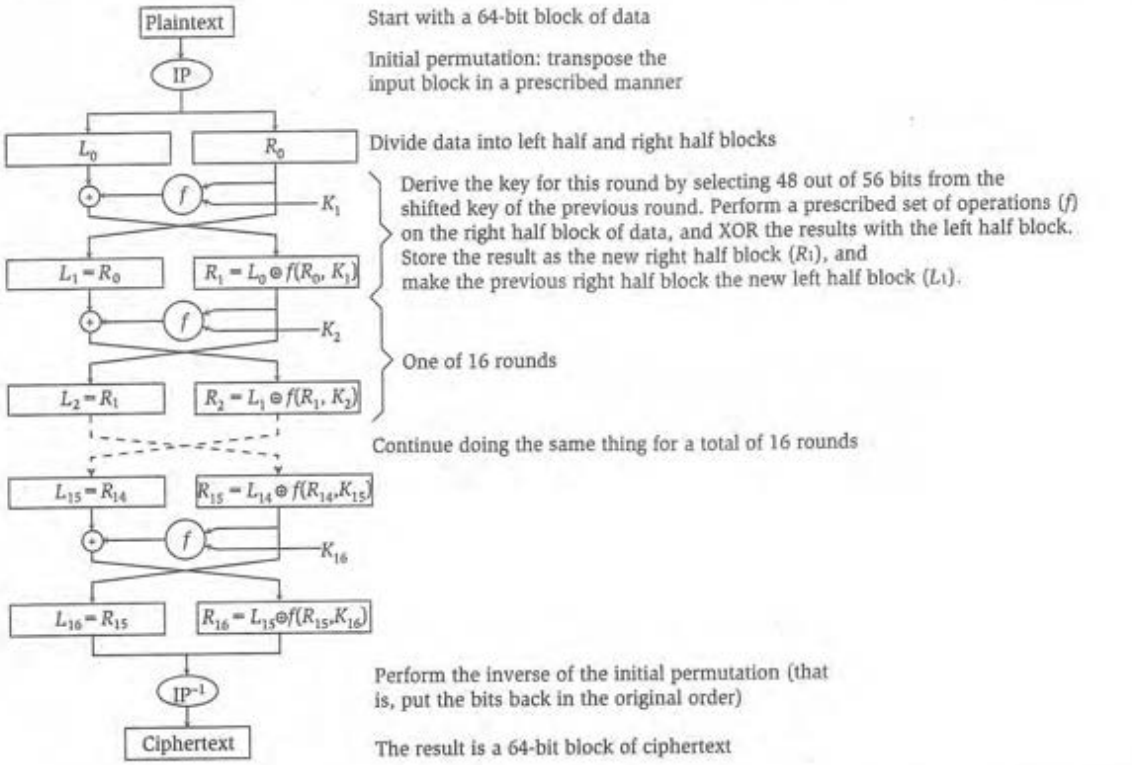


Figure 22.13 Operation of DES.

make sure that every bit in the block of ciphertext depends on every bit of the input and every bit of the key. The algorithm is divided into 16 nearly identical *rounds*, each involving all the bits of the data. Only one key is used, but, as we shall see, the bits in the key are changed for each round.

As shown at the top of Figure 22.13, DES starts by performing an *initial permutation* on the block of plaintext. This simply means that the order of the bits is changed in a prescribed manner. The 58th bit becomes the first bit, the 50th bit becomes the second bit, the 42nd bit becomes the third bit, and so on (the referenced text for this section shows all of the permutations involved). The initial permutation has nothing to do with the security of the algorithm. As best anyone can tell today, its primary purpose was to make it easier to load plaintext into a DES chip in byte-sized pieces. Next, the data is divided into a 32-bit left half and a 32-bit right half. Each half will be operated on separately in each round; at the conclusion of each round, the left and right halves are swapped. The left half is referred to as L_n and the right half as R_n, where n is the round number.

For each round, the right half is operated on by a function *f*, which we shall define shorthy. Function *f* involves expanding the data from 32 to 48 bits by duplicating certain bits, XORing the expanded data with 48 bits derived from the key, and performing certain other operations. The result is then XORed with the left half of the bits and the result is stored as the right half for the next round. The right half is stored as the left half for the next round.

After 16 rounds, an inverse of the initial permutation is performed; the resulting 64-bit block is the ciphertext that is transmitted. There is defined a triple DES, in which this algorithm is performed three times using two or three different keys.

Figure 22.14 shows what goes on during each round of DES. On the right we show what happens to the key during each round, and on the left we show what happens to the data during that round. The input key is the key from the previous round or, for the first round, is the agreed-upon key. The actual key is 56 bits long, but it is often expressed as a 64-bit number. The difference is the inclusion of

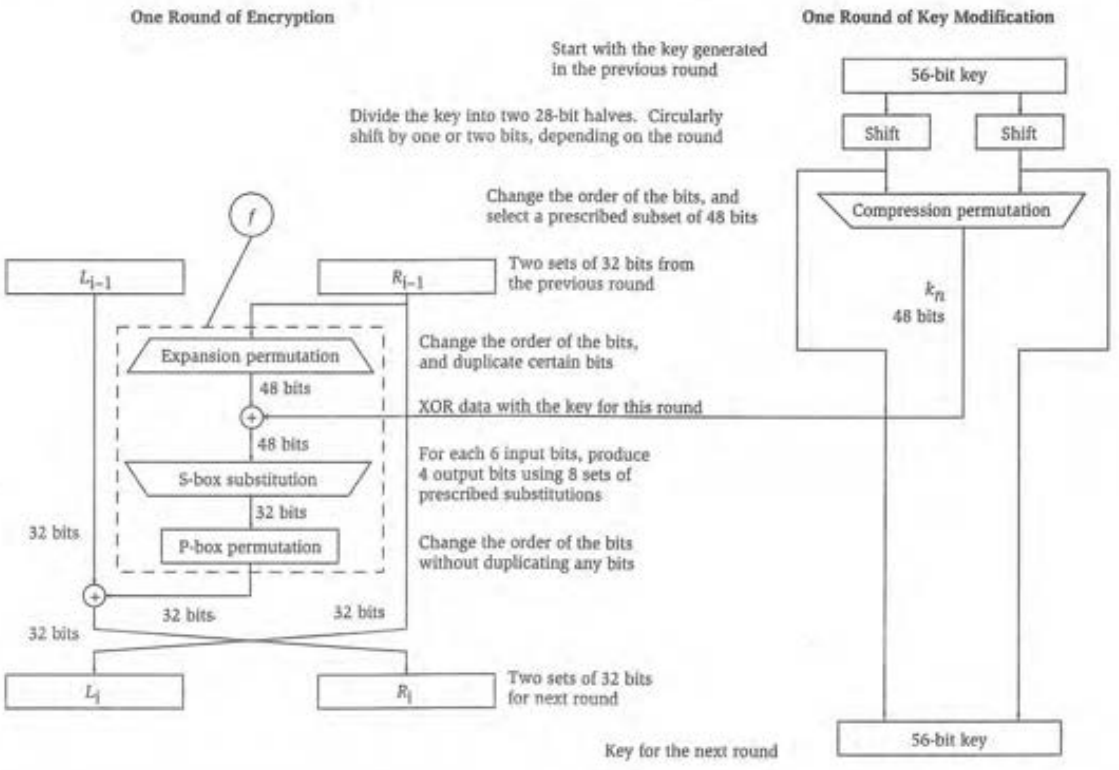


Figure 22.14 Operations performed in one round of DES.

8 parity bits, each derived from 7 bits of the key. The key is divided into halves, and each half is shifted by one or two bits depending on the round. The shifting is done circularly, meaning that bits are shifted in a shift register, with the bit shifted out of one end being shifted into the other end of the shift register.

Next a compression permutation is performed. *Permutation* means that the order of bits is changed according to a prescribed ordering. *Compression* means that certain bits are removed from the key to bring the number of bits remaining down to 48. Again, the bits removed are carefully prescribed. The 48 bits taken from this operation make up the key used in the round. The shifted halves of the key before the compression permutation become the 56-bit key for the next round.

The data encryption performed in this round is shown on the left. The right half block of data (R_{i-1}) becomes the left half for the following round (L_i). Also, the right half is subjected to a set of operations that define the function *f*. What goes on in function *f* is shown inside the dashed lines. The first operation in *f* is an expansion permutation, in which certain bits are repeated to bring the 32 bits up to 48 bits (the same length as the key for the round), and the order of bits is changed. The output of the expansion permutation is XORed with the key.

The next operation is the so-called S-box substitution. A *substitution* is an early form of a cipher. A simple substitution might be to build a table of substitute letters, as shown in Table 22.4. In this simple example, A becomes Z, B becomes P, and so on through the alphabet. A person intercepting such an encrypted message would not be able to make sense out of it unless he had the decoding table. This is the principle behind the S-box substitution in DES. Recall that the input to the S-box substitution is 48 bits, but only 32 of those bits carry the data. The other bits are discarded simultaneously with the substitution. For each six input bits, four output bits are produced according to a prescribed set of eight substitution tables following the same principle as Table 22.4 (the inputs and outputs are sets of bits rather than letters).

The output of the S-box substitution is the input to the P-box permutation. This is a simple permutation of the order of the bits, according to yet another prescribed ordering. The output of the P-box permutation is the output of the function *f* shown in Figure 22.13. It is XORed with the left half bits and placed in the right half block for the next round of the permutation.

Table 22.4 Substitution cipher

In	*Out*
A	Z
B	P
C	J
D	A

The set of actions depicted in Figure 22.14 is repeated through the 16th round, after which the inverse of the initial permutation is performed, as shown in Figure 22.13. The output of the inverse initial permutation is the 64 bits of ciphertext, which are ready to be transmitted, and a new block of 64 bits enters the encryption as the process starts over.

In order to decrypt the data, the same steps are repeated, with the only difference being that the same key is used but the order of processing the key is reversed. Note that we have not proved DES works, nor have we proved the security of DES. For those wishing to study the subject and having several years of free time on their hands, many books have been written on the subject.[23]

22.5.4 Digital Video Copy Protection

Having established the basics of encryption, we can now describe the several copy protection methods used in host devices. Both DVI and the IEEE 1394 interface have associated copy protection mechanisms, and a third protects content coming out of a POD module. (Yet another encryption is applied to the scrambled data transmitted over the cable plant, to be decrypted on the POD module.)

DVI and HDCP

This interface is intended to operate between a host and a display only. Recording the signal is not an appropriate use of the interface. The interface, as described earlier, transfers decoded red, green, and blue pixel information, each serially on three balanced lines. HDCP, the copy protection mechanism, is governed by the specification *high-bandwidth digital copy protection system*, currently revision 1.0, 17 February 2000.[24] Implementation of the specification requires a license from Digital Content Protection, LLC. The system allows for use of *repeaters*, that is, devices that take in the signal and also put it out. A repeater might be a digital video distribution amplifier, or it might be a monitor that supplies an output to another monitor. It is possible to have seven levels of repeaters and as many as 128 devices in total. When repeaters are used, the source (host in our terminology) is made aware of all downstream devices before transmissions commence.

There are three elements of the content protection system. The first is *authentication*, in which the host verifies that all devices connected are legitimate, licensed devices. Shared secrets are established during authentication and form the encryption key for *program transmission*, the second element. Finally, *renewability* allows the host to identify a compromised device and to desist from sending data to it.

Authentication, the first element of the content protection system, also precedes in three steps. It is done slightly differently than what we showed earlier. Each licensed device is given an array of 40 secret keys, each 56 bits

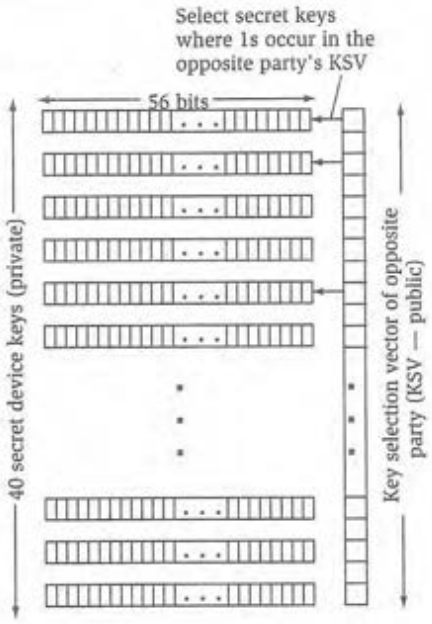


Figure 22.15 HDCP keys.

long. These come from Digital Content Protection, LLC. The device also receives a *key selection vector*, KSV. All are programmed when the device is manufactured. Figure 22.15 shows the secret keys and the KSV of the *opposite* party. The KSV is functionally equivalent to a public key, and the 40 secret keys are just that: secret keys. Authentication is initiated by the host sending an initiation message to the display. This message includes the host's KSV and a 64-bit pseudo-random number, *An*. The receiver responds with a message including its KSV and a repeater bit if the device is a repeater. The host checks the KSV to make sure it has not been revoked and that it includes 20 1s and 20 0s.

A shared secret is derived as follows, in the first step of authentication. From its secret keys, the host selects the ones corresponding to 1s in the receiver's KSV. All selected secret keys are added together, modulo (mod) 56, producing a 56-bit key. Call this result *km* in the host and *km'* in the receiver. The secret keys and the KSVs are chosen such that the host and the receiver will compute $km = km'$, so each end will have the same secret key. If either device has an invalid set of secret keys or a bad KSV, then the two keys will not be the same.

At this point, both devices have several pieces of information in common. They know *km*, they know the random number generated at the Host, *An*, and they know the repeater bit. These values are used to generate three values, *Ks*, M_0,

and R_0. These are generated by a block cipher algorithm defined in the specification. *Ks* is a 56-bit session key. M_0 is a 64-bit secret value used in the second part of the authorization protocol. R_0 is a 16-bit response value that the video receiver returns to the host as in indication as to the success of the authentication exchange. It must be returned within 100 ms. If R_0 returned from the receiver matches that generated by the host, then authentication was successful.

In the second step in the authentication process, the host (transmitter) and any repeaters gather information regarding all the devices that are attached. The host confirms that none of the devices attached are on the revocation list. This step is omitted if there are no repeaters (for instance, there is only one host sending pictures to one display). The step can take up to about five seconds to complete. Finally, the third step occurs in the vertical blanking interval (VBI) about every two seconds while pictures are being transferred. New cipher initialization values are calculated to establish a new 56-bit key for encrypting video content. The integrity of the link is also confirmed.

The second element of the HDCP protocol is encryption of the red, green, and blue data using a pseudo-random sequence generated from secrets exchanged in the authentication process. The pseudo-random sequence is reset frequently. The pseudo-random sequence is exclusive ORed with the data, as shown earlier.

The third and final element of the HDCP protocol is renewability, the process of revocation of devices that have been compromised. The devices are identified by their 40-bit KSV, as described in Figure 22.15. *System renewability messages* (SRMs) containing one or more KSV revocation lists, called *vector revocation lists* (VRLs), are distributed with programs as described above. They are checked by the transmitter (host) when they are available. The SRM carries a digital signature, as described earlier, so that the host can determine that the data did indeed come from Digital Content Protection, LLC.

IEEE 1394 and DTCP

The *digital transmission content protection* (DTCP) standard for secure transmission over IEEE 1394 physical links was developed by five companies — Intel, Matsushita (Panasonic), Toshiba, Sony, and Hitachi — and is thus sometimes called the "5C" content protection scheme. 1394 is intended as a broader-use standard than is DVI, being intended not only for transmission to a display or displays, but also for transfer to recording devices, computers, and whatever other devices you wish. DVI transfers uncompressed (hence very wideband) red, green, and blue data for immediate display; IEEE 1394 transmits compressed (therefore lower bandwidth) data.

DTCP, the associated verification and encryption standard, includes the same elements as does HDCP, with one more. It includes *authentication and key exchange* (AKE), transmission, referred to as *content encryption*, and renewability. It also includes a new element, *copy control information*.

DTCP includes instructions to the receiving device as to how it should treat the data. This is called *copy control information*, CCI, which is embedded in the content. The actual data in the CCI is known as the *encryption mode indicator* (EMI). It may take on any of four states. The material may be used for unlimited copying, copy free, and the opposite state, copy never. Another state is copy once, such as to allow for time shifting a TV program. When the material is played, the EMI is changed to copy no more. These last two values of the EMI are intended for devices with limited computational ability. The associated limited, or restricted, authentication required involves fewer computations than does full authentication.

Obviously, the same concerns apply here as to other systems. Every device licensed to receive material protected by DTCP must obey these rules. Thus, the same need exists for a trusted authority to certify each device. DTCP is administered by the Digital Transmission Licensing Administrator (DTLA), established by the five companies. The DTLA acts as the certificate authority (CA) for conformant devices. Besides the CCI, DTCP includes authentication and key exchange, content encryption, and renewability.

Two levels of authentication and key exchange are specified: Full authentication using public key cryptography as described earlier is used for all types of content. For copy once and copy no more content, a more limited, restricted authentication using common key cryptography is permitted.

For full authentication, the method is similar to, but uses different formats than, that described in Section 22.5.3.[25] A certificate is embedded in each device and is verified as described earlier. Both the source device (supplying the content) and the sink device (receiving the content) check the other device's certificate. A form of Diffie–Hellman key exchange is used for derivation of a secret key used to encrypt program material. Restricted authentication may use a short form of authentication, where the sink device proves that it shares a secret with the source device.[26]

After authentication, the source device (the host) sends an exchange key, encrypted with the authentication key, to the sink device. This exchange key is sent as a seed, a random number generated by the source, from which both devices can compute the decryption key. The key is changed periodically, every 30–120 seconds.

In order to provide for system renewability, *system renewability messages* (SRMs) are sent to each IEEE 1394-compliant device. These may be sent as part of a program stream from a headend, or they may be incorporated into other distribution media, such as DVDs. The SRMs carry identification of any device that has had its certificate revoked. As material containing the identification propagates throughout the nation, devices eventually learn of any certificate revocations.

POD Module Output

The POD module shown in Figure 22.1 is responsible for descrambling the output of the transmission from the headend. POD, an acronym for *point of*

deployment, is the term used for renewable security modules. The POD module is built into a standard PCMCIA (*Personal Computer Memory Card International Association*) card, the familiar card used for modems, Ethernet connections, and many other added functions in portable computers. It uses a 68-pin connector. The POD module output is also copy protected. We can easily explain the first stage of the copy protection algorithm used at this interface. It is the X.509 certificate and encryption algorithm described in Section 22.5.3. The certificate exchange is, however, just the first phase of having the POD module and the host authenticate each other.

In phase 2 of the authentication process, the host and the POD module identify themselves to the headend. The headend holds the DTLA (see preceding subsection) certification revocation list (CRL), showing the identification of all hosts and POD modules that have had their certification revoked. Note the contrast with the general case we showed earlier for revocation of certificates. Previously we showed the CRL being contained with the content. But POD modules and hosts identify themselves back to the headend, which then checks its CRL and will refuse to authorize descrambling if either device is on the CRL.

Identification to the headend is done differently depending on whether a two-way communications path is available and or unavailable. If a communications path is unavailable, then the host displays a message on the monitor, telling the subscriber to call a number and to give the identification numbers of the POD module and the host. The identification numbers are numerical, so it is possible to automate the process of relaying the information to the headend. Also, when the subscriber purchases his or her equipment, it is possible for the seller to perform identification. If two-way communications are available, the POD module can relay the information back to the headend conditional authorization system without subscriber action.

The third phase of authentication consists of computation of a long-term *authentication key* by both the POD module and the host. This authentication key is stored in nonvolatile memory (that is, the content remains even when the device is not powered). It is used to confirm authentication after power has been removed and is reapplied. Upon a subsequent power-up, the POD module and the host confirm that they have the same authentication key; if so, no further action is required before they can begin exchanging programming information.

The complete process of authentication is referred to as a *binding* between the POD module and the host. A POD module may bind to only one host at a time, but the POD module may be moved from one host to another, with the authentication process starting all over when that happens; that is, the POD module will restart the binding procedure. In two-way systems, the process is unseen by the subscriber. But in a one-way system, the subscriber will have to telephone the POD module and host identification numbers to the service provider. Regardless of successful authentication, programs sent in the clear (no scrambling) may be processed.[27]

22.6 Using Two Conditional Access Systems

Operators consider it desirable to be able to use more than one STT system in order to gain possible price and feature advantages. However, each manufacturer would like to use his own conditional access system, to which he ascribes unique benefits. A simplistic approach is to duplicate programming using two scrambling systems. However, this approach doubles bandwidth requirements, an obvious impracticality. In 1997 the so-called *harmony agreement* was negotiated in the industry, whereby one encrypted datastream could be used along with more than one conditional access system.

Besides addressing scrambling, the harmony agreement includes an agreement on the use of open standards for transmitting digital video. These include MPEG-2 video, Dolby Digital audio, MPEG-2 transport, ATSC system information, and ITU-J83B modulation (64- and 256-QAM).[28]

Figure 22.16 illustrates the principle of this harmony agreement. The agreement starts in the headend with clear video (called *plaintext*). This plaintext is encrypted in the common service encryptor, which scrambles the signal using an agreed-upon algorithm, the DES encryption standard described earlier in this chapter. A common (shared) control word, or encryption key, is generated in the control word generator and supplied to the common service encryptor and also to the conditional access system for each vendor whose STTs are being used. The *service encryptor* is the device that scrambles the program data. It is called "common" because the same scrambling is applied for use in both systems.

The control word is also supplied to each vendor's conditional access system, which then distributes that control word as it sees fit. The control word may be encrypted for distribution according to the individual manufacturer's proprietary facility. Control words are embedded in so-called entitlement control messages (ECMs), which are transmitted as part of the MPEG transport stream.[29]

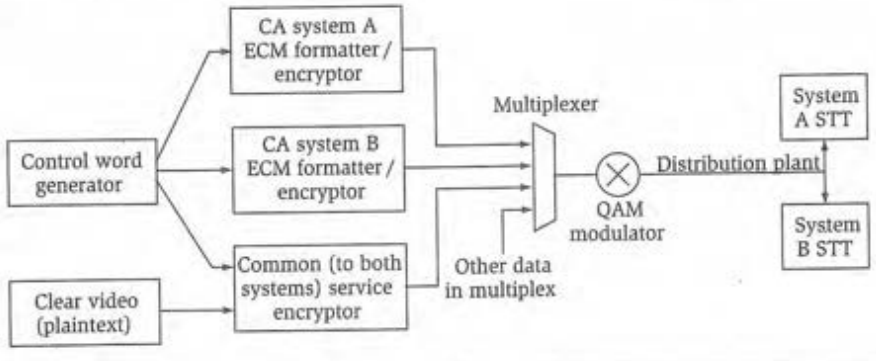


Figure 22.16 Principle of the harmony agreement.

As this book goes to press, another means of sharing conditional access has been demonstrated but not yet commercialized. This method extracts critical data from the MPEG-2 datastream. This data is required for reconstructed video. The critical data is encrypted using any number of encryption methods desired by different vendors. The remaining portion of the MPEG signal is not encrypted. The decoder extracts the critical data encrypted with its encryption system and combines it with the majority of the video data that was not encrypted.

Other means of encrypting the signal are available and are practiced in Europe and other localities.

22.7 Optional Personal Video Recorder

An optional personal video recorder (PVR) function is shown in Figure 22.1. PVR capability allows a subscriber to record and play back material from a hard drive. He or she can use it for time shifting programs and also to generate an "instant replay" of a scene. If the manufacturer wants to allow the PVR to be used with analog signals, the manufacturer must include the optional real-time MPEG encoder shown in Figure 22.1.

Of concern if the subscriber records a program in a PVR, is the possibility of exchanging the program with someone else, either by transmitting the program on to the other person or by physically removing the hard drive from the host and putting it in another host or in a computer. For this reason, the content recorded on the hard drive must also be encrypted, unless it is marked for unlimited recording. As of this writing, an encryption standard has not been specified for this interface. The host manufacturer chooses one.

22.8 Middleware

A modern STT or host is as much software as hardware. In the beginning, STTs were limited to the manufacturer's software. Several proprietary sets of so-called middleware were developed; then standards began to emerge for certain levels of software. In particular, it is important to allow downloaded programs that will run on any STT to which they are downloaded so that an operator or subscriber may select from any of several STTs on the market. This can be achieved by a standardized *middleware* solution. The CableLabs/SCTE-standardized middleware is known as *OpenCable applications platform* (OCAP). As of this writing, two OCAP specifications exist, the OCAP 1 and 2 profiles.[30] OCAP is based on an earlier European initiative called *Multimedia Home Platform*, MHP.[31] There are a lot of similarities between OCAP, MHP, and DASE, the ATSC's *DTV Applications Software Environment*.[32] As of this writing, there are efforts to harmonize DASE, used in North American digital off-air broadcast, and OCAP, used in cable TV.

OCAP 1 closely follows MHP 1.0.2, but there are differences, some of which we will mention in this description.

22.8.1 Layers of Software

You may think of at least three layers of software running on the microprocessor(s) in a host. The lowest layer is the operating system, just as a desktop computer has an operating system. It is specific to the hardware being used and is typically purchased from a third-party vendor by the host manufacturer for the processor(s) chosen to be used. (This is consistent with the way OCAP is explained in the standards, but note that in some cases there is a layer of software below the operating system, call the *hardware abstraction layer* (HAL). The HAL allows the operating system to be platform independent.) The highest layer of software is the application software, which may, for example, provide an EPG in the form that the operator wishes to display it. Many other applications software packages are possible, from communications to games to messaging to VOD and others. Usually, application software is downloaded to a particular host based on what the user purchases and what the operator wants to provide. Application software is usually downloaded from a central server and may change from time to time.

The problem is that the operator needs to be able to download one software package for one application and expect it to run the same way on all hosts and STTs in the system. This must obtain even if there are different STTs and hosts made by many different manufacturers at different times, using different processors and operating systems, and having different capabilities. If a capability is supported by the hardware in a box, then the same software should enable it in all boxes. Making this happen is the role of *middleware*. The host manufacturer is responsible for providing middleware for his products that conforms to the appropriate OCAP specification, 1.0 or 2.0, as specified by the cable operator. It is possible to use a host that supports 1.0 but not 2.0. Each specification details the services to be supported and how relevant data is communicated to the host.

22.8.2 OCAP

OCAP 1 is an application interface that includes the *application program interfaces* (APIs) necessary to implement certain specified applications. It is intended to define a minimum software capability for OCAP-certified hosts sold at retail. Software providers may assume the existence of this level of capability when developing host applications. API is the set of routines, protocols, and tools for building software applications. A good API makes it easier to develop a program by providing all the building blocks. An application programmer puts the blocks together.[33] Thus, OCAP provides the programmer with all the tools needed to put together a certain application, which may then be downloaded to all STTs

and hosts, without regard to who made the device. The performance will meet some minimum standard of consistency on all devices.

OCAP 1 is based on a Java virtual machine model, a popular standard for the so-called *execution engine*. That is, you write Java programs to run on the host. Related terms you will encounter include *Hypertext Markup Language* (HTML), the authoring language used in creating Web pages. *Hypertext Transfer Protocol* (HTTP) defines how messages are formatted and transmitted to Web browsers and what actions Web servers and browsers should take in response to various commands. HTML is not a part of OCAP 1 (it is part of OCAP 2), but it is possible to use Java to write a Web display application if desired. Java can leverage the graphics capabilities supported by the hardware.

All applications in MHP are *bound*. That is, they are tied to the channel to which the host is tuned. If the viewer tunes off the channel, then the application stops unless it is also associated with the newly tuned channel. On the other hand, OCAP recognizes both bound and *unbound* applications. An unbound application continues to run even when the channel is changed.

Besides bound and unbound applications, and unlike MHP, OCAP allows native applications. Any function that can be programmed in Java (meaning just about any function) may be programmed and downloaded to an OCAP host. However, unlike bound and unbound applications, native applications are not guaranteed to run on any host. Native applications are written for a particular host.

Because it will not always be the case that the cable operator can control what applications are downloaded to a host, OCAP includes a *monitor application*, which decides what other applications are allowed to run on the host. This way, the operator has some control of what the subscriber does with his host and what the host might do to the network. An application can be downloaded on the FAT, the high-bandwidth channels on which video is downloaded. But the FAT channel is often passed through the headend without the operator's making any changes, opening up the system for other applications forwarded by the program supplier.

Figure 22.17 illustrates the relationship between OCAP 1 and other software and hardware in the host.[34] At the lowest layer is the OCAP host device hardware, what we traditionally think of as the host or STT. This is the collection of hardware on which the software runs. The lowest layer of software is the operating system/middleware layer. This includes a lot of traditional operating system functions, such as scheduling different activities, handling interrupts (such as when the subscriber presses a key), software to handle different interfaces, and memory management.

The second layer of the software is the execution engine, whose functionality and upward-bound interfaces are specified in OCAP 1. It provides prescribed interfaces with the application software above, through OCAP APIs. The upward-bound interfaces must be device independent. That is, any host, implemented with any software operating system, must respond in the same way when an application issues some command or supplies some piece of data.

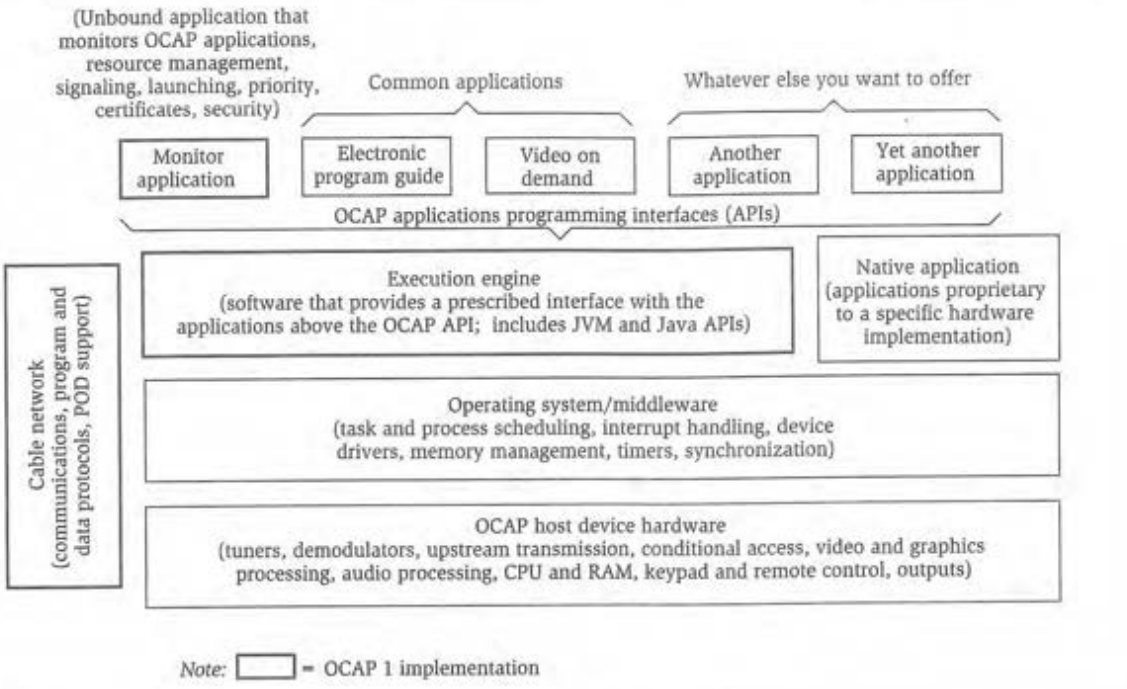


Figure 22.17 OCAP 1 software architecture.

Above the execution engine are the various applications, which might come from the host vendor but which might also come from other vendors. Common applications include EPG and VOD, but many other applications are possible. Some examples are listed later. The monitor application is included in OCAP 1. It controls what other applications may reside on the host, and it controls resource management between the applications and other functions that involve one or more applications. This is typically operator specific, because each operator will want to make his or her own decision about what runs on the hosts in the system.

The cable network is generally defined in SCTE 40, formerly DVS/313. It defines the levels and characteristics of the input and return signals interfacing on the left of Figure 22.1.

Native applications are any applications that run on a specific host and are not conformant to OCAP. Such applications may be written when the execution speed of an OCAP application is not fast enough (some games fall into this category). If an operator wants to enable such applications on several different hosts, he or she may have to write the application several times, once for each different host (assuming they are not compatible). In contrast, an OCAP application need be written only once, and it will run on all OCAP 1–compliant platforms.

OCAP 1 Representative Applications

A number of representative applications are presented, which OCAP 1-compliant systems are required to support. Typically these applications are written on top of OCAP 1 (using its services). This is not an exhaustive list — other applications can also be supported. There is no requirement that all of these applications be present on a particular host, but it is required that the box support software that implements these applications as appropriate. An OCAP 1–compliant box will be capable of having application software downloaded to support these services, but it is not a given that, just because a box is OCAP compliant, these applications exist on it. They exist when client software is downloaded. Also, every client might not have the hardware required to support every application listed.

Electronic program guide: Electronic program guide (EPG) software must be capable of being written using OCAP APIs, but OCAP does not specify a particular EPG. This is up to the software chosen by the operator. EPGs typically offer a grid showing programs available on the system at the present time and for some length of time into the future. The user can move through the grid, with current location being highlighted. If he or she presses a "SELECT" or "OK" key when on a particular program listing, the host tunes to that program. If the program is not on yet, the EPG may set the host to tune to that channel when the program does come on, or it may present a reminder to the viewer that the program is starting. The EPG may also be used for purchasing pay services such as pay-per-view (PPV) programs, which are presented at preset times and billed as an extra service. It can also be used for purchasing video on demand (VOD), a similar service except that the subscriber can order it to start when he or she wants it and can exercise VCR-like control (stop, play, fast-forward, rewind).

Watching TV: Though this might not be thought of as a software function, there are certain aspects of watching TV that might be controlled by applications written on middleware, and OCAP must support them. These applications support such functions as selecting standard- or high-definition viewing and, if in high definition, the format (aspect ratio, scan format, and resolution). The subscriber would want to control whether audio is delivered as mono, stereo, or 5.1-channel surround (based of course on what formats are transmitted). He or she might also want to select an alternate language or to select closed captioning information to be overlaid on the picture. Picture-in-picture, time shifting, emergency alert, and content advisory ("V-chip") are other services related to watching TV. Software to implement these services must be supported by OCAP.

Pay-per-view: Pay-per-view (PPV) is a traditional way of ordering pay programming. The subscriber contacts the system operator in advance to arrange for purchase of a particular event. The operator then downloads

authorization to the host. At the time of the event, the EPG may provide a reminder to the subscriber. Related services for delivery of pay programming include Impulse Pay-Per-View (IPPV), video on demand (VOD), and near video on demand (NVOD). These are described in Section 21.4. Additional services may be provided using the EPG capability of the host.

Email: Email clients are supported by OCAP. This would be an example of an unbound application, because it is not tied to any one broadcast channel. The operator is free to download an email client of his or her choice to an OCAP-compliant host.

Chat/conferencing: A chat or conferencing application would most likely be another unbound application, available no matter to what channel the subscriber was tuned. It might appear as a floating box over the video, in which the viewer and others could send instant messages to one or more other participants to discuss a show or anything else people can think of to chat about.

IP telephony: OCAP 1 specifically lists IP telephony support, though the host may not have the interfaces required for telephone service. IP telephony is covered in Chapter 6.

Games: OCAP 1 provides certain APIs required for game playing. These include windowed video and playing sound effects and animation. OCAP does not provide support for multiplayer networked games, but this capability can be provided in native application software. Some games may require faster response than provided under OCAP 1, in which case the game client may include native-code interfaces to support certain hardware.

Music/radio: OCAP provides for decoding audio without decoding associated video to permit you to transmit digital audio broadcasts. If you want to download audio in any other format, this could be done by downloading a software decoder. An audio decoding application would most likely be bound, because when the subscriber tunes away from the channel, the audio would stop.

E-commerce: These applications are usually broken into shopping and banking applications. Shopping benefits from high-quality full-screen graphics and flashy transitions and other effects to allow a merchant to present wares in the most pleasing way. Shopping also involves exchange of credit card information, so security is of importance. OCAP does not support a secure connection, leaving this to the application software. The operator must address privacy concerns. Banking services don't require the same high-quality video, but they do require extremely good security and user identification. Again, the application software must provide these services.

Minimum Functionality

OCAP also specifies certain minimal functionality that any host must support. Any OCAP host must support background, video, and graphics resolution of at

least 480 × 640 pixels, which is roughly equivalent to the monochrome resolution of an NTSC signal with a 4:3 aspect ratio (4/3 × 480 = 640). This resolution must be supported for both 4:3 and 16:9 aspect ratios.

The minimum host must support certain decoder format conversions (DFCs). That is, when the signal arrives in one display format, the host must support conversion to another format supported by the display device. For example, if a picture arrives in a wide-screen (16 × 9) format and the display cannot handle wide screen, the host must be able to convert to a conventional 4:3 format. There are several ways this is to be done. A central portion of the picture may be sent to the display, or a letterbox format (black bands above and below the picture) may be sent. A number of specific processing options are required, as specified in the MHP documentation.[35] In addition, and different from MHP, an OCAP 1-compliant host must support a JMF (Java Media Framework) player for presenting subtitling or closed captioning information.

There is a rather lengthy list of remote control commands that are to be accepted. Support for both keyboards and remote control transmitters is recommended. A certain minimum amount of memory and the recommended architecture of that memory are included in the OCAP 1 specification. STTs and other hosts usually are running even when the user perceives power to be off. This is a standby mode, and OCAP recommends that the middleware recognize the difference between "standby" and "on" modes, because this is important to some applications.

22.8.3 OCAP 2

OCAP 2 is a second, related specification that builds on OCAP 1. Besides the functionality built into OCAP 1, OCAP 2 adds content formats based on Web technologies, such as XML, DOM, and ECMAScript. It is possible to provide complete functionality in an OCAP 1 device, but OCAP 2 opens additional options for providing that functionality.

Extensible Markup Language (XML) was developed by the World Wide Web Consortium,[36] an organization that develops technologies to expand the potential of the Web. XML allows designers to create their own customized tags, enabling the definition, transmission, validation, and interpretation of data between applications and between organizations.

The document object model (DOM) is the specification for how objects in a Web page are represented. The DOM defines what attributes are associated with each object and how the objects and attributes can be manipulated.

The European Computer Manufacturers Association (ECMA) has created a standard (ECMA-335) that allows applications to be written in a variety of high-level programming languages and executed in different system environments.[37]

22.9 Summary

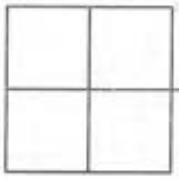
In this chapter we have presented a view of modern techniques used for digital video reception in the home. As this is written, many of the concepts and standards are new and have not been completely implemented in commercial equipment. The standard now is so-called *thin client* STTs, having limited functionality within themselves and depending on the headend for many functions. *Thick client* STTs, which include more processing power, have been used. The standards we have summarized can cover both. These standards are intended to support a wide range of functionality, to provide the cable operator with a rich choice of models through which to sell advanced services to subscribers. Of great importance is satisfying video content owners that their content will not be pirated. The standards anticipate the availability of devices besides STTs. For example, it is possible to subsume security needs into TV sets by using POD modules provided by the operator. Time will tell how the market for advanced services progresses.

In the next chapter, we discuss interfaces between consumer and cable TV equipment.

Endnotes

1. Cable Television Laboratories, *OpenCable™ Host Device Core Functional Requirements*, OC-SP-HOST-CFR-I11-021126, November 26, 2002. As of this writing, the standard is available at http://www.opencable.com.
2. www.scte.org.
3. Charles E. Spurgeon, *Ethernet: The Definitive Guide*. Boolder, CO: Cablelubs 2000.
4. Charles Poynton, *HDTV Video Short Course*, tutorial presented at the 2002 IEEE Consumer Electronics Society's International Conference on Consumer Electronics.
5. http://www.ddwg.org.
6. http://www.webopedia.com.
7. Another good reference is Luke Chang and Joe Goodart, *Digital Visual Interface*, May 2000. Available at www.dell.com — search for DVI.
8. Available from http://global.ihs.com.
9. Mark Eyer, EIA/CEA 775-A: The IEEE 1394–Based Digital Interface for DTV Receivers, *IEEE Consumer Electronics Society Newsletter*, August 2002, No. 3.
10. IEEE 1394-1995, *IEEE Standard for a High Performance Serial Bus*, available from www.ieee.org.
11. IEEE 1394a-2000, *IEEE Standard for a High Performance Serial Bus*, Amendment to IEEE 1394-1995, available from www.ieee.org.

12. Adam Kunzman and Alan Wetzel, 1394 High-Performance Serial Bus: The Digital Interface for ATV, *IEEE Transactions on Consumer Electronics*, Vol. 41, No. 3, August 1995.
13. Information for this section was provided by Jeffrey Riedmiller of Dolby Laboratories.
14. One asymmetric algorithm, RSA, is described in Justin Junkus, *Digipoints*, Vol. 2. San Francisco: SCTE, 2000, p. 169.
15. Bruce Schneier, *Applied Cryptography*. New York: Wiley, 1996.
16. Larry L. Peterson and Bruce S. Davie, *Computer Networks* San Francisco: Morgan Kaufmann, 1996.
17. Stephen Thomas, *SSL and TLS Essentials*. New York: Wiley, 2000, chap. 2.
18. http://www.digital-cp.com.
19. http://www.dtcp.com.
20. Bill Pearson, *Digital Transmission Content Protection* (slides), Intel Corp. 1999. Available at http://www.dtcp.com. Look for the DTCP tutorial.
21. Bruce Schneier, op. cit., section 16.2.
22. Bruce Schneier, op. cit., section 22.1.
23. Bruce Schneier, op. cit., chap. 12.
24. http://www.digital-cp.com.
25. The cryptographic schemes, primitives, and encoding methods are described in IEEE P1363.
26. DTLA, *Digital Transmission Content Protection White Paper*. Available at http://www.dtcp.com.
27. SCTE, *ANSI/SCTE 41 2001, POD Copy Protection System*. Available at www.scte.org.
28. Justin Junkus, *DigiPoints*, Vol. 2. Exton, PA: Society of Cable Telecommunications Engineers, 2000, p. 166.
29. Michael Adams and Tony Wasilewski, Multiple Conditional Access Systems, *Communications Technology*, November 1997.
30. See http://www.opencable.com. At some point these specifications may be transferred to the SCTE Web site.
31. See http://www.mhp.org.
32. See http://www.atsc.org.
33. http://www.webopedia.com.
34. OpenCable™ Application Platform Specification OC-SP-OCAP1.0-I04-021028.
35. ETSI TS 101 812 V1.2.1 (2002–06).
36. http://www.w3.org.
37. http://www.webopedia.com.



Chapter 23
Consumer Electronics Interface

23.1 Introduction

Cable services are enjoyed via the subscriber's consumer electronics products, the most important of which are television receivers and videocassette recorders. Stereo and surround-sound systems are also important to home theater installations. Coordinating the use of all the features of these products with cable services causes complications. Add-on devices such as video games, Web appliances, digital video disc (DVD) players, personal computers, home security and control systems, and other consumer electronics products add still more challenges. The cable–consumer electronics interface issues are complex, and satisfactory solutions have been elusive even with extensive committee work and cooperation between the industries. This chapter will survey the issues and some of the attempts to alleviate the problem.

There are two principal aspects to the issues concerning the interface between consumer electronics products and cable systems. One aspect is the way in which this interface affects the usage of cable services and of consumer electronics products and their features. The other aspect involves matters of law and politics. Both aspects will be considered, but separately.

Five major constituents have interests in these issues: (1) the subscriber/consumer, (2) the service provider, (3) the service provider's equipment manufacturer, (4) the consumer electronics and related industries, and (5) government agencies creating laws and rules related to these issues.

23.2 The Nature of the Interface

23.2.1 Cable Services

For purposes of this discussion, cable services can be categorized as (1) basic, (2) expanded basic, (3) pay, (4) pay per view, and (5) other.

Basic services are not scrambled as a matter of law. This is a requirement of the 1992 Cable Act, which we discuss later. Signal delivery is accomplished

merely by connection — usually by a technician physically connecting the drop cable to the cable plant. In a few cases, a switched tap is used to connect the subscriber under electronic control without the need for a technician to visit the tap location. Signal denial for those who no longer subscribe is done by disconnecting the subscriber from the cable plant. The subscriber who takes only the basic service and has a consumer electronics product that can tune all the channels of interest experiences the same amount of complexity as found when connected to an antenna.[1]

A converter with appropriate tuning capability is used in situations where (1) the subscriber's consumer electronics product does not tune all the channels of interest, (2) the product suffers from inadequate tuner performance, or (3) the product does not include remote control and the subscriber desires that feature. A converter is an extended capability tuner with a channel 3 or 4 output. It usually also includes a remote control. Converters were once commonly leased by cable operators. That is much less common today because converters are rarely needed since few very old receivers are still in service. Converters are available at retail.

Expanded basic services involve additional channels and an additional price. Signal denial for nonsubscribers is usually accomplished by traps or by scrambling. Alternative signal denial possibilities include interdiction. Interdiction has not found wide acceptance. All these technologies are described in the previous two chapters. When traps or interdiction is used, the subscriber experiences the same amount of complexity as found when the TV and VCR are merely connected to an antenna in the usual way recommended by the TV and VCR manufacturers. This assumes that the consumer electronics products tune all the channels of interest; otherwise, a converter is required.

Pay services, also called premium services, are sold by the month on a channel-by-channel basis. Pay services feature movies and special events and require a substantial additional fee. Analog signal denial for nonsubscribers is usually accomplished with traps for services with stable acceptance, that is, that have subscribers who keep their subscriptions for extended periods. There are two major types of traps: negative and positive. Negative traps allow subscribers to the premium service to receive it "in the clear" without any signal processing. With this technique, basic subscribers who do not take popular pay services have the negative trap installed in their connection to the cable system. The negative trap filters out the channel that is not taken. Popular services are those where more than half of basic subscribers take the service. Positive traps are installed at premium service subscriber locations when less than half of the basic subscribers take the service. A factor in making the choice between positive traps and negative traps is the need to minimize the number of units required. Alternative signal denial methods include interdiction and broadband descrambling. Conventional descrambling is used for services with higher rates of turnover or where additional security is desired. When traps, interdiction, or broadband descrambling is used, the subscriber experiences the same amount

of complexity as when the TV and VCR are connected to an antenna (assuming the consumer electronics products tune all the channels of interest).

Pay-per-view services are sold one at a time and almost always utilize an addressable descrambler. In the past, disposable traps that timed-out after the pay-per-view event were used, but with only moderate success.

Other services are generally taken by more sophisticated subscribers who have little or no difficulty with the interface between their equipment and the cable system.

23.2.2 Relevant Cable Set-Top Terminal Features

Cable set top terminals have a couple of important features that have been designed specifically to help with the interface between cable systems and consumer electronics products.[2]

The *auto bypass switch* is a feature that bypasses the entire cable spectrum to the TV (or VCR) when the set top terminal is turned off. This is an option that either is built into set top terminals or is available as an add-on. It is intended to allow the TV's (or VCR's) remote control to operate the TV's (or VCR's) tuner when the set top terminal is not used for scrambled channels. Though the auto bypass switch was available before the 1992 Cable Act's Consumer Electronics Interface amendment, the act motivated wider availability of the feature. Cable operators are required to make these switches available to subscribers but are allowed to charge for their costs. This feature has had little consumer demand.

The *dual descrambler* feature puts two tuners and descramblers in one package so that subscribers are not limited to unscrambled channels for some of the modes of operation of consumer electronics products. Cable operators are required to make these units available to subscribers and are allowed to charge for their costs. This feature also has had very little consumer demand.

The cable industry and independent entrepreneurs have developed a number of switching appliances to facilitate configuration of the interconnections for various purposes. Other devices intended to ameliorate these difficulties have been offered for sale. None of them has found long-lasting marketplace acceptance.

23.2.3 Relevant Consumer Electronics Characteristics and Features

TVs and VCRs are the primary consumer electronics equipment affected by cable service. These products come with a wide variety of features. The most important features for these issues relate to the remote control and the tuner. The remote control is used to turn the device on and off, change channels, and operate other features. Another feature of the VCR is its timer, which is used to commence and terminate recording and to change channels at appropriate times to capture programs of interest. Other TV and VCR features of interest include parental control, last channel tuning, sleep timers, and favorite channel access.[3]

Premium TV receivers include a feature called *picture-in-picture*, (PIP), which allows a small picture to be inserted into the large picture. Most often, the signal for the second picture comes from another source such as the tuner in a VCR or a television camera that might be placed in a baby's room or at the door to monitor visitors. Often the two signals may be interchanged with the press of a button. Another variation of the PIP feature includes the scanning of the channels with the capture of a set of miniature still frames providing a glimpse of what is on the various channels.

Most modern television receivers and VCRs have a feature, called *auto channel setup*, that automatically searches for available channels when the unit is first plugged in (and after every extensive power outage). The TV or VCR takes several minutes to search all the channels it can tune and loads into memory those that contain programming. Many (but not all) of these receivers will skip over most scrambled channels. When the viewer uses the channel up or down button, these channels will be skipped.

The *antenna and cable/converter switch* feature (usually operated by remote control) switches between two or more inputs to the TV. One input is intended to be from the converter and the other directly from the cable (or an antenna). The intention here is to allow the TV's tuner to do the work when nonscrambled channels are watched and just use the converter/descrambler's tuner when scrambled channels are watched. This feature is similar in function and purpose to the auto bypass switch option in converter/descramblers.

The *TV/VCR switch* feature on a VCR determines whether the entire cable spectrum is presented to the TV or just the channel tuned by the VCR or the tape playback.

23.2.4 Consumer Frustration

The two principal aspects of consumer frustration are the installation and the utilization of consumer electronics products with cable services. Both of these frustrations come from the variety of services and features found in these products.

23.3 Connecting Cable, TVs, VCRs, and Set-Top Terminals

In 1987, the NCTA Engineering Committee formed a Subcommittee on Consumer Interconnection and developed a set of 27 diagrams indicating the most likely methods of connecting a TV, VCR, and one or more set-top terminals along with an assortment of switches and signal splitters. Many of these diagrams are also found in consumer's manuals that come with new TVs and VCRs. All the interconnection schemes have advantages and disadvantages. None solves all of the problems. Reviewing some of these diagrams results in a better understanding of the complexity of the problem.[4]

When the original diagrams were developed, most TVs had 300-ohm balanced RF inputs, whereas today they have 75-ohm coaxial inputs. In the 300-ohm balanced input, you had to use a *balanced-to-unbalanced* transformer (balun) with a 4:1 impedance step-down ratio. This is not necessary with 75-ohm coaxial inputs, but we have retained the balun in Figure 23.1 since it is in the original document.

Where a VCR is shown, a *personal video recorder*, PVR, may be substituted. A PVR has an internal hard drive and saves programs on that rather than on a tape. Operation of a PVR is somewhat more problematical than that of a VCR because a PVR has added intelligence to allow it to select programs for later viewing, based on the subscriber's past actions. Having to impose an STT for tuning disrupts operation of this feature. Some STTs include PVR functions, so this becomes a nonissue in those cases.

Where digital services are to be received, an STT is required unless the TV has a cable interface to receive 64- and 256-QAM broadcasts directly without an STT. If so, then the STT may be omitted for nonscrambled digital programs. If the TV has a POD module interface as described later and in Chapter 22, it may be used without an STT in all cases. As of this writing, such TVs are not available. If the PVR has a POD module, it can be used on all channels without the STT. Again, such devices are not available as of this writing.

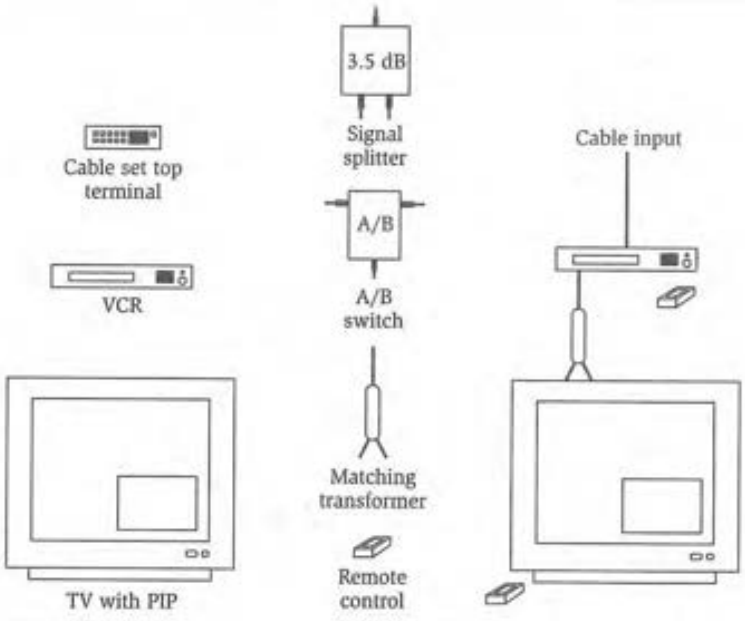


Figure 23.1 In-home equipment and simplest connection without set-top terminal.

Network-based PVRs may be implemented by the cable operator if desired, in which case broadcasts are saved on a hard drive at the headend and played out to subscribers as a VOD program. These would need an STT with appropriate software to control the network-based PVR.

23.3.1 Basic Cable and Trapped Premium Service Connection

Figure 23.1 is illustration 12 of the NCTA connection diagram set. The cable is connected to the VCR, which is connected to the TV. This diagram is for the basic-service-only subscriber and the subscriber who has premium services controlled by traps. No set-top terminal is used. Since less than half of cable subscribers have set-top terminals and since most subscribers connect more than one TV or VCR to cable, the majority of cable connections use this simple configuration.

Advantages

1. Least expensive.
2. TV and VCR operate as with antenna connections.
3. Can consecutively record different unscrambled channels automatically.
4. Can watch one unscrambled channel while recording a different unscrambled channel.
5. PIP works with unscrambled channels when VCR's baseband output is connected to TV's baseband input.

Disadvantages

1. No scrambled services available for either watching or recording.
2. No interactive services available.
3. Have to remember to put TV to channel 3 when watching recorded tapes.
4. Need to use two remote controls or a "universal" remote control.
5. TV's parental control is inoperative when using VCR's tuner.
6. Consumer's tuner subject to overload and damage from lightning strikes on cable.
7. Signal leakage possible from inadequately shielded tuners.

23.3.2 A Simple Start

Figure 23.2 is illustration 1 of the NCTA connection diagram set. The cable is connected to the set-top terminal, which is connected to the VCR, which is connected to the TV. This is by far the most popular connection method

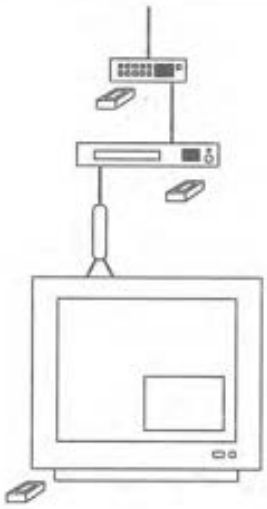


Figure 23.2 Most common cable connection with set-top terminal.

involving a set-top terminal because it is the least expensive and simplest to implement and use.

Advantages

1. Inexpensive and easy to install.
2. Easy to understand.
3. Can record scrambled channels.

Disadvantages

1. Can't watch one channel and record a different channel.
2. Can't consecutively tape different channels automatically because VCR tuner does not do channel selection.
3. Neither the TV tuner nor the VCR tuner has any control.
   (a) TV's parental control, favorite channels, and last channel recall features are inoperative. The TV is just tuned to channel 3 (or 4).
   (b) VCR's timer, parental control, favorite channels, and last channel recall features are inoperative. The VCR is just tuned to channel 3 (or 4).
4. Need to use three remote controls or a universal remote control to turn on the three devices and put the TV and VCR to channel 3 (or 4).
5. Makes the TV's picture-in-picture feature useless.

Special Case

1. If the converter/descrambler has an auto bypass switch, then when the converter is turned off, the entire cable spectrum passes through to the VCR, and the situation is exactly the same as in Figure 23.1, which is the case where no set-top terminal is used. When again using a scrambled channel, the subscriber must remember to put the VCR to channel 3 (or 4). If the set-top terminal has the channel-mapping feature and the TV or VCR does not, the same programming may appear on two different channel numbers. One channel number will be found on the set-top terminal and the other on the TV or VCR. This may be confusing.

23.3.3 Two Set-Top Terminal Case

Figure 23.3 is illustration 2 of the NCTA connection diagram set. The cable signal is split and connected to two set-top terminals. The left set-top terminal is connected to a VCR, which is connected to one input of an A/B switch, which is connected to the TV. The right set-top terminal is connected to the other input of the A/B switch.

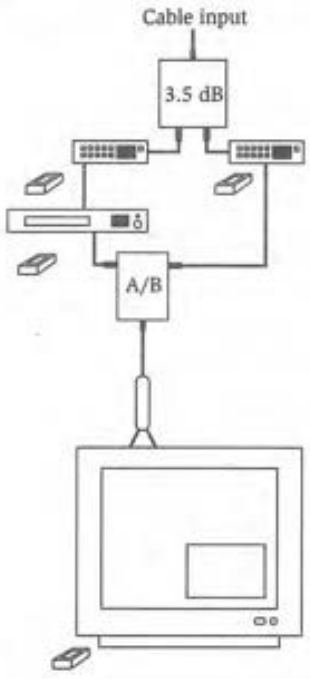


Figure 23.3 Connection with two set-top terminals.

This is one of the least popular installations because of its expense and complexity. It would be used by "techies" who want the most flexible system and can both visualize how it works and handle multiple remotes with ease. It will be very frustrating to others. The dual descrambler feature could be used to simplify both the installation and operation of this connection. Onscreen displays could further assist the subscriber. Despite the flexibility of this connection and the potential to make it relatively easy to use with onscreen displays, its usage is very limited.

Advantages

1. Most flexible.
2. Can watch any authorized scrambled or unscrambled program.
3. Can record any (other or same) authorized scrambled or unscrambled program.
4. PIP can be used with both channels if the TV has baseband inputs that are connected to the VCR's baseband outputs.
5. Useful in systems that have both analog scrambling and digital television.

Disadvantages

1. Can't consecutively tape different channels automatically because VCR tuner does not do channel selection.
2. Expensive hardware and monthly fees.
3. Complex to install, many cables, many opportunities for signal leakage.
4. Complex to operate; need to use at least three remote controls.
5. Neither the TV tuner nor the VCR tuner has any control.
   (a) TV's parental control, favorite channels, and last channel recall features are inoperative. The TV is just tuned to channel 3 (or 4).
   (b) VCR's timer, parental control, favorite channels, and last channel recall features are inoperative. The VCR is just tuned to channel 3 (or 4).
6. Need to use three of four remote controls or a universal remote control.
7. Most brands of set-top terminals will operate simultaneously from one remote control. Consequently, the subscriber may have to operate one terminal as manual. If the set-top terminal supports two control codes, four remote controls may be used with this configuration.
8. If both analog scrambled and digital channels are being used, may need to interchange set-top terminals depending on which is to be recorded and which is to be watched.

9. Switch at TV is usually manual, requiring getting out of the chair unless the switch is built into the TV and remote controllable.
10. Splitter cuts signal in half and may introduce noticeable noise.

Special Cases

1. If the left converter/descrambler has an auto bypass switch, then when the converter is turned off, the entire cable spectrum passes through to the VCR, and the situation is exactly the same as in Figure 23.1.
2. If the right converter/descrambler has an auto bypass switch, then when the converter is turned off, the entire cable spectrum passes through to the TV, and the situation is similar to that shown in Figure 23.1 but without the VCR.
3. If the set-top terminal has the channel-mapping feature and the TV or VCR does not, the same programming may appear on two different channel numbers. One channel number will be found on the set-top terminal and the other on the TV or VCR. This may be confusing.

23.3.4 Minimal Scrambled Channel Usage

Figure 23.4 is illustration 5 of the NCTA connection diagram set. The cable signal is split into two paths. The left path is connected to a set-top terminal, which is connected to an A/B switch, which is connected to a TV. The right path is connected to a VCR, which is connected to the other input of the A/B switch.

This connection could be used by someone who wishes to continue the use of the features of their TV and VCR as enjoyed before cable, and only occasionally watches, but never records, scrambled channels. This is especially convenient for someone with only one scrambled channel or who uses the converter/descrambler only for pay per view but doesn't record that programming. The right-hand path is identical to Figure 23.1. Alternatively, since the path with the converter can be used to watch either scrambled or unscrambled channels, the only purpose for the switch is to watch recorded tapes. That could also be accomplished with the VCR's baseband output and the TV's baseband input.

Advantages

1. Easy to use if most of the usage is for unscrambled channels.
2. Can watch any channel while recording any unscrambled channel.
3. Can consecutively record unscrambled channels automatically.
4. Especially easy if only one scrambled channel is taken.
5. PIP can be made to work as long as the second picture is unscrambled.

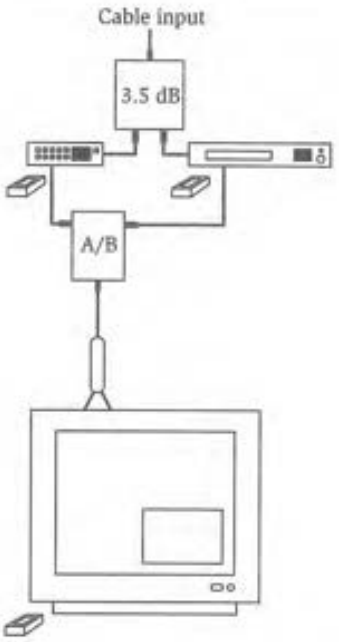


Figure 23.4 Connection for few scrambled channels.

Disadvantages

1. Can't record scrambled channels.
2. On the left side, TV's parental control, favorite channels, and last channel recall features are inoperative. The TV is just tuned to channel 3 (or 4).
3. Switch must be manually operated unless it is built into the TV.
4. Need to use three remote controls or a universal remote control.
5. Splitter cuts signal in half and may introduce noticeable noise.

Special Cases

1. If the converter/descrambler has an auto bypass switch, then when the converter is turned off, the entire cable spectrum passes through to the TV, and the situation is similar to that shown in Figure 23.1 but without the VCR.

2. If the set-top terminal has the channel-mapping feature and the TV or VCR does not, the same programming may appear on two different channel numbers. One channel number will be found on the set-top terminal and the other on the TV or VCR. This may be confusing.

23.3.5 Improved Simple Connection

Figure 23.5 is illustration 4 of the NCTA connection diagram set. This configuration inexpensively overcomes some of the difficulties of Figure 23.2 and especially makes sense if the TV has a built-in A/B switch.

Advantages

1. Inexpensive and easy to install.
2. Easy to understand.
3. Can record scrambled channels.
4. PIP feature can be used when one of the channels is always unscrambled and the VCR's baseband output is connected to TV's baseband input.

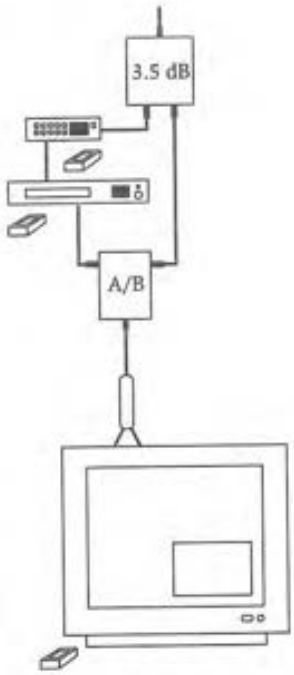


Figure 23.5 Modified most common connection.

Disadvantages

1. Can't consecutively tape different channels automatically because VCR tuner does not do channel selection.
2. Need to use three remote controls or a universal remote control.
3. On the left side, neither the TV tuner nor the VCR tuner has any control.
   (a) TV's parental control, favorite channels, and last channel recall features are inoperative. The TV is just tuned to channel 3 (or 4).
   (b) VCR's timer, parental control, favorite channels, and last channel recall features are inoperative. The VCR is just tuned to channel 3 (or 4).
4. Switch at TV is usually manual, requiring getting out of the chair unless the switch is built into the TV and remote controllable.
5. Splitter cuts signal in half and may introduce noticeable noise.

Special Cases

1. If the converter/descrambler has an auto bypass switch, then when the converter is turned off, the entire cable spectrum passes through to the VCR, and the situation on the left side is exactly the same as in Figure 23.1.
2. If the set top terminal has the channel-mapping feature and the TV or VCR does not, the same programming may appear on two different channel numbers. One channel number will be found on the set top terminal and the other on the TV or VCR. This may be confusing.

23.3.6 Problematic Connection Example

Figure 23.6 is illustration 7 of the NCTA connection diagram set. This figure is almost never used with modern set-top terminals because they usually require a separate frequency for receiving control signals. The VCR would block those signals, and the set-top terminal would turn itself off after a preset time. Also, this configuration would block upstream signals, making impulse pay-per-view (IPPV) and other two-way services impossible.

Advantages

1. Attempts similar functionality to Figure 23.4 but without the need for the A/B switch.
2. PIP can be used with the second picture unscrambled if the TV has baseband inputs connected to baseband outputs from the VCR.

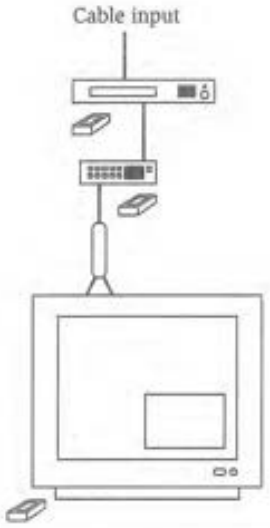


Figure 23.6 Connection with set-top terminal difficulties.

Disadvantages

1. Many set-top terminals can't be used in this configuration.
2. Impulse pay-per-view and other two-way services are impossible.
3. Can't record scrambled programs.
4. TV's parental control, favorite channels, and last channel recall features are inoperative. The TV is just tuned to channel 3 (or 4).
5. Need to use three remote channels or a universal remote control.
6. Converter/descrambler's parental control is inoperative when using VCR's channel 3 (or 4) output.

Special Cases

1. If the converter/descrambler has an auto bypass switch, then when the set top terminal is off, the entire output of the VCR passes through to the TV, and the situation is exactly the same as in Figure 23.1.
2. If the set-top terminal has the channel-mapping feature and the TV or VCR does not, the same programming may appear on two different channel numbers. One channel number will be found on the set-top terminal and the other on the TV or VCR. This may be confusing.

23.3.7 Improvements

One of the main disadvantages of the configurations shown in Figures 23.2, 23.3, and 23.5 is that the VCR is not able to consecutively record different programs except if the set-top terminal has an auto bypass switch and no scrambled programming is to be recorded.

This disadvantage can be overcome, at least in theory, with set-top terminals that include a programming function similar to that found in VCRs. This programming function can be set up to turn on and change channels at appropriate times. Usually, an onscreen menu is used for this purpose. The subscriber then uses the same principles to program the set-top terminal as are used to program the VCR, with the VCR programmed only to turn on and record the output channel of the set-top terminal at the appropriate time. In practice, this procedure meets with limited success by the occasional user since there are additional opportunities to make mistakes and thereby fail to record the desired programs. Nonetheless, the careful subscriber can master these techniques and successfully record different channels consecutively.

An alternative approach to overcoming the deficiency that inhibits the recording of consecutive programs on different channels is found in certain premium VCRs. These products have an infrared (IR) light emitter on a short wire. This device is sometimes called an IR blaster. The IR control is set up to use the IR control codes of the brand set-top terminal employed by the cable operator. The VCR can then change the channels of the set-top terminal and thereby record different channels, scrambled or unscrambled, consecutively. Since the set-top terminal is now controlled by the VCR, the set-top terminal's remote control is not necessary. Only two remote controls are needed by the subscriber, one for the TV and one for the VCR.

In the case of Figure 23.3, this type of VCR overcomes yet another difficulty. That is, if the two set-top terminals respond to the same IR codes, the VCR's IR controller can be placed in such a way that the set-top terminal used with the VCR is isolated from the IR signals used with the set-top terminal connected to the TV. Then only three remote controls are necessary, and both set-top terminals can be controlled remotely.

It is possible to use the same IR blaster technique with the TV controlling the set-top terminal (if the TV is so equipped). Also, the set-top terminal may use an IR blaster to control the VCR or TV. In that case, the subscriber programs the set-top terminal and uses its remote control.

An unfortunate complication with the IR blaster stems from the fact that the usual on–off control on consumer electronics products and set-top terminals is a toggle function. That is, there is not a separate on and a separate off control. Rather, if the unit is on, the control turns it off, and vice versa. This uncertainty over the current state is a source of error in setup and usage of these techniques.

There are numerous complications in these configurations. Most subscribers have difficulty visualizing what is happening with the multiple remote controls or the multiple modes of use of the universal remote control.[5]

23.4 Consumer Electronics Interface Issues

23.4.1 The Importance of Compatibility

The cable–consumer electronics interface establishes the manner in which a cable system, a consumer, and consumer electronics products function together to provide a useful result. Compatibility means the convenient use of the features of in-home equipment that are of interest to the subscriber. It also means a minimum of hurdles to subscribing to advanced cable services. From the perspective of the cable operator, compatibility is an issue of customer service and satisfaction. Effective compatibility can be a strong competitive advantage over other service providers. It can mean the difference between subscriber retention and the loss of subscribers to other media.

23.4.2 The Participants and Their Priorities

To better understand the consumer electronics interface issues, it is helpful to consider the problems from the perspectives of the various participants. These participants have differing goals and desires that are not always in harmony.

There are five major participants: (1) the subscriber/consumer, (2) the service provider, (3) the service provider's equipment manufacturer, (4) the consumer electronics industry, and (5) government entities making laws and rules.

The subscriber/consumer is the ultimate objective of all the efforts. If the subscriber/consumer is not happy and if there are alternative entertainment and information choices, the result will be a loss of business. Of course, there always are alternatives. The simplest alternative is to simply not subscribe to premium services and just use an off-air antenna with videotape and video disc rentals as supplementary video.

The next participant to be considered is the service provider who has a love–hate relationship with the in-home hardware. Following is the service provider's equipment manufacturer who has legitimate business concerns. The participant that is most anxious to become an active player in this arena is the consumer electronics industry. Lastly, the Congress has established laws that require the FCC to regulate these matters.

23.4.3 The Subscriber/Consumer

The subscriber/consumer participates in two ways. The subscriber/consumer buys equipment from the consumer electronics industry and subscribes to services that play on that equipment. The ultimate goal is the services. Additional programming choices make the TV and VCR even more useful.

The best reason for a large screen and a surround-sound system with multiple speakers and high-wattage amplifiers is the premium programming that comes

with pay services and pay-per-view events. Likewise, having a quality receiver and sound system makes movies and premium services more enjoyable. Given all this, it is clear that the cable and consumer electronics industries are highly motivated to work closely together to reinforce this symbiotic relationship.

The consumer wants good programming, good pictures and sound, cost-effectiveness, and simplicity. The subscriber does not care which medium makes the programming available. The consumer is frustrated with product features that are difficult to use and understand. The blinking "12:00" on the VCR has become emblematic of consumer technophobia — many consumers do not even know how to set the correct time. The maze of wires behind the TV and VCR are bad enough with just off-air reception, but they become beyond comprehension for many consumers when a premium service provider adds a set-top terminal of some sort.

23.4.4 The Service Provider

There are two elements to the service provider participant: the cable operator and the program supplier. Although this is an interdependent relationship, there are differences in their priorities.

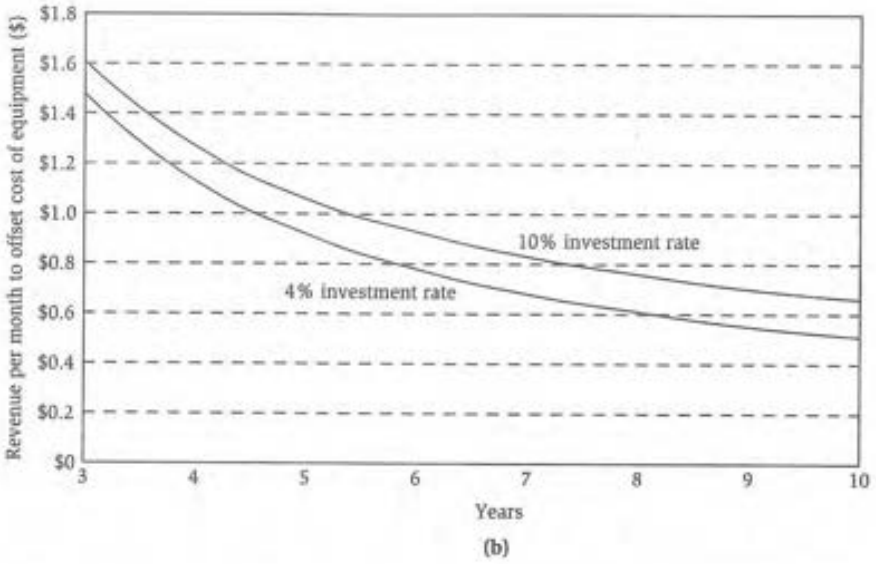
The service provider's love–hate relationship with the in-home terminal concentrates on its major disadvantage, the cost, and its major advantage, the facilitation of new services. The capital cost of the in-home terminal is a significant portion of the total capital investment required for service provision. This is especially burdensome with digital set-top terminals.

A rule of thumb used in the cable industry for a decade and a half is that for every $50 of in-home hardware, $1 a month is required just to cover the cost of the equipment. This rule of thumb is based on the assumption that the money to buy the set top terminal is borrowed from some capital source that requires repayment. The capital source may be internal or external, but capital is always a scarce commodity with multiple demands, each of which must justify its use. The capital must be repaid over the period during which the set-top terminal remains functional, and the interest must meet corporate hurdle rates. Over the years that set-top terminals have been used, the interest rate has varied over a wide range. Using the usual financial equations that are available in spreadsheet programs, Figure 23.7(a) shows a table of the amounts that must be paid back monthly to cover a $50 loan for various interest rates over a number of years. Figure 23.7(b) graphs this data with the number of years across the top, the dollars of outflow on the vertical axis, and a separate curve for each percentage value.

In the graph, the curves for the various percentages cluster closely. Clearly, the number of years over which the loan must be paid back is the more important parameter. Going from 3 years at 5% interest to 3 years at 10% interest increases the monthly cost by only 7%. Going from 3 years at 5% to 4 years at 5% reduces the monthly outlay by 23%. Long life of the equipment is

Years	Interest rate in percentage						
	4	5	6	7	8	9	10
3	$(1.48)	$(1.50)	$(1.52)	$(1.54)	$(1.57)	$(1.59)	$(1.61)
4	$(1.13)	$(1.15)	$(1.17)	$(1.20)	$(1.22)	$(1.24)	$(1.27)
5	$(0.92)	$(0.94)	$(0.97)	$(0.99)	$(1.01)	$(1.04)	$(1.06)
6	$(0.78)	$(0.81)	$(0.83)	$(0.85)	$(0.88)	$(0.90)	$(0.93)
7	$(0.68)	$(0.71)	$(0.73)	$(0.75)	$(0.78)	$(0.80)	$(0.83)
8	$(0.61)	$(0.63)	$(0.66)	$(0.68)	$(0.71)	$(0.73)	$(0.76)
9	$(0.55)	$(0.58)	$(0.60)	$(0.63)	$(0.65)	$(0.68)	$(0.70)
10	$(0.51)	$(0.53)	$(0.56)	$(0.58)	$(0.61)	$(0.63)	$(0.66)

(a)



(b)

Figure 23.7 Set-top costs. (a) Monthly cost of $50 loan at various rates paid back in multiple years. (b) Graph of data.

important to low monthly costs. Conversely, high-tech equipment that is soon obsolete is very expensive on a monthly basis. In addition to repaying the capital for the set-top terminal, there is a monthly maintenance cost.

The benefit of the in-home terminal to the service provider is that it facilitates services. The in-home terminal does this in two ways. First, it provides the electronics necessary for the service. For example, a Web appliance provides

the signaling, memory, and display electronics needed to browse the Web and display the results on the TV. Second, the in-home terminal is the instrument of conditional access. Unless access can be restricted, a price cannot be demanded. A wide variety of ways of protecting signals were discussed in the previous two chapters. Nearly all of them require some in-home device.

Experience shows that consumers respond to the ability to make impulse purchases of programming. The easier the mechanism of purchase, the more is bought. This capability also requires some sort of in-home equipment.

The in-home terminal is one of the fundamental determinants of the quality of the video and audio experienced by the subscriber. Technical specifications such as noise figure, linearity, and the tightness of the shielding are critical to good pictures and sound. The need for a quality terminal is paramount.

Lastly, there is a need to be able to experiment with new services. In the past, this has been done by the system operator providing the in-home hardware necessary for the marketing experiment. If the service proves compelling, additional equipment is procured, and the service is rolled out across the population. The cable industry is very good at copying success and applying it widely. If the experimental service is disappointing, the equipment is abandoned and forgotten. In neither case is the consumer required to purchase new hardware that might be later orphaned. This is just an extension of one of cable's major marketing advantages. The salesperson can offer a low-cost or even a free trial hookup to the consumer without asking for the purchase of any equipment. Consumers know that if they are dissatisfied with the service, they may discontinue at any time without getting stuck with hardware that has no other use. This ability to offer a risk-free trial is an important advantage.

The program supplier's principal concern is protection against theft. Though this is important to the cable operator, it is even more critical to the program suppliers. The framers of the U.S. Constitution realized the importance of protecting the rights of authors, artists, and other creative workers when they provided for the copyright. Protection encourages the creative contributor. Without the fruits of these contributors, there would be no cable industry and no consumer electronics industry. The consumer would have minimal interest in what little would be left. Although the public may view theft of programming as a "victimless crime" or one of little significance, this is a misperception. All suffer from the loss of incentive to produce quality programming. Legitimate consumers pay higher prices because those who steal do not contribute their fair share to covering the costs of operation. Fortunately, the 1992 Cable Act, the 1996 Telecommunications Act, and the actions of the FCC have recognized the importance of preventing signal theft.

The program supplier is also interested in minimizing the impact of the copying of programs on other means of distribution. Businesses in the rental and sale of videotapes and video discs are undermined by easy copying of programming delivered on cable.

23.4.5 The Service Provider's Equipment Manufacturer

The priorities of the service provider's equipment manufacturer are perhaps easiest to understand. Its goal is to profitably sell systems involving large quantities of in-home hardware and the headend equipment needed to supply signals. Since cable systems are individual closed systems of limited geographic area and since the in-home hardware has traditionally been leased, adjacent cable systems can utilize totally incompatible equipment and still well serve the TVs and VCRs owned by subscribers. The subscriber who moves from system to system is not inconvenienced by these incompatibilities since the system that serves the new residence will supply the necessary in-home equipment while the system serving the previous residence takes back its equipment.

A major advantage of this approach is that new technology is easily accommodated. When a cable system is ready to upgrade its hardware, it can purchase the best of the newest equipment available without concern about making the consumer's equipment obsolete. From the manufacturer's perspective, this keeps the market from becoming saturated. New services and new technologies can recapitalize the consumer's set top while bringing excitement to the service. Imprudently implemented consumer ownership of this equipment would interfere with this process of innovation.

23.4.6 The Consumer Electronics Industry

The consumer electronics industry has two main elements: the manufacture and the retail sale of products.

The nature of the consumer electronics business is one of high competitiveness. The manufacturer is interested in making more products and in selling upscale models. Almost 100 million residences receive television signals in one form or another in the United States. There are more than 250 million television receivers and 150 million VCRs in those homes. About 25 million new TVs and 15 million new VCRs are sold annually. The consumer electronics industry has the capacity to produce even more. Each year, new efficiencies make possible the production of more units from existing production lines. There simply is massive production overcapacity. The consequence of this is that it is impossible to maintain prices, much less raise them. Margins are tight. Only the full-featured products make money. The simple products now sell for something under $10 per inch of screen size. The cable set-top terminal has the potential to interfere with attractive features in the TV and VCR. This drives the consumer toward the profitless low end of the product line. Clearly, this is a survival problem for the consumer electronics industry.

The retailer does not always have the option of selling set-top terminals to cable operators. STT sales are traditionally large-scale industrial sales that are seen by the retailers as a bypass of their business. The retailer's desire is either to totally prevent the use of set-top terminals or to get into the business of selling

them. After a few years of attempting to outlaw set-top terminals, the retailers have come to realize that the age of interactive television, convergence, and advanced services requires set-top terminals. The emphasis is now on requiring these devices to be available at retail.

23.4.7 The Two Extremes

After the 1992 Cable Act, discussions between the cable and consumer electronics representatives began in earnest. Some on each side adopted extreme and simple-minded positions. The extremists in the consumer electronics industry said, "Why don't they just not scramble?" Whereas the extremists in the cable industry said, "Why don't they just make monitors?"

Both of these outlooks are life threatening to the industries concerned. Without scrambling, cable cannot protect signals and offer premium services. This would spell the end of the cable business. Without premium features, the monitor business would degenerate into a profitless commodity business that would cause chaos to the consumer electronics sector. Neither alternative is workable. Neither alternative considers the consumer. Something else is required.

23.4.8 The Origins of the Problem

The history of the situation provides some help in understanding its complexity. The first "cable-ready" TV was almost an accident. A couple of decades ago, television tuners were analog devices with mechanical rotary turrets. A multitude of contacts switched coils, changing the frequency tuned and the local oscillator frequency. Remote control was implemented with a motor driving the shaft of the tuner. The tuner contacts wore out or oxidized, and the tuner became the major source of service problems. The digitally controlled varactor tuner was an innovation that tuned the receiver with electrically variable capacitors (the varactors) instead of screwing a ferrite core into a coil of wire. The tuning voltage for the varactors was derived from a digital-to-analog converter driven by digital electronics. The voltage data was stored in semiconductor memory. The same digital electronics served to implement the remote control. The advantage to the consumer was a significant increase in reliability because of the elimination of moving parts and the vulnerable electrical contacts. However, the cost of TVs with electronic tuners was around $100 more than comparable units with mechanical tuners, making these advanced receivers difficult to sell.

An engineer recognized that memory comes in fixed sizes and that most applications do not use all the memory available on the chip. He proposed filling the memory chip with additional data that allowed the varactor tuner to tune some nonbroadcast frequencies used on cable. With the addition of a cable television F connector to accommodate connection to a cable, the product could be sold as "cable ready." The additional cost was near zero—just better

utilization of the memory that was already there. This was a feature consumers could understand. They were willing to pay more for "cable ready" because it promised to free them of the set-top terminal.

As time went on, subscribers noticed that though the TV gave them access to some cable channels, it did not give them access to all. A "horsepower race" developed where the number of channels tuned was part of the advertising. As more channels were added, a larger size memory was needed; this added a small increment of cost. Further increases in channel capacity required slightly more complex tuners. But this still was a minor cost increment.

23.4.9 Other "Cable-Ready" Issues

There are a number of important design features included in most cable set-top terminals that very few "cable-ready" TVs and VCRs incorporate. Most cable set-top terminals have a feature called *channel mapping*. Channel mapping decouples the channel display number from the frequency tuned. Thus, if there is a direct pickup problem in the cable system because of the presence of a strong broadcast signal, the broadcast channel can be placed at a different frequency to avoid this interference while allowing the subscriber to request it by its familiar channel number. The remote control command and the channel indicator all provide the familiar number while the tuner is "mapped" to another frequency. The set-top terminal has downloadable control commands to change the mapping as channel lineups change and to suit the needs of the cable system using the unit. A consequence of the absence of this important feature in "cable-ready" TVs and VCRs is the annoying need for multiple channel numbers in many local program listings. The familiar channel number is used for those subscribers using a cable operator-provided set-top terminal. An alternative channel number is used for "cable-ready" TVs and VCRs. This problem exists in the same home when some TVs have set-top terminals and others do not or when the auto bypass feature is used with a set-top terminal. Receivers built in compliance with the EIA-608A standard may include automatic channel mapping as an option.

A more important application of channel mapping is the manner in which scrambled channels are handled in a set-top terminal. When a subscriber using a set-top terminal attempts to tune a channel for which he or she is not authorized, the set-top terminal avoids that channel. Either it goes to another channel that has an appropriate message or it simply presents a blue (or blank) screen and muted audio. The alternative channel may include advertising for premium cable services. In that case, it is referred to as a *barker channel*. The term comes from the circus "barker" who hawked sideshows and attempted to attract an audience. If a blue screen is used, it may include a text or graphics message indicating how the subscriber could upgrade the subscription. In either case, the video and audio are kept from the receiver. This feature allows for practical video scrambling that has relaxed constraints. The video scrambling does not need to completely hide the video. There is no concern that "objectionable"

video might "bleed" through to offend sensitive viewers. Additionally, audio does not need to be scrambled since, in most cases, scrambling the video is sufficient to require a subscription. Children are protected from unintentional exposure to the audio of objectionable programming and do not accidentally see the "random lockup" of a scrambled video because of the set-top tuner's mapping to another channel.

The analog television audio signal is complex. It includes a stereo component, a second audio program (SAP), usually for monaural second language, and auxiliary signals. Since the human ear is very sensitive to audio distortion, scrambling the complex analog television audio signal in a manner that results in quality sound when descrambled is expensive and technically challenging. This expense would be unnecessary if channel mapping were universally used.

Set-top terminals have other important features as well. Parental control features are required by some franchises. These franchises often specify the manner in which the parental control feature must function and the level of security it must provide. "Cable-ready" TVs and VCRs manufactured since 2000 must have parental control (the "V-chip"). Additionally, set-top terminals often have an emergency alert feature that allows local government officials to alert citizens of hazardous conditions. "Cable-ready" TVs and VCRs do not include this capability, though it can be made available through use of a POD module.

23.5 Frustrated Product Features

Adding a set-top terminal either makes certain features and functions of TVs and VCRs redundant or interferes with their operation. In most situations, the feature or function is not blocked, but its use requires additional effort that the consumer finds inconvenient. It is this frustration that motivated the compatibility section of the 1992 Cable Act.

23.5.1 Remote Control

Remote controls were included in even early set-top terminals. The first remote controls were "wired." A long wire led to a separate unit positioned near the viewing chair and allowed channel changing without having to go to the TV. Cordless remotes based on ultrasonic tones, infrared light waves, or radio links were introduced later on. Consumers were able to add the remote control feature to existing non-remote-controlled TVs by utilizing the cable set-top terminal. This was seen as an important ancillary advantage to a cable subscription. The consumer electronics industry was not pleased with the slowing impact this had on replacement sales. However, a benefit accrued in that consumers were presold on the advantages of remote control. It became almost impossible to sell a new TV or VCR without remote control.

A problem arises when both the TV and the set-top terminal are remote controlled. The customer may feel he or she paid for a remote control feature on the TV that is not needed because the set-top terminal has the remote control. Worse yet, the consumer may need to juggle two remote controls. He or she may have to use the TV's remote to turn it on, put it to the output channel of the set-top terminal, and adjust volume. The consumer then uses the set-top terminal's remote to change channels. The set-top terminal often includes a switched convenience power outlet. This works well with older, non-remote-controlled TVs. But there are problems with remote-controlled TVs. These products usually have a toggle on–off push button, which means that simply powering the line cord does not result in the TV going on. Worse yet, if the TV has an automatic channel setup feature, it will go through the setup procedure every time it is repowered. Since this takes several minutes, the result is unacceptable. The net result is the juggling of two remote controls. Note that the same situation exists in a noncable application where the VCR is connected between the antenna and the TV.

23.5.2 Timers

When a set-top terminal is used with a VCR, further complications are encountered. Since the VCR timer controls only the VCR, it cannot turn the set-top terminal on and off, and it cannot change the terminal's channel. Early on, this was recognized, and set-top terminals with built-in timers have been available for well over a decade. The very same techniques used for programming the VCR are used for programming the set-top terminal. The consumer who can do one can do the other as well. Unfortunately, most consumers do this function so infrequently that they must relearn it every time they wish to record a program. The chances for error are great, and the success rate is low. This leads to frustration. The majority of consumers — cable and noncable — use VCRs for playback of prerecorded tapes and just avoid these headaches.

Some set-top terminals have the ability to set their clocks using signals coming from the cable headend. Thus, their time is always correct, even in spring and fall, when daylight savings time changes occur. Some newer VCRs implement this feature with signals from the broadcaster. Another common timer found in TVs is the sleep timer. Although a cable set-top terminal does not interfere with the TV's sleep timer operation, some set-top terminals have a sleep timer that makes one of the two timers redundant.

23.5.3 Picture-in-Picture

Picture-in-picture (PIP) is the insertion of a small picture within the larger picture. The purpose is to be able to monitor a second program while watching and listening to another program. If the action in the second program becomes compelling, the two images can switch places with the push of a button. Most TVs above the basic models include PIP.

There are a number of implementations of PIP. The most common embodiment has only one tuner, which is used for the main picture; the image for the second picture comes from an external device — usually a VCR's tuner. Other video source alternatives include a baby-monitor camera or a front-door camera. More expensive TV receivers have two tuners. An advanced form of PIP can be operated to scan through channels and put up a mosaic of small still pictures from a number of channels. As such, it is a kind of program guide that allows owners to scan the channels prior to making a selection.

In a TV with two built-in tuners, a problem arises if an internal splitter is used and the set-top terminal is placed ahead of the single antenna input. In that case, the output of the set-top terminal appears on both the large and small pictures. If on the other hand, two antenna inputs are provided, the use of the set-top terminal is noninterfering except when the customer wishes both pictures to come from scrambled channels. In that case, two descramblers are required.

The least common PIP implementation is the one that presents a mosaic of pictures from various channels. This system is interfered with since the program selector is in the set-top terminal and is not under the control of the TV receiver.

23.6 Early Attempts

There were a number of early attempts to solve or mitigate the compatibility problems between consumer electronics and cable systems. All proposed solutions were partial. The 80/20 rule applies. Eighty percent of the problem can be solved with 20% of the effort (and expense). However, solving the remaining 20% of the problem takes 80% of the effort.

23.6.1 Universal Remotes

Universal remote controls sought to alleviate the need for using two or more remote controls with the set-top terminal and TV or VCR. The universal remote control is configured for the brand TV, the brand VCR, and the brand set-top terminal. Simple universal remote controls have a hidden row of switches that are configured according to printed instructions that come with the device. An alternative approach is to enter appropriate setup codes from the printed instructions. Most models direct the channel change commands to the set top terminal, the volume commands to the TV, and the on–off commands to both units. VCR control can also be accommodated. Either a slide switch or a mode button governs which device is primarily controlled.

A problem arises with simple universal remote controls when a new model of TV, VCR, or set-top terminal becomes available or if a favorite feature of the TV or VCR is not included in the universal remote control. "Learning" universal

remote controls solve this problem by allowing the universal remote control to learn the signals emitted by the remote control that comes with the product. A disadvantage is the complexity; some consumers seem unable to master the instructions. Another type of learning universal remote control uses acoustic signals from the telephone to enter the data requested by the customer.

Advanced universal remote controls include *macros*. A macro is a collection of commands emitted in response to the pressing of just one button. For example, a common macro is the *advanced on* command. First, the set-top terminal is turned on, then the TV receiver is turned on, then the TV receives a command to tune to the output channel of the set-top terminal. It is important that all the devices "see" the complete IR emission sequence and respond appropriately, or the result may be disappointing.

Another model of advanced universal remote control includes a clock and a timer. The remote control can be positioned so that it simultaneously communicates with both the VCR and the set-top terminal. The universal remote control is programmed with the VCR control instructions. The remote turns on the VCR at the appropriate time, ensures that the VCR is tuned to the output channel of the set-top terminal, then tunes the set-top terminal to the desired channel, and finally starts the VCR recording. Consecutive recording of different channels is made possible.

A few models of two-way remote controls have a small screen on the remote control that displays signals received from the set-top terminal. Titles of songs, purchase order forms, interactive games, and other instructions can be viewed.

23.6.2 Electronic Program Guides

Electronic program guides (EPGs) have been introduced with features that alleviate some of the interface problems. The EPGs are available as stand-alone devices, built into cable set-top terminals, or built into TVs or VCRs. Several of these devices include one or more infrared light-emitting diodes (LEDs) on short wires that are meant to be attached near the remote control window of the VCR, the cable set-top terminal, or the TV. The device with the EPG controls the other devices to accomplish the desired result. If the EPG is built into the cable set-top terminal, the LED controls the VCR so that it turns on and off at the correct times and is tuned to the correct channel. If the EPG is built into the VCR, the LED controls the channel on the cable set-top terminal. Of course, none of these approaches can ensure that sufficient blank tape has been put into the VCR!

A related device, called VCR Plus, uses numeric codes published in printed guides (or conveyed electronically in a later version) to the VCR or to a form of universal remote control. The codes are decoded using a proprietary algorithm to derive the start and stop times and the channel number. Infrared emissions control the VCR and cable set-top terminal.

23.6.3 The TV Integrating Device

In the mid-1980s, the research department of a cable operator, American Television and Communications in Englewood, Colorado, began development of an approach the engineers called the TV integrating device (TVID). It was intended to be included in a cable set-top terminal and to use IR LEDs to control the TV and one or more VCRs. It had a built-in EPG and an on-screen display. The goal was to provide a simple menu that would allow subscribers to indicate their desires. The subscriber would indicate what he or she wanted recorded and what he or she wanted to watch. The TVID would take care of switching the descrambler into the correct part of the system and turn things on and off at the right time to implement the subscriber's desires. If a conflict arose, such as wanting to watch a scrambled channel while recording another scrambled channel, the subscriber would be advised of his or her choices and allowed to act upon them.

The TVID included a "mood guide," which gave the subscriber a way of indicating his or her mood. Those programs that were appropriate to his or her mood would then be displayed for selection to watch or record. The microprocessor could keep track of the programs watched and deduce the subscriber's interests. The subscriber would then be alerted to future programs that fit his or her previous viewing habits. The project reached the early prototype stage but was not completed because of other priorities. Some of the concepts, however, have become part of some PVRs.

23.7 Management of Expectations

Most of the subscriber's problems have solutions. Some solutions are more complex than others. The crux of the compatibility problem seems to be the expectation that the term "cable ready" brings. To consumers, this means that the TV or VCR can be connected to cable and all features and all services are readily available. When operation of features is complicated or unexpected results are obtained, frustration and anger set in. If instead, expectations were properly managed and the subscriber/consumer had a correct understanding of what was possible, the likelihood of disappointment would be reduced. The fundamental problem is the management of expectations.

23.7.1 Complexity Overload

A major contributor to the difficulties of the consumer electronics interface is the wide variety of choices available in features and functions, the way in which the features and functions are operated and controlled, and the diversity of services offered by cable operators. The number of combinations and permutations is almost infinite. Just the simple matter of entering channel numbers has several methods. Some products use an enter button. Others prescribe the way in which

the digits must be entered. For example, on some products, requesting channel 7 requires pressing 07. Others can require waiting for a fixed time after entering the command. The names for common functions are brand specific. The number of complex features boggles the mind. It may simply not be possible to preserve all the product functions and features while offering access to all the available cable services.

23.8 Legal Aspects of Cable–Consumer Electronics Compatibility

23.8.1 Compatibility Definitions

The word *compatibility* should mean that two things work well with each other. In the case of compatibility between cable service and consumer electronics, *compatible* should mean that the subscriber should be able to enjoy the cable services of interest and the features of the consumer electronics that he or she finds appealing. The subscriber should be able to conveniently do both.

Note that compatible is a situational definition. If a subscriber takes only basic (unscrambled) service from a cable system with a modest number of analog channels and is located far from strong interfering broadcast signals, he or she may enjoy all the functions of his or her "cable-ready" TV and VCR. If the subscriber then decides to add a scrambled service or moves near a broadcast transmitter, the situation changes significantly. He or she may lose compatibility between the TV or VCR and the cable system.

23.8.2 The 1992 Cable Act's Definition of Compatibility

Senator Patrick J. Leahy (Democrat from Vermont) drafted and introduced a bill in 1991 pertaining to the operation of cable systems and the design of consumer electronics products so as to improve compatibility. The bill did not pass. In the following year, it was introduced as Section 17 to the 1992 Cable Act.[6] Section 17 requires the Federal Communications Commission (FCC) to create rules to improve the compatibility between cable systems and consumer electronics products. The effect of this section of the law was to define compatibility as having three principal aspects:[7]

(i) to watch a program on one channel while simultaneously using a video cassette recorder to tape a program on another channel;

(ii) to use a video cassette recorder to tape two consecutive programs that appear on different channels; and

(iii) to use advanced television picture generation and display features

Although this definition of compatibility may not be complete, it does define a requirement for the FCC's rules.

The bulk of the 1992 Cable Act pertains to economic issues such as rate regulation, the rules for buying and selling systems, and the permissible business relationships between programmers and operators. A small, but very important part of the Cable Act concerned the interoperability of cable systems and consumer electronics products. Though the 1992 Cable Act gave the FCC a difficult assignment, it also laid down some guidelines. The FCC was told to consider the costs and benefits to consumers of its rules and the provision of effective protection against theft. But the ultimate goal was clear: the minimization of interference with certain special functions in consumer electronics devices when connected to cable systems.

The FCC was given some new authority to help it achieve these goals. The FCC's primary influence over the cable industry was granted in the following clause: "the Commission shall determine whether and, if so, under what circumstances to *permit* cable systems to *scramble or encrypt* signals or to *restrict* cable systems in the manner in which they encrypt or scramble signals."

The authority given the FCC over the consumer electronics industry is much more limited: "to specify the technical requirements with which a television receiver or video cassette recorder must comply in order to be sold as 'cable compatible' or 'cable ready.'"

To ensure that the goals of compatibility continued to be pursued in the future, the FCC was given continuing responsibility:

> The Commission shall periodically review and, if necessary, modify the regulations issued pursuant to this section in light of any actions taken in response to such regulations and to reflect improvements and changes in cable systems, television receivers, video cassette recorders, and similar technology.

Further requirements imposed on the operation of cable systems included obligations for the FCC to:

> promote the commercial availability . . . of converter boxes and of remote control devices to the extent technically and economically feasible, to offer subscribers the option of having all other channels delivered directly to the subscribers' television receivers or video cassette recorders without passing through the converter box.

The FCC was required to provide a report to Congress within a year of the passage of the act. This voluminous report was based on the extensive responses to the FCC's initial notice of inquiry on the subject. The FCC issued a notice of proposed rule making on November 10, 1993, and released its first Report & Order on May 4, 1994. In those documents, the FCC indicated its directions:

> [W]e clarify that the compatibility regulations being adopted herein apply only to cable systems and do not apply to . . . video dialtone services.

> Signal thieves [have been] notoriously successful.... We will not require cable systems to allow their subscribers to own descrambling equipment.
>
> [W]e expect that the Decoder Interface connector will be flexible enough to accommodate most, if not all, of the attributes of cable digital service.

The FCC realized that there is an extensive installed base of more than 11,000 cable systems, more than 250 million television receivers, and more than 150 million VCRs. In addition, approximately 100 million TV households purchase around 25 million color TVs and about 15 million VCRs each year. No law could bring compatibility to this situation in a short time. The FCC divided its approach into three periods: short-term measures, intermediate-term measures, and long-term measures.

In the short term, cable operators were required to avoid scrambling the basic services, which consisted of mostly the broadcast channels carried on cable. In addition, the cable operators were required to provide certain equipment such as switches to bypass the entire spectrum around the set-top terminal when the terminal is not required for descrambling, and timers in set-top terminals to change channels at specific times. Cable operators were given a year to make available multiple tuners and descramblers in set-top terminals. They were allowed to charge for the costs of this equipment and its installation. Additionally, the consumer electronics industry was given a year to discontinue the use of the terms "cable ready" and "cable compatible" or any similar terminology in the sale of TVs or VCRs that did not fully comply with the FCC's promised technical definition of "cable ready." A rule was included that would have required a label on devices that tuned cable channels but did not fully comply with the FCC's definition of "cable ready," notifying consumers that this device may under some circumstances fail to provide satisfactory performance when connected to cable. The consumer electronics industry managed to get this rule reversed and gained a relaxation of other rules in a subsequent petition for reconsideration submitted to the FCC. The FCC imposed a bizarre rule prohibiting the cable industry from changing infrared codes in each set top terminal installed in each home in order to protect the home owner's "investment" in purchased remote controls. This rule was all but impossible to follow because of the large number of different codes. It was removed in response to a subsequent petition for reconsideration. Also, the FCC recognized the aggressive work of those who would steal signals:

> [W]e are clarifying that the availability of set-top devices from retailers applies only to basic converters without descrambling or other access control function.
>
> [W]e do not believe it is desirable to require cable operators to make their security technology available for ownership by the public.

The FCC then separately considered longer-term measures centered around the decoder interface (described in the previous chapter) being developed by the

Joint Engineering Committee of the National Cable Television Association and the Electronic Industries Association: "We... emphasize... the Decoder Interface connector and associated component descrambler/decoders to be an important part of... equipment compatibility."

The FCC envisioned a digital world that would ultimately be based on standards and methods that avoided the problems of compatibility. The FCC intends to consider that environment separately in future actions.

23.8.3 The 1996 Telecommunications Act

The Communications Act of 1934 created the Federal Communications Commission and set the original public policy on communications. There have been multiple amendments to that act over the years, making it a quilt of patches. Additionally, technology had changed so dramatically that a complete rewrite became necessary. The result is the 1996 Telecommunications Act.[8] Although the act has a wide range of impact on cable television practice, only the portions pertaining to compatibility will be discussed here.

Probably the most important provision of the 1996 Telecommunications Act pertaining to compatibility is Section 304, "Competitive Availability of Navigation Devices," by Representative Thomas J. Bliley Jr., of Virginia:[9]

> (a) The Commission... in consultation with... industry standard-setting organizations... [shall] assure the commercial availability, to consumers of... services offered over multichannel video programming systems, of converter boxes, interactive communications equipment, and other equipment... from manufacturers, retailers, and other vendors not affiliated with any multichannel video programming distributor. Such regulations shall not prohibit any multichannel video programming distributor from also offering... equipment... if the system operator's charges... are separately stated and not subsidized.

The national headquarters of Circuit City stores is in Congressman Thomas Bliley's district, and his efforts to further the interests of an important constituent are clearly evident in this legislation.

The term *navigation devices* is a strange one. It has not been used in the cable or other related business except in reference to electronic program guides and related means of helping consumers find their way — navigate — through the wide variety of services available on systems with hundreds of channels. The law uses this term to mean any converter or set-top terminal intended to go between the service provider's terminal and the TV or VCR.

Several important points should be noted. Although the 1992 Cable Act was restricted to cable operators, the 1996 Telecommunications Act applies to any multichannel video programming distributor. Additionally, the services covered are not limited to video.

Limits were placed on what must be sold in order to protect the interests of programmers and system operators.

> (b) The Commission shall not prescribe regulations . . . which would jeopardize security of multichannel video programming and other services . . . or impede the legal rights of a provider of such services to prevent theft.

In order to accommodate experimentation with new services and technologies, the law provides for waivers allowing a period in which hardware can be under trial.

> (c) The Commission shall waive a regulation adopted under subsection (a) for a limited time upon an appropriate showing . . . that such waiver is necessary to assist the development . . . of a new or improved . . . service.

A number of significant hazards are brought about by this legislation. The most important hazard is that the other users of cable systems and of the over-the-air spectrum could suffer interference and degradation of service when consumers purchase their own navigation devices. Second, consumers are subject to confusion with these complex devices. Money may be wasted on features that are unusable or not needed. When technology changes, consumers will be stuck with orphaned hardware. In some cases, the ownership of equipment will create a reluctance to try new services that require additional hardware that partially duplicates what has already been purchased.

It is likely that many, if not most, subscribers will continue to lease equipment because of the many benefits that come from such an arrangement. The equipment is matched to the services provided. New equipment replaces old when technology evolves, service of equipment is included in the lease, and assistance on its use is readily available.

The other major portion of the 1996 Telecommunications Act pertaining to compatibility is Section 301(f), the "Eschoo Amendment" by California Congresswoman Anna Eschoo. This amendment modifies the compatibility rules of the 1992 Cable Act by putting restrictions on the FCC's actions. Specifically, it calls for mandating "a minimum degree of common design and operation, leaving all features, functions, protocols, and other product and service options for selection through open competition in the market." The goal of this amendment is to prevent the FCC's rules from picking winners and losers in a competitive environment where a variety of different approaches — some proprietary and others public — are vying for the same customer. Though that is a laudatory objective, it does make a difficult and complex task more challenging.

23.9 Standards

23.9.1 The Need for Standards

The cable industry has traditionally had little need for additional standards because it provides signals usable on the consumer's ordinary television or radio.

The goal has always been to allow the consumer to take cable service without having to purchase additional equipment. A major attraction to cable is the ability to try a subscription with little investment and the assurance that it is easy to drop the subscription if it is disappointing. So the most important standard is NTSC. When the cable system provides services beyond the ability of the consumer's TV, an adapter is provided to allow access to the additional services. The earliest adapters were just tuners that gave access to channels beyond the range of the TV receiver. Since a cable system is a closed universe unto itself, there is no need for interoperability with adjacent cable systems.

Conditional access and scrambling are deliberately not standardized. Standardization of these functions would give the signal thief a larger market to sell "pirate boxes." By having multiple types of conditional access and scrambling, the garage manufacturer of pirate boxes is frustrated. Since the service provider has ownership of the conditional access equipment, he or she can decide when the economic losses due to a breach of the security system are of sufficient magnitude to require replacing the equipment. When the service provider replaces the equipment, the consumers are not adversely affected.

The 1996 Telecommunications Act includes a section that requires the FCC to make rules that will regulate how multichannel video service providers will make in-home equipment available at retail. Now, for the first time, there is a need for standards. Consumers who purchase equipment for connection to cable need to be assured that it will work when they move to another cable system. Fortunately, the law and the FCC have recognized the problem of signal theft and have included a clarification that security systems do not have to be made available for subscriber ownership. There are three types of standards: mandatory, official voluntary, and de facto.

23.9.2 Mandatory Standards

Mandatory standards are those imposed by law or as a result of law. In the former case, the law itself contains the standards; in the latter, more common case, the law directs a government agency to do the difficult work and create rules. The FCC is the source of rules that govern the cable industry. There are two major sets of rules: (1) Part 78, the Community Antenna Relay Service (CARS) rules, which regulate the use of microwave links to convey signals across natural or man-made barriers or simply to transport the signals from one location to another where the geography in between does not have the cable service, and (2) Part 76, which governs the operation of the cable system itself. The items regulated include safety, signal leakage, and minimum performance requirements. These were first put into place in 1972 and then immediately stayed while further deliberations took place regarding the details. The performance rules cover frequency stability, carrier-to-noise ratio, allowable interference, in-band frequency response, hum content, adjacent channel response, cross modulation, must-carry and retransmission consent,

and signal leakage. The rules (with the exception of signal leakage) apply only to Class 1 signals, which are defined as signals originally broadcast. However, the rules cover only additional performance deficiencies. The cable operator is not responsible for problems in the signal as delivered to the cable headend. Class 2 signals are satellite-delivered programming, and Class 3 signals are locally originated. The rules were suspended by the 1984 Cable Act, which deregulated cable. At that time, they were converted into a recommended practice. Local governments were permitted to create their own rules, but those rules could not be more severe than the recommended practice. This was challenged by the League of Cities. A consequence of this was that a panel of representatives from the cable industry, the FCC, and the League of Cities negotiated a set of rules that has become the current FCC Part 76 in 1991.

23.9.3 Official Voluntary Standards

Voluntary standards do not have the force of law. The main motivation for compliance with the standard is the ability to use the fact of compliance in marketing efforts. Official voluntary standards have the sanction of a standards-setting organization. The primary standards-setting organization in the United States is the American National Standards Institute (ANSI). ANSI has very rigorous rules for the procedures involved in setting these standards. The rules are intended to ensure that participation is available to all interested parties and that due process is followed. ANSI certifies organizations as complying with these rules. Currently, the primary ANSI-certified cable organization is the Society Cable of Telecommunication Engineers (SCTE). Another organization that has worked to develop standards that govern both the cable industry and the consumer electronics industry is the Consumer Electronics Manufacturers Association (CEMA) sector of the Electronic Industries Association (EIA). Although CEMA is the official ANSI standards-setting organization, its efforts are under the auspices of the Joint Engineering Committee (JEC) of the EIA and the National Cable Television Association (NCTA) Engineering Committee.

The SCTE has about 100 standards under development at the time of this writing. Five of those standards are in a condition where they are suitable for submission to ANSI. One of those standards has been approved as ANSI/SCTE PS SP 400 1996 F-Port (Female Outdoor) Physical Dimensions. Next will be the 5/8–24 RF Outdoor Fitting Physical Dimensions standard. The Digital Video Transmission Standards for Cable Television is expected to be submitted before the end of 1998.

The JEC efforts began in the very early 1980s and have not been nearly as successful in completing standards. This lack of results is an indication of the difficulty involved in the interface between cable and consumer electronics. Its first standard, the ANSI/EIA-563 Decoder Interface Standard,[10,11] was in place for several years and then officially withdrawn. Another standard, EIA-542 Cable Television Channel Identification Plan,[12] lists channel numbers

and frequencies for three different channel plans: the standard plan related to the broadcast frequency plan, the harmonically related carrier plan, and the incrementally related carrier plan. Though this standard has been slow to make its way through the official process, it has been used by both the cable and the consumer electronics industry as a de facto standard for more than a decade. Both industries have found it a useful means of cooperation.

One of the most important JEC standards is the EIA-IS-105[13] Decoder Interface Standard described in the previous chapter. Other standards in the works include the National Renewable Security Standard (NRSS)[14] and the RF Interface Standard IS-23.[15] An effort to create a digital television standard has been transferred to the SCTE. The IS-23 standard governs such parameters as the physical connection to cable, the frequency response, the signal levels, and direct pickup by inadequately shielded components.

23.9.4 De Facto Standards

De facto standards are the result of marketplace forces where the cable industry has settled on a particular approach because it works well. The types of connectors, cables, amplifiers, and other equipment used in cable have been sorted out by the marketplace, and a limited set of models have evolved on a de facto basis. Some of these de facto standards are being converted into official voluntary standards by the SCTE. This approach has served the industry well and will continue to be in use.

23.10 OpenCable

Cable Television Laboratories, or CableLabs, has undertaken a program intended to provide options for alleviating many of the interface problems. The goals of the program, called OpenCable, are to provide sufficient standardization of features and functions so that a multivendor interoperable environment can be created. CableLabs will certify compliance with OpenCable. OpenCable functions are described in Chapter 22.

23.11 National Renewable Security System

The Joint Engineering Committee formed a subcommittee called the NRSS subcommittee to consider methods that could overcome a security breach. The motivation is to facilitate the sale of set top terminals or consumer electronics products while protecting the conditional access system by having it renewable in the event the system is defeated. The initial approach involved the Smart Card system, which was used for conditional access and for credit and debit cards. The Smart Card was intended to include a digital key that would decrypt the digital signal.

After some discussion, it was decided that the only safe approach is to have every transistor involved in the security function replaceable in a module that plugged into the set-top terminal. A wide variety of proposals were considered. One of the proposals used the Smart Card physical format but put a lot more electronics on it. This NRSS Type A card complied with the International Standards Organization (ISO) 7816 standard. Several limitations flow from this. There are only eight contact pads, forcing the MPEG signal to be transported serially. There is no mechanism for handling cable's out-of-band signaling. Eventually, the committee settled on an optional Type B approach based on the Personal Computer Manufacturers Communications Interface Adapter (PCMCIA) Type II standard card. This card has 68 pins, which is adequate to allow the set-top terminal to convey out-of-band signals received by a built-in modem.

The NRSS standard was designated the EIA-679. NRSS provided the basic idea for the POD module concept described in Chapter 22.

23.12 Summary

The interface between cable service and consumer electronics products affects the quality of service and the ease of use of a service. In severe cases, the interface prevents access to a service. This same problem interferes with the ease of use of some of the functions and features of consumer electronics products. These problems are complex because of the wide variety of services, functions, and features possible with modern methods.

Endnotes

1. Walter Ciciora, Consumer Electronics Friendly Cable, *1992 NCTA Convention Technical Papers*. NCTA, Washington, DC.
2. David Large, Options for Friendly Delivery of Cable Services. *CED*, April 1993, pp. 62–74, 83.
3. Walter Ciciora, The Realities of the Consumer Electronic Compatibility Interface Regulations, *1995 NCTA Convention Technical Papers*. NCTA, Washington, DC.
4. David Large, *Connecting Cable Systems to Subscribers' TVs and VCRs—Guidelines for the Cable Television Industry*. NCTA Engineering Committee, NCTA, Washington, DC, January 1987.
5. Walter Ciciora, Regulation and the Continuing Struggle for Consumer Electronic Compatibility, *1995 Montreux International Television Symposium Technical Record*. Montreux, Switzerland, June 1995, pp. 597–611.

6. Walter Ciciora, The 1992 Cable Act and the Consumer Electronics Interface, *1994 NCTA Convention Technical Papers*. NCTA, Washington, DC.
7. Cable Television Consumer Protection and Competition Act of 1992, House of Representatives Report 102–862, 102nd Congress, 2nd Session.
8. Telecommunications Act of 1996, House of Representatives Report 104–458, 104th Congress, 2nd Session.
9. Walter Ciciora, The Realities of the Retail Sale of "Navigation Devices," *1996 NCTA Convention Technical Papers*. NCTA, Washington, DC.
10. *EIA/IS-15-A, Standard Baseband (Audio/Video) Interface Between NTSC Television Receiving Devices and Peripheral Devices*, March 1989. Electronic Industries Association, Engineering Department, Arlington, VA.
11. *The EIA-653 Decoder Interface Standard*, 1990. Electronic Industries Association, Arlington, VA. Available from Global Engineering Documents, Englewood, CO, 800–854–7179.
12. *EIA-542 Cable Television Channel Identification Plan*, 1997. Electronic Industries Association, Engineering Department, Arlington, VA. Available from Global Engineering Documents, Englewood, CO, 800–854–7179.
13. *EIA/IS-105 Decoder Interface Standard*, 1997. Electronic Industries Association, Engineering Department, Arlington, VA. Available from Global Engineering Documents, Englewood, CO, 800–854–7179.
14. *EIA/IS-679 National Renewable Security Standard*, 1997. Electronic Industries Association, Engineering Department, Arlington, VA. Available from Global Engineering Documents, Englewood, CO, 800–854–7179.
15. *EIA/IS-23 RF Interface Specification for Television Receiving Devices and Cable Television Systems*, 1997. Electronic Industries Association, Engineering Department, Arlington, VA. Available from Global Engineering Documents, Englewood, CO, 800–854–7179.

Other References of Interest

David Large, Defining "Cable Ready" Consumer Equipment. *CED*, April 1993, pp. 84–90.

EIA-608-A Recommended Practice for Line 21 Data Service, 1994. Electronic Industries Association, Engineering Department, Arlington, VA. Available from Global Engineering Documents, Englewood, CO, 800–854–7179.

Several feature articles on the decoder interface. *Communications Technology*, August 1989.

The Consumer Interface Problem, A Status Report, CableLabs, Louisville, CO, February 1990.

Customer Premises Equipment Performance and Compatibility Testing, CableLabs, Louisville, CO, 1993.

Subjective Assessment of Cable Impairments on Television Picture Quality, CableLabs, Louisville, CO, 1992.

Chapter 24
Equipment Compatibility

24.1 Introduction

The previous chapter described the nature of the challenges regarding the interface between cable and consumer electronics equipment. That chapter concentrated on the way the equipment is connected and used, as well as on the impact on usability of features and services. The legal requirements and FCC rules pertaining to usage of the equipment were also covered.

This chapter will cover the minimum technical requirements for consumer-owned equipment that is to be connected to cable. There are two broad aspects to this issue: (1) the quality of the pictures and sound delivered to the subscriber, and (2) the potential for interference with other devices connected to cable and with other users of the radio spectrum. The first aspect is considered a marketplace issue to be determined by the purchaser of the equipment. But this aspect of compatibility unfortunately also affects the enjoyment of cable-delivered services and, to some extent, the likelihood that the consumer will subscribe to premium services. Often, the consumer purchases his or her equipment in an environment where its performance on cable is impossible to judge and factor into the purchase decision.

24.1.1 Broadcast Versus Broadband Receivers

The technical requirements of tuners for broadcast channels versus broadband cable channels are significantly different. It is possible to design an excellent broadcast tuner that performs poorly on cable channels, and vice versa. Designing a cost-effective tuner that performs well in both environments is a significant challenge.

The broadcast environment consists of channels with gaps. Local channels are almost never in adjacent frequencies. However, the viewer may wish to receive a weak signal with frequency between two strong local signals. The weak signal may be from a distant location. This is called *fringe reception* and was an important selling feature before cable became popular. The tuner must function well in that environment. Note that the lower channel numbers, the

VHF channels, are designated by the FCC so that the frequency of no one channel is a second harmonic of any other. This reduces the requirements on linearity in the tuner circuits.

Broadcast channels are not located on an "image" channel of another broadcast channel. The image channel problem will be described later in this chapter. Image interference problems are avoided in the broadcast environment. An inspection of the frequencies assigned to the VHF channels reveals that no channels are assigned so that they are images of valid channel frequencies. The higher channel numbers, the UHF channels, have an elaborate set of "taboos" in the FCC's rules that are intended to relax the design requirements of practical consumer electronics tuners. Finally, in the broadcast environment, there is no particular need to prevent direct pickup of the broadcast channel. Direct pickup simply adds to the signal brought in from the antenna and causes no harm. Also, since only broadcast channel frequencies are found in the tuner, there is no need to prevent their radiation from the tuner terminals or from the tuner itself.

The cable tuner is very different. The cable spectrum is fully loaded with no significant gaps up to the highest frequency. The signal levels are strong and relatively uniform across the frequency band. The challenge is in handling a lot of energy and in avoiding distortion and cross modulation of the signals.

A consequence of these differences is that a properly designed cable tuner will probably not perform well in the broadcast environment, and a properly designed broadcast tuner will probably not perform well in the cable environment.

Essentially, all cable converter tuners are double conversion tuners. That is, they apply the superheterodyne principle twice. They have two local oscillators and two intermediate frequency amplifiers and filters. This gives improved selectivity and blocks the image channel. Also, the first local oscillator frequency is usually above the highest channel tuned, and the second local oscillator signal is isolated from the cable by the first conversion stage. So backfeed of internally generated spurious signals is not a problem with cable converters. The use of a double conversion tuner in a "cable-ready" TV or VCR is extremely rare. Nearly all consumer electronics tuners are single conversion devices.

The deficiencies of "cable-ready" tuners will become more apparent as cable systems expand bandwidths beyond 450 MHz on to 750 MHz and even 1.0 GHz and beyond.

24.2 Levels of Interference

Four kinds of interference have been considered. In order of severity, they are (1) interference with the product's own performance, (2) interference with other products in the same home, (3) interference with another subscriber's reception, and (4) interference with other users of the electromagnetic spectrum.[1]

Interference with the product's own performance results in a dissatisfied customer for the product and probably for the services delivered. Unfortunately,

in many cases, the poor performance of the equipment can be confused with a problem with the service. This results in unnecessary and unproductive service calls, missed work while waiting for the service call, and the potential for finger-pointing between the cable service provider and the retailer of the consumer product. Too often the consumer gives up. The problem is not resolved, and the consumer is suspicious of both the consumer electronics product and the cable operator.

The next level of severity occurs when the product backfeeds interfering signals into the cable, causing problems with another TV or VCR in the same home. This is often a transitory problem occurring only when certain combinations of channels are tuned by the two devices. This is likely to be blamed on the cable operator. Yet one step worse is when another subscriber suffers interference from a product not even in the same home. This will almost always be blamed on the cable operator.

The most severe situation is when interference is caused to other users of the radio frequency spectrum, particularly safety-related users such as aircraft communications and navigation, fire, police, and emergency communications. This level of interference goes beyond merely reducing the entertainment value of cable and can become dangerous to others' well-being.

24.3 Direct Pickup Interference

One of the most troublesome problems is direct pickup (DPU) interference.[2] DPU in equipment connected to cable is caused by inadequate shielding of tuner and related circuits and loose connections that allow signal ingress on channel frequencies. A common DPU occurrence is the location of a receiver in the strong field of a transmitter whose signal is also carried on cable at or near its broadcast frequency. Since the propagation time in cable is more than the propagation time in free space, the over-the-air signal gets to the receive site first. A short while later, the cable signal arrives. The two mix in the receiver, and a deteriorated picture results. The consequences can range from a usable but degraded picture to a loss of the channel. The degradation can take the form of a ghost of the channel and/or beats between offset visual carriers. If the interference is relatively weak, only the sync pulse will be visible as a vertical bar in the picture. If the broadcast signal is not carried on the same frequency in the cable, the cable video will be different from the ingress signal. The result can be a moving background pattern. When the pattern is weak, the sync bar of the interferer wipes through the picture like a windshield wiper. When the interfering signal is strong, both pictures may be visible moving past each other. Depending on the phase of the ghost's visual carrier, the image may appear as a distorted color negative.

DPU is not limited to the simultaneous reception of a broadcast television signal over the air and through cable. In other parts of the spectrum, signal such as from amateur radio or pagers can make its way into the receiver and cause

problems. In these cases, the interference manifests itself as alternating dark and light bands that usually move through the picture.

It is important to appreciate that a receiver properly designed for broadcast reception may have no need for extensive shielding. It may perform well in the broadcast environment. When connected to cable, the result might be unacceptable. The consumer may then draw the wrong conclusion. He or she may deduce that since antenna operation yields good performance and the connection to cable has difficulties, there must be something wrong with the cable system. A technically sophisticated observer is not likely to fall prey to such a failure in logic. An intelligent, but nonexpert, consumer is likely to make such a mistake and complain to the cable operator.

24.3.1 Estimate of Exposure

CableLabs and the EIA funded research by consultants to study the problem. Stern Telecommunications Corporation of New York City did a study of the potential impact of DPU. The study included a literature search, measurement of representative recent "cable-ready" receivers, and a computer model of transmitter power levels at various distances from the transmitter.

The receiver testing was done at the Carl T. Jones Corporation in Springfield, Virginia, in July 1992. Five receivers were tested. The testing revealed that there is a range of susceptibility to DPU. This susceptibility is a complex phenomenon that depends on the orientation of the receiver relative to the signal's field, the way in which the coaxial cable and the power cord are oriented with respect to the signal's field, and the channel involved.

The computer modeling considered the 10 top broadcast-relevant geographic areas, called areas of dominant influence (ADIs). The ADI terminology is important in considering advertising reach and effectiveness. For the purposes of this study, the ADIs have well-documented data on the distribution and density of households. This geographic data is important to the estimates of exposure to DPU. The 10 ADIs include about 30% of U.S. households. The field strength contours were computed with a commercially available program by Communications Data Service, Inc., which is based on the FCC's propagation curves. The study considered only UHF channels up to channel 27 (visual carrier at 549.25 MHz) on the assumption that most cable systems are limited to 550 MHz. As such, the study is a little out of date, but the general trend is clear. Field strengths of 80 dBu (10 mV/m), 90 dBu (31.6 mV/m), 100 dBu (100 mV/m), and 120 dBu (1 V/m) were considered.

Data was presented for each of the 10 top ADIs and extrapolated to all U.S. households. Table 24.1 lists the percentage of all U.S. households that may be subject to at least the field strengths listed from stations operating up to 550 MHz.

This data shows that the previously used Canadian specification of 100 mV/m will leave a significant number of receivers vulnerable to DPU. Even at an

Table 24.1 Exposure to various levels of fields

Frequency Band	*80 dBu 10 mV/m*	*90 dBu 31.6 mV/m*	*100 dBu 100 mV/m*	*120 dBu 1 V/m*
VHF	61.5%	41.0%	19.9%	0.9%
VHF and UHF	65.4%	54.8%	40.8%	6.0%

immunity level set at 1 V/m, a modest number of subscribers will still either have to accept DPU or take a set top terminal with better shielding.

24.3.2 Estimate of Susceptibility

In December 1992, CableLabs solicited proposals to develop test methods for measuring DPU susceptibility and to measure a representative grouping of current products used in the home. The Carl T. Jones Corporation was hired for both of these tasks. Thirty-five television receivers, 8 VCRs, and 13 cable set-top terminals were tested for DPU susceptibility. The products were subjected to fields of increasing strength until a DPU 55 dB below the desired channel's sync tip was achieved. The 55-dB figure was confirmed to be reasonable by perception of interference studies. The discussion in Chapter 2 on the W curve and DPU indicated the 1993 tests generally confirmed the results of the 1986 CBS experiments. Across the three channels, four frequencies, two carrier-to-noise ratios, and two desired signal levels, the average value of the sine wave strength attenuation relative to the carrier strength to achieve "just perceptible" impairment levels was 56 dB.

Table 24.2 summarizes the results for the sample set of products. The mV/m figures required to breach the −55-dB threshold for slightly worse than "just perceptible" interference are listed for the best, worst, and median products. As can be seen, for the 1993 model products tested, a wide range of results were obtained.

An important difference between set top terminals supplied by cable operators and consumer electronics products is that the cable technician has several

Table 24.2 Performance of units tested

Susceptibility	*TV*	*VCR*	*Set-Top Terminal*
Best	193.4 mV/m	162.73 mV/m	732.33 mV/m
Median	51.57 mV/m	57.27 mV/m	292.73 mV/m
Worst	1.17 mV/m	29.9 mV/m	77.23 mV/m
Best/Worst	165	5.4	9.5
Sample Size	35	8	14

units with him or her and will keep trying them until the DPU problem is solved. Since the consumer purchases a TV or VCR and takes it home, the ability to pick the best from a collection is not available. Also, when the consumer makes a purchase decision, that selection will usually be in a different neighborhood from his or her residence. The retailer may not be using a cable feed for display purposes. Often, a laser disc is used as the source in order to provide the best possible pictures to encourage a purchase.

24.4 Receiver Performance

The Joint Engineering Committee (JEC) of the National Cable Television Association (NCTA) Engineering Committee and Consumer Electronics Manufacturers Association (CEMA) of the Electronic Industries Association (EIA) developed an RF Interface Specification for Television Receiving Devices and Cable Television Systems. This specification has the designation EIA-IS-23.[3] Eight parameters of customer premises equipment are important to this specification: (1) reradiation of cable signals, (2) local oscillator leakage and backfeed, (3) A/B switch isolation, (4) DPU backfeed, (5) VCR through-loss, (6) adjacent channel rejection, (7) image rejection, and (8) tuner overload distortions.

Carl T. Jones Corporation developed tests for these eight parameters. The test procedures were reviewed by the JEC in October 1993. Although no formal acceptance of the test procedures was solicited, comments were noted, and the tests adjusted where appropriate. There were two goals for this work: to develop test procedures for these parameters, and to determine the current state of the art.[4]

24.4.1 Reradiation of Cable Signals

Cable systems operate in an enclosed spectral environment. If this environment is sealed, the same frequencies that are used in the over-the-air spectrum for other purposes can be reused for cable channels. In fact, multiple enclosed cable systems can be simultaneously operated to further increase the number of signals delivered to subscribers. As long as the cable system remains sealed, no harm is done to other users of the over-the-air spectrum.

A potential source of signal leakage exists in any equipment connected to the cable system. This is because the entire cable spectrum is conveyed to the input terminals of the connected device. If leakage of these signals occurs because of inadequate shielding of the circuits, the potential for interference with other users of the spectrum exists. This is especially serious in the case of frequencies used for aircraft navigation and communications and for emergency messages. Cable systems occupy these frequencies in their closed systems and cause no harm as long as the integrity of the system remains intact.

Equipment connected to the cable system must also be designed to prevent signal leakage. If this equipment leaks signals, interference may result. Set top

terminals purchased by cable operators are specifically designed to prevent such leakage. The output of these set top terminals is just one television channel that is on a frequency not used locally for over-the-air service, usually channel 3 or 4. If leakage of this channel occurs from the TV or VCR connected to it, no harm is done since this is an unused local frequency. Consumer electronics products designed for broadcast applications do not have this requirement. When used for broadcast reception, they do not tune frequencies that are used for other purposes; they just tune the broadcast channels. Because of these differences, it is important to impose additional requirements on products connected to cable and prevent the connection of equipment that is not specifically designed to avoid leakage.

The FCC rules, Part 76.605 (a)12, limit the leakage or egress of RF energy. Cable systems must not emit energy that causes more than 15 μV per meter at a distance of 30 meters for frequencies less than 54 MHz or greater than 216 MHz. Between these frequencies, the limit is 20 μV per meter at 3 meters distance.

Of the units tested, only two television receivers failed.

24.4.2 Local Oscillator Leakage and Backfeed

Television receivers generate internal signals that are used to process the received spectrum. In addition, other internal signals are generated to display information on the screen and to operate functions and features of the product. Switch mode power supplies are also a source of interference. If these signals backfeed into the cable system, they can cause interference and degradation of performance in other TVs and VCRs connected to the cable system. They can add to the background noise, which makes two-way cable operation difficult. The other TVs and VCRs on the same signal splitter in the same residence are most at risk since there is a minimum of attenuation of the interfering signals. However, neighbors' receivers on the same cable system tap are also subject to this interference. Consequently, signal backfeed is carefully specified in equipment used by the service provider. Consumer equipment requires the same degree of care if problems are to be avoided.

The local oscillator in the single conversion tuners commonly used in TVs and VCRs operates at 45.75 MHz above the desired channel's visual carrier frequency. This frequency is in the frequency range of the cable system for most of the channels tuned. If this signal leaks back into the cable system, interference will result to other receivers according to the W curve discussed in Chapter 2. Fortunately, the splitters used by cable operators are required to have a minimum of 18-dB isolation between ports and usually have 20 dB or more. Thus if the local oscillator or other signal backfeed is no worse than 35 dB below the visual carrier, the 20 dB of the splitter will increase the signal attenuation to 55 dB below the visual carrier when it reaches the other receiver connected to the same splitter. At this level, it will be "just perceptible," but not annoying. Any further losses in the connecting cable will also help. Cable set-top terminals avoid this problem by using double conversion tuners, with the first local oscillator frequency higher than the cable spectrum tuned.

24.4.3 A/B Switch Isolation

A/B switches are part of the input of many receivers and VCRs, and allow remote-controlled selection of input sources. These sources include external antennas; cable signals; and the outputs of set top terminals, video games, and other playback devices. If insufficient isolation exists between ports of the switch, cross feed of signals will occur. The FCC regulations dated 1987 in Part 15.606 require a minimum of 80-dB isolation for frequencies between 54 and 216 MHz; 60 dB is required above that to 550 MHz. Since 1987, cable systems have gone beyond 550 MHz. Switches that have inadequate isolation at these higher frequencies will be a source of difficulty.

When an antenna is used, two kinds of problems may result. Antenna input signals can exceed +25 dBmV, whereas cable signals can be as low as 0 dBmV — lower if the subscriber does his or her own wiring and has split the signal excessively without appropriate amplification. In those situations, the isolation must be adequate to prevent interference. This explains why the cable industry uses switches that generally exceed 90 dB and often 100 dB at all used frequencies. The other problem arises when strong cable signals are not sufficiently isolated from a rooftop antenna. If cross feed of the cable signal occurred in many homes, the cumulative effect of efficient radiators all pointing in the same direction, that is, toward the broadcaster's transmitting tower, could yield unacceptable interference to aircraft traversing that path.

Although most of the devices tested had no problems with this specification, there were three instances that failed.

24.4.4 DPU Backfeed

The DPU backfeed issue is related to the DPU phenomenon but is not the same problem. A receiver that has a DPU problem may also backfeed such signals into the cable to cause difficulties with other devices. Again, there is a benefit from the 20-dB isolation between signal splitter ports. DPU backfeed of less than −35 dB will be at the perceptible but not annoying point.

Unfortunately, 37% of TVs had DPU backfeed above the perceptibility point on channel 6, 34% on channel 12, 26% on channel 78, and 17% on channel 59. Understandably, units with DPU problems had DPU backfeed problems as well.

24.4.5 VCR Through-Loss

Designing VCRs with adequate playback signal quality is a difficult technical challenge. A technique used in early VCRs to mitigate the problem was to unsymmetrically split the signal at the input terminals of the device, giving the larger share to the recording circuits and shortchanging the "through" side of the splitter. The through side of the splitter provides an attenuated version of the entire input spectrum to the RF output terminal of the VCR when neither playback nor the VCR tuner is used. The attenuation is due to the use of a splitter

without first amplifying the signal. This trick helped the noise performance of the recording circuits while providing a less disadvantageous comparison for the viewer when watching unrecorded materials.

As the last chapter discussed, the most common configuration of set-top terminal, VCR, and TV connects the cable system to the set-top terminal, which is connected to the VCR, which is connected to the TV. The TV suffers the additional through-loss of the VCR. This through-loss should be no greater than 5 dB. All eight tested VCRs passed this hurdle.

24.4.6 Adjacent Channel Rejection

The broadcast environment avoids adjacent channels. Cable utilizes as much of the spectrum as possible with nearly all channels having adjacent neighbors. An important issue for cable tuners is the ability to separate the desired channel from the upper adjacent visual carrier and the lower adjacent aural and chroma carriers. Cable practice is to operate the aural carriers some 17 dB below the visual carriers. This additional attenuation relaxes the difficulty somewhat. The criteria again is the 55 dB used to bring the signal to the not-annoying level.

The passband of the receiver is a combination of the IF response and the tuner response, which differs slightly from channel to channel and is under the influence of the automatic gain control (AGC) strategy. Consequently, devices may have isolated points of failure on some channels but not on others.

Of the devices tested, three television receivers and one VCR failed to sufficiently reject the lower adjacent color subcarrier. Two television receivers failed the aural rejection test on all channels, and two others had problems on just two channels. Of the eight VCRs, three had no adjacent aural problems, three others failed on the four channels tested, one failed on three channels, and one unit had difficulties with just one channel. All TVs and VCRs behaved adequately with the upper channel's visual carrier. Not surprisingly, cable set-top terminals passed all tests since they are specifically designed for such service.

24.4.7 Image Rejection

Image channels are a result of the superheterodyne receiver design, which uses a balanced mixer with two inputs followed by a filter for the purpose of separating out the desired signal from the spectrum at its input terminals. The inputs to the balanced mixer are the received spectrum and a local oscillator. The output of the mixer consists of the sum and differences of the input frequencies.

TV local oscillators for VHF are 45.75 MHz above the desired visual carrier. The difference between the local oscillator frequency and the desired frequency is the intermediate frequency (IF), which passes through the filter characteristic of the IF amplifier. The signal is then detected and becomes the source for the picture and sound selected by the viewer. The sum of the local oscillator and the desired channel is blocked by the filtering of the IF amplifier.

There is another difference frequency equal to 45.75 MHz. Signals that are 45.75 MHz above the local oscillator frequency will also have a difference of 45.75 MHz relative to the local oscillator frequency. This signal, if present at the input of the mixer, will be heterodyned to a frequency that will pass through the filter characteristic of the IF amplifier and provide interference to the picture and/or sound of the desired program. The visual carrier of the desired channel is 45.75 MHz below the local oscillator frequency. The image signal will be 45.75 MHz above the local oscillator frequency or 91.5 MHz above the visual carrier of the desired channel, about 15 6-MHz channels away. The visual carrier is the strongest part of the interfering channel. Visual carriers are spaced 6 MHz apart in cable practice. The visual carrier 15 channels above the desired carrier will be 90 MHz higher in frequency. When local oscillator frequency is subtracted from it, the visual carrier in the image region will be 1.5 MHz above the desired visual carrier and produce an interfering beat in the picture. The aural carrier of the channel that is 14 channels above the desired channel will also fall into the video bandwidth of the desired channel.

Since the tuner determines the rejection at the image frequency, the level of rejection varies with the channel. All receivers did well on channel 3, and only four had problems with channel 12. Half the receivers had interference on channel 53, and 67% failed on channel 74.

24.4.8 Tuner Overload Performance

Tuner problems and other deficiencies may result in impaired performance on the TV or VCR itself. In addition to the DPU problem previously described, nonlinearity in the receiver's circuits can give rise to the production of new frequencies not present in the original signal. The signals thus generated produce either diagonal bars in the picture or even a discernable image from another channel.

Tuner overload is a serious problem that will become more serious as cable bandwidth expands. Of the TVs, 66% had problems on at least one channel, 50% of VCRs suffered as well. Surprisingly, even 36% of the set-top terminals had troubles.

24.4.9 Rejection of Signals in the Upstream Band

Two-way services, once rare, are now commonly offered by cable operators. High-speed data service, two-way video set-top boxes, sometimes telephone interface units, and various types of status monitoring combined fill up a bandwidth whose upper limit has increased from 30 to about 42 MHz in recent years to meet the demand for additional upstream capacity. The result is that signals are generated at fairly high levels within customers' homes and transmitted toward the headend. Due to the finite isolation offered by taps and in-house splitters, some of that energy is directed toward the input terminals of television receiving devices.

The problems are twofold. First, the amplitudes may be sufficient to cause input stages to distort, producing intermodulation products that may fall within the normally tuned channels. Second, the upper frequencies fall within the common television intermediate frequency band of 41–47 MHz and may leak directly into sensitive amplifier stages if not sufficiently rejected by the input circuits.

The magnitude of the reflected signals can be estimated as follows. The power level of transmitted signals from standard cable modems are controlled by the headend and must be adjustable to as high as +58 dBmV to meet the DOCSIS specification (though some models can go higher). For reasons discussed in Chapter 16, higher levels are consistent with reduced average levels of ingress, and cable operators design their plants so that as many modems as possible are transmitting near their maximum capabilities. Using the assumption in Section 24.4.2 that the sum of cable loss and tap or splitter isolation will be at least 20 dB, the maximum reflected power at the antenna terminals of a television receiver in the same or a neighboring house can be as high as +38 dBm.

The presence of signals of that magnitude falling below 30 MHz should not be a problem because television receivers are designed to reject signals below that frequency with magnitudes as high as −7 dBm (+ 42 dBmV). At the other extreme, some receivers offer little protection for signals falling at or above 41 MHz, with levels as low as +10 dBmV causing visible interference in some models. IS-23 is silent on the issue of receiver protection above 30 MHz and also silent regarding maximum levels in that range that may be delivered by cable operators. Clearly, however, without some form of additional protection or avoidance of at least the frequencies above 41 MHz, many receivers will generate unacceptable levels of picture degradation when nearby two-way devices are transmitting near the upper end of the upstream band. The testing described earlier did not include this critical parameter.

24.5 Regulatory Status

The issue of consumer equipment compatibility was addressed in Section 17 of the 1992 Cable Act, which directed the Federal Communications Commission to create rules[5,6] "to specify the technical requirements with which a television receiver or video cassette recorder must comply in order to be sold as 'cable compatible' or 'cable ready.'"

In response, the FCC issued a Notice of Inquiry in January 1993 requesting input on a variety of consumer equipment interface issues, including the issue of "cable-ready" receivers.[7] Almost a full year later (December 1993), it issued a Notice of Proposed Rulemaking (NPRM) covering the same issues. This time, the key elements of draft EIA-23 were jointly submitted by the Electronic Industries Association and the NCTA,[8] supported by the CableLabs compatibility testing results (submitted by InterMedia Partners).

Finally, in May 1994 the FCC issued rules[9] defining the minimum performance characteristics of equipment marketed as "cable ready" or "cable compatible." These were amended in April 1996. While largely adopting some of the provisions in the interindustry agreement, the commission made several critical changes. First, they declined to adopt a shielding standard that would ensure that reradiation of cable system signals from connected receivers would be below the limits allowed cable systems. Second, they declined to control any aspects of the quality of devices with extended-range tuners specifically designed to tune cable channels, unless they were labeled specifically as "cable ready" or "cable compatible," even though some of the requested quality provisions existed purely to ensure the integrity of the system to which they were attached. As a practical matter, there has been no enforcement of even these weakened provisions.

24.6 Summary

There is a substantial difference in the technical requirements for good performance in the broadcast environment and in the cable environment. Receivers that perform well in one environment will likely have difficulties in the other. Designing a cost-effective product to do well with both types of signals is very difficult. We have described the difficulties and the consequence of failure. Testing reveals that much of the equipment available at the time of the tests did not make the grade in several aspects.

In summary, the relevant parameters for equipment connected to cable are at a minimum:

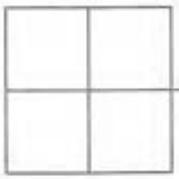
- Immunity to external fields of at least 1 V/m.
- No more reradiation of cable signals than is allowed by FCC leakage rules.
- At least 80-dB isolation between switch ports.
- DPU backfeed into the cable system at least 35 dB below the desired visual carrier.
- Through-switch attenuation of no more than 5 dB.
- Rejection of adjacent channel carriers to levels at least 55 dB below the desired visual carrier.
- Image signal rejection of at least 55 dB below the desired visual carrier.
- Tuner overload performance so that spurious signals are at least 55 dB below the desired visual carrier.

The challenge of making products that function well in both environments is a heavy one. There is no enforced set of standards for awarding the title "cable ready." There is no compliance testing service. Much of the consumer's

difficulty stems from not appreciating that extended tuning range television receivers are not necessarily fully compatible with properly designed and maintained cable television systems.

Endnotes

1. *The Consumer Interface Problem, A Status Report*, CableLabs, Louisville, CO, February 1990.
2. *Customer Premises Equipment Performance and Compatibility Testing*, CableLabs, Louisville, CO, 1993.
3. *EIA/IS-23 RF Interface Specification for Television Receiving Devices and Cable Television Systems*, 1997. Electronic Industries Association, Engineering Department, Arlington, VA. Available from Global Engineering Documents, Englewood, CO, 800-854-7179.
4. David Large, Defining "Cable-Ready" Consumer Equipment. *CED*, April 1993, pp. 84–90.
5. *Cable Television Consumer Protection and Competition Act of 1992*, House of Representatives Report 102-862, 102nd Congress, 2nd Session.
6. Walter Ciciora, The 1992 Cable Act and the Consumer Electronics Interface, *1994 NCTA Convention Technical Papers*. NCTA, Washington, DC.
7. The response of Intermedia Partners, authored by David Large, cochair of the IS-23 (released as EIA-23) Subcommittee, is summarized in: David Large, op. cit.
8. Summarized in: Joseph Van Loan and David Large, Tuner Characteristics of Cable-Ready Receivers, *1994 NCTA Technical Papers*, NCTA, Washington, DC.
9. CFR 47 §15.118.



Appendix A
Channel Allocation

The channel allocation shown here conforms to the requirements of CEA 542-B, which specifies the correspondence between channel number and frequency for North America. This standard was developed by the joint EIA (now CEA)/NCTA engineering committee, and the original version has been adopted by the FCC as the standard for channel allocations in the United States. (As of this writing the FCC Rules refer to the standard by its old working designation, IS-6.) Picture carrier frequencies are shown, even for QAM signals, which have but do not transmit, a carrier. The channel extends from 1.25 MHz below the analog picture carrier, to 4.75 MHz above. For QAM signals, the carrier is in the center of the channel. For 8-VSB signals, the carrier is 310 kHz above the low band edge. For all digital modulation, there is a single carrier, not separate carriers for picture and sound. Three channelization plans are shown. They are explained in section 9.2.

The complete standard is available from the Consumer Electronics Association's duplication contractor, Global Engineering, at 800-854-7179, http://global.ihs.com/. This is reprinted with permission of the Consumer Electronics Association.

	Channel Edge			Analog Picture Carrier			Digital Carrier					
Channel	Low (MHz) (Standard and IRC)	High (MHz)	Note	STD (MHz)	IRC (MHz)	HRC (MHz)	STD 8/16VSB (MHz)	STD QAM (MHz)	IRC 8/16VSB (MHz)	IRC QAM (MHz)	HRC 8/16VSB (MHz)	HRC QAM
1	72.0000	76.0000	+++		73.2625	72.0036			72.3100	75.0000	71.0600	73.7500
2	54.0000	60.0000		55.2500	55.2625	54.0027	54.3100	57.0000	54.3100	57.0000	53.0600	55.7500
3	60.0000	66.0000		61.2500	61.2625	60.0030	60.3100	63.0000	60.3100	63.0000	59.0600	61.7500
4	66.0000	72.0000		67.2500	67.2625	66.0033	66.3100	69.0000	66.3100	69.0000	65.0600	67.7500
5	76.0000	82.0000		77.2500	79.2625	78.0039	76.3100	79.0000	78.3100	81.0000	77.0600	79.7500
6	82.0000	88.0000		83.2500	85.2625	84.0042	82.3100	85.0000	84.3100	87.0000	83.0600	85.7500
7	174.0000	180.0000		175.2500	175.2625	174.0087	174.3100	177.0000	174.3100	177.0000	173.0600	175.7500
8	180.0000	186.0000		181.2500	181.2625	180.0090	180.3100	183.0000	180.3100	183.0000	179.0600	181.7500
9	186.0000	192.0000		187.2500	187.2625	186.0093	186.3100	189.0000	186.3100	189.0000	185.0600	187.7500
10	192.0000	198.0000		193.2500	193.2625	192.0096	192.3100	195.0000	192.3100	195.0000	191.0600	193.7500
11	198.0000	204.0000		199.2500	199.2625	198.0099	198.3100	201.0000	198.3100	201.0000	197.0600	199.7500
12	204.0000	210.0000		205.2500	205.2625	204.0102	204.3100	207.0000	204.3100	207.0000	203.0600	205.7500
13	210.0000	216.0000		211.2500	211.2625	210.0105	210.3100	213.0000	210.3100	213.0000	209.0600	211.7500
14	120.0000	126.0000		121.2625	121.2625	120.0060	120.3100	123.0000	120.3100	123.0000	119.0600	121.7500
15	126.0000	132.0000		127.2625	127.2625	126.0063	126.3100	129.0000	126.3100	129.0000	125.0600	127.7500
16	132.0000	138.0000		133.2625	133.2625	132.0066	132.3100	135.0000	132.3100	135.0000	131.0600	133.7500
17	138.0000	144.0000		139.2500	139.2625	138.0069	138.3100	141.0000	138.3100	141.0000	137.0600	139.7500
18	144.0000	150.0000		145.2500	145.2625	144.0072	144.3100	147.0000	144.3100	147.0000	143.0600	145.7500
19	150.0000	156.0000		151.2500	151.2625	150.0075	150.3100	153.0000	150.3100	153.0000	149.0600	151.7500
20	156.0000	162.0000		157.2500	157.2625	156.0078	156.3100	159.0000	156.3100	159.0000	155.0600	157.7500
21	162.0000	168.0000		163.2500	163.2625	162.0081	162.3100	165.0000	162.3100	165.0000	161.0600	163.7500
22	168.0000	174.0000		169.2500	169.2625	168.0084	168.3100	171.0000	168.3100	171.0000	167.0600	169.7500
23	216.0000	222.0000		217.2500	217.2625	216.0108	216.3100	219.0000	216.3100	219.0000	215.0600	217.7500
24	222.0000	228.0000		223.2500	223.2625	222.0111	222.3100	225.0000	222.3100	225.0000	221.0600	223.7500
25	228.0000	234.0000		229.2625	229.2625	228.0114	228.3100	231.0000	228.3100	231.0000	227.0600	229.7500
26	234.0000	240.0000		235.2625	235.2625	234.0117	234.3100	237.0000	234.3100	237.0000	233.0600	235.7500
27	240.0000	246.0000		241.2625	241.2625	240.0120	240.3100	243.0000	240.3100	243.0000	239.0600	241.7500
28	246.0000	252.0000		247.2625	247.2625	246.0123	246.3100	249.0000	246.3100	249.0000	245.0600	247.7500
29	252.0000	258.0000		253.2625	253.2625	252.0126	252.3100	255.0000	252.3100	255.0000	251.0600	253.7500
30	258.0000	264.0000		259.2625	259.2625	258.0129	258.3100	261.0000	258.3100	261.0000	257.0600	259.7500

31	264.0000	270.0000	265.2625	265.2625	264.0132	264.3100	267.0000	264.3100	267.0000	263.0600	265.7500
32	270.0000	276.0000	271.2625	271.2625	270.0135	270.3100	273.0000	270.3100	273.0000	269.0600	271.7500
33	276.0000	282.0000	277.2625	277.2625	276.0138	276.3100	279.0000	276.3100	279.0000	275.0600	277.7500
34	282.0000	288.0000	283.2625	283.2625	282.0141	282.3100	285.0000	282.3100	285.0000	281.0600	283.7500
35	288.0000	294.0000	289.2625	289.2625	288.0144	288.3100	291.0000	288.3100	291.0000	287.0600	289.7500
36	294.0000	300.0000	295.2625	295.2625	294.0147	294.3100	297.0000	294.3100	297.0000	293.0600	295.7500
37	300.0000	306.0000	301.2625	301.2625	300.0150	300.3100	303.0000	300.3100	303.0000	299.0600	301.7500
38	306.0000	312.0000	307.2625	307.2625	306.0153	306.3100	309.0000	306.3100	309.0000	305.0600	307.7500
39	312.0000	318.0000	313.2625	313.2625	312.0156	312.3100	315.0000	312.3100	315.0000	311.0600	313.7500
40	318.0000	324.0000	319.2625	319.2625	318.0159	318.3100	321.0000	318.3100	321.0000	317.0600	319.7500
41	324.0000	330.0000	325.2625	325.2625	324.0162	324.3100	327.0000	324.3100	327.0000	323.0600	325.7500
42	330.0000	336.0000*	331.2750	331.2750	330.0165	330.3100	333.0000	330.3100	333.0000	329.0600	331.7500
43	336.0000	342.0000	337.2625	337.2625	336.0168	336.3100	339.0000	336.3100	339.0000	335.0600	337.7500
44	342.0000	348.0000	343.2625	343.2625	342.0171	342.3100	345.0000	342.3100	345.0000	341.0600	343.7500
45	348.0000	354.0000	349.2625	349.2625	348.0174	348.3100	351.0000	348.3100	351.0000	347.0600	349.7500
46	354.0000	360.0000	355.2625	355.2625	354.0177	354.3100	357.0000	354.3100	357.0000	353.0600	355.7500
47	360.0000	366.0000	361.2625	361.2625	360.0180	360.3100	363.0000	360.3100	363.0000	359.0600	361.7500
48	366.0000	372.0000	367.2625	367.2625	366.0183	366.3100	369.0000	366.3100	369.0000	365.0600	367.7500
49	372.0000	378.0000	373.2625	373.2625	372.0186	372.3100	375.0000	372.3100	375.0000	371.0600	373.7500
50	378.0000	384.0000	379.2625	379.2625	378.0189	378.3100	381.0000	378.3100	381.0000	377.0600	379.7500
51	384.0000	390.0000	385.2625	385.2625	384.0192	384.3100	387.0000	384.3100	387.0000	383.0600	385.7500
52	390.0000	396.0000	391.2625	391.2625	390.0195	390.3100	393.0000	390.3100	393.0000	389.0600	391.7500
53	396.0000	402.0000	397.2625	397.2625	396.0198	396.3100	399.0000	396.3100	399.0000	395.0600	397.7500
54	402.0000	408.0000	403.2500	403.2625	402.0201	402.3100	405.0000	402.3100	405.0000	401.0600	403.7500
55	408.0000	414.0000	409.2500	409.2625	408.0204	408.3100	411.0000	408.3100	411.0000	407.0600	409.7500
56	414.0000	420.0000	415.2500	415.2625	414.0207	414.3100	417.0000	414.3100	417.0000	413.0600	415.7500
57	420.0000	426.0000	421.2500	421.2625	420.0210	420.3100	423.0000	420.3100	423.0000	419.0600	421.7500
58	426.0000	432.0000	427.2500	427.2625	426.0213	426.3100	429.0000	426.3100	429.0000	425.0600	427.7500
59	432.0000	438.0000	433.2500	433.2625	432.0216	432.3100	435.0000	432.3100	435.0000	431.0600	433.7500
60	438.0000	444.0000	439.2500	439.2625	438.0219	438.3100	441.0000	438.3100	441.0000	437.0600	439.7500
61	444.0000	450.0000	445.2500	445.2625	444.0222	444.3100	447.0000	444.3100	447.0000	443.0600	445.7500
62	450.0000	456.0000	451.2500	451.2625	450.0225	450.3100	453.0000	450.3100	453.0000	449.0600	451.7500
63	456.0000	462.0000	457.2500	457.2625	456.0228	456.3100	459.0000	456.3100	459.0000	455.0600	457.7500
64	462.0000	468.0000	463.2500	463.2625	462.0231	462.3100	465.0000	462.3100	465.0000	461.0600	463.7500
65	468.0000	474.0000	469.2500	469.2625	468.0234	468.3100	471.0000	468.3100	471.0000	467.0600	469.7500

Channel	Channel Edge Low (MHz) (Standard and IRC)	Channel Edge High (MHz) (Standard and IRC)	Note	Analog Picture Carrier STD (MHz)	Analog Picture Carrier IRC (MHz)	Analog Picture Carrier HRC (MHz)	Digital Carrier STD 8/16VSB (MHz)	Digital Carrier STD QAM (MHz)	Digital Carrier IRC 8/16VSB (MHz)	Digital Carrier IRC QAM (MHz)	Digital Carrier HRC 8/16VSB (MHz)	Digital Carrier HRC QAM
66	474.0000	480.0000		475.2500	475.2625	474.0237	474.3100	477.0000	474.3100	477.0000	473.0600	475.7500
67	480.0000	486.0000		481.2500	481.2625	480.0240	480.3100	483.0000	480.3100	483.0000	479.0600	481.7500
68	486.0000	492.0000		487.2500	487.2625	486.0243	486.3100	489.0000	486.3100	489.0000	485.0600	487.7500
69	492.0000	498.0000		493.2500	493.2625	492.0246	492.3100	495.0000	492.3100	495.0000	491.0600	493.7500
70	498.0000	504.0000		499.2500	499.2625	498.0249	498.3100	501.0000	498.3100	501.0000	497.0600	499.7500
71	504.0000	510.0000		505.2500	505.2625	504.0252	504.3100	507.0000	504.3100	507.0000	503.0600	505.7500
72	510.0000	516.0000		511.2500	511.2625	510.0255	510.3100	513.0000	510.3100	513.0000	509.0600	511.7500
73	516.0000	522.0000		517.2500	517.2625	516.0258	516.3100	519.0000	516.3100	519.0000	515.0600	517.7500
74	522.0000	528.0000		523.2500	523.2625	522.0261	522.3100	525.0000	522.3100	525.0000	521.0600	523.7500
75	528.0000	534.0000		529.2500	529.2625	528.0264	528.3100	531.0000	528.3100	531.0000	527.0600	529.7500
76	534.0000	540.0000		535.2500	535.2625	534.0267	534.3100	537.0000	534.3100	537.0000	533.0600	535.7500
77	540.0000	546.0000		541.2500	541.2625	540.0270	540.3100	543.0000	540.3100	543.0000	539.0600	541.7500
78	546.0000	552.0000		547.2500	547.2625	546.0273	546.3100	549.0000	546.3100	549.0000	545.0600	547.7500
79	552.0000	558.0000		553.2500	553.2625	552.0276	552.3100	555.0000	552.3100	555.0000	551.0600	553.7500
80	558.0000	564.0000		559.2500	559.2625	558.0279	558.3100	561.0000	558.3100	561.0000	557.0600	559.7500
81	564.0000	570.0000		565.2500	565.2625	564.0282	564.3100	567.0000	564.3100	567.0000	563.0600	565.7500
82	570.0000	576.0000		571.2500	571.2625	570.0285	570.3100	573.0000	570.3100	573.0000	569.0600	571.7500
83	576.0000	582.0000		577.2500	577.2625	576.0288	576.3100	579.0000	576.3100	579.0000	575.0600	577.7500
84	582.0000	588.0000		583.2500	583.2625	582.0291	582.3100	585.0000	582.3100	585.0000	581.0600	583.7500
85	588.0000	594.0000		589.2500	589.2625	588.0294	588.3100	591.0000	588.3100	591.0000	587.0600	589.7500
86	594.0000	600.0000		595.2500	595.2625	594.0297	594.3100	597.0000	594.3100	597.0000	593.0600	595.7500
87	600.0000	606.0000		601.2500	601.2625	600.0300	600.3100	603.0000	600.3100	603.0000	599.0600	601.7500
88	606.0000	612.0000		607.2500	607.2625	606.0303	606.3100	609.0000	606.3100	609.0000	605.0600	607.7500
89	612.0000	618.0000		613.2500	613.2625	612.0306	612.3100	615.0000	612.3100	615.0000	611.0600	613.7500
90	618.0000	624.0000		619.2500	619.2625	618.0309	618.3100	621.0000	618.3100	621.0000	617.0600	619.7500
91	624.0000	630.0000		625.2500	625.2625	624.0312	624.3100	627.0000	624.3100	627.0000	623.0600	625.7500
92	630.0000	636.0000		631.2500	631.2625	630.0315	630.3100	633.0000	630.3100	633.0000	629.0600	631.7500
93	636.0000	642.0000		637.2500	637.2625	636.0318	636.3100	639.0000	636.3100	639.0000	635.0600	637.7500
94	642.0000	648.0000		643.2500	643.2625	642.0321	642.3100	645.0000	642.3100	645.0000	641.0600	643.7500
95	90.0000	96.0000		91.2500	91.2625	90.0045	90.3100	93.0000	90.3100	93.0000	89.0600	91.7500

96	96.0000	102.0000	97.2500	97.2625	96.0048	96.3100	99.0000	96.3100	99.0000	95.0600	97.7500
97	102.0000	108.0000	103.2500	103.2625	102.0051	102.3100	105.0000	102.3100	105.0000	101.0600	103.7500
98	108.0000	114.0000*	109.2750	109.2750	108.0250	108.3100	111.0000	108.3100	111.0000	107.0600	109.7500
99	114.0000	120.0000*	115.2750	115.2750	114.0250	114.3100	117.0000	114.3100	117.0000	113.0600	115.7500
100	648.0000	654.0000	649.2500	649.2625	648.0324	648.3100	651.0000	648.3100	651.0000	647.0600	649.7500
101	654.0000	660.0000	655.2500	655.2625	654.0327	654.3100	657.0000	654.3100	657.0000	653.0600	655.7500
102	660.0000	666.0000	661.2500	661.2625	660.0330	660.3100	663.0000	660.3100	663.0000	659.0600	661.7500
103	666.0000	672.0000	667.2500	667.2625	666.0333	666.3100	669.0000	666.3100	669.0000	665.0600	667.7500
104	672.0000	678.0000	673.2500	673.2625	672.0336	672.3100	675.0000	672.3100	675.0000	671.0600	673.7500
105	678.0000	684.0000	679.2500	679.2625	678.0339	678.3100	681.0000	678.3100	681.0000	677.0600	679.7500
106	684.0000	690.0000	685.2500	685.2625	684.0342	684.3100	687.0000	684.3100	687.0000	683.0600	685.7500
107	690.0000	696.0000	691.2500	691.2625	690.0345	690.3100	693.0000	690.3100	693.0000	689.0600	691.7500
108	696.0000	702.0000	697.2500	697.2625	696.0348	696.3100	699.0000	696.3100	699.0000	695.0600	697.7500
109	702.0000	708.0000	703.2500	703.2625	702.0351	702.3100	705.0000	702.3100	705.0000	701.0600	703.7500
110	708.0000	714.0000	709.2500	709.2625	708.0354	708.3100	711.0000	708.3100	711.0000	707.0600	709.7500
111	714.0000	720.0000	715.2500	715.2625	714.0357	714.3100	717.0000	714.3100	717.0000	713.0600	715.7500
112	720.0000	726.0000	721.2500	721.2625	720.0360	720.3100	723.0000	720.3100	723.0000	719.0600	721.7500
113	726.0000	732.0000	727.2500	727.2625	726.0363	726.3100	729.0000	726.3100	729.0000	725.0600	727.7500
114	732.0000	738.0000	733.2500	733.2625	732.0366	732.3100	735.0000	732.3100	735.0000	731.0600	733.7500
115	738.0000	744.0000	739.2500	739.2625	738.0369	738.3100	741.0000	738.3100	741.0000	737.0600	739.7500
116	744.0000	750.0000	745.2500	745.2625	744.0372	744.3100	747.0000	744.3100	747.0000	743.0600	745.7500
117	750.0000	756.0000	751.2500	751.2625	750.0375	750.3100	753.0000	750.3100	753.0000	749.0600	751.7500
118	756.0000	762.0000	757.2500	757.2625	756.0378	756.3100	759.0000	756.3100	759.0000	755.0600	757.7500
119	762.0000	768.0000	763.2500	763.2625	762.0381	762.3100	765.0000	762.3100	765.0000	761.0600	763.7500
120	768.0000	774.0000	769.2500	769.2625	768.0384	768.3100	771.0000	768.3100	771.0000	767.0600	769.7500
121	774.0000	780.0000	775.2500	775.2625	774.0387	774.3100	777.0000	774.3100	777.0000	773.0600	775.7500
122	780.0000	786.0000	781.2500	781.2625	780.0390	780.3100	783.0000	780.3100	783.0000	779.0600	781.7500
123	786.0000	792.0000	787.2500	787.2625	786.0393	786.3100	789.0000	786.3100	789.0000	785.0600	787.7500
124	792.0000	798.0000	793.2500	793.2625	792.0396	792.3100	795.0000	792.3100	795.0000	791.0600	793.7500
125	798.0000	804.0000	799.2500	799.2625	798.0399	798.3100	801.0000	798.3100	801.0000	797.0600	799.7500
126	804.0000	810.0000	805.2500	805.2625	804.0402	804.3100	807.0000	804.3100	807.0000	803.0600	805.7500
127	810.0000	816.0000	811.2500	811.2625	810.0405	810.3100	813.0000	810.3100	813.0000	809.0600	811.7500
128	816.0000	822.0000	817.2500	817.2625	816.0408	816.3100	819.0000	816.3100	819.0000	815.0600	817.7500
129	822.0000	828.0000	823.2500	823.2625	822.0411	822.3100	825.0000	822.3100	825.0000	821.0600	823.7500
130	828.0000	834.0000	829.2500	829.2625	828.0414	828.3100	831.0000	828.3100	831.0000	827.0600	829.7500

	Channel Edge			Analog Picture Carrier			Digital Carrier					
Channel	Low (MHz) (Standard and IRC)	High (MHz)	Note	STD (MHz)	IRC (MHz)	HRC (MHz)	STD 8/16VSB (MHz)	STD QAM (MHz)	IRC 8/16VSB (MHz)	IRC QAM (MHz)	HRC 8/16VSB (MHz)	HRC QAM
131	834.0000	840.0000		835.2500	835.2625	834.0417	834.3100	837.0000	834.3100	837.0000	833.0600	835.7500
132	840.0000	846.0000		841.2500	841.2625	840.0420	840.3100	843.0000	840.3100	843.0000	839.0600	841.7500
133	846.0000	852.0000		847.2500	847.2625	846.0423	846.3100	849.0000	846.3100	849.0000	845.0600	847.7500
134	852.0000	858.0000		853.2500	853.2625	852.0426	852.3100	855.0000	852.3100	855.0000	851.0600	853.7500
135	858.0000	864.0000		859.2500	859.2625	858.0429	858.3100	861.0000	858.3100	861.0000	857.0600	859.7500
136	864.0000	870.0000		865.2500	865.2625	864.0432	864.3100	867.0000	864.3100	867.0000	863.0600	865.7500
137	870.0000	876.0000		871.2500	871.2625	870.0435	870.3100	873.0000	870.3100	873.0000	869.0600	871.7500
138	876.0000	882.0000		877.2500	877.2625	876.0438	876.3100	879.0000	876.3100	879.0000	875.0600	877.7500
139	882.0000	888.0000		883.2500	883.2625	882.0441	882.3100	885.0000	882.3100	885.0000	881.0600	883.7500
140	888.0000	894.0000		889.2500	889.2625	888.0444	888.3100	891.0000	888.3100	891.0000	887.0600	889.7500
141	894.0000	900.0000		895.2500	895.2625	894.0447	894.3100	897.0000	894.3100	897.0000	893.0600	895.7500
142	900.0000	906.0000		901.2500	901.2625	900.0450	900.3100	903.0000	900.3100	903.0000	899.0600	901.7500
143	906.0000	912.0000		907.2500	907.2625	906.0453	906.3100	909.0000	906.3100	909.0000	905.0600	907.7500
144	912.0000	918.0000		913.2500	913.2625	912.0456	912.3100	915.0000	912.3100	915.0000	911.0600	913.7500
145	918.0000	924.0000		919.2500	919.2625	918.0459	918.3100	921.0000	918.3100	921.0000	917.0600	919.7500
146	924.0000	930.0000		925.2500	925.2625	924.0462	924.3100	927.0000	924.3100	927.0000	923.0600	925.7500
147	930.0000	936.0000		931.2500	931.2625	930.0465	930.3100	933.0000	930.3100	933.0000	929.0600	931.7500
148	936.0000	942.0000		937.2500	937.2625	936.0468	936.3100	939.0000	936.3100	939.0000	935.0600	937.7500
149	942.0000	948.0000		943.2500	943.2625	942.0471	942.3100	945.0000	942.3100	945.0000	941.0600	943.7500
150	948.0000	954.0000		949.2500	949.2625	948.0474	948.3100	951.0000	948.3100	951.0000	947.0600	949.7500
151	954.0000	960.0000	++	955.2500	955.2625	954.0477	954.3100	957.0000	954.3100	957.0000	953.0600	955.7500
152	960.0000	966.0000	++	961.2500	961.2625	960.0480	960.3100	963.0000	960.3100	963.0000	959.0600	961.7500
153	966.0000	972.0000	++	967.2500	967.2625	966.0483	966.3100	969.0000	966.3100	969.0000	965.0600	967.7500
154	972.0000	978.0000		973.2500	973.2625	972.0486	972.3100	975.0000	972.3100	975.0000	971.0600	973.7500
155	978.0000	984.0000		979.2500	979.2625	978.0489	978.3100	981.0000	978.3100	981.0000	977.0600	979.7500
156	984.0000	990.0000		985.2500	985.2625	984.0492	984.3100	987.0000	984.3100	987.0000	983.0600	985.7500
157	990.0000	996.0000		991.2500	991.2625	990.0495	990.3100	993.0000	990.3100	993.0000	989.0600	991.7500
158	996.0000	1002.0000		997.2500	997.2625	996.0498	996.3100	999.0000	996.3100	999.0000	995.0600	997.7500

++ These channels have been used as the 1st IF in a few TVs. The use of these frequencies may result in interference on these TVs.
+++ Channel 1 is the gap between channels 4 and 5. It is only 4 MHz wide in the standard plan and is not used for television signals. In the HRC and IRC plants it opens to 6 MHz and may be used for television transmissions.
*Excluded from comb due to FCC offset in analog HRC.