Chapter

14

Satellite Access

14.1 Introduction

A transponder channel aboard a satellite may be fully loaded by a single transmission from an earth station. This is referred to as a *single-access* mode of operation. It is also possible, and more common, for a transponder to be loaded by a number of carriers. These may originate from a number of earth stations geographically separate, and each earth station may transmit one or more of the carriers. This mode of operation is termed *multiple access.* The need for multiple access arises because more than two earth stations, in general, will be within the service area of a satellite. Even so-called spot beams from satellite antennas cover areas several hundred miles across.

The two most commonly used methods of multiple access are *frequency-division multiple access* (FDMA) and *time-division multiple access* (TDMA). These are analogous to frequency-division multiplexing (FDM) and time-division multiplexing (TDM) described in Chaps. 9 and 10. However, multiple access and multiplexing are different concepts, and as pointed out in CCIR Report 708 (1982), modulation (and hence multiplexing) is essentially a transmission feature, whereas multiple access is essentially a traffic feature.

A third category of multiple access is *code-division multiple access* (CDMA). In this method each signal is associated with a particular code that is used to spread the signal in frequency and/or time. All such signals will be received simultaneously at an earth station, but by using the key to the code, the station can recover the desired signal by means of correlation. The other signals occupying the transponder channel appear very much like random noise to the correlation decoder.

Multiple access also may be classified by the way in which circuits are assigned to users (*circuits* in this context implies one communication channel through the multiple-access transponder). Circuits may be *preassigned,* which means they are allocated on a fixed or partially fixed basis to certain users. These circuits are therefore not available for general use. Preassignment is simple to implement but is efficient only for circuits with *continuous heavy* traffic.

An alternative to preassignment is *demand-assigned multiple access* (DAMA). In this method, all circuits are available to all users and are assigned according to the demand. DAMA results in more efficient overall use of the circuits but is more costly and complicated to implement.

Both FDMA and TDMA can be operated as preassigned or demand-assigned systems. CDMA is a random-access system, there being no control over the timing of the access or of the frequency slots accessed.

These multiple-access methods refer to the way in which a single *transponder* channel is utilized. A satellite carries a number of transponders, and normally each covers a different frequency channel, as shown in Fig. 7.13. This provides a form of frequency-division multiple access to the whole satellite. It is also possible for transponders to operate at the same frequency but to be connected to different spot-beam antennas. These allow the satellite as a whole to be accessed by earth stations widely separated geographically but transmitting on the same frequency. This is termed *frequency reuse.* This method of access is referred to as *space-division multiple access* (SDMA). It should be kept in mind that each spot beam may itself be carrying signals in one of the other multiple-access formats.

14.2 Single Access

With single access, a single modulated carrier occupies the whole of the available bandwidth of a transponder. Single-access operation is used on heavy-traffic routes and requires large earth station antennas such as the class A antenna shown in Fig. 8.7. As an example, Telesat Canada provides heavy route message facilities, with each transponder channel being capable of carrying 960 one-way voice circuits on an FDM/FM carrier, as illustrated in Fig. 14.1. The earth station employs a 30-m-diameter antenna and a parametric amplifier, which together provide a minimum [G/T] of 37.5 dB/K.

14.3 Preassigned FDMA

Frequency slots may be preassigned to analog and digital signals, and to illustrate the method, analog signals in the FDM/FM/FDMA format will be considered first. As the acronyms indicate, the signals are fre-

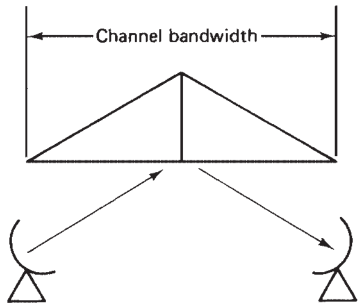


Figure 14.1 Heavy route message (frequency modulation—single access). (*From Telesat Canada, 1983.*)

quency-division multiplexed, frequency modulated (FM), with frequency-division multiple access to the satellite. In Chap. 9, FDM/FM signals are discussed. It will be recalled that the voice-frequency (telephone) signals are first SSBSC amplitude modulated onto voice carriers in order to generate the single sidebands needed for the frequency-division multiplexing. For the purpose of illustration, each earth station will be assumed to transmit a 60-channel supergroup. Each 60-channel supergroup is then frequency modulated onto a carrier which is then upconverted to a frequency in the satellite uplink band.

Figure 14.2 shows the situation for three earth stations: one in Ottawa, one in New York, and one in London. All three earth stations access a single satellite transponder channel simultaneously, and each communicates with both of the others. Thus it is assumed that the satellite receive and transmit antenna beams are *global,* encompassing all three earth stations. Each earth station transmits one uplink carrier modulated with a 60-channel supergroup and receives two similar downlink carriers.

The earth station at New York is shown in more detail. One transmit chain is used, and this carries telephone traffic for both Ottawa and London. On the receive side, two receive chains must be provided, one for the Ottawa-originated carrier and one for the London-originated carrier. Each of these carriers will have a mixture of traffic, and in the demultiplexing unit, only those telephone channels intended for New York are passed through. These are remultiplexed into an FDM/FM format which is transmitted out along the terrestrial line to the New York switching office. This earth station arrangement should be compared with that shown in Fig. 8.6.

Figure 14.3 shows a hypothetical frequency assignment scheme for the hypothetical network of Fig. 14.2. Uplink carrier frequencies of 6253, 6273, and 6278 MHz are shown for illustration purposes. For the satellite transponder arrangement of Fig. 7.13, these carriers would be translated down to frequencies of 4028, 4048, and 4053 MHz (i.e., the

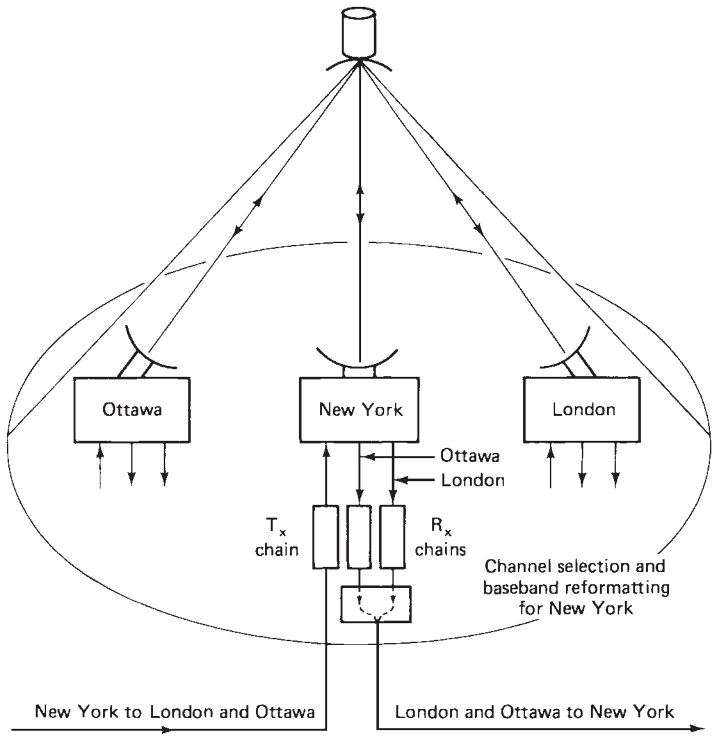


Figure 14.2 Three earth stations transmitting and receiving simultaneously through the same satellite transponder, using fixed-assignment FDMA.

corresponding 4-GHz-band downlink frequencies) and sent to transponder 9 of the satellite. Typically, a 60-channel FDM/FM carrier occupies 5 MHz of transponder bandwidth, including guardbands. A total frequency allowance of 15 MHz is therefore required for the three stations, and each station receives all the traffic. The remainder of the transponder bandwidth may be unused, or it may be occupied by other carriers, which are not shown.

As an example of preassignment, suppose that each station can transmit up to 60 voice circuits and that 40 of these are preassigned to the New York–London route. If these 40 circuits are fully loaded, additional calls on the New York–London route will be blocked even though there may be idle circuits on the other preassigned routes.

Telesat Canada operates medium-route message facilities utilizing FDM/FM/FDMA. Figure 14.4 shows how five carriers may be used to support 168 voice channels. The earth station that carries the full load has a [G/T] of 37.5 dB/K, and the other four have [G/T]'s of 28 dB/K.

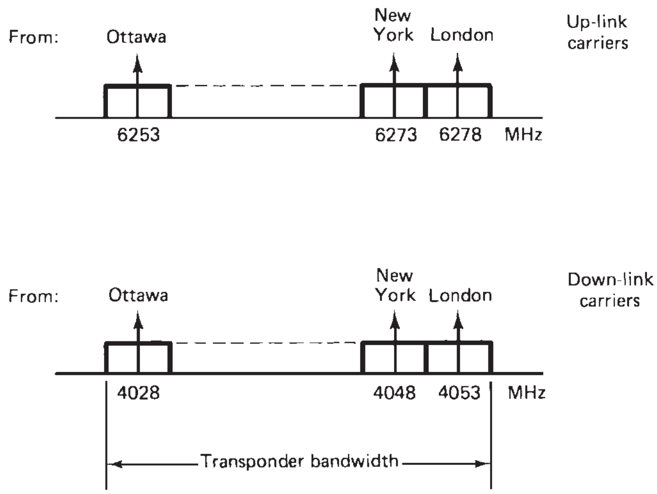


Figure 14.3 Transponder channel assignments for the earth stations shown in Fig. 14.2.

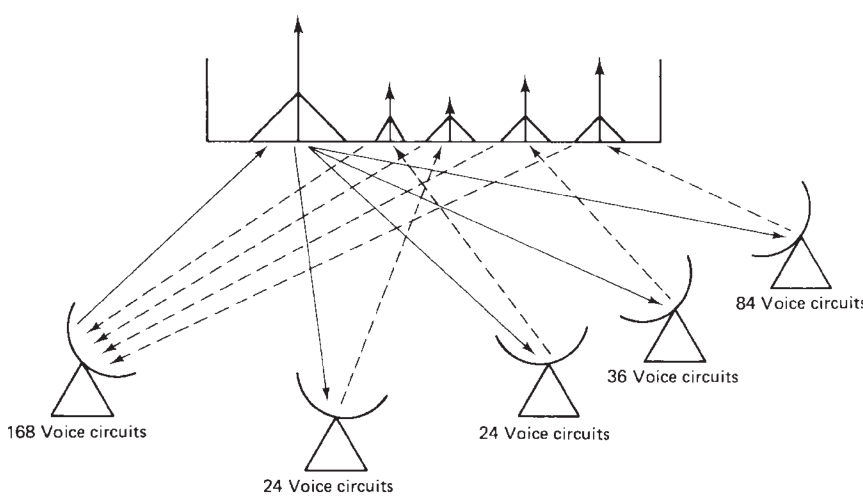


Figure 14.4 Medium route message traffic (frequency-division multiple access, FM/FDMA). (*From Telesat Canada, 1983.*)

Preassignment also may be made on the basis of a single channel per carrier (SCPC). This refers to a single voice (or data) channel per carrier, not a transponder channel, which may in fact carry some hundreds of voice channels by this method. The carriers may be frequency modulated or phase-shift modulated, and an earth station may be capable of transmitting one or more SCPC signals simultaneously.

Figure 14.5 shows the INTELSAT SCPC channeling scheme for a 36-MHz transponder. The transponder bandwidth is subdivided into 800 channels each 45 kHz wide. The 45 kHz, which includes a guardband, is required for each digitized voice channel, which utilizes QPSK modulation. The channel information signal may be digital data or PCM voice signals (see Chap. 10). A pilot frequency is transmitted for the purpose of frequency control, and the adjacent channel slots on either side of the pilot are left vacant to avoid interference. The scheme therefore provides a total of 798 one-way channels or up to 399 full-duplex voice circuits. In duplex operation, the frequency pairs are separated by 18.045 MHz, as shown in Fig. 14.5.

The frequency tolerance relative to the assigned values is within ±1 kHz for the received SCPC carrier and must be within ±250 Hz for the transmitted SCPC carrier (Miya, 1981). The pilot frequency is transmitted by one of the earth stations designated as a primary station. This provides a reference for automatic frequency control (AFC) (usually through the use of phase-locked loops) of the transmitter frequency synthesizers and receiver local oscillators. In the event of failure of the primary station, the pilot frequency is transmitted from a designated backup station.

An important feature of the INTELSAT SCPC system is that each channel is voice-activated. This means that on a two-way telephone conversation, only one carrier is operative at any one time. Also, in long pauses between speech, the carriers are switched off. It has been estimated that for telephone calls, the one-way utilization time is 40 percent of the call duration. Using voice activation, the average number of carriers being amplified at any one time by the transponder traveling-wave tube (TWT) is reduced. For a given level of intermodulation distortion (see Secs. 7.7.3 and 12.10), the TWT power output per FDMA carrier therefore can be increased.

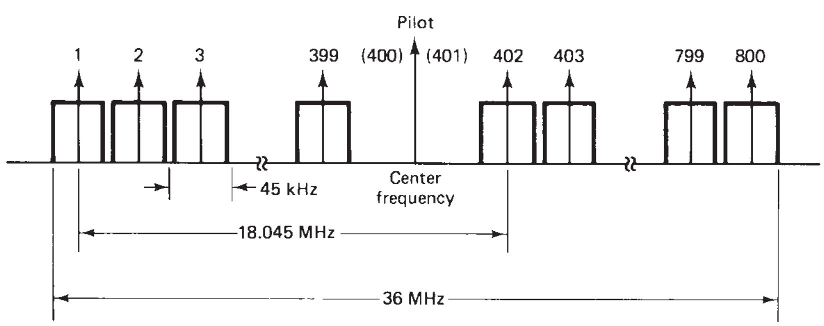


Figure 14.5 Channeling arrangement for Intelsat SCPC system.

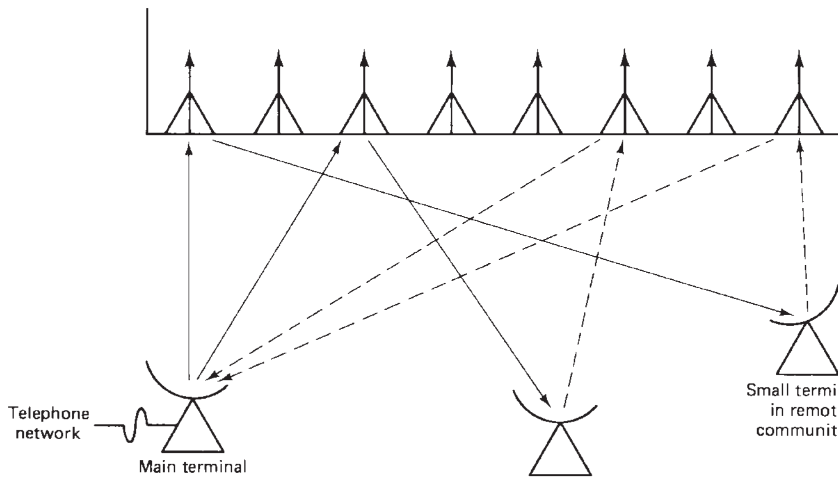


Figure 14.6 Thin route message traffic (single channel per carrier, SCPC/FDMA). (*From Telesat Canada, 1983.*)

SCPC systems are widely used on lightly loaded routes, this type of service being referred to as a *thin route service.* It enables remote earth stations in sparsely populated areas to connect into the national telephone network in a reasonably economical way. A main earth station is used to make the connection to the telephone network, as illustrated in Fig. 14.6. The Telesat Canada Thin Route Message Facilities provide up to 360 two-way circuits using PSK/SCPC (PSK = phase-shift keying). The remote terminals operate with 4.6-m-diameter antennas with [G/T] values of 19.5 or 21 dB/K. Transportable terminals are also available, one of these being shown in Fig. 14.7. This is a single-channel station that uses a 3.6-m antenna and comes complete with a desktop electronics package which can be installed on the customers' premises.

14.4 Demand-Assigned FDMA

In the demand-assigned mode of operation, the transponder frequency bandwidth is subdivided into a number of channels. A channel is assigned to each carrier in use, giving rise to the single-channel-per-carrier mode of operation discussed in the preceding section. As in the preassigned access mode, carriers may be frequency modulated with analog information signals, these being designated FM/SCPC, or they may be phase modulated with digital information signals, these being designated as PSK/SCPC.

Figure 14.7 Transportable message station. (*From Telesat Canada, 1983.*)

Demand assignment may be carried out in a number of ways. In the polling method, a master earth station continuously polls all the earth stations in sequence, and if a *call request* is encountered, frequency slots are assigned from the pool of available frequencies. The polling delay with such a system tends to become excessive as the number of participating earth stations increases.

Instead of using a polling sequence, earth stations may request calls through the master earth station as the need arises. This is referred to as *centrally controlled random access.* The requests go over a digital orderwire, which is a narrowband digital radio link or a circuit through a satellite transponder reserved for this purpose. Frequencies are assigned, if available, by the master station, and when the call is completed, the frequencies are returned to the pool. If no frequencies are available, the blocked call requests may be placed in a queue, or a second call attempt may be initiated by the requesting station.

As an alternative to centrally controlled random access, control may be exercised at each earth station, this being known as *distributed-control random access.* A good illustration of such a system is provided by the Spade system operated by INTELSAT on some of its satellites. This is described in the next section.

14.5 Spade System

The word *Spade* is a loose acronym for single-channel-per-carrier pulse-code-modulated multiple-access demand-assignment equipment. Spade was developed by Comsat for use on the INTELSAT satellites (see, e.g., Martin, 1978) and is compatible with the INTELSAT SCPC preassigned

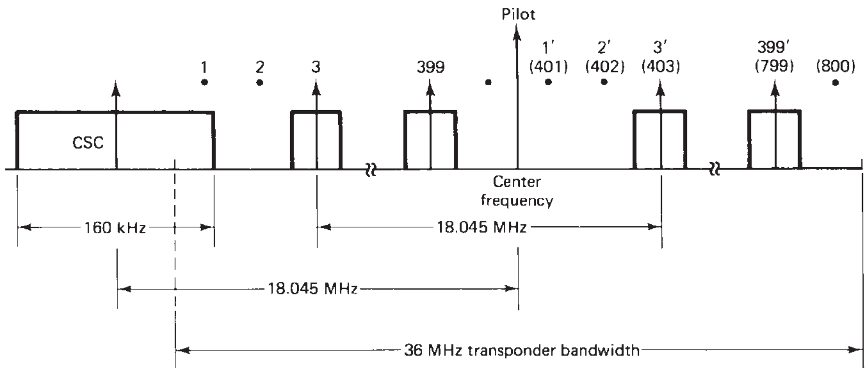


Figure 14.8 Channeling scheme for the Spade system.

system described in Sec. 14.3. However, the distributed-demand assignment facility requires a common signaling channel (CSC). This is shown in Fig. 14.8. The CSC bandwidth is 160 kHz, and its center frequency is 18.045 MHz below the pilot frequency, as shown in Fig. 14.8. To avoid interference with the CSC, voice channels 1 and 2 are left vacant, and to maintain duplex matching, the corresponding channels 1′ and 2′ are also left vacant. Recalling from Fig. 14.5 that channel 400 also must be left vacant, this requires that channel 800 be left vacant for duplex matching. Thus six channels are removed from the total of 800, leaving a total of 794 one-way or 397 full-duplex voice circuits, the frequencies in any pair being separated by 18.045 MHz, as shown in Fig. 14.8. (An alternative arrangement is shown in Freeman, 1981.)

All the earth stations are permanently connected through the common signaling channel (CSC). This is shown diagrammatically in Fig. 14.9 for six earth stations *A, B, C, D, E,* and *F.* Each earth station has the facility for generating any one of the 794 carrier frequencies using frequency synthesizers. Furthermore, each earth station has a memory containing a list of the frequencies currently available, and this list is continuously updated through the CSC. To illustrate the procedure, suppose that a call to station *F* is initiated from station *C* in Fig. 14.9. Station *C* will first select a frequency pair at random from those currently available on the list and signal this information to station *F* through the CSC. Station *F* must acknowledge, through the CSC, that it can complete the circuit. Once the circuit is established, the other earth stations are instructed, through the CSC, to remove this frequency pair from the list.

The round-trip time between station *C* initiating the call and station *F* acknowledging it is about 600 ms. During this time, the two frequencies chosen at station *C* may be assigned to another circuit. In this

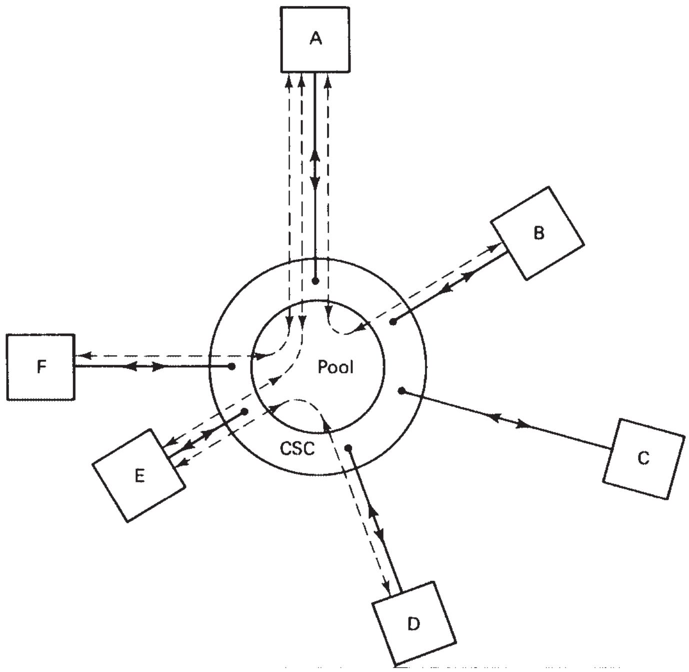


Figure 14.9 Diagrammatic representation of a Spade communications system.

event, station *C* will receive the information on the CSC update and will immediately choose another pair at random, even before hearing back from station *F*.

Once a call has been completed and the circuit disconnected, the two frequencies are returned to the pool, the information again being transmitted through the CSC to all the earth stations.

As well as establishing the connection through the satellite, the CSC passes signaling information from the calling station to the destination station, in the example above from station *C* to station *F*. Signaling information in the Spade system is routed through the CSC rather than being sent over a voice channel. Each earth station has equipment called the *demand assignment signaling and switching* (DASS) *unit* which performs the functions required by the CSC.

Some type of multiple access to the CSC must be provided for all the earth stations using the Spade system. This is quite separate from the SCPC multiple access of the network's voice circuits. Time-division multiple access, described in Sec. 14.7.8, is used for this purpose, allowing up to 49 earth stations to access the common signaling channel.

14.6 Bandwidth-Limited and Power-Limited TWT Amplifier Operation

A transponder will have a total bandwidth B_{TR}, and it is apparent that this can impose a limitation on the number of carriers which can access the transponder in an FDMA mode. For example, if there are K carriers each of bandwidth B, then the best that can be achieved is $K = B_{TR}/B$. Any increase in the transponder EIRP will not improve on this, and the system is said to be *bandwidth-limited.* Likewise, for digital systems, the bit rate is determined by the bandwidth, which again will be limited to some maximum value by B_{TR}.

Power limitation occurs where the EIRP is insufficient to meet the $[C/N]$ requirements, as shown by Eq. (12.34). The signal bandwidth will be approximately equal to the noise bandwidth, and if the EIRP is below a certain level, the bandwidth will have to be correspondingly reduced to maintain the $[C/N]$ at the required value. These limitations are discussed in more detail in the next two sections.

14.6.1 FDMA downlink analysis

To see the effects of intermodulation noise which results with FDMA operation, consider the overall carrier-to-noise ratio as given by Eq. (12.62). In terms of noise power rather than noise power density, Eq. (12.62) states

$$\left(\frac{N}{C}\right) = \left(\frac{N}{C}\right)_U + \left(\frac{N}{C}\right)_D + \left(\frac{N}{C}\right)_{\mathrm{IM}} \tag{14.1}$$

A certain value of carrier-to-noise ratio will be needed, as specified in the system design, and this will be denoted by the subscript REQ. The overall C/N must be at least as great as the required value, a condition which can therefore be stated as

$$\left(\frac{N}{C}\right)_{\mathrm{REQ}} \geq \left(\frac{N}{C}\right) \tag{14.2}$$

Note that because the noise-to-carrier ratio rather than the carrier-to-noise ratio is involved, the actual value is equal to or less than the required value. Using Eq. (14.1), the condition can be rewritten as

$$\left(\frac{N}{C}\right)_{\mathrm{REQ}} \geq \left(\frac{N}{C}\right)_U + \left(\frac{N}{C}\right)_D + \left(\frac{N}{C}\right)_{\mathrm{IM}} \tag{14.3}$$

The right-hand side of Eq. (14.3) is usually dominated by the downlink ratio. With FDMA, backoff is utilized to reduce the intermodulation noise to an acceptable level, and as shown in Sec. 12.10, the uplink noise contribution is usually negligible. Thus the expression can be approximated by

$$\left(\frac{N}{C}\right)_{\mathrm{REQ}} \geq \left(\frac{N}{C}\right)_D$$

or

$$\left(\frac{C}{N}\right)_{\mathrm{REQ}} \leq \left(\frac{C}{N}\right)_D \tag{14.4}$$

Consider the situation where each carrier of the FDMA system occupies a bandwidth B and has a downlink power denoted by $[\mathrm{EIRP}]_D$. Equation (12.54) gives

$$\left[\frac{C}{N}\right]_D = [\mathrm{EIRP}]_D + \left[\frac{G}{T}\right]_D - [\mathrm{LOSSES}] - [k] - [B] \tag{14.5}$$

where it is assumed that $B_N \cong B$. This can be written in terms of the required carrier-to-noise ratio as

$$\left[\frac{C}{N}\right]_{\mathrm{REQ}} \leq [\mathrm{EIRP}]_D + \left[\frac{G}{T}\right]_D - [\mathrm{LOSSES}] - [k] - [B] \tag{14.6}$$

To set up a reference level, consider first single-carrier operation. The satellite will have a saturation value of EIRP and a transponder bandwidth B_{TR}, both of which are assumed fixed. With single-carrier access, no backoff is needed, and Eq. (14.6) becomes

$$\left[\frac{C}{N}\right]_{\mathrm{REQ}} \leq [\mathrm{EIRP}_S] + \left[\frac{G}{T}\right]_D - [\mathrm{LOSSES}] - [k] - [B_{\mathrm{TR}}] \tag{14.7}$$

or

$$\left[\frac{C}{N}\right]_{\mathrm{REQ}} - [\mathrm{EIRP}_S] - \left[\frac{G}{T}\right]_D + [\mathrm{LOSSES}] + [k] + [B_{\mathrm{TR}}] \leq 0 \tag{14.8}$$

If the system is designed for single-carrier operation, then the equality sign applies and the reference condition is

$$\left[\frac{C}{N}\right]_{\mathrm{REQ}} - [\mathrm{EIRP}_S] - \left[\frac{G}{T}\right]_D + [\mathrm{LOSSES}] + [k] + [B_{\mathrm{TR}}] = 0 \tag{14.9}$$

Consider now the effect of power limitation imposed by the need for backoff. Suppose the FDMA access provides for K carriers which share the output power equally, and each requires a bandwidth B. The output power for each of the FDMA carriers is

$$[\text{EIRP}]_D = [\text{EIRP}_S] - [\text{BO}]_O - [K] \tag{14.10}$$

The transponder bandwidth B_{TR} will be shared between the carriers, but not all of B_{TR} can be utilized because of the power limitation. Let α represent the fraction of the total bandwidth actually occupied, such that $KB = \alpha B_{\text{TR}}$, or in terms of decilogs

$$[B] = [\alpha] + [B_{\text{TR}}] - [K] \tag{14.11}$$

Substituting these relationships in Eq. (14.6) gives

$$\left[\frac{C}{N}\right]_{\text{REQ}} \leq [\text{EIRP}_S] - [\text{BO}]_O + \left[\frac{G}{T}\right]_D - [\text{LOSSES}] - [k] - [B_{\text{TR}}] - [\alpha] \tag{14.12}$$

It will be noted that the $[K]$ term cancels out. The expression can be rearranged as

$$\left[\frac{C}{N}\right]_{\text{REQ}} - [\text{EIRP}_S] - \left[\frac{G}{T}\right]_{\text{D}} + [\text{LOSSES}] + [k] + [B_{\text{TR}}] \leq -[BO]_O - [\alpha] \tag{14.13}$$

But as shown by Eq. (14.9), the left-hand side is equal to zero if the single carrier access is used as reference, and hence

$$0 \leq -[\text{BO}]_O - [\alpha] \qquad \text{or} \qquad [\alpha] \leq -[\text{BO}]_O \tag{14.14}$$

The best that can be achieved is to make $[\alpha] = -[\text{BO}]_O$, and since the backoff is a positive number of decibels, $[\alpha]$ must be negative, or equivalently, α is fractional. The following example illustrates the limitation imposed by backoff.

Example 14.1 A satellite transponder has a bandwidth of 36 MHz and a saturation EIRP of 27 dBW. The earth station receiver has a G/T ratio of 30 dB/K, and the total link losses are 196 dB. The transponder is accessed by FDMA carriers each of 3-MHz bandwidth, and 6-dB output backoff is employed. Calculate the downlink carrier-to-noise ratio for single-carrier operation and the number of carriers which can be accommodated in the FDMA system. Compare this with the number which could be accommodated if no backoff were needed. The carrier-to-noise ratio determined for single-carrier operation may be taken as the reference value, and it may be assumed that the uplink noise and intermodulation noise are negligible.

solution *Note:* For convenience in the Mathcad solution, decibel or decilog values will be indicated by dB. For example, the output backoff in decibels is shown as $\mathrm{BOdB_O}$.

Transponder bandwidth:

$$\mathrm{B_{TR}} := 36 \cdot \mathrm{MHz} \qquad \mathrm{BdB_{TR}} := 10 \cdot \log\left(\frac{\mathrm{B_{TR}}}{\mathrm{Hz}}\right)$$

Carrier bandwidth:

$$\mathrm{B} := 3 \cdot \mathrm{MHz} \qquad \mathrm{BdB} := 10 \cdot \log\left(\frac{\mathrm{B}}{\mathrm{Hz}}\right)$$

Saturation eirp:

$$\mathrm{eirpdBW_S} := 27$$

Output backoff:

$$\mathrm{BOdB_O} := 6$$

Total losses:

$$\mathrm{LOSSESdB} := 196$$

Ground station *G*/*T*:

$$\mathrm{GTRdB} := 30$$

$$\mathrm{CNRdB_D} := \mathrm{eirpdBW_S} + \mathrm{GTRdB} - \mathrm{LOSSESdB} + 228.6 - \mathrm{BdB_{TR}} \quad \text{Eq. (12.54)}$$

$$\mathrm{CNRdB_D} = 14$$

========

$$\alpha\mathrm{dB} := -\mathrm{BOdB_O} \qquad \text{Eq. (14.14)}$$

$$\mathrm{KdB} := \alpha\mathrm{dB} + \mathrm{BdB_{TR}} - \mathrm{BdB} \qquad \text{Eq. (14.11)}$$

$$\mathrm{K} := 10^{\frac{\mathrm{KdB}}{10}} \qquad \mathrm{K} = 3$$

======

If backoff was not required, the number of carriers which could be accommodated would be

$$\frac{B_{TR}}{B} = 12$$

14.7 TDMA

With time-division multiple access, only one carrier uses the transponder at any one time, and therefore, intermodulation products, which result from the nonlinear amplification of multiple carriers, are absent. This leads to one of the most significant advantages of TDMA, which is that the transponder traveling-wave tube (TWT) can be operated at maximum power output or saturation level.

Because the signal information is transmitted in bursts, TDMA is only suited to digital signals. Digital data can be assembled into burst format for transmission and reassembled from the received bursts through the use of digital buffer memories.

Figure 14.10 illustrates the basic TDMA concept, in which the stations transmit bursts in sequence. Burst synchronization is required, and in the system illustrated in Fig. 14.10, one station is assigned sole-

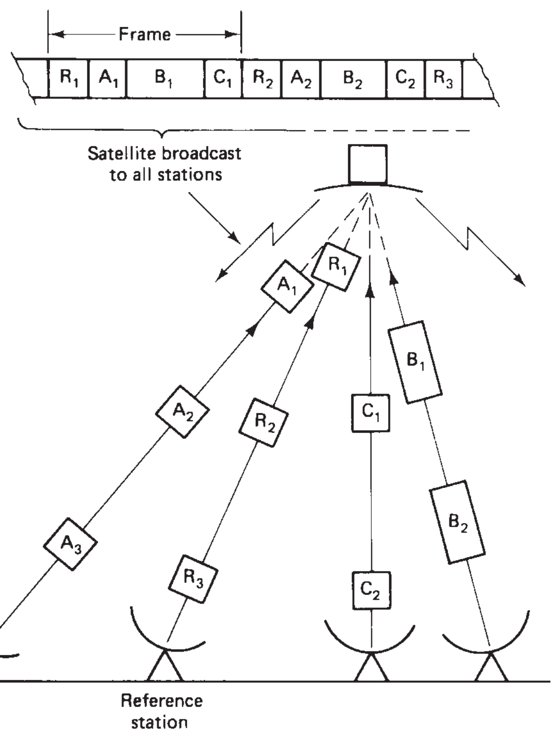


Figure 14.10 Time-division multiple access (TDMA) using a reference station for burst synchronization.

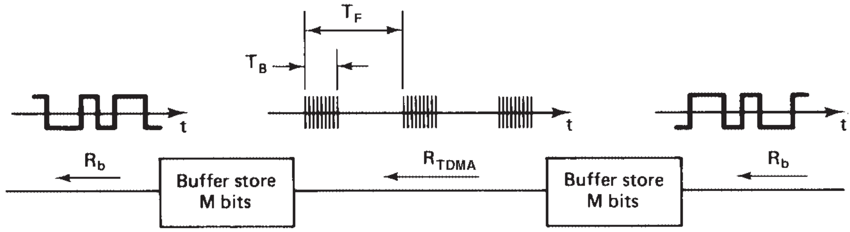


Figure 14.11 Burst-mode transmission linking two continuous-mode streams.

ly for the purpose of transmitting *reference bursts* to which the others can be synchronized. The time interval from the start of one reference burst to the next is termed a *frame.* A frame contains the reference burst R and the bursts from the other earth stations, these being shown as A, B, and C in Fig. 14.10.

Figure 14.11 illustrates the basic principles of burst transmission for a single channel. Overall, the transmission appears continuous because the input and output bit rates are continuous and equal. However, within the transmission channel, input bits are temporarily stored and transmitted in bursts. Since the time interval between bursts is the frame time T_F, the required buffer capacity is

$$M = R_b T_F \tag{14.15}$$

The buffer memory fills up at the input bit rate R_b during the frame time interval. These M bits are transmitted as a burst in the next frame without any break in continuity of the input. The M bits are transmitted in the burst time T_B, and the *transmission rate,* which is equal to the burst bit rate, is

$$\begin{aligned} \mathrm{R}_{\mathrm{TDMA}} &= \frac{M}{T_B} \\ &= R_b \frac{T_F}{T_B} \end{aligned} \tag{14.16}$$

This is also referred to as the *burst rate,* but note that this means the instantaneous bit rate within a burst (not the number of bursts per second, which is simply equal to the frame rate). It will be seen that the *average* bit rate for the burst mode is simply M/T_F, which is equal to the input and output rates.

The frame time T_F will be seen to add to the overall propagation delay. For example, in the simple system illustrated in Fig. 14.11, even if the actual propagation delay between transmit and receive buffers is assumed to be zero, the receiving side would still have to wait a time

T_F before receiving the first transmitted burst. In a geostationary satellite system, the actual propagation delay is a significant fraction of a second, and excessive delays from other causes must be avoided. This sets an upper limit to the frame time, although with current technology other factors restrict the frame time to well below this limit. The frame period is usually chosen to be a multiple of 125 μs, which is the standard sampling period used in pulse-code modulation (PCM) telephony systems, since this ensures that the PCM samples can be distributed across successive frames at the PCM sampling rate.

Figure 14.12 shows some of the basic units in a TDMA ground station, which for discussion purposes is labeled earth station *A*. Terrestrial links coming into earth station *A* carry digital traffic addressed to destination stations, labeled *B, C, X.* It is assumed that the bit rate is the same for the digital traffic on each terrestrial link. In the units labeled *terrestrial interface modules* (TIMs), the incoming continuous-bit-rate signals are converted into the intermittent-burst-rate mode. These individual burst-mode signals are *time-division multiplexed* in the time-division multiplexer (MUX) so that the traffic for each destination station appears in its assigned time slot within a burst.

Certain time slots at the beginning of each burst are used to carry timing and synchronizing information. These time slots collectively are referred to as the *preamble.* The complete burst containing the preamble and the traffic data is used to phase modulate the radiofrequency (rf) carrier. Thus the composite burst which is transmitted at rf consists of a number of time slots, as shown in Fig. 14.13. These will be described in more detail shortly.

The received signal at an earth station consists of bursts from all transmitting stations arranged in the frame format shown in Fig. 14.13. The rf carrier is converted to intermediate frequency (IF), which is then demodulated. A separate preamble detector provides timing information for transmitter and receiver along with a carrier synchronizing signal for the phase demodulator, as described in the next section. In many systems, a station receives its own transmission along with the others in the frame, which can then be used for burst-timing purposes.

A reference burst is required at the beginning of each frame to provide timing information for the *acquisition* and *synchronization* of bursts (these functions are described further in Sec. 14.7.4). In the INTELSAT international network, at least two reference stations are used, one in the East and one in the West. These are designated *primary* reference stations, one of which is further selected as the *master primary.* Each primary station is duplicated by a *secondary* reference station, making four reference stations in all. The fact that all the reference stations are identical means that any one can become the master primary. All the

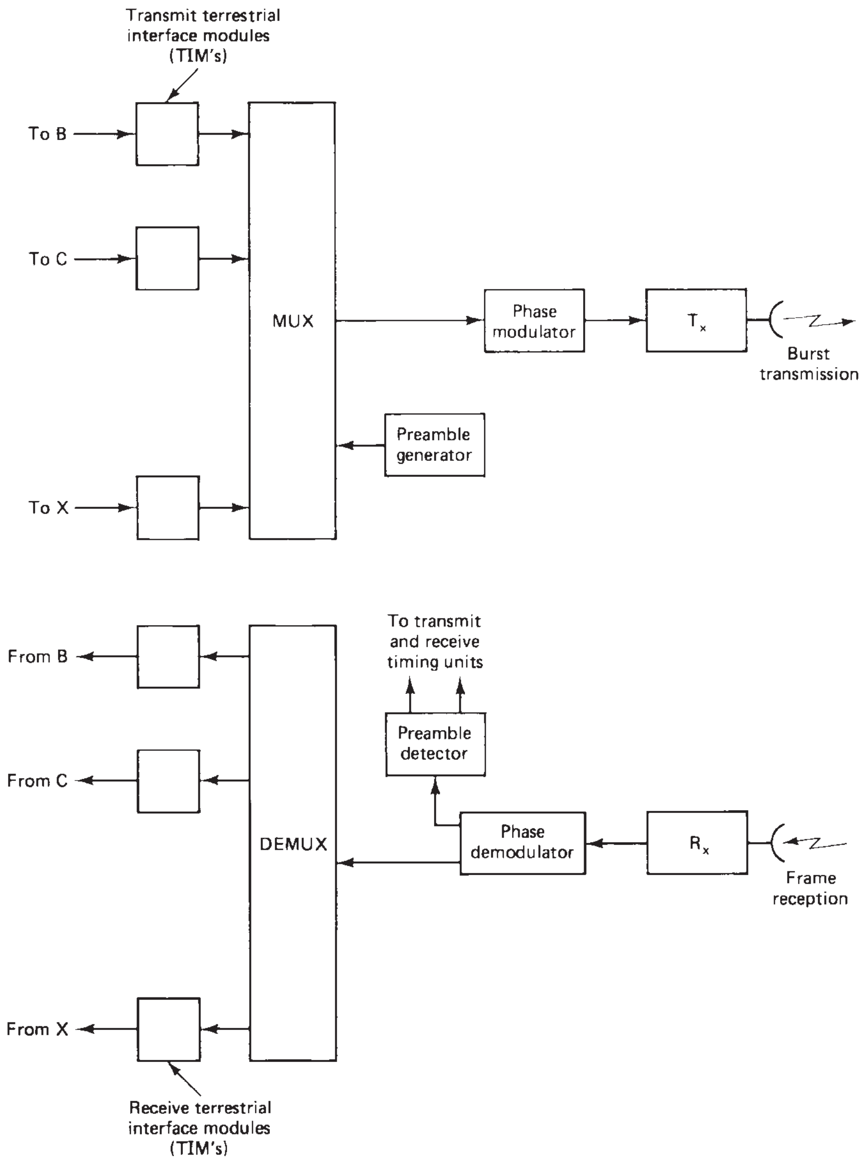


Figure 14.12 Some of the basic equipment blocks in a TDMA system.

system timing is derived from the high-stability clock in the master primary, which is accurate to 1 part in 10^{11} (Lewis, 1982). A clock on the satellite is locked to the master primary, and this acts as the clock for the other participating earth stations. The satellite clock will provide a constant frame time, but the participating earth stations must make

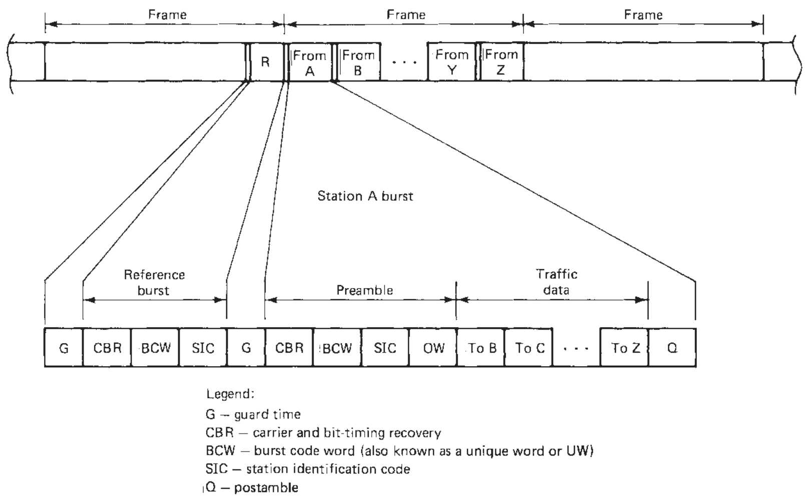


Figure 14.13 Frame and burst formats for a TDMA system.

corrections for variations in the satellite range, since the transmitted bursts from all the participating earth stations must reach the satellite in synchronism. Details of the timing requirements will be found in Spilker (1977).

In the INTELSAT system, two reference bursts are transmitted in each frame. The first reference burst, which marks the beginning of a frame, is transmitted by a master primary (or a primary) reference station and contains the timing information needed for the acquisition and synchronization of bursts. The second reference burst, which is transmitted by a secondary reference station, provides synchronization but not acquisition information. The secondary reference burst is ignored by the receiving earth stations unless the primary or master primary station fails.

14.7.1 Reference burst

The reference burst that marks the beginning of a frame is subdivided into time slots or channels used for various functions. These will differ in detail for different networks, but Fig. 14.13 shows some of the basic channels that are usually provided. These can be summarized as follows:

Guard time (*G*). A guard time is necessary between bursts to prevent the bursts from overlapping. The guard time will vary from burst to burst depending on the accuracy with which the various bursts can be positioned within each frame.

Carrier and bit-timing recovery (*CBR*). To perform coherent demodulation of the phase-modulated carrier as described in Secs. 10.7 and 10.8, a coherent carrier signal must first be recovered from the burst. An unmodulated carrier wave is provided during the first part of the CBR time slot. This is used as a synchronizing signal for a local oscillator at the detector, which then produces an output coherent with the carrier wave. The carrier in the subsequent part of the CBR time slot is modulated by a known phase-change sequence which enables the bit timing to be recovered. Accurate bit timing is needed for the operation of the sample-and-hold function in the detector circuit (see Figs. 10.13 and 10.23). Carrier recovery is described in more detail in Sec. 14.7.3.

Burst code word (*BCW*). (Also known as a *unique word.*) This is a binary word, a copy of which is stored at each earth station. By comparing the incoming bits in a burst with the stored version of the BCW, the receiver can detect when a group of received bits matches the BCW, and this in turn provides an accurate time reference for the burst position in the frame. A known bit sequence is also carried in the BCW, which enables the phase ambiguity associated with coherent detection to be resolved.

Station identification code (*SIC*). This identifies the transmitting station.

Figure 14.14 shows the makeup of the reference bursts used in certain of the INTELSAT networks. The numbers of symbols and the corresponding time intervals allocated to the various functions are shown. In addition to the channels already described, a *coordination and delay channel* (sometimes referred to as the *control and delay channel*) is provided. This channel carries the identification number of the earth station being addressed and various codes used in connection with the acquisition and synchronization of bursts at the addressed earth station. It is also necessary for an earth station to know the propagation time delay to the satellite to implement burst acquisition and synchronization. In the INTELSAT system, the propagation delay is computed from measurements made at the reference station and transmitted to the earth station in question through the coordination and delay channel.

The other channels in the INTELSAT reference burst are the following:

TTY: telegraph order-wire channel, used to provide telegraph communications between earth stations.

SC: service channel which carries various network protocol and alarm messages.

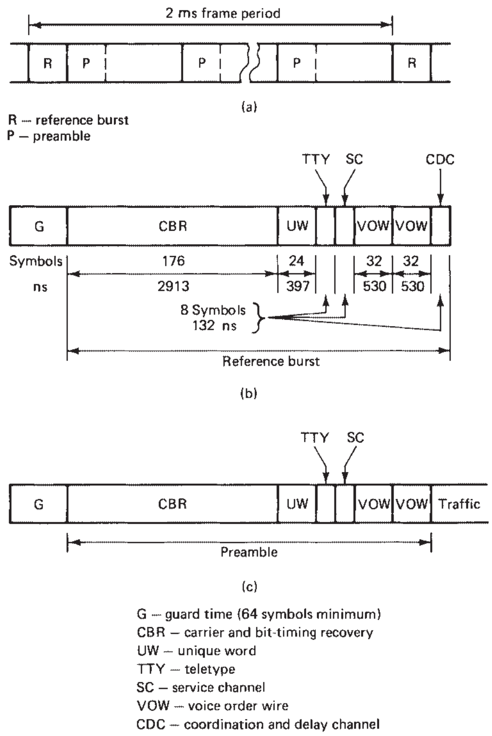


Figure 14.14 (*a*) Intelsat 2-ms frame; (*b*) composition of the reference burst *R*; (*c*) composition of the preamble *P*. (QPSK modulation is used, giving 2 bits per symbol. Approximate time intervals are shown.)

VOW: voice-order-wire channel used to provide voice communications between earth stations. Two VOW channels are provided.

14.7.2 Preamble and postamble

The *preamble* is the initial portion of a traffic burst which carries information similar to that carried in the reference burst. In some systems the channel allocations in the reference bursts and the preambles are identical. No traffic is carried in the preamble. In Fig. 14.13, the only difference between the preamble and the reference burst is that the preamble provides an orderwire (OW) channel.

For the INTELSAT format shown in Fig. 14.14, the preamble differs from the reference burst in that it does not provide a coordination and delay channel (CDC). Otherwise, the two are identical.

As with the reference bursts, the preamble provides a carrier and bit-timing recovery channel and also a burst-code-word channel for burst-timing purposes. The burst code word in the preamble of a traffic burst is different from the burst code word in the reference bursts, which enables the two types of bursts to be identified.

In certain phase detection systems, the phase detector must be allowed time to recover from one burst before the next burst is received by it. This is termed *decoder quenching,* and a time slot, referred to as a *postamble,* is allowed for this function. The postamble is shown as Q in Fig. 14.13. Many systems are designed to operate without a postamble.

14.7.3 Carrier recovery

A factor which must be taken into account with TDMA is that the various bursts in a frame lack coherence so that carrier recovery must be repeated for each burst. This applies to the traffic as well as the reference bursts. Where the carrier recovery circuit employs a phase-locked loop such as shown in Fig. 10.20, a problem known as *hangup* can occur. This arises when the loop moves to an unstable region of its operating characteristic. The loop operation is such that it eventually returns to a stable operating point, but the time required to do this may be unacceptably long for burst-type signals.

One alternative method utilizes a narrowband tuned circuit filter to recover the carrier. An example of such a circuit for quadrature phase-shift keying (QPSK), taken from Miya (1981), is shown in Fig. 14.15. The QPSK signal, which has been downconverted to a standard IF of 140 MHz, is quadrupled in frequency to remove the modulation, as described in Sec. 10.7. The input frequency must be maintained at the resonant frequency of the tuned circuit, which requires some form of automatic frequency control. Because of the difficulties inherent in working with high frequencies, the output frequency of the quadrupler is downconverted from 560 to 40 MHz, and the AFC is applied to the voltage-controlled oscillator (VCO) used to make the frequency conversion. The AFC circuit is a form of phase-locked loop (PLL) in which the phase difference between input and output of the single-tuned circuit is held at zero, which ensures that the 40-MHz input remains at the center of the tuned circuit response curve. Any deviation of the phase difference from zero generates a control voltage which is applied to the VCO in such a way as to bring the frequency back to the required value.

Interburst interference may be a problem with the tuned-circuit method because of the energy stored in the tuned circuit for any given burst. Avoidance of interburst interference requires careful design of the tuned circuit (Miya, 1981) and possibly the use of a postamble, as mentioned in the previous section.

Other methods of carrier recovery are discussed in Gagliardi (1991).

14.7.4 Network synchronization

Network synchronization is required to ensure that all bursts arrive at the satellite in their correct time slots. As mentioned previously, tim-

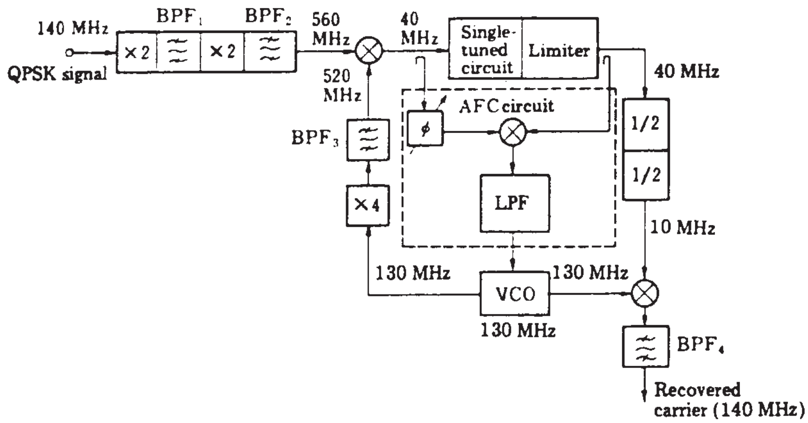


Figure 14.15 An example of carrier recovery circuit with a single-tuned circuit and AFC. (*From Miya, 1981.*)

ing markers are provided by the reference bursts, which are tied to a highly stable clock at the reference station and transmitted through the satellite link to the traffic stations. At any given traffic station, detection of the unique word (or burst code word) in the reference burst signals the *start of receiving frame* (SORF), the marker coinciding with the last bit in the unique word.

It would be desirable to have the highly stable clock located aboard the satellite because this would eliminate the variations in propagation delay arising from the uplink for the reference station, but this is not practical because of weight and space limitations. However, the reference bursts retransmitted from the satellite can be treated, for timing purposes, as if they originated from the satellite (Spilker, 1977).

The network operates what is termed a *burst time plan,* a copy of which is stored at each earth station. The burst time plan shows each earth station where the receive bursts intended for it are relative to the SORF marker. This is illustrated in Fig. 14.16. At earth station A the SORF marker is received after some propagation delay t_A, and the burst time plan tells station A that a burst intended for it follows at time T_A after the SORF marker received by it. Likewise, for station B, the propagation delay is t_B, and the received bursts start at T_B after the SORF markers received at station B. The propagation delays for each station will differ, but typically they are in the region of 120 ms each.

The burst time plan also shows a station when it must transmit its bursts in order to reach the satellite in the correct time slots. A major advantage of the TDMA mode of operation is that the burst time plan is essentially under software control so that changes in traffic patterns can be accommodated much more readily than is the case with FDMA, where modifications to hardware are required. Against this,

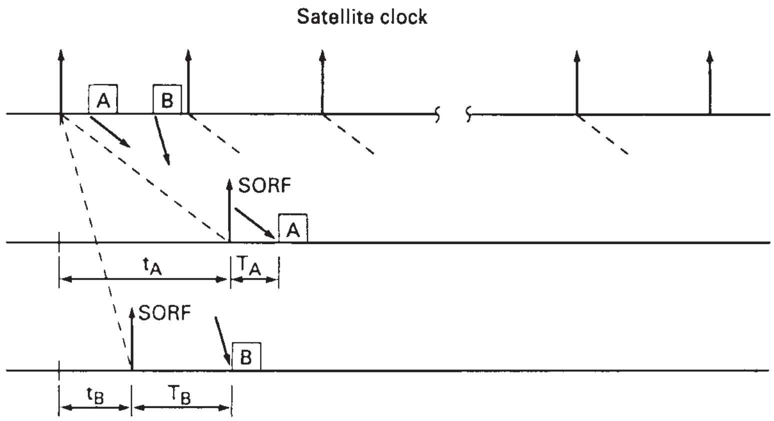


Figure 14.16 Start of receive frame (SORF) marker in a time burst plan.

implementation of the synchronization is a complicated process. Corrections must be included for changes in propagation delay which result from the slowly varying position of the satellite (see Sec. 7.4). In general, the procedure for transmit timing control has two stages. First, there is the need for a station just entering, or reentering after a long delay, to acquire its correct slot position, this being referred to as *burst position acquisition.* Once the time slot has been acquired, the traffic station must maintain the correct position, this being known as *burst position synchronization.*

Open-loop timing control. This is the simplest method of transmit timing. A station transmits at a fixed interval following reception of the timing markers, according to the burst time plan, and sufficient guard time is allowed to absorb the variations in propagation delay. The burst position error can be large with this method, and longer guard times are necessary, which reduces frame efficiency (see Sec. 14.7.7). However, for frame times longer than about 45 ms, the loss of efficiency is less than 10 percent. In a modified version of the open-loop method known as *adaptive open-loop timing,* the range is computed at the traffic station from orbital data or from measurements, and the traffic earth station makes its own corrections in timing to allow for the variations in the range. It should be noted that with open-loop timing, no special acquisition procedure is required.

Loopback timing control. *Loopback* refers to the fact that an earth station receives its own transmission, from which it can determine range. It follows that the loopback method can only be used where the satellite transmits a global or regional beam encompassing all the earth stations in the network. A number of methods are available for the acquisition process (see, for example, Gagliardi, 1991), but basically, these all require some form of ranging to be carried out so that a close estimate of the slot position can be acquired. In one method, the traffic station transmits a low-level burst consisting of the preamble only. The power level is 20 to 25 dB below the normal operating level (Ha, 1990) to prevent interference with other bursts, and the short burst is swept through the frame until it is observed to fall within the assigned time slot for the station. The short burst is then increased to full power, and fine adjustments in timing are made to bring it to the beginning of the time slot. Acquisition can take up to about 3 s in some cases. Following acquisition, the traffic data can be added, and synchronization can be maintained by continuously monitoring the position of the loopback transmission with reference to the SORF marker. The timing positions are reckoned from the last bit of the unique word in the preamble (as is also the case for the reference burst). The loopback method is also known as *direct closed-loop feedback*.

Feedback timing control. Where a traffic station lies outside the satellite beam containing its own transmission, loopback of the transmission does not of course occur, and some other method must be used for the station to receive ranging information. Where the synchronization information is transmitted back to an earth station from a distant station, this is termed *feedback closed-loop control*. The distant station may be a reference station, as in the INTELSAT network, or it may be another traffic station which is a designated *partner*. During the acquisition stage, the distant station can feed back information to guide the positioning of the short burst, and once the correct time slot is acquired, the necessary synchronizing information can be fed back on a continuous basis.

Figure 14.17 illustrates the feedback closed-loop control method for two earth stations A and B. The SORF marker is used as a reference point for the burst transmissions. However, the reference point which denotes the start of transmit frame (SOTF) has to be delayed by a certain amount, shown as D_A for earth station A and D_B for earth station B. This is necessary so that the SOTF reference points for each earth station coincide at the satellite transponder, and the traffic bursts, which are transmitted at their designated times after the SOTF, arrive in their correct relative positions at the transponder, as shown in Fig. 14.17. The total time delay between any given satellite clock pulse and the corresponding SOTF is a constant, shown as C in Fig. 14.17. C is equal to $2t_A + D_A$ for station A and $2t_B + D_B$ for station B. In general, for earth station i, the delay D_i is determined by

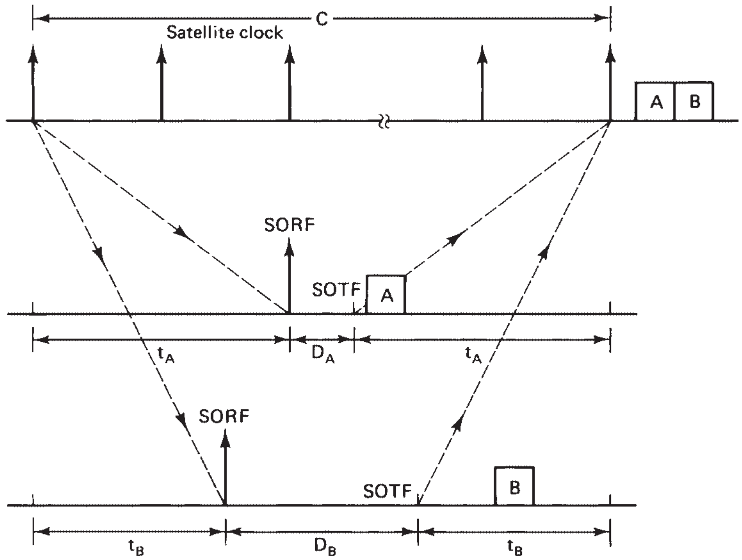


Figure 14.17 Timing relationships in a TDMA system. SORF, start of receive frame; SOTF, start of transmit frame.

$$2t_i + D_i = C \tag{14.17}$$

In the INTELSAT network, $C = 288$ ms.

For a truly geostationary satellite, the propagation delay t_i would be constant. However, as shown in Sec. 7.4, station-keeping maneuvers are required to keep a geostationary satellite at its assigned orbital position, and hence this position can be held only within certain tolerances. For example, in the INTELSAT network, the variation in satellite position can lead to a variation of up to ±0.55 ms in the propagation delay (INTELSAT, 1980). In order to minimize the guard time needed between bursts, this variation in propagation delay must be taken into account in determining the delay D_i required at each traffic station. In the INTELSAT network, the D_i numbers are updated every 512 frames, which is a period of 1.024 s, based on measurements and calculations of the propagation delay times made at the reference station. The D_i numbers are transmitted to the earth stations through the CDC channel in the reference bursts. (It should be noted that the open-loop synchronization described previously amounts to using a constant D_i value.)

The use of traffic burst preambles along with reference bursts to achieve synchronization is the most common method, but at least one other method, not requiring preambles, has been proposed by Nuspl and de Buda (1974). It also should be noted that there are certain types of "packet satellite networks," for example, the basic Aloha system (Rosner, 1982), which are closely related to TDMA, in which synchronization is not used.

14.7.5 Unique word detection

The *unique word* (UW) or *burst code word* (BCW) is used to establish burst timing in TDMA. Figure 14.18 shows the basic arrangement for detecting the UW. The received bit stream is passed through a shift register which forms part of a correlator. As the bit stream moves through the register, the sequence is continuously compared with a stored version of the UW. When correlation is achieved, indicated by a high output from the threshold detector, the last bit of the UW provides the reference point for timing purposes. It is important therefore to know the probability of error in detecting the UW. Two possibilities have to be considered. One, termed the *miss probability,* is the probability of the correlation detector failing to detect the UW even though it is present in the bit stream. The other, termed the *probability of false alarm,* is the probability that the correlation detector misreads a sequence as the UW. Both of these will be examined in turn.

Miss probability. Let E represent the maximum number of errors allowed in the UW of length N bits, and let I represent the actual number of errors in the UW as received. The following conditions apply:

When $I \leq E,$ the detected sequence is declared to be the UW.

When $I > E,$ the detected sequence N is declared not to be the UW; that is, the unique word is missed.

Let p represent the average probability of error in transmission (the BER). The probability of receiving a sequence N containing I errors in any one particular arrangement is

$$p_I = p^I (1 - p)^{N-I} \tag{14.18}$$

The number of combinations of N things taken I at a time, usually written as ${}_NC_I,$ is given by

$${}_NC_I = \frac{N!}{I!\,(N-I)!} \tag{14.19}$$

The probability of receiving a sequence of N bits containing I errors is therefore

$$P_I = {}_NC_I p_I \tag{14.20}$$

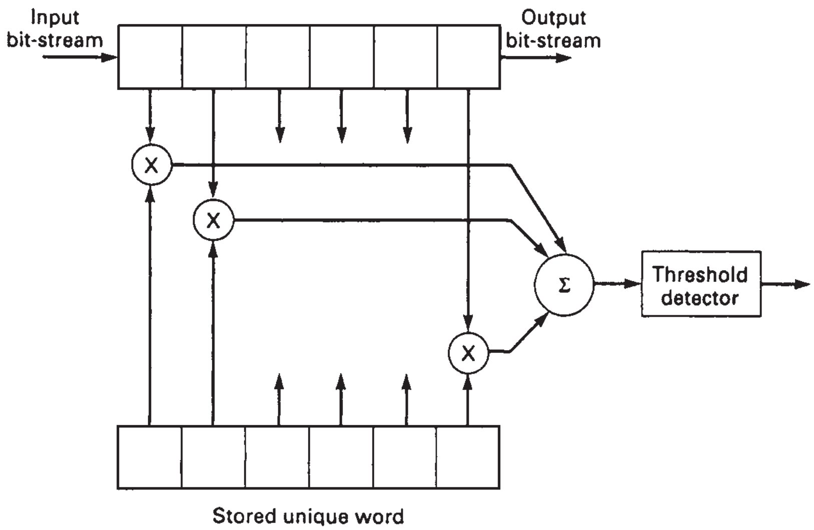


Figure 14.18 Basic arrangement for detection of the unique word (UW).

Now since the UW is just such a sequence, Eq. (14.20) gives the probability of a UW containing I errors. The condition for a miss occurring is that $I > E$, and therefore, the miss probability is

$$P_{\text{miss}} = \sum_{I=E+1}^{N} P_I \tag{14.21}$$

Written out in full, this is

$$P_{\text{miss}} = \sum_{I=E+1}^{N} \frac{N!}{I!\,(N-I)!}\, p^I\,(1-p)^{N-I} \tag{14.22}$$

Equation (14.22) gives the *average* probability of missing the UW even though it is present in the shift register of the correlator. Note that because this is an average probability, it is not necessary to know any specific value of I. Example 14.2 shows this worked in Mathcad.

Example 14.2 Determine the miss probability for the following values:

$$\mathrm{N} := 40 \qquad \mathrm{E} := 5 \qquad \mathrm{p} := 10^{-3}$$

solution

$$\mathrm{P_{miss}} := \sum_{\mathrm{I}=\mathrm{E}+1}^{\mathrm{N}} \frac{\mathrm{N}!}{\mathrm{I}! \cdot (\mathrm{N}-\mathrm{I})!} \cdot \mathrm{p}^{\mathrm{I}} \cdot (1-\mathrm{p})^{\mathrm{N}-\mathrm{I}} \qquad \mathrm{P_{miss}} = 3.7 \cdot 10^{-12}$$

======

False detection probability. Consider now a sequence of N which is not the UW but which would be interpreted as the UW even if it differs from it in some number of bit positions E, and let I represent the number of bit positions by which the random sequence actually does differ from the UW. Thus E represents the number of acceptable "bit errors" considered from the point of view of the UW, although they may not be errors in the message they represent. Likewise, I represents the actual number of "bit errors" considered from the point of view of the UW, although they may not be errors in the message they represent. As before, the number of combinations of N things taken I at a time is given by Eq. (14.19), and hence the number of words acceptable as the UW is

$$W = \sum_{I=0}^{E} {}_{N}C_{I} \tag{14.23}$$

The number of words which can be formed from a random sequence of N bits is 2^N, and on the assumption that all such words are equiprobable, the probability of receiving any one particular word is 2^{-N}. Hence the probability of a false detection is

$$P_F = 2^{-N}W \tag{14.24}$$

Written out in full, this is

$$P_F = 2^{-N} \sum_{I=0}^{E} \frac{N!}{I!\,(N-I)!} \tag{14.25}$$

Again it will be noticed that because this is an average probability, it is not necessary to know a specific value of I. Also, in this case, the BER does not enter into the calculation. A Mathcad calculation is given in Example 14.3.

Example 14.3 Determine the probability of false detection for the following values:

$$\mathrm{N} := 40 \qquad \mathrm{E} := 5$$

solution

$$\mathrm{P_F} := 2^{-\mathrm{N}} \cdot \sum_{\mathrm{I}=0}^{\mathrm{E}} \frac{\mathrm{N}!}{\mathrm{I}! \cdot (\mathrm{N}-\mathrm{I})!} \qquad \mathrm{P_F} = 6.9 \cdot 10^{-7}$$

======

From Examples 14.2 and 14.3 it is seen that the probability of a false detection is much higher than the probability of a miss, and this is true in general. In practice, once frame synchronization has been established, a time window can be formed around the expected time of arrival for the UW such that the correlation detector is only in operation for the window period. This greatly reduces the probability of false detection.

14.7.6 Traffic data

The traffic data immediately follow the preamble in a burst. As shown in Fig. 14.13, the traffic data subburst is further subdivided into time slots addressed to the individual destination stations. Any given destination station selects only the data in the time slots intended for that station. As with FDMA networks, TDMA networks can be operated with both preassigned and demand assigned channels, and examples of both types will be given shortly.

The greater the fraction of frame time that is given over to traffic, the higher is the efficiency. The concept of *frame efficiency* is discussed in the next section.

14.7.7 Frame efficiency and channel capacity

The frame efficiency is a measure of the fraction of frame time used for the transmission of traffic. *Frame efficiency* may be defined as

$$\text{Frame efficiency} = \eta_F = \frac{\text{traffic bits}}{\text{total bits}} \tag{14.26}$$

Alternatively, this can be written as

$$\eta_F = 1 - \frac{\text{overhead bits}}{\text{total bits}} \tag{14.27}$$

In these equations, bits per frame are implied. The overhead bits consist of the sum of the preamble, the postamble, the guard intervals, and the reference-burst bits per frame. The equations may be stated in terms of symbols rather than bits, or the actual times may be used.

For a fixed overhead, Eq. (14.27) shows that a longer frame, or greater number of total bits, results in higher efficiency. However, longer frames require larger buffer memories and also add to the propagation delay. Synchronization also may be made more difficult, keeping in mind that the satellite position is varying with time. It is clear that a lower overhead also leads to higher efficiency, but again, reducing synchronizing and guard times may result in more complex equipment being required.

Example 14.4 Calculate the frame efficiency for an INTELSAT frame given the following information:

Total frame length = 120,832 symbols

Traffic bursts per frame = 14

Reference bursts per frame = 2

Guard interval = 103 symbols

solution From Fig. 14.14, the preamble symbols add up to

$$\begin{aligned} P &= 176 + 24 + 8 + 8 + 32 + 32 \\ &= 280 \end{aligned}$$

With addition of the CDC channel, the reference channel symbols add up to

$$\begin{aligned} R &= 280 + 8 \\ &= 288 \end{aligned}$$

Therefore, the overhead symbols are

$$\begin{aligned} OH &= 2 \times (103 + 288) + 14 \times (103 + 280) \\ &= 6144 \text{ symbols} \end{aligned}$$

Therefore, from Eq. (14.27),

$$\eta_F = 1 - \frac{6144}{120{,}832} = 0.949$$

The voice-channel capacity of a frame, which is also the voice-channel capacity of the transponder being accessed by the frame, can be found from a knowledge of the frame efficiency and the bit rates. Let R_b be the bit rate of a voice channel, and let there be a total of n voice channels shared between all the earth stations accessing the transponder. The total incoming *traffic* bit rate to a frame is nR_b. The traffic bit rate of the frame is $\eta_F R_T$, and therefore

$$nR_b = \eta_F R_{\text{TDMA}}$$

or

$$n = \frac{\eta_F R_{\text{TDMA}}}{R_b} \tag{14.28}$$

Example 14.5 Calculate the voice-channel capacity for the INTELSAT frame in Example 14.2, given that the voice-channel bit rate is 64 kb/s and that QPSK modulation is used. The frame period is 2 ms.

solution The number of symbols per frame is 120,832, and the frame period is 2 ms. Therefore, the symbol rate is 120,832/2 ms = 60.416 megasymbols/s. QPSK modulation utilizes 2 bits per symbol, and therefore, the transmission rate is $R_T = 60.416 \times 2 = 120.832$ Mb/s.

Using Eq. (14.28) and the efficiency as calculated in Example 14.4,

$$n = 0.949 \times 120.832 \times \frac{10^3}{64} = 1792$$

14.7.8 Preassigned TDMA

An example of a preassigned TDMA network is the common signaling channel (CSC) for the Spade network described in Sec. 14.5. The frame and burst formats are shown in Fig. 14.19. The CSC can accommodate up to 49 earth stations in the network plus one reference station, making a maximum of 50 bursts in a frame.

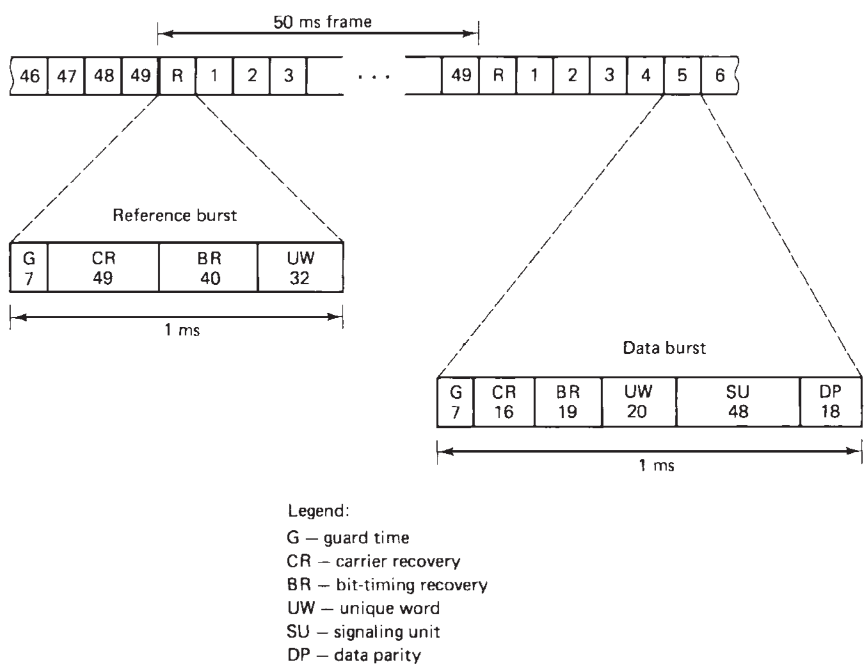


Figure 14.19 Frame and bit formats for the common signaling channel (CSC) used with the Spade system. (*Data from Miya, 1981.*)

All the bursts are of equal length. Each burst contains 128 bits and occupies a 1-ms time slot. Thus the bit rate is 128 kb/s. As discussed in Sec. 14.5, the frequency bandwidth required for the CSC is 160 kHz.

The *signaling unit* (SU) shown in Fig. 14.19 is that section of the data burst which is used to update the other stations on the status of the frequencies available for the SCPC calls. It also carries the signaling information, as described in Sec. 14.5.

Another example of a preassigned TDMA frame format is the INTELSAT frame shown in simplified form in Fig. 14.20. In the INTELSAT system, preassigned and demand-assigned voice channels are carried together, but for clarity, only a preassigned traffic burst is shown. The traffic burst is subdivided into time slots, termed *satellite*

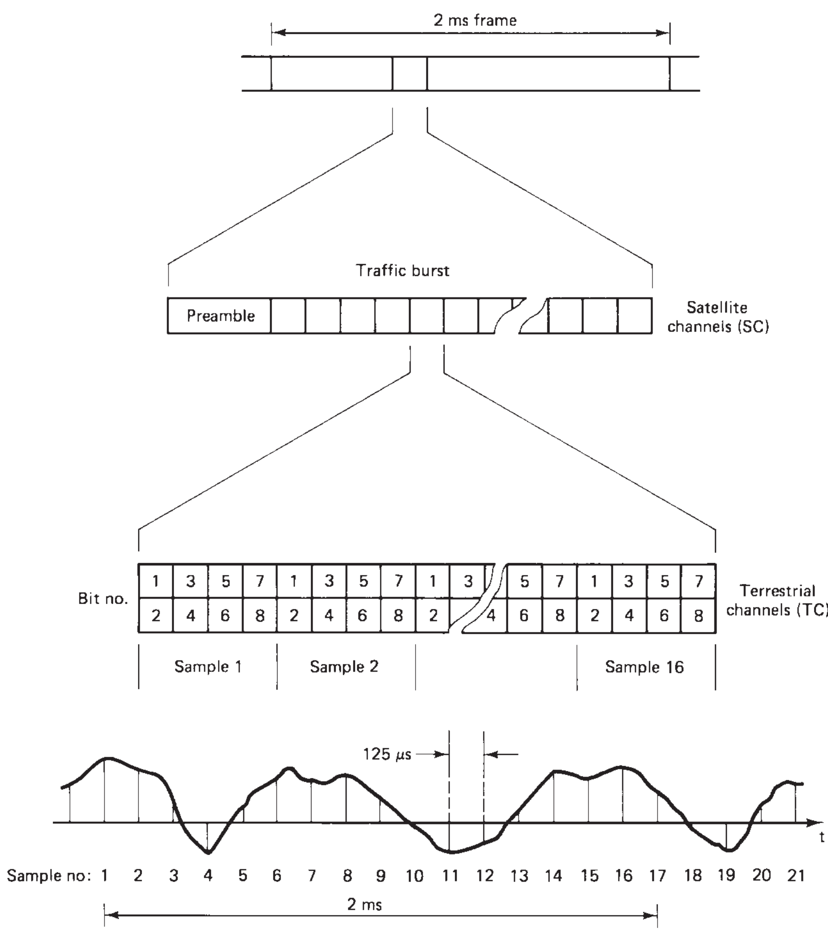


Figure 14.20 Preassigned TDMA frame in the Intelsat system.

channels in the INTELSAT terminology, and there can be up to 128 of these in a traffic burst. Each satellite channel is further subdivided into 16 time slots termed *terrestrial channels,* each terrestrial channel carrying one PCM sample of an analog telephone signal. QPSK modulation is used, and therefore, there are 2 bits per symbol as shown. Thus each terrestrial channel carries 4 symbols (or 8 bits). Each satellite channel carries $4 \times 16 = 64$ symbols, and at its maximum of 128 satellite channels, the traffic burst carries 8192 symbols.

As discussed in Sec. 10.3, the PCM sampling rate is 8 kHz, and with 8 bits per sample, the PCM bit rate is 64 kb/s. Each satellite channel can accommodate this bit rate. Where input data at a higher rate must be transmitted, multiple satellite channels are used. The maximum input data rate which can be handled is 128(SC) $\times$ 64 kb/s = 8.192 Mb/s.

The INTELSAT frame is 120,832 symbols or 241,664 bits long. The frame period is 2 ms, and therefore, the burst bit rate is 120.832 Mb/s.

As mentioned previously, preassigned and demand-assigned voice channels can be accommodated together in the INTELSAT frame format. The demand-assigned channels utilize a technique known as *digital speech interpolation* (DSI), which is described in the next section. The preassigned channels are referred to as *digital noninterpolated* (DNI) *channels.*

14.7.9 Demand-assigned TDMA

With TDMA, the burst and subburst assignments are under software control, compared with hardware control of the carrier frequency assignments in FDMA. Consequently, compared with FDMA networks, TDMA networks have more flexibility in reassigning channels, and the changes can be made more quickly and easily.

A number of methods are available for providing traffic flexibility with TDMA. The burst length assigned to a station may be varied as the traffic demand varies. A central control station may be employed by the network to control the assignment of burst lengths to each participating station. Alternatively, each station may determine its own burst-length requirements and assign these in accordance with a prearranged network discipline.

As an alternative to burst-length variation, the burst length may be kept constant and the number of bursts per frame used by a given station varied as demand requires. In one proposed system (CCIR Report 708, 1982), the frame length is fixed at 13.5 ms. The basic burst time slot is 62.5 μs, and stations in the network transmit information bursts varying in discrete steps over the range 0.5 ms (8 basic bursts) to 4.5 ms (72 basic bursts) per frame. Demand assignment for speech channels takes advantage of the intermittent nature of speech, as described in the next section.

14.7.10 Speech interpolation and prediction

Because of the intermittent nature of speech, a speech transmission channel lies inactive for a considerable fraction of the time it is in use. A number of factors contribute to this. The talk-listen nature of a normal two-way telephone conversation means that transmission in any one direction occurs only about 50 percent of the time. In addition, the pauses between words and phrases may further decrease this to about 33 percent. If further allowance is made for "end party" delays such as the time required for a party to answer a call, the average fraction of the total connect time may drop to as low as 25 percent. The fraction of time a transmission channel is active is known as the *telephone load activity factor,* and for system design studies, the value of 0.25 is recommended by Comité Consutatif Internationale Télégraphique et Téléphonique (CCITT), although higher values are also used (Pratt and Bostian, 1986). The point is that for a significant fraction of the time the channel is available for other transmissions, and advantage is taken of this in a form of demand assignment known as *digital speech interpolation.*

Digital speech interpolation may be implemented in one of two ways, these being digital *time assignment speech interpolation* (digital TASI) and *speech predictive encoded communications* (SPEC).

Digital TASI. The traffic-burst format for an INTELSAT burst carrying demand-assigned channels and preassigned channels is shown in Fig. 14.21. As mentioned previously, the demand-assigned channels utilize digital TASI, or what is referred to in the INTELSAT nomenclature as

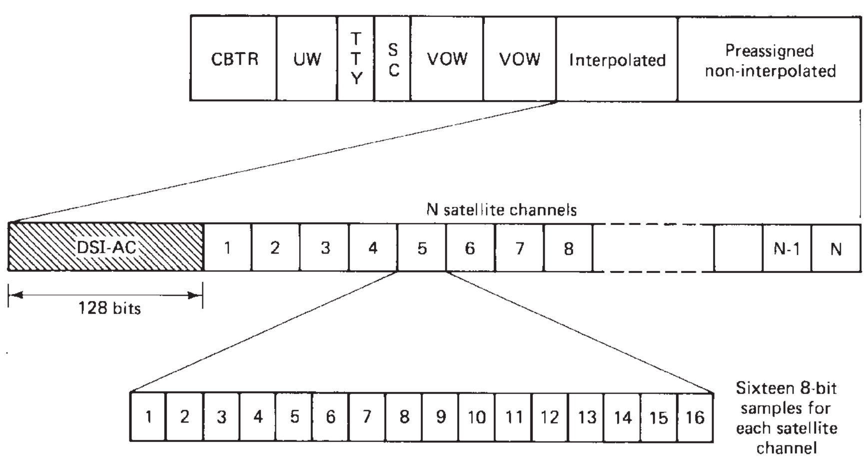


Figure 14.21 Intelsat traffic burst structure. (*From Intelsat, 1983. With permission.*)

DSI, for *digital speech interpolation.* These are shown by the block labeled "interpolated" in Fig. 14.21. The first satellite channel (channel 0) in this block is an assignment channel, labeled *DSI-AC.* No traffic is carried in the assignment channel; it is used to transmit channel assignment information as will be described shortly.

Figure 14.22 shows in outline the DSI system. Basically, the system allows N terrestrial channels to be carried by M satellite channels, where $N > M$. For example, in the INTELSAT arrangement, $N = 240$ and $M = 127$.

On each incoming terrestrial channel, a speech detector senses when speech is present, the intermittent speech signals being referred to as *speech spurts.* A speech spurt lasts on average about 1.5 seconds (Miya, 1981). A control signal from the speech detector is sent to the channel assignment unit, which searches for an empty TDMA buffer. Assuming that one is found, the terrestrial channel is assigned to this satellite channel, and the speech spurt is stored in the buffer, ready for transmission in the DSI subburst. A delay is inserted in the speech circuit, as shown in Fig. 14.22, to allow some time for the assignment process to be completed. However, this delay cannot exactly compensate for the assignment delay, and the initial part of the speech spurt may be lost. This is termed a *connect clip.*

In the INTELSAT system an intermediate step occurs where the terrestrial channels are renamed *international channels* before being assigned to a satellite channel (Pratt and Bostian, 1986). For clarity, this step is not shown in Fig. 14.22.

At the same time as an assignment is made, an assignment message is stored in the assignment channel buffer, which informs the receive stations which terrestrial channel is assigned to which satellite channel. Once an assignment is made, it is not interrupted, even during pauses between spurts, unless the pause times are required for another DSI channel. This reduces the amount of information needed to be transmitted over the assignment channel.

At the receive side, the traffic messages are stored in their respective satellite-channel buffers. The assignment information ensures that the correct buffer is read out to the corresponding terrestrial channel during its sampling time slot. During speech pauses when the channel has been reassigned, a low-level noise signal is introduced at the receiver to simulate a continuous connection.

It has been assumed that a free satellite channel will be found for any incoming speech spurt, but of course there is a finite probability that all channels will be occupied and the speech spurt lost. Losing a speech spurt in this manner is referred to as *freeze-out,* and the freeze-out fraction is the ratio of the time the speech is lost to the average spurt duration. It is found that a design objective of 0.5 percent for a

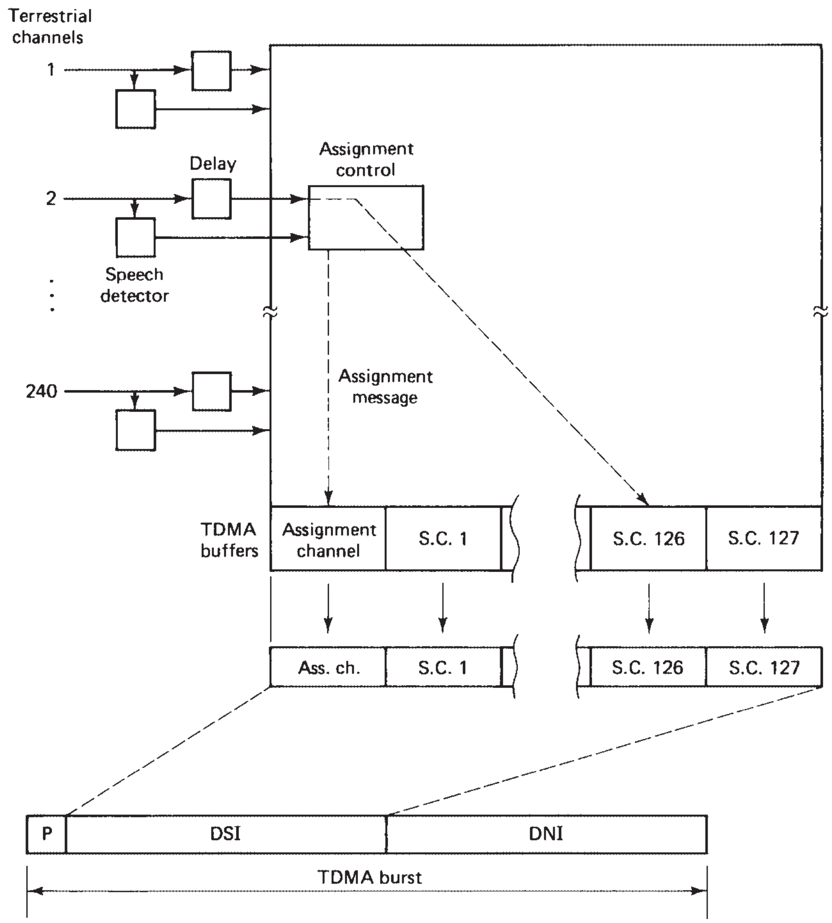


Figure 14.22 Digital speech interpolation. DSI = digital speech interpolation; DNI = digital noninterpolation.

freeze-out fraction is satisfactory in practice. This means that the probability of a freeze-out occurring is 0.005.

Another source of signal mutilation is the *connect clip* mentioned earlier. Again, it is found in practice that clips longer than about 50 ms are very annoying to the listener. An acceptable design objective is to limit the fraction of clips which are equal to or greater than 50 ms to a maximum of 2 percent of the total clips. In other words, the probability of encountering a clip that exceeds 50 ms is 0.02.

The *DSI gain* is the ratio of the number of terrestrial channels to number of satellite channels, or *N/M*. The DSI gain depends on the number of satellite channels provided as well as the design objectives stated above. Typically, DSI gains somewhat greater than 2 can be achieved in practice.

Speech predictive encoded communications (SPEC). The block diagram for the SPEC system is shown in Fig. 14.23 (Sciulli and Campanella, 1973). In this method, the incoming speech signals are converted to a PCM multiplexed signal using 8 bits per sample quantization. With 64 input lines and sampling at 125 μs, the output bit rate from the multiplexer is $8 \times 64/125 = 4.096$ Mb/s.

The digital voice switch following the PCM multiplexer is timeshared between the input signals. It is voice-activated to prevent transmission of noise during silent intervals. When the zero-order predictor receives a new sample, it compares it with the previous sample for that voice channel, which it has stored, and transmits the new sample only if it differs from the preceding one by a predetermined amount. These new samples are labeled *unpredictable PCM samples* in Fig. 14.23*a*.

For the 64 channels a 64-bit assignment word is also sent. A logic 1 in the channel for the assignment word means that a new sample was sent for that channel, and a logic 0 means that the sample was unchanged. At the receiver, the sample assignment word either directs the new (unpredictable) sample into the correct channel slot, or it results in the previous sample being regenerated in the reconstruction decoder. The output from this is a 4.096-Mb/s PCM multiplexed signal which is demultiplexed in the PCM decoder.

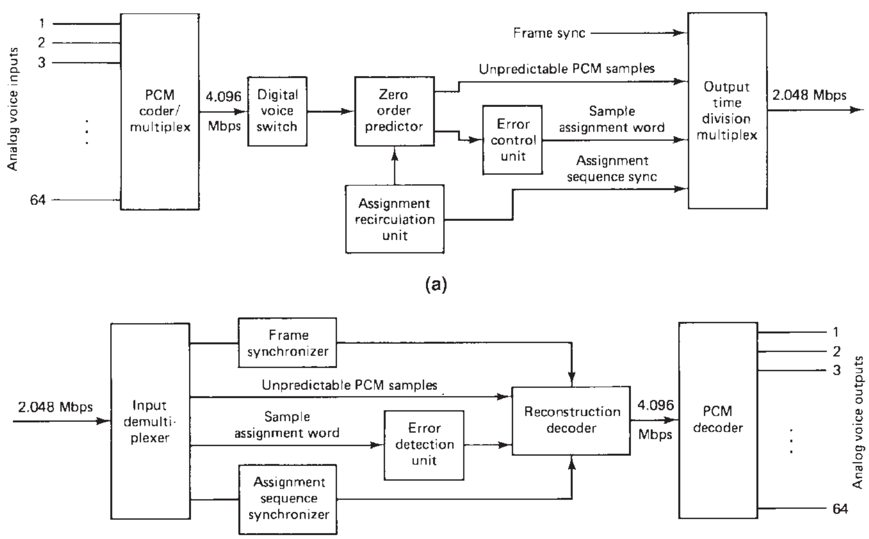


Figure 14.23 (*a*) SPEC transmitter, (*b*) SPEC receiver. (*From Sciulli and Campanella, 1973. © 1973—IEEE.*)

By removing the redundant speech samples and silent periods from the transmission link, a doubling in channel capacity is achieved. As shown in Fig. 14.23, the transmission is at 2.048 Mb/s for an input-output rate of 4.096 Mb/s.

An advantage of the SPEC method over the DSI method is that freeze-out does not occur during overload conditions. During overload, sample values which should change may not. This effectively leads to a coarser quantization and therefore an increase in quantization noise. This is subjectively more tolerable than freeze-out.

14.7.11 Downlink analysis for digital transmission

As mentioned in Sec. 14.6, the transponder power output and bandwidth both impose limits on the system transmission capacity. With TDMA, TWT backoff is not generally required, which allows the transponder to operate at saturation. One drawback arising from this is that the uplink station must be capable of saturating the transponder, which means that even a low-traffic-capacity station requires comparatively large power output compared with what would be required for FDMA. This point is considered further in Sec. 14.7.12.

As with the FDM/FDMA system analysis, it will be assumed that the overall carrier-to-noise ratio is essentially equal to the downlink carrier-to-noise ratio. With a power-limited system this C/N ratio is one of the factors that determines the maximum digital rate, as shown by Eq. (10.24). Equation (10.24) can be rewritten as

$$[R_b] = \left[\frac{C}{N_0}\right] - \left[\frac{E_b}{N_0}\right] \tag{14.29}$$

The $[E_b/N_0]$ ratio is determined by the required bit-error rate, as shown in Fig. 10.17 and described in Sec. 10.6.4. For example, for a BER of 10^{-5} an $[E_b/N_0]$ of 9.6 dB is required. If the rate R_b is specified, then the $[C/N_0]$ ratio is determined, as shown by Eq. (14.29), and this value is used in the link budget calculations as required by Eq. (12.53). Alternatively, if the $[C/N_0]$ ratio is fixed by the link budget parameters as given by Eq. (12.53), the bit rate is then determined by Eq. (14.29).

The bit rate is also constrained by the IF bandwidth. As shown in Sec. 10.6.3, the ratio of bit rate to IF bandwidth is given by

$$\frac{R_b}{B_{\mathrm{IF}}} = \frac{m}{1 + \rho}$$

where $m = 1$ for BPSK and $m = 2$ for QPSK and ρ is the rolloff factor. The value of 0.2 is commonly used for the rolloff factor, and therefore, the bit rate for a given bandwidth becomes

$$R_b = \frac{mB_{\mathrm{IF}}}{1.2} \tag{14.30}$$

Example 14.6 Using Eq. (12.53), a downlink $[C/N_0]$ of 87.3 dBHz is calculated for a TDMA circuit that uses QPSK modulation. A BER of 10^{-5} is required. Calculate the maximum transmission rate. Calculate also the IF bandwidth required assuming a rolloff factor of 0.2.

solution Figure 10.17 is applicable for QPSK and BPSK. From this figure, $[E_b/N_0] = 9.5$ dB for a BER of 10^{-5}. Hence

$$[R_b] = 87.3 - 9.5 = 77.8 \text{ dBb/s}$$

This is equal to 60.25 Mb/s.

For QPSK $m = 2$ and using Eq. (14.30), we have

$$B_{\mathrm{IF}} = 60.25 \times \frac{1.2}{2} = 36.15 \text{ MHz}$$

From Example 14.6 it will be seen that if the satellite transponder has a bandwidth of 36 MHz, and an EIRP that results in a $[C/N_0]$ of 87.3 dBHz at the receiving ground station, the system is optimum in that the power and the bandwidth limits are reached together.

14.7.12 Comparison of uplink power requirements for FDMA and TDMA

With frequency-division multiple access, the modulated carriers at the input to the satellite are retransmitted from the satellite as a combined frequency-division-multiplexed signal. Each carrier retains its modulation, which may be analog or digital. For this comparison, digital modulation will be assumed. The modulation bit rate for each carrier is equal to the input bit rate [adjusted as necessary for forward error correction (FEC)]. The situation is illustrated in Fig. 14.24*a,* where for simplicity, the input bit rate R_b is assumed to be the same for each earth station. The [EIRP] is also assumed to be the same for each earth station.

With time-division multiple access, the uplink bursts which are displaced in time from one another are retransmitted from the satellite as a combined time-division-multiplexed signal. The uplink bit rate is equal to the downlink bit rate in this case, as illustrated in Fig. 14.24*b*. As described in Sec. 14.7, compression buffers are needed in order to convert the input bit rate R_b to the transmitted bit rate R_{TDMA}.

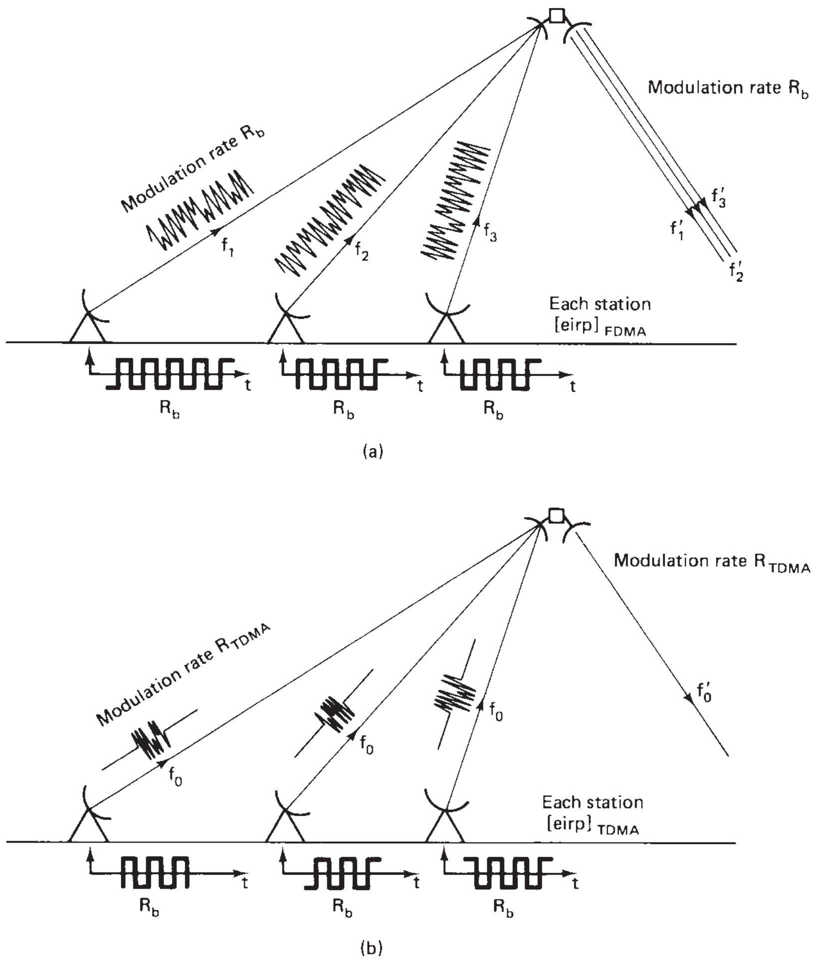


Figure 14.24 (*a*) FDMA network; (*b*) TDMA network.

Because the TDMA earth stations have to transmit at a higher bit rate compared with FDMA, a higher [EIRP] is required, as can be deduced from Eq. (10.24). Equation (10.24) states that

$$\left[\frac{C}{N_0}\right] = \left[\frac{E_b}{N_0}\right] + [R]$$

where $[R]$ is equal to $[R_b]$ for an FDMA uplink and $[R_{\mathrm{TDMA}}]$ for a TDMA uplink.

For a given bit error rate (BER) the $[E_b/N_0]$ ratio is fixed as shown by Fig. 10.17. Hence, assuming that $[E_b/N_0]$ is the same for the TDMA and the FDMA uplinks, an increase in $[R]$ requires a corresponding increase in $[C/N_0]$. Assuming that the TDMA and FDMA uplinks operate with the same [LOSSES] and satellite $[G/T]$, Eq. (12.39) shows that the increase in $[C/N_0]$ can be achieved only through an increase in the earth station [EIRP], and therefore

$$[\text{EIRP}]_{\text{TDMA}} - [\text{EIRP}]_{\text{FDMA}} = [R_{\text{TDMA}}] - [R_b] \tag{14.31}$$

For the same earth station antenna gain in each case, the decibel increase in earth station transmit power for TDMA compared with FDMA is

$$[P]_{\text{TDMA}} - [P]_{\text{FDMA}} = [R_{\text{TDMA}}] - [R_b] \tag{14.32}$$

Example 14.7 A 14-GHz uplink operates with transmission losses and margins totaling 212 dB and a satellite $[G/T]$ = 10 dB/K. The required uplink $[E_b/N_0]$ is 12 dB. (*a*) Assuming FDMA operation and an earth station uplink antenna gain of 46 dB, calculate the earth station transmitter power needed for transmission of a T1 baseband signal. (*b*) If the downlink transmission rate is fixed at 74 dBb/s, calculate the uplink power increase required for TDMA operation.

solution (*a*) From Sec. 10.4 the T1 bit rate is 1.544 Mb/s or $[R]$ = 62 dBb/s. Using the $[E_b/N_0]$ = 12-dB value specified, Eq. (10.24) gives

$$\left[\frac{C}{N_0}\right] = 12 + 62 = 74 \text{ dBHz}$$

From Eq. (12.39),

$$\begin{aligned}[\text{EIRP}] &= \left[\frac{C}{N_0}\right] - \left[\frac{G}{T}\right] + [\text{LOSSES}] - 228.6 \\ &= 74 - 10 + 212 - 228.6 \\ &= 47.4 \text{ dBW}\end{aligned}$$

Hence the transmitter power required is

$$[P] = 47.4 - 46 = 1.4 \text{ dBW} \qquad \text{or} \qquad 1.38 \text{ W}$$

(*b*) With TDMA operation the rate increase is 74 − 62 = 12 dB. All other factors being equal, the earth station [EIRP] must be increased by this amount, and hence

$$[P] = 1.4 + 12 = 13.4 \text{ dBW} \qquad \text{or} \qquad 21.9 \text{ W}$$

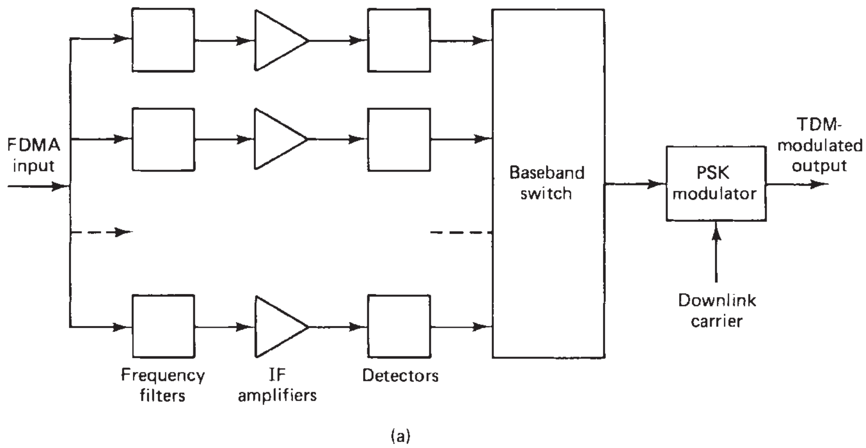
For small satellite business systems it is desirable to be able to operate with relatively small earth stations, which suggests that FDMA should be the mode of operation. On the other hand, TDMA permits more efficient use of the satellite transponder by eliminating the need for backoff. This suggests that it might be worthwhile to operate a hybrid system in which FDMA is the uplink mode of operation, with the individual signals converted to a time-division-multiplexed format in the transponder before being amplified by the TWTA. This would allow the transponder to be operated at saturation as in TDMA. Such a hybrid mode of operation would require the use of a signal-processing transponder as discussed in the next section.

14.8 On-Board Signal Processing for FDMA/TDM Operation

As seen in the preceding section, for small earth stations carrying digital signals at relatively low data rates, there is an advantage to be gained in terms of earth station power requirements by using FDMA. On the other hand, TDMA signals make more efficient use of the transponder because back-off is not required.

Market studies show that what is termed *customer premises services* (CPS) will make up a significant portion of the satellite demand over the decade 1990–2000 (Stevenson et al., 1984). Multiplexed digital transmission will be used, most likely at the T1 rate. This bit rate provides for most of the popular services, such as voice, data, and videoconferencing, but specifically excludes standard television signals. Customer premises services is an ideal candidate for the FDMA/TDM mode of operation mentioned in the preceding section. To operate in this mode requires the use of *signal-processing transponders,* in which the FDMA uplink signals are converted to the TDM format for retransmission on the downlink. It also should be noted that the use of signal-processing transponders "decouples" the uplink from the downlink. This is important because it allows the performance of each link to be optimized independently of the other.

A number of signal-processing methods have been proposed. One conventional approach is illustrated in the simplified block schematic of Fig. 14.25*a*. Here the individual uplink carriers at the satellite are selected by frequency filters and detected in the normal manner. The baseband signals are then combined in the baseband processor, where they are converted to a time-division-multiplexed format for remodulation onto a downlink carrier. More than one downlink carrier may be provided, but only one is shown for simplicity. The disadvantages of the conventional approach are those of excessive size, weight, and



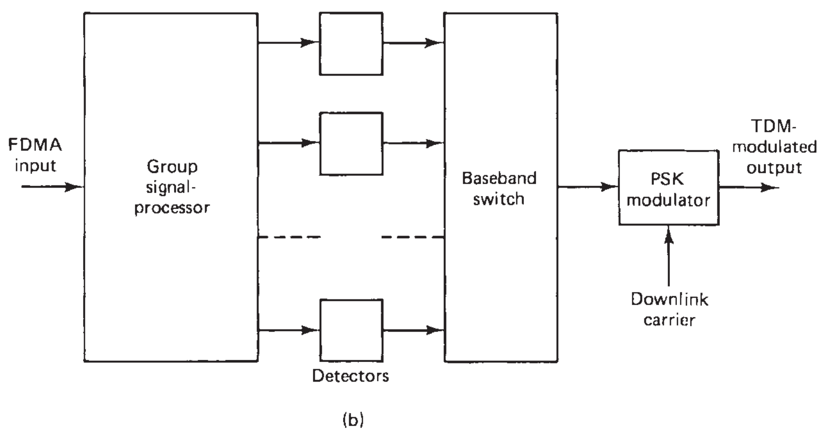


Figure 14.25 On-board signal processing for FDMA/TDM operation; (*a*) conventional approach; (*b*) group signal processing.

power consumption, since the circuitry must be duplicated for each input carrier.

The disadvantages associated with processing each carrier separately can be avoided by means of *group processing,* in which the input FDMA signals are demultiplexed as a group in a single processing circuit, illustrated in Fig. 14.25*b*. Feasibility studies are being conducted into the use of digital-type group processors, although it would appear that these may require very high speed integrated circuits (VHSICs) not presently available. A different approach to the problem of group processing has been proposed, which makes use of an analog device known as a surface acoustic wave (SAW) Fourier transformer (Atzeni et al., 1975; Hays et al., 1975; Hays and Hartmann, 1976; Maines and Paige, 1976; Nud and Otto, 1975).

In its basic form, the surface acoustic wave device consists of two electrodes deposited on the surface of a piezoelectric dielectric. An electrical signal applied to the input electrode sets up a surface acoustic wave which induces a corresponding signal in the output electrode. In effect, the SAW device is a coupled circuit in which the coupling mechanism is the surface acoustic wave.

Because the propagation velocity of the acoustic wave is much lower than that of an electromagnetic wave, the SAW device exhibits useful delay characteristics. In addition, the electrodes are readily shaped to provide a wide range of useful transfer characteristics. These two features, along with the fact that the device is small, rugged, and passive, make it a powerful signal-processing component. SAW devices may be used conventionally as delay lines, as bandpass or bandstop filters, and they are the key component in a unit known as a Fourier transformer.

The Fourier transformer, like any other transformer, works with input and output signals which are functions of time. The unique property of the Fourier transformer is that the output signal is a time analog of the frequency spectrum of the input signal. When the input is a group of FDMA carriers, the output in the ideal case would be an analog of the FDMA frequency spectrum. This allows the FDMA signals to be demultiplexed in real time by means of a commutator switch, which eliminates the need for the separate frequency filters required in the conventional analog approach. Once the signals have been separated in this way, the original modulated-carrier waveforms may be recovered through the use of SAW inverse Fourier transformers.

In a practical transformer, continuous operation can only be achieved by repetitive cycling of the transformation process. As a result, the output is periodic, and the observation interval has to be chosen to correspond to the desired spectral interval. Also, the periodic interval over which transformation takes place results in a broadening of the output "pulses" which represent the FDMA spectra.

Repetitive operation of the Fourier transformer at a rate equal to the data bit rate will produce a suitable repetitive output. The relative positions of the output pulses will remain unchanged, fixed by the frequencies of the FDMA carriers. The PSK modulation on the individual FDMA carriers appears in the phase of the carriers within each output pulse. Thus the FDMA carriers have been converted to a pulsed TDM signal. Further signal processing is required before this can be retransmitted as a TDM signal.

Figure 14.26 shows the output obtained from a practical Fourier transformer for various input signals. For Fig. 14.26*a* the input was

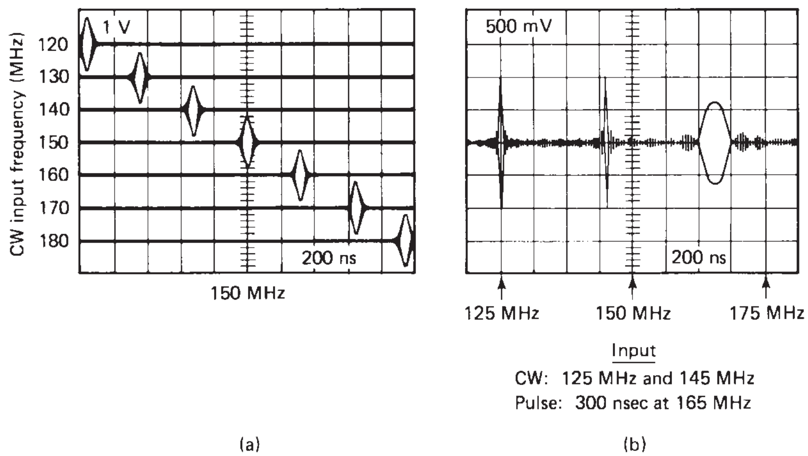


Figure 14.26 Prototype chirp transform of (*a*) seven successive CW input signals and (*b*) three simultaneous input signals, including CW and pulsed RF. 200 ns/div; 31.5 MHz./μs chirp rate. (*From Hays and Hartmann, 1976. © 1976—IEEE.*)

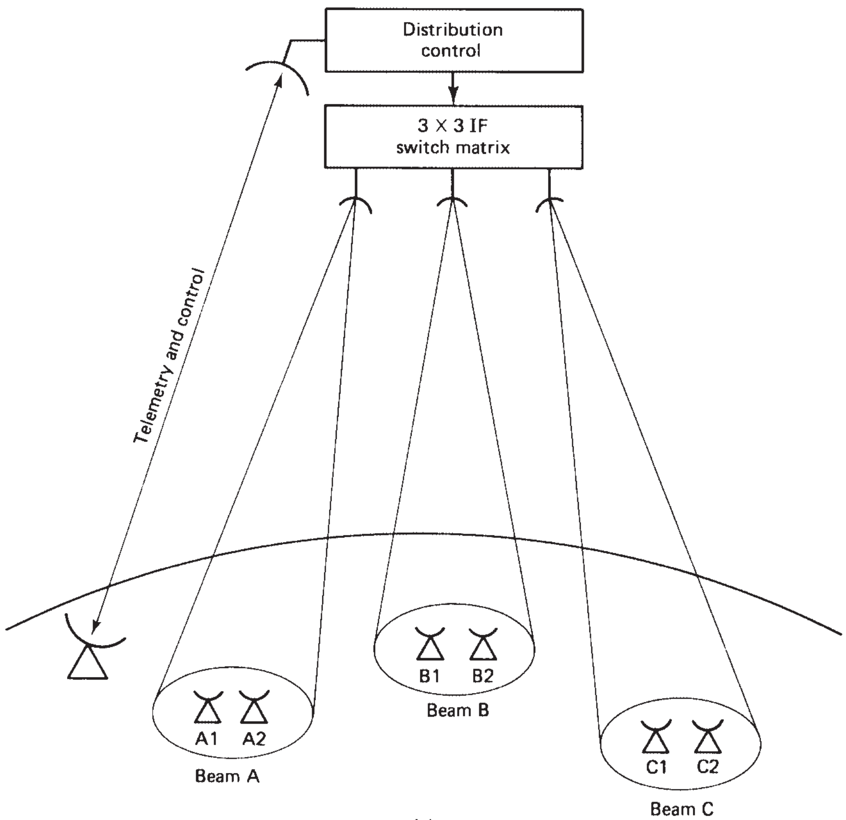
seven continuous-wave (CW) signals applied in succession. The output is seen to be pulses corresponding to the line spectra for these waves. The broadening of the lines is a result of the finite time gate over which the Fourier transformer operates. It is important to note that the horizontal axis in Fig. 14.26 is a time axis on which the equivalent frequency points are indicated.

Figure 14.26*b* shows the output obtained with three simultaneous inputs, two CW waves and one pulsed carrier wave. Again, the output contains two pulses corresponding to the CW signals and a time function which has the shape of the spectrum for the pulsed wave (Hays and Hartmann, 1976). A detailed account of SAW devices will be found in Morgan (1985) and in the *IEEE Proceedings* (1976).

14.9 Satellite-Switched TDMA

More efficient utilization of satellites in the geostationary orbit can be achieved through the use of antenna spot beams. The use of spot beams is also referred to as *space-division multiplexing*. Further improvements can be realized by switching the antenna interconnections in synchronism with the TDMA frame rate, this being known as *satellite-switched TDMA* (SS/TDMA).

Figure 14.27*a* shows in simplified form the SS/TDMA concept (Scarcella and Abbott, 1983). Three antenna beams are used, each beam serving two earth stations. A 3 × 3 satellite switch matrix is shown. This is the key component that permits the antenna interconnections to be made on a switched basis. A *switch mode* is a connectivity arrangement. With three beams, six modes would be required for full interconnectivity, as shown in Fig. 14.27*b,* and in general with N beams, $N!$ modes are required for full interconnectivity. Full interconnectivity means that the signals carried in each beam are transferred to each of the other beams at some time in the switching sequence. This includes the loopback connection, where signals are returned



(a)

Input	Output					
	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
A	A	A	B	C	B	C
B	B	C	A	A	C	B
C	C	B	C	B	A	A

(b)

Figure 14.27 (*a*) Satellite switching of three spot beams; (*b*) connectivities or modes.

along the same beam, enabling intercommunications between stations within a beam. Of course, the uplink and downlink microwave frequencies are different.

Because of beam isolation, one frequency can be used for all uplinks, and a different frequency for all downlinks (e.g., 14 and 12 GHz in the Ku band). To simplify the satellite switch design, the switching is carried out at the intermediate frequency that is common to uplinks and downlinks. The basic block schematic for the 3 × 3 system is shown in Fig. 14.28.

A *mode pattern* is a repetitive sequence of satellite switch modes, also referred to as *SS/TDMA frames.* Successive SS/TDMA frames need not be identical, since there is some redundancy between modes. For example, in Fig. 14.27*b,* beam *A* interconnects with beam *B* in modes 3 and 5, and thus not all modes need be transmitted during each SS/TDMA frame. However, for full interconnectivity, the *mode pattern* must contain all modes.

All stations within a beam receive all the TDM frames transmitted in the downlink beam. Each frame is a normal TDMA frame consisting of bursts, addressed to different stations in general. As mentioned, successive frames may originate from different transmitting stations and therefore have different burst formats. The receiving station in a beam recovers the bursts addressed to it in each frame.

The two basic types of switch matrix are the *crossbar matrix* and the *rearrangeable network.* The crossbar matrix is easily configured for the *broadcast mode,* in which one station transmits to all stations. The broadcast mode with the rearrangeable network-type switch is more complex, and this can be a deciding factor in favor of the crossbar matrix (Watt, 1986). The schematic for a 3 × 3 crossbar matrix is shown in Fig. 14.29, which also shows input beam *B* connected in the broadcast mode.

The switching elements may be ferrites, diodes, or transistors. The dual-gate FET appears to offer significant advantages over the other

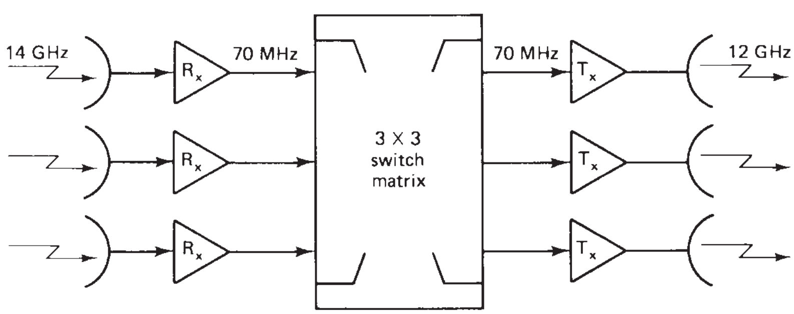


Figure 14.28 Switch matrix in the R.F. link.

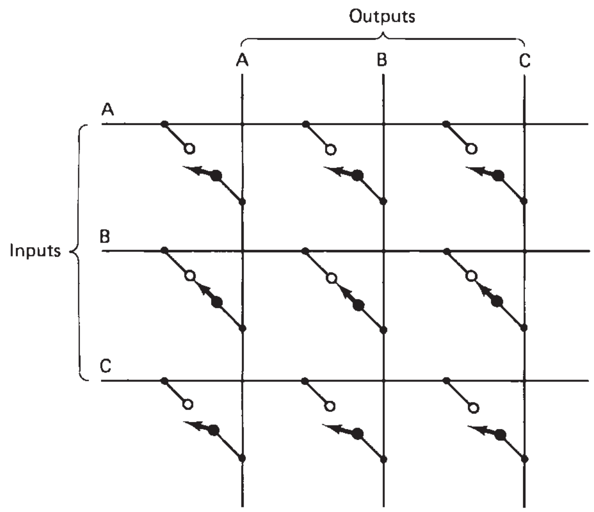


Figure 14.29 3×3 crossbar matrix switch, showing input B connected in the broadcast mode.

types and is considered by some to be the most promising technology (Scarcella and Abbott, 1983).

Figure 14.30 shows how a 3×3 matrix switch may be used to reroute traffic. Each of the ground stations *U, V,* and *W* accesses a separate antenna on the satellite and carries traffic destined for the downlink beams *X, Y,* and *Z.* The switch is controlled from a ground control station, and the switching sequence for the frame labeled with subscript 1 is shown in the lower part of the figure.

The schematic for a 4×4 matrix switch as used on the European Olympus satellite is shown in Fig.14.31 (Watt, 1986). This arrangement is derived from the crossbar matrix. It permits broadcast mode operation but does not allow more than one input to be connected to one output. Diodes are used as switching elements, and as shown, diode quads are used which provide redundancy against diode failure. It is clear that satellite-switched TDMA adds to the complexity of the on-board equipment and to the synchronization requirements.

14.10 Code-Division Multiple Access

With code-division multiple access (CDMA) the individual carriers may be present simultaneously within the same rf bandwidth, but each carrier carries a unique code waveform (in addition to the information signal) that allows it to be separated from all the others at the receiver. The carrier is modulated in the normal way by the information waveform and then is further modulated by the code waveform to

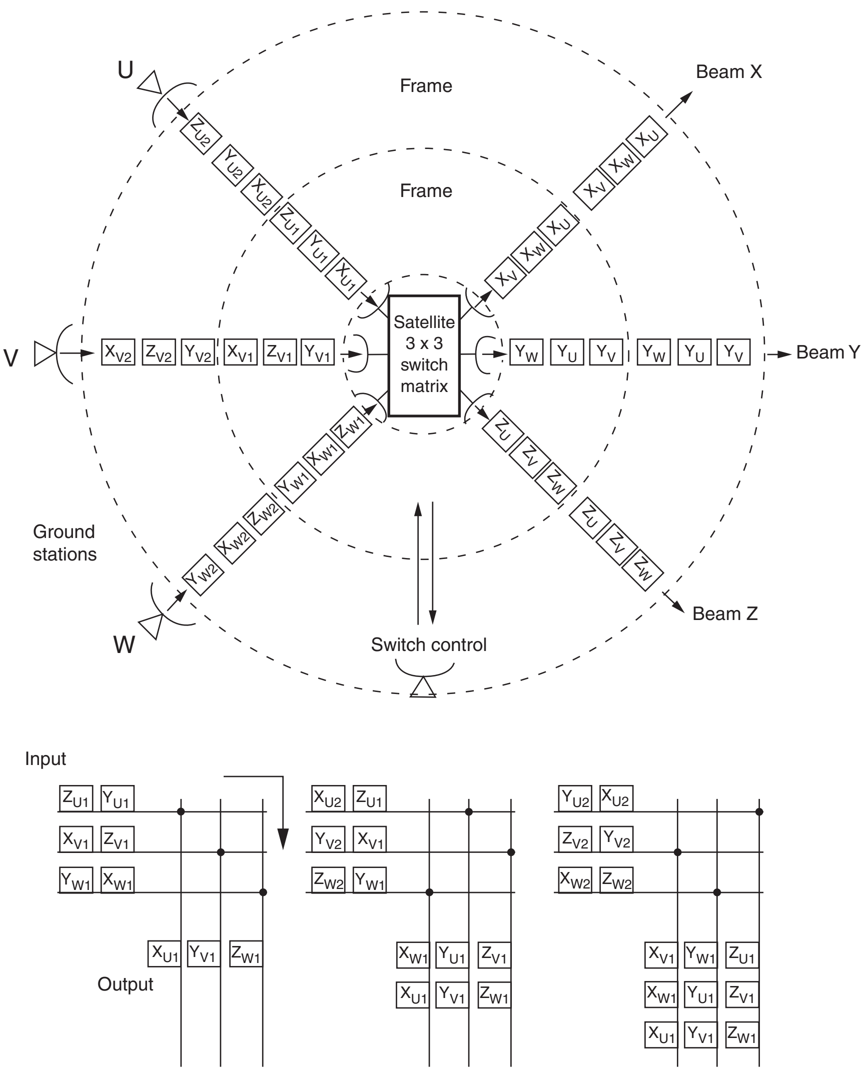


Figure 14.30 Traffic from earth stations U, V, W rerouted into designated beams X, Y, Z. The lower diagrams show part of the switching sequence.

spread the spectrum over the available rf bandwidth. Many of the key properties of CDMA rely on this spectrum spreading, and the systems employing CDMA are also known as *spread-spectrum multiple access* (SSMA). Care must be taken not to confuse the SS here with that for satellite switched (SS/TDMA) used in the previous section.

CDMA can be used with analog and digital signals (see, for example, Dixon, 1984), but only digital systems will be described here. For illustration purposes, a polar NRZ waveform denoted by $p(t)$ (see

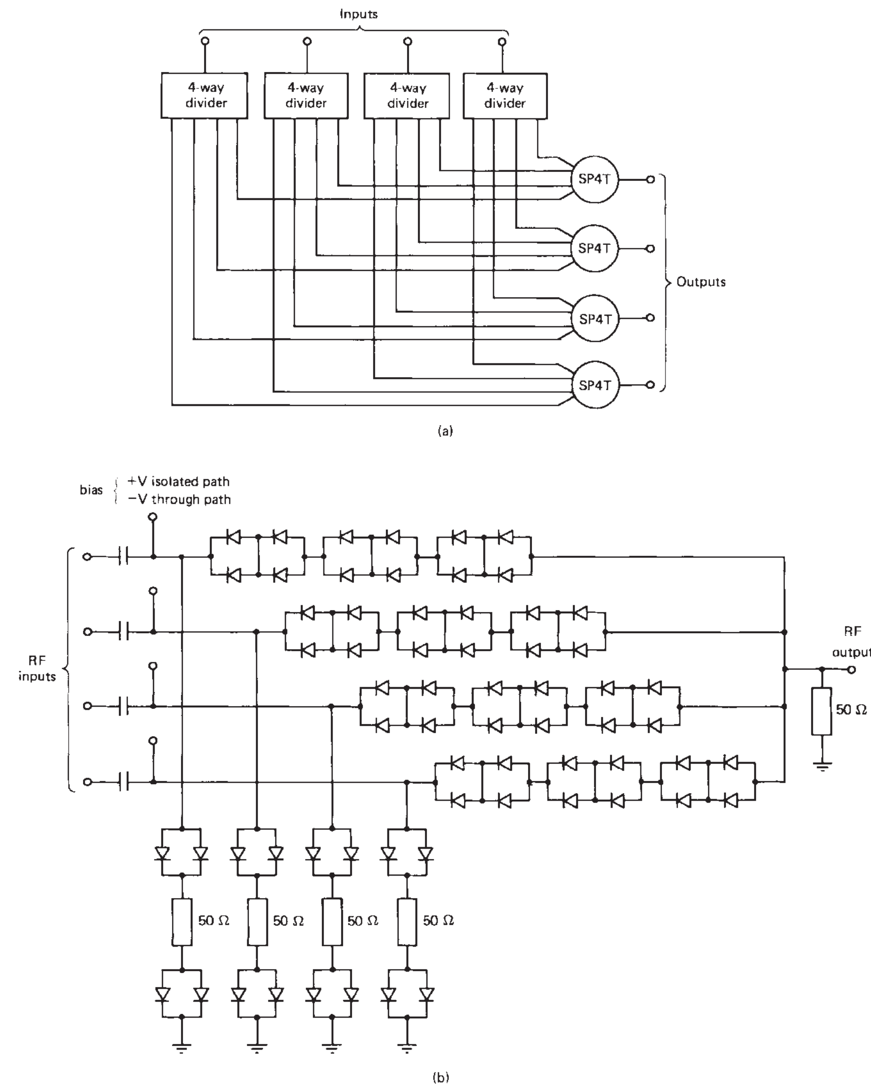


Figure 14.31 (*a*) 4 × 4 switch matrix; (*b*) circuit diagram of redundant SP4T switch element. (*From Watt, 1986; reprinted with permission of IEEE, London.*)

Fig.10.2) will be used for the information signal, and BPSK modulation (see Sec.10.6.1) will be assumed. The code waveform $c(t)$ is also a polar NRZ signal, as sketched in Fig.14.32. What would be called *bits* in an information waveform are called *chips* for the code waveform, and in most practical systems the chip rate is much greater than the information bit rate. The pulses (chips) in the code wave-

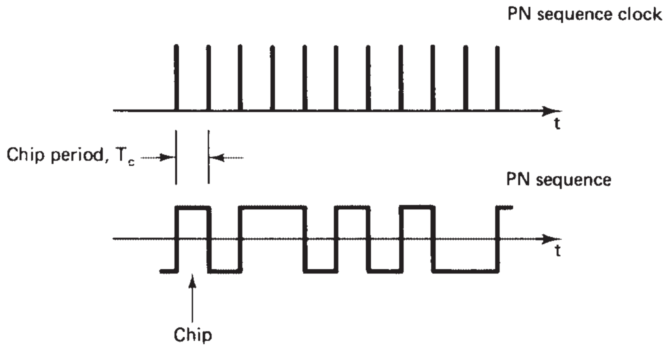


Figure 14.32 PN binar sequence. One element is known as a chip.

form vary randomly between $+V$ and $-V$. The randomness is an essential feature of spread-spectrum systems, and more will be said about this shortly. The code signal may be applied as modulation in exactly the same way as the information signal so that the BPSK signal carries both the information signal $p(t)$ and the code signal $c(t)$. This method is referred to as *direct-sequence spread spectrum* (DS/SS). Other techniques are also used to spread the spectrum, such as frequency hopping, but the discussion here will be limited to the DS/SS method.

14.10.1 Direct-sequence spread spectrum

In Fig.14.33, $p(t)$ is an NRZ binary information signal, and $c(t)$ is a NRZ binary code signal. These two signals form the inputs to a multiplier (balanced modulator), the output of which is proportional to the product $p(t)c(t)$. This product signal is applied to a second balanced modulator, the output of which is a BPSK signal at the carrier fre-

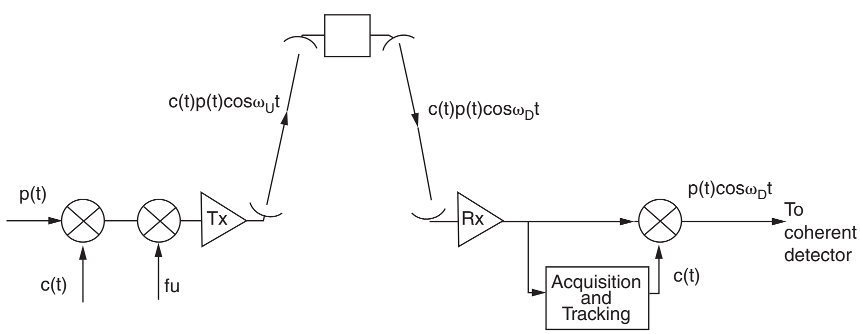


Figure 14.33 A basic CDMA system.

quency. For clarity, it is assumed that the carrier is the uplink frequency, and hence the uplink carrier is described by

$$e_U(t) = c(t)p(t) \cos \omega_U t \tag{14.33}$$

The corresponding downlink carrier is

$$e_D(t) = c(t)p(t) \cos \omega_D t \tag{14.34}$$

At the receiver, an identical $c(t)$ generator is synchronized to the $c(t)$ of the downlink carrier. This synchronization is carried out in the *acquisition and tracking block.* With $c(t)$ a polar NRZ type waveform, and with the locally generated $c(t)$ exactly in synchronism with the transmitted $c(t)$, the product $c^2(t) = 1$. Thus the output from the multiplier is

$$\begin{aligned} c(t)\, e_D(t) &= c^2(t)p(t) \cos \omega_D t \\ &= p(t) \cos \omega_D t \end{aligned} \tag{14.35}$$

This is identical to the conventional BPSK signal given by Eq.(10.14), and hence detection proceeds in the normal manner.

14.10.2 The code signal $c(t)$

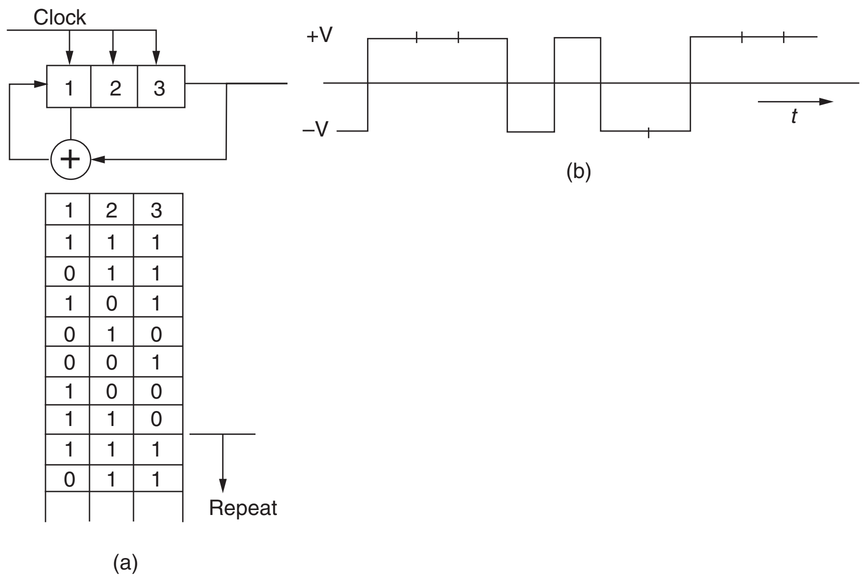
The code signal $c(t)$ carries a binary code that has special properties needed for successful implementation of CDMA. The binary symbols used in the codes are referred to as *chips* rather than *bits* to avoid confusion with the information bits that also will be present. Chip generation is controlled by a clock, and the chip rate, in chips per second, is given by the clock speed. Denoting the clock speed by R_{ch}, the chip period is the reciprocal of the clock speed:

$$T_{ch} = \frac{1}{R_{ch}} \tag{14.36}$$

The waveform $c(t)$ is periodic, in that each period is a repetition of a given sequence of N chips. The sequence itself exhibits random properties, which will be described shortly. The periodic time for the waveform is

$$T_N = NT_{ch} \tag{14.37}$$

The codes are generated using binary shift registers and associated linear logic circuits. The circuit for a three-stage shift register that generates a sequence of $N = 7$ chips is shown in Fig.14.34*a*. Feedback occurs from stages 1 and 3 as inputs to the exclusive OR gate. This pro-



1	2	3
1	1	1
0	1	1
1	0	1
0	1	0
0	0	1
1	0	0
1	1	0
1	1	1
0	1	1

Figure 14.34 Generation of a 7-chip maximal sequence code.

vides the input to the shift register, and the chips are clocked through at the clock rate R_{ch}. The generator starts with all stages holding binary 1s, and the following states are as shown in the table in Fig. 14.34. Stage 3 also provides the binary output sequence. The code waveform generated from this code is shown in Fig. 14.34*b*.

Such codes are known as *maximal sequence* or *m-sequence* codes because they utilize the maximum length sequence that can be generated. For Fig. 14.34*a* the maximum length sequence is 7 chips as shown. In general, the shift register passes through all states (all combinations of 1s and 0s in the register) except the all-zero state when generating a maximal sequence code. Therefore, a code generator employing an n-stage shift register can generate a maximum sequence of N chips, where

$$N = 2^n - 1 \tag{14.38}$$

The binary 1s and 0s are randomly distributed such that the code exhibits noiselike properties. However, there are certain deterministic features described below, and the codes are more generally known as *PN codes,* which stands for *pseudo-noise codes.*

1. The number of binary 1s is given by

$$No.\ of\ 1s = \frac{2^n}{2} \tag{14.39}$$

and the number of binary 0s is given by

$$No.\ of\ 0s = \frac{2^n}{2} - 1 \tag{14.40}$$

The importance of this relationship is that when the code uses $+V$ volts for a binary 1 and $-V$ volts for a binary 0, the dc offset is close to zero. Since there is always one more positive chip than negative, the dc offset will be given by

$$dc\ offset = \frac{V}{N} \tag{14.41}$$

The dc offset determines the carrier level relative to the peak value; that is, the carrier is suppressed by amount $1/N$ for BPSK. For example, using a code with $n = 8$ with BPSK modulation, the carrier will be suppressed by 1/255 or 48 dB.

2. The total number of maximal sequences that can be generated by an n-stage shift register (and its associated logic circuits) is given by

$$S_{\max} = \frac{\phi(N)}{n} \tag{14.42}$$

Here, $\phi(N)$ is known as *Euler's* ϕ-function, which gives the number of integers in the range 1, 2, 3…, $N - 1$, that are relatively prime to N [N is given by Eq. (14.38)]. Two numbers are relatively prime when their greatest common divisor is 1. A general formula for finding $\phi(N)$ is (see Ore, 1988)

$$\phi\ (N) = N\left(\frac{p_1 - 1}{p_1}\right)\cdots\left(\frac{p_r - 1}{p_r}\right) \tag{14.43}$$

where $p_1,\dots, p_r$ are the prime factors of N. For example, for $n = 8$, $N = 2^8 - 1 = 255$. The prime factors of 255 are 3, 5, and 17, and hence

$$\phi\ (255) = 255\left(\frac{2}{3}\right)\left(\frac{4}{5}\right)\left(\frac{16}{17}\right)$$

$$= 128$$

The total number of maximal sequences that can be generated by an eight-stage code generator is therefore

$$S_{\max} = \frac{128}{8}$$

$$= 16$$

As a somewhat simpler example, consider the case when $n = 3$. In this instance, $N = 7$. There is only one prime factor, 7 itself, and therefore

$$\phi(7) = 7 \cdot \frac{6}{7}$$

$$= 6$$

$$S_{\max} = \frac{\phi(7)}{n}$$

$$= 2$$

In this case there are only two distinct maximal sequences.

14.10.3 The autocorrelation function for $c(t)$

One of the most important properties of $c(t)$ is its *autocorrelation function.* The autocorrelation function is a measure of how well a time-shifted version of the waveform compares with the unshifted version. Figure 14.35*a* shows how the comparison may be made. The $c(t)$ waveform is multiplied with a shifted version of itself, $c(t - \tau)$, and the output is averaged (shown by the integrator). The average, of course, is independent of time t (the integrator integrates out the time-t dependence), but it will depend on the time lead or lag introduced by τ. When the waveforms are coincident, $\tau = 0$, and the average output is a maximum, which for convenience will be normalized to 1. Any shift in time, advance or delay, away from the $\tau = 0$ position will result in a decrease in output voltage. A property of m-sequence code waveforms is that the autocorrelation function decreases linearly from the maximum value (unity in this case) to a negative level $1/N$, as shown in Fig. 14.35*b*. The very pronounced peak in the autocorrelation function provides the chief means for acquiring acquisition and tracking so that the locally generated m-sequence code can be synchronized with the transmitted version.

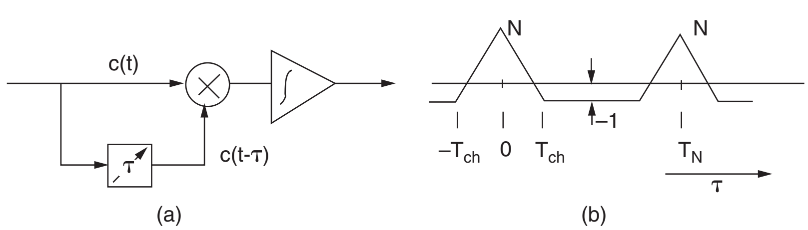


Figure 14.35 (a) Generating the autocorrelation function; (b) the autocorrelation waveform.

14.10.3 Acquisition and tracking

One form of acquisition circuit that makes use of the autocorrelation function is shown in Fig. 14.36. The output from the first multiplier is

$$\begin{aligned} e(t) &= c(t - \tau)\, c(t)\, p(t) \cos \omega_D t \\ &= c(t - \tau)\, c(t) \cos [\omega_D t + \phi(t)] \end{aligned} \tag{14.44}$$

Here, the information modulation, which is BPSK, is shown as $\phi(t)$ so that the effect of the following bandpass filter (BPF) on the amplitude can be more clearly seen. The BPF has a passband centered on ω_D, wide with respect to the information modulation but narrow with respect to the code signal. It performs the amplitude-averaging function on the code signal product (see Maral and Bousquet, 1998). The averaging process can be illustrated as follows. Consider the product of two cosine terms and its expansion:

$$\begin{aligned} \cos \omega t \cos (\omega t - \delta) &= \frac{1}{2} \{ \cos [\omega t + (\omega t - \delta)] + \cos [\omega t - (\omega t - \delta)] \} \\ &= \frac{1}{2} [\cos (2\omega t - \delta) + \cos (\delta)] \end{aligned} \tag{14.45}$$

The BPF will reject the high-frequency component, leaving only the average component $^1/_2 \cos(\delta)$. This signal may be considered analogous to the $c(t)c(t - \tau)$ term in Eq. (14.44). The envelope detector following the BPF produces an output proportional to the envelope of the signal, that is, to the average value of $c(t)c(t - \tau)$. This is a direct measure of the autocorrelation function. When it is less than the predetermined threshold V_T required for synchronism, the time shift τ incremented.

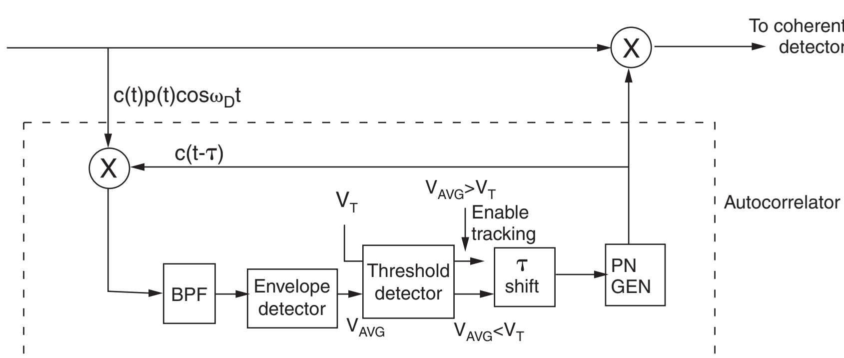


Figure 14.36 Acquisition of a carrier in a CDMA system.

Once the threshold has been reached or exceeded, the system switches from acquisition mode to tracking mode.

One form of tracking circuit, the *delay lock loop,* is shown in Fig. 14.37. Here, two correlators are used, but the local signal to one is advanced by half a chip period relative to the desired code waveform, and the other is delayed by the same amount. The outputs from the correlators are subtracted, and this difference signal provides the control voltage for the voltage-controlled oscillator (VCO) that drives the shift register clock. With the control voltage at the zero crossover point, the locally generated code signal is in phase with the received code signal. Any tendency to drift out of phase changes the VCO in such a way as to bring the control voltage back to the zero crossover point, thus maintaining synchronism.

The acquisition and tracking circuits also will attempt to correlate the stored version of $c(t)$ at the receiver with all the other waveforms being received. Such correlations are termed *cross-correlations.* It is essential that the cross-correlation function not show a similar peak as

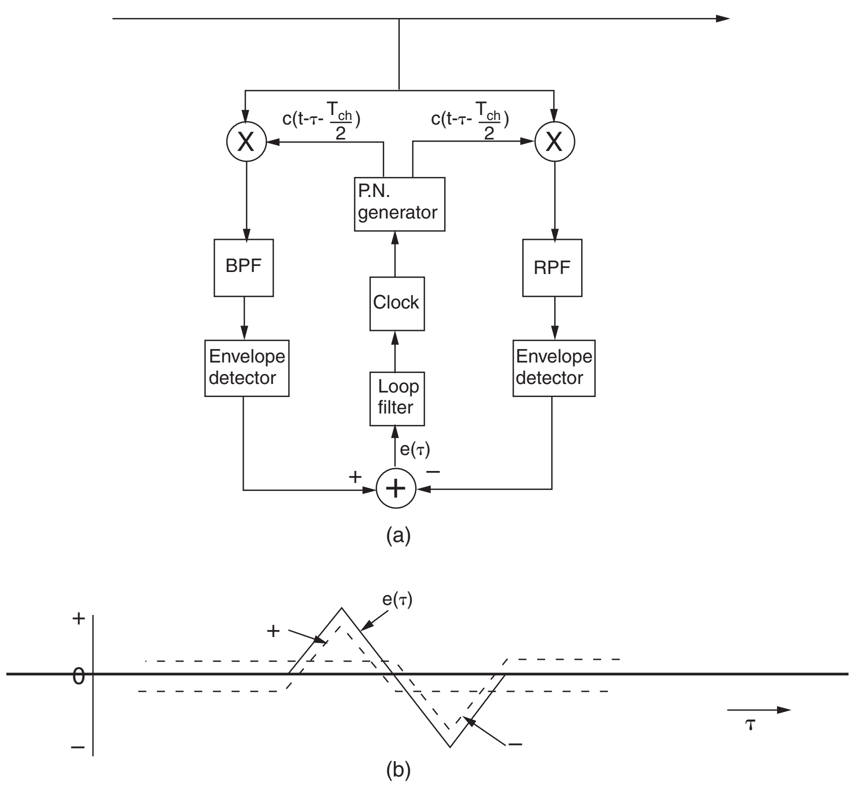


Figure 14.37 (a) The delay lock loop; (b) the waveform at the adder.

the autocorrelation, and this requires careful selection of the spreading functions used in the overall system (see, for example, Dixon, 1984).

14.10.4 Spectrum spreading and despreading

In Sec. 10.6.3 the idea of bandwidth for PSK modulation was introduced. In general, for a BPSK signal at a bit rate R_b, the main lobe of the power-density spectrum occupies a bandwidth extending from $f_c - R_b$ to $f_c + R_b$. This is sketched in Fig. 14.38*a*. A similar result applies when the modulation signal is $c(t)$, the power density spectrum being as sketched in Fig. 14.38*b*. It should be mentioned here that because $c(t)$ exhibits periodicity, the spectrum density will be a line function, and Fig. 14.38*b* shows the envelope of the spectrum. The spectrum shows the *power density* (watts per hertz) in the signal. For constant carrier power, it follows that if a signal is forced to occupy a wider bandwidth, its spectrum density will be reduced. This is a key result in CDMA systems. In all direct-sequence spread-spectrum systems, the chip rate is very much greater than the information bit rate, or $R_{ch} \gg R_b$. The bandwidth is determined mainly by R_{ch} so that the power density of the signal described by Eq. (14.34) is spread over the bandwidth determined by R_{ch}. The power density will be reduced approximately in the ratio of R_{ch} to R_b.

Assuming then that acquisition and tracking have been accomplished, $c(t)$ in the receiver (Fig. 14.33) performs in effect a *despreading function* in that it restores the spectrum of the wanted signal to what it was before the spreading operation in the transmitter. This is

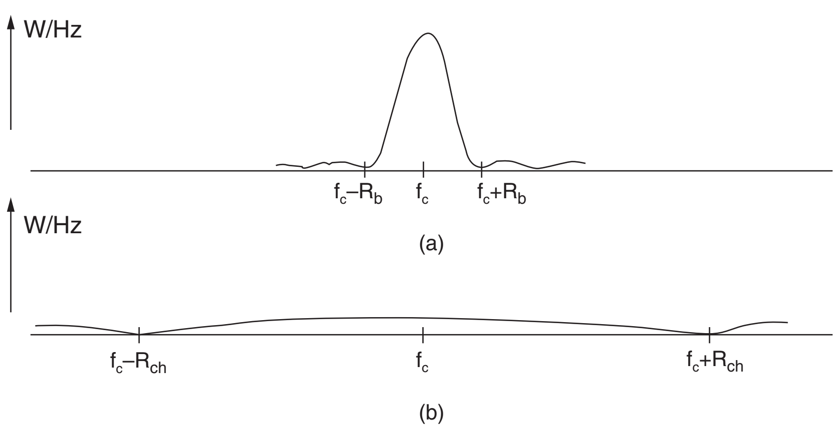


Figure 14.38 Spectrum for a BPSK signal: (a) without spreading, (b) with spreading.

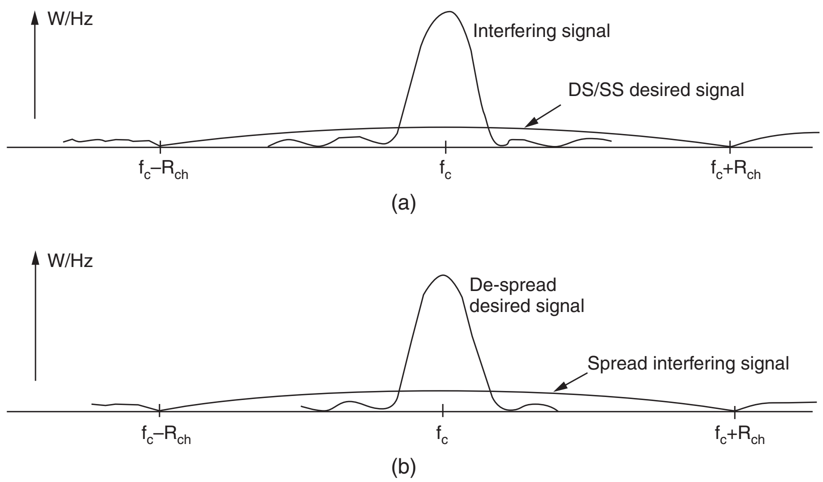


Figure 14.39 (a) Spectrum of an interfering, nonspread signal along with the spread desired signal; (b) the effect of the de-spreading operation on the desired signal resulting in spread of the interfer.

also how the spread-spectrum technique can reduce interference. Figure 14.39*a* shows the spectra of two signals, an interfering signal that is not part of the CDMA system and that has not been spread, and the desired DS/SS received signal. Following the despreading operation for the desired signal, its spectrum is restored as described previously. The interfering signal, however, is simply multiplied by the $c(t)$ signal, which results in it being spread.

14.10.5 CDMA throughput

The maximum number of channels in a CDMA system can be estimated as follows: It is assumed that the thermal noise is negligible compared with the noise resulting from the overlapping channels, and also for comparison purposes, it will be assumed that each channel introduces equal power P_R into the receiver. For a total of K channels, $K - 1$ of these will produce noise, and assuming that this is evenly spread over the noise bandwidth B_N of the receiver, the noise density, in watts per hertz, is

$$N_o = \frac{(K-1)\,P_R}{} \tag{14.46}$$

Let the information rate of the wanted channel be R_b; then, from Eq. (10.22),

$$E_b = \frac{P_R}{R_b} \tag{14.47}$$

Hence the bit energy to noise density ratio is

$$\frac{E_b}{N_o} = \frac{B_N}{(K - 1)\, R_b} \tag{14.48}$$

The noise bandwidth at the BPSK detector will be approximately equal to the IF bandwidth as given by Eq. (10.15), but using the chip rate

$$\begin{aligned} B_N &\cong B_{\mathrm{IF}} \\ &= (1 + \rho)\, R_{ch} \end{aligned} \tag{14.49}$$

where ρ is the rolloff factor of the filter. Hence the bit energy to noise density ratio becomes

$$\frac{E_b}{N_o} = \frac{(1 + \rho)\, R_{\mathrm{ch}}}{(K - 1)\, R_b} \tag{14.50}$$

As pointed out in Chap. 10, the probability of bit error is usually a specified objective, and this determines the E_b/N_o ratio, for example, through Fig.10.17. The number of channels is therefore

$$K = 1 + (1 + \rho)\, \frac{R_{\mathrm{ch}}}{R_b}\, \frac{N_o}{E_b} \tag{14.51}$$

The *processing gain* G_p is basically the ratio of power density in the unspread signal to that in the spread signal. Since the power density is inversely proportional to bandwidth, an approximate expression for the processing gain is

$$G_p = \frac{R_{\mathrm{ch}}}{R_b} \tag{14.52}$$

Hence

$$K = 1 + (1 + \rho)\, G_P\, \frac{N_o}{E_b} \tag{14.53}$$

Example 14.8 The code waveform in a CDMA system spreads the carriers over the full 36 MHz bandwidth of the channel, and the rolloff factor for the filtering is 0.4. The information bit rate is 64 kb/s, and the system uses BPSK. Calculate the processing gain in decibels. Given that the bit error rate must not exceed 10^{-5}, give an estimate of the maximum number of channels that can access the system.

solution

$$R_{\mathrm{ch}} = \frac{B_{\mathrm{IF}}}{(1 + \rho)}$$

$$= \frac{36 \times 10^6}{1.4}$$

$$= 25.7 \times 10^6 \text{ chips/s}$$

Hence the processing gain is

$$G_p = \frac{25.7 \times 10^6}{64 \times 10^3}$$

$$= 401.56$$

$$[G_p] = 10 \log (401.56)$$

$$= 26 \text{ dB}$$

From Fig. 10.18 for $P_e = 10^{-5}$, $[E_b/N_o] = 9.6$ dB approximately. This is a power ratio of 9.12, and from Eq. (14.53),

$$K = 1 + \frac{1.4 \times 401.56}{9.12}$$

$$\cong 62 \text{ (rounded down)}$$

The *throughput efficiency* is defined as the ratio of the total number of bits per unit time that can be transmitted with CDMA to the total number of bits per unit time that could be transmitted with single access and no spreading. For K accesses as determined above, each at bit rate R_b, the total bits per unit time is KR_b. A single access could utilize the full bandwidth, and hence its transmission rate as determined by Eq. (10.15) as

$$R_T = \frac{B_{\mathrm{IF}}}{(1 + \rho)} \tag{14.54}$$

This is the same as the chip rate, and hence the throughput is

$$\eta = \frac{KR_b}{R_T}$$

$$= \frac{KR_b}{R_{\mathrm{ch}}}$$

$$= \frac{K}{G_p} \tag{14.55}$$

Using the values obtained in Example 14.8 gives a throughput of 0.15, or 15 percent. This should be compared with the frame efficiency for

TDMA (see Example 14.4), where it is seen that the throughput efficiency can exceed 90 percent.

CDMA offers several advantages for satellite networking, especially where VSAT-type terminals are involved. These are

1. The beamwidth for VSAT antennas is comparatively broad and therefore could be subject to interference from adjacent satellites. The interference rejection properties of CDMA through spreading are of considerable help here.
2. Multipath interference, for example, that resulting from reflections, can be avoided provided the time delay of the reflected signal is greater than a chip period and the receiver locks onto the direct wave.
3. Synchronization between stations in the system is not required (unlike TDMA, where synchronization is a critical feature of the system). This means that a station can access the system at any time.
4. Degradation of the system (reduction in E_b/N_o) is gradual with an increase in number of users. Thus additional traffic could be accommodated if some reduction in performance was acceptable.

The main disadvantage is the low throughput efficiency.

14.11 Problems

14.1. Explain what is meant by a *single access* in relation to a satellite communications network. Give an example of the type of traffic route where single access would be used.

14.2. Distinguish between *preassigned* and *demand-assigned traffic* in relation to a satellite communications network.

14.3. Explain what is meant by *frequency-division multiple access,* and show how this differs from frequency-division multiplexing.

14.4. Explain what the acronym *SCPC* stands for. Explain in detail the operation of a preassigned SCPC network.

14.5. Explain what is meant by *thin route service.* What type of satellite access is most suited for this type of service?

14.6. Briefly describe the ways in which demand assignment may be carried out in an FDMA network.

14.7. Explain in detail the operation of the Spade system of demand assignment. What is the function of the common signaling channel?

14.8. Explain what is meant by *power-limited* and *bandwidth-limited operation* as applied to an FDMA network. In an FDMA scheme the carriers utilize equal powers and equal bandwidths, the bandwidth in each case being 5 MHz. The transponder bandwidth is 36 MHz. The saturation EIRP for the downlink is 34 dBW, and an output backoff of 6 dB is employed. The downlink losses are 201 dB, and the destination earth station has a G/T ratio of 35 dBK^{-1}. Determine the [C/N] value assuming this is set by single carrier operation. Determine also the number of carriers which can access the system, and state, with reasons, whether the system is power limited or bandwidth limited.

14.9. Distinguish between *bandwidth-limited* and *power-limited operation* as applied to an FDMA network.

14.10. In some situations it is convenient to work in terms of the carrier-to-noise temperature. Show that $[C/T] = [C/N_0] + [k]$. The downlink losses for a satellite circuit are 196 dB. The earth station [G/T] ratio is 35 dB/K, and the received [C/T] ratio is -138 dBW/K. Calculate the satellite [EIRP].

14.11. The earth station receiver in a satellite downlink has an FM detector threshold level of 10 dB and operates with a 3-dB threshold margin. The emphasis improvement figure is 4 dB, and the noise-weighting improvement figure is 2.5 dB. The required [S/N] ratio at the receiver output is 46 dB. Calculate the receiver processing gain. Explain how the processing gain determines the IF bandwidth.

14.12. A 252-channel FM/FDM telephony carrier is transmitted on the downlink specified in Prob. 14.10. The peak/rms ratio factor is 10 dB, and the baseband bandwidth extends from 12 to 1052 kHz. The voice-channel bandwidth is 3.1 kHz. Calculate the peak deviation, and hence, using Carson's rule, calculate the IF bandwidth.

14.13. Given that the IF bandwidth for a 252-channel FM/FDM telephony carrier is 7.52 MHz and that the required [C/N] ratio at the earth station receiver is 13 dB, calculate (*a*) the [C/T] ratio and (*b*) the satellite [EIRP] required if the total losses amount to 200 dB and the earth station [G/T] ratio is 37.5 dB/K.

14.14. Determine how many carriers can access an 80-MHz transponder in the FDMA mode, given that each carrier requires a bandwidth of 6 MHz, allowing for 6.5-dB output backoff. Compare this number with the number of carriers possible without backoff.

14.15. (*a*) Analog television transmissions may be classified as *full-transponder* or *half-transponder transmissions.* State what this means in terms of transponder access. (*b*) A composite TV signal (video plus audio) has a top baseband frequency of 6.8 MHz. Determine for a 36-MHz transponder the peak frequency deviation limit set by (1) half-transponder and (2) full-transponder transmission.

14.16. Describe the general operating principles of a *time-division multiple-access network*. Show how the transmission bit rate is related to the input bit rate.

14.17. Explain the need for a reference burst in a TDMA system.

14.18. Explain the function of the preamble in a TDMA traffic burst. Describe and compare the channels carried in a preamble with those carried in a reference burst.

14.19. What is the function of (*a*) the burst-code word and (*b*) the carrier and bit-timing recovery channel in a TDMA burst?

14.20. Explain what is meant by (*a*) *initial acquisition* and (*b*) *burst synchronization* in a TDMA network. (*c*) The nominal range to a geostationary satellite is 42,000 km. Using the station-keeping tolerances stated in Sec. 7.4 in connection with Fig. 7.10, determine the variation expected in the propagation delay.

14.21. (*a*) Define and explain what is meant by *frame efficiency* in relation to TDMA operation. (*b*) In a TDMA network the reference burst and the preamble each requires 560 bits, and the nominal guard interval between bursts is equivalent to 120 bits. Given that there are eight traffic bursts and one reference burst per frame and the total frame length is equivalent to 40,800 bits, calculate the frame efficiency.

14.22. Given that the frame period is 2 ms and the voice-channel bit rate is 64 kb/s, calculate the equivalent number of voice channels that can be carried by the TDMA network specified in Prob. 14.21.

14.23. Calculate the frame efficiency for the CSC shown in Fig. 14.19.

14.24. (*a*) Explain why the frame period in a TDMA system is normally chosen to be an integer multiple of 125 μs. (*b*) Referring to Fig. 14.20 for the INTELSAT preassigned frame format, show that there is no break in the timing interval for sample 18 when this is transferred to a burst.

14.25. Show that, all other factors being equal, the ratio of uplink power to bit rate is the same for FDMA and TDMA. In a TDMA system the preamble consists of the following slots, assigned in terms of number of bits: bit timing recovery 304, unique word 48, station identification channel 8, order wire 64. The guard slot is 120 bits, the frame reference burst is identical to the preamble, and the burst traffic is 8192 bits. Given that the frame accommodates 8 traffic bursts, calculate the frame efficiency. The traffic is preassigned PCM voice channels for which the bit rate is 64 kb/s, and the satellite transmission rate is nominally 60 Mb/s. Calculate the number of voice channels which can be carried.

14.26. In comparing design proposals for multiple access, the two following possibilities were considered: (1) uplink FDMA with downlink TDM and (2) uplink TDMA with downlink TDM. The incoming baseband signal is at 1.544 Mb/s in each case and the following table shows values in decilogs:

	Uplink	Downlink
$[E_b/N_0]$	12	12
$[G/T]$	10	19.5
[Losses]	212	210
[EIRP]	—	48
Transmit antenna gain $[G_T]$	45.8	—

Determine (*a*) the downlink TDM bit rate and (*b*) the transmit power required at the uplink earth station for each proposal.

14.27. A TDMA network utilizes QPSK modulation and has the following symbol allocations: guard slot 32, carrier and bit timing recovery 180, burst code word (unique word) 24, station identification channel 8, order wire 32, management channel (reference bursts only) 12, service channel (traffic bursts only) 8. The total number of traffic symbols per frame is 115,010, and a frame consists of two reference bursts and 14 traffic bursts. The frame period is 2 ms. The input consists of PCM channels each with a bit rate of 64 kb/s. Calculate the frame efficiency and the number of voice channels that can be accommodated.

14.28. For the network specified in Prob. 14.27 the BER must be at most 10^{-5}. Given that the receiving earth station $[G/T]$ value is 30 decilogs and total losses are 200 dB, calculate the satellite [EIRP] required.

14.29. Discuss briefly how demand assignment may be implemented in a TDMA network. What is the advantage of TDMA over FDMA in this respect?

14.30. Define and explain what is meant by the terms *telephone load activity factor* and *digital speech interpolation.* How is advantage taken of the load activity factor in implementing digital speech interpolation?

14.31. Define and explain the terms *connect clip* and *freeze-out* used in connection with digital speech interpolation.

14.32. Describe the principles of operation of a speech-predictive encoded communications (SPEC) system, and state how this compares with digital speech interpolation.

14.33. Determine the bit rate that can be transmitted through a 36-MHz transponder, assuming a rolloff factor of 0.2 and QPSK modulation.

14.34. On a satellite downlink, the $[C/N_0]$ ratio is 86 dBHz and an $[E_b/N_o]$ of 12 dB is required at the earth station. Calculate the maximum bit rate that can be transmitted.

14.35. FDMA is used for uplink access in a satellite digital network, with each earth station transmitting at the T1 bit rate of 1.544 Mb/s. Calculate (a) the uplink [C/N_0] ratio required to provide a [E_b/N_0] = 14 dB ratio at the satellite and (b) the earth-station [EIRP] needed to realize the [C/N_0] value. The satellite [G/T] value is 8 dB/K, and total uplink losses amount to 210 dB.

14.36. In the satellite network of Prob. 14.35, the downlink bit rate is limited to a maximum of 74.1 dBb/s, with the satellite TWT operating at saturation. A 5-dB output backoff is required to reduce intermodulation products to an acceptable level. Calculate the number of earth stations that can access the satellite on the uplink.

14.37. The [EIRP] of each earth station in an FDMA network is 47 dBW, and the input data are at the T1 bit rate with 7/8 FEC added. The downlink bit rate is limited to a maximum of 60 Mb/s with 6-dB output backoff applied. Compare the [EIRP] needed for the earth stations in a TDMA network utilizing the same transponder.

14.38. (a) Describe the general features of an on-board signal processing transponder that would allow a network to operate with FDMA uplinks and a TDMA downlink. (b) In such a network, the overall BER must not exceed 10^{-5}. Calculate the maximum permissible BER of each link, assuming that each link contributes equally to the overall value.

14.39. Explain what is meant by *full interconnectivity* in connection with satellite switched TDMA. With four beams, how many switch modes would be required for full interconnectivity?

14.40. Identify all the redundant modes in Fig. 14.27.

14.41. The shift register in an m-sequence generator has 7 stages. Calculate the number of binary 1s and 0s. The code is used to generate a NRZ polar waveform at levels +1 V and −1 V. Calculate the dc offset and the carrier suppression in decibels that can be achieved when BPSK is used.

14.42. The shift register in an m-sequence generator has 10 stages. Calculate the length of the m-sequences. Determine the prime factors for N and hence the total number of maximal length sequences that can be produced.

14.43. As shown in Sec. 14.10.2, an m-sequence generator having a 3-stage shift register is capable of generating a total of 2 maximal sequences, and Fig. 14.34 shows one of these. Draw the corresponding circuit for the other sequence, and the waveform.

14.44. Draw accurately to scale the autocorrelation function over one complete cycle for the waveform shown in Fig. 14.34. Assume $V = 1$ V and $T_{ch} = 1$ ms.

14.45. Draw accurately to scale the autocorrelation function over one complete cycle for the waveform determined in Prob. 14.43. Assume $V = 1$ V, and $T_{\mathrm{ch}} = 1$ ms.

14.46. Describe in your own words how signal acquisition and tracking are achieved in a DS/SS system.

14.47. An m-sequence generator having a 3-stage shift register is capable of generating a total of 2 maximal sequences. Neatly sketch the cross-correlation function for the two m-sequences.

14.48. Explain the principle behind spectrum spreading and despreading and how this is used to minimize interference in a CDMA system.

14.49. The IF bandwidth for a CDMA system is 3 MHz, the rolloff factor for the filter being 1. The information bit rate is 2.4 kb/s, and an $[E_b/N_o]$ of 11 dB is required for each channel accessing the CDMA system. Calculate the maximum number of accesses permitted.

14.50. Determine the throughput efficiency for the system in Prob. 14.49.

14.51. Show that when K is large such that the first term, unity, on the right hand side of Eq. (14.53) can be neglected, the throughput efficiency is independent of the processing gain. Hence plot the throughput efficiency as a function of $[E_b/N_o]$ for the range 7 to 11 dB.

Chapter

15

Satellite Services and the Internet

15.1 Introduction

On October 24, 1995, the Federal Networking Council (FNC) in the United States passed a resolution defining the *Internet* as a global information system that

(I) is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions/follow-ons;

(II) is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions/follow-ons, and/or other IP-compatible protocols; and

(III) provides, uses or makes accessible, either publicly or privately, high level services layered on the communications and related infrastructure described herein.

This formal description of the Internet summarizes what in fact was many years of evolutionary growth and change (see Leiner et al., 2000). The key elements in this definition are the *Transmission Control Protocol* (TCP) and the *Internet Protocol* (IP), both of which are described shortly. These protocols are usually lumped together as TCP/IP and are embedded in the software for operating systems and browsers such as Windows and Netscape.

The Internet does not have its own physical structure. It makes use of existing physical plant, the copper wires, optical fibers, satellite links, etc. owned by companies such as AT&T, MCI, Sprint, etc. Although there is no identifiable structure, access to the Internet follows well-defined rules. Users connect to *Internet service providers* (ISPs), who in turn connect to *network service providers* (NSPs), who complete the connections to other users and to *servers*. Servers are computers dedicated to the purpose of providing information to the Internet. They run specialized software for each type of Internet application. These include email, discussion groups, long-distance computing, and file transfers (Corbett, undated; Sterling, 1993). *Routers* are computers that form part of the communications net and that route or direct the data along the best available paths in the network.

Although there is no central management or authority for the Internet, its extraordinarily rapid growth has meant that some control has to be exercised over what is permitted. A summary of the controlling groups is shown in Fig. 15.1*a*. A description of the groups will be found in Leiner et al. (2000) and Mackenzie (1998).

The World Wide Web (WWW) is probably the most widely used application on the Internet. The evolution and growth of the WWW has been rather similar to that of the Internet itself, with no central authority but still with a structure that attempts to regulate what happens. The World Wide Web Consortium, referred to as W3C, was founded in October 1994 (Jacobs, 2000). W3C oversees a number of special interest groups, as shown in Fig. 15.1*b,* and coordinates its efforts with the IETF and with other standards bodies. Details of the W3C will be found in Jacobs (2000).

15.2 Network Layers

The uplink and downlink between satellite and earth stations form what is known as the *physical layer* in a data communication system. By *data communications* is meant communications between computers and peripheral equipment. The signals are digital, and although digital signals are covered in Chap. 10, the satellite links must be able to accommodate the special requirements imposed by networks. The terminology used in networks is highly specialized, and some of these terms are explained here to provide the background needed to understand the satellite aspects. The Internet, of course, is a data communication system (although there is presently a move to incorporate voice communications along with data in what is known as *voice over Internet Protocol,* or *VoIP*).

The data are transmitted in *packets.* Many separate functions have to be performed in packet transmission, such as packet addressing, routing, and coping with packet congestion. The modern approach is to assign each function to a layer in what is termed the *network architecture.* The layers are conceptual in the sense that they may consist of software or some combination of software and hardware. In the case of the Internet, the network architecture is referred to the *TCP/IP model,* although there are protocols other than TCP/IP con-

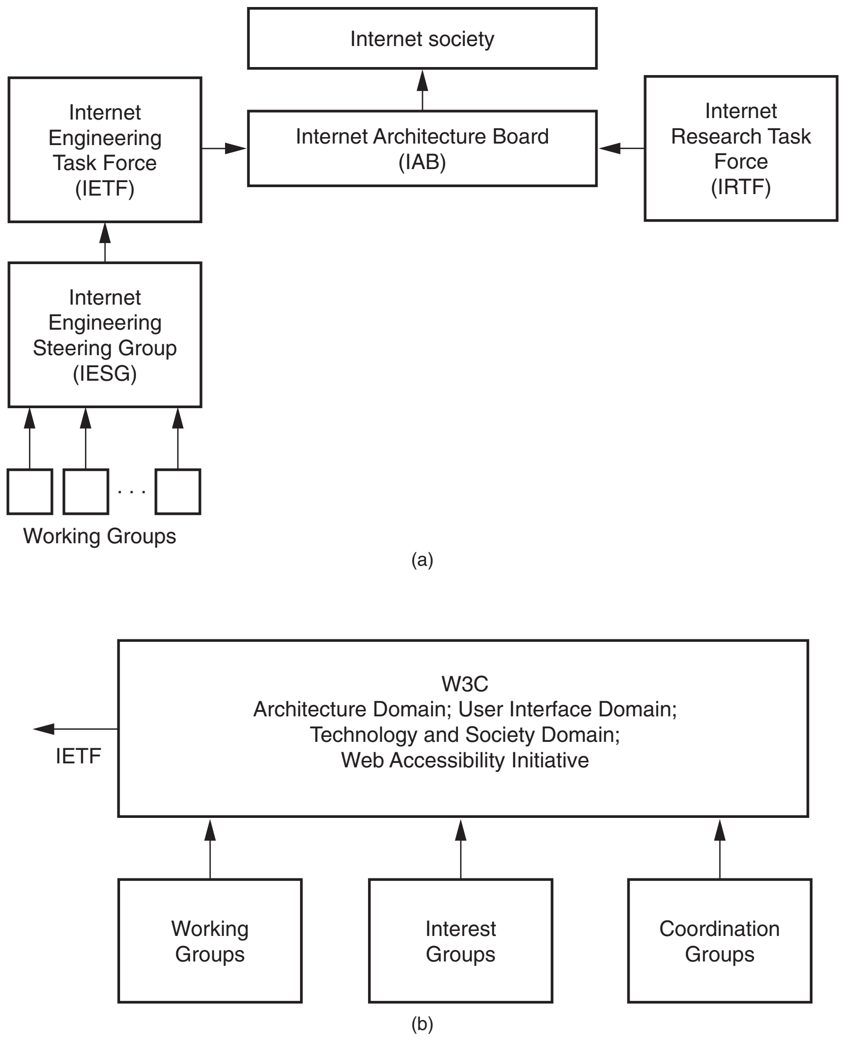


Figure 15.1 *(a)* Internet groups; *(b)* World Wide Web groups.

tained in the model. The layered structure is shown in Fig. 15.2. A brief description of these layers is included to familiarize the reader with some of the terms used in network communications, although the TCP layer is of most interest in this chapter.

- *Physical layer.* This covers such items as the physical connectors, signal format, modulation, and the uplink and downlink in a satellite communications system.

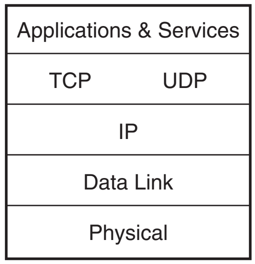


Figure 15.2 Layered structure for TCP/IP. (*After Feit, 1997.*)

- *Data-link layer.* The function of this layer is to organize the digital data into blocks as required by the physical layer. For example, if the physical layer uses Asynchronous Transfer Mode (ATM) technology, the data are organized into cells. Digital transmission by satellite frequently uses TDMA, as described in Chap. 14, and satellite systems are being developed which transmit Internet data over ATM. Thus the data-link layer has to organize the data into a suitable format to suit the physical layer technology. In the terrestrial Internet, the data link converts the data into *frames.* The data-link layer and the physical layer are closely interrelated, and it can be difficult sometimes to identify the interface between these two layers (Mackenzie, 1998).
- *Network layer.* This is strictly an Internet Protocol (IP) layer. The packets are passed along the Internet from router to router and to the host stations. No exact path is laid out beforehand, and the IP layers in the routers must provide the destination address for the next leg of the journey so to speak. This destination address is part of the IP header attached to the packet. The source address is also included as part of the IP header. The problems of lost packets or packets arriving out of sequence are not a concern of the IP layer, and for this reason, the IP layer is called *connectionless* (i.e., it does not require a connection to be established before sending a packet on). These problems are taken care of by the transport layer.
- *Transport layer.* Two sets of protocol are provided in this layer. With the Transmission Control Protocol (TCP), information is passed back and forth between transport layers, which controls the information flow. This includes such functions as the correct sequencing of packets, replacement of lost packets, and adjusting the transmission rate of packets to prevent congestion. In the early days of the Internet when traffic was comparatively light, these problems could be handled even where satellite transmissions were involved. With the enormous increase in traffic on the present-day Internet, these prob-

lems require special solutions where satellite systems are used, which are discussed in later sections. The TCP layer is termed *connection-oriented* (compared with the connectionless service mentioned above) because sender and receiver must be in communication with each other to implement the protocol. There are situations where a simple standalone message may need to be sent which does not require the more complex TCP. For these types of message, another transport layer protocol called the *User Datagram Protocol* (UDP) is used. The UDP provides a connectionless service, similar to IP. The UDP header adds the port numbers for the source and destination applications.

The term *packet* has been used somewhat loosely up to this point. A more precise terminology is used for packets at the various layers, and this is shown in Fig. 15.3. At the application level the packet is simply referred to as *data*. The packet comprising the TCP header, and the data are a *TCP segment*. The packet comprising the UDP header and the data is a *UDP message*. The packet comprising the IP header, the TCP or UDP header, and the data is an *IP datagram*. Finally, the packet comprising the data-link frame header, the frame trailer (used for error control), and the IP datagram is a *frame*.

It should be noted here that the preceding definitions are those used in version 4 of the Internet Protocol (IPv4). Internet Protocol version 6 (IPv6) is a more recent version being brought on-stream in which the IP datagram is in fact called an *IP packet*.

Some of the units used in data transmission are

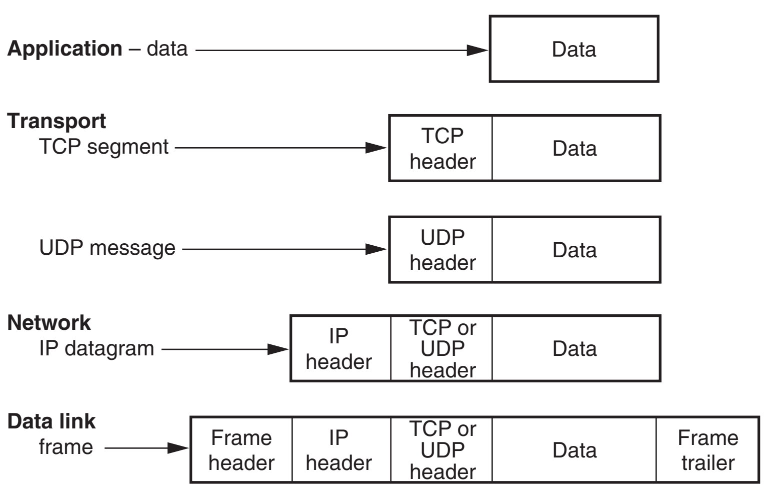


Figure 15.3 Packet terminology. (*After Feit, 1997.*)

- *Byte.* Common usage has established the byte (symbol B) as a unit of 8 bits, and this practice will be followed here. It should be noted, however, that in computer terminology, a byte can mean a unit other than 8 bits, and the 8-bit unit may be called an *octet.*
- *Kilobyte.* The kilobyte (symbol kB) is 1024 bytes. Transmission rates may be stated in kilobytes per second, or kB/s.
- *Megabyte.* The megabyte (symbol MB) is 1024 kilobytes. Transmission rates may be stated in megabytes per second. or MB/s.

The TCP/IP suite is shown in Fig.15.4, and an excellent detailed description of these protocols will be found in Feit (1997). The present text will be concerned more with the special enhancements needed on TCP/IP for successful satellite transmission.

15.3 The TCP Link

A *virtual communications link* exists between corresponding layers in a network. The header in the TCP segment (see Fig. 15.3) carries instructions that enable communication between the send and receive TCP layers. Of course, the communication has to pass through the other layers and along the physical link, but only the TCP layers act on the Transmission Control Protocols (TCPs) contained in the segment header. There is no direct physical link between the TCP layers, and for this reason, it is called a *virtual link.*

The send and receive TCP layers have buffer memories (usually just called *buffers*). The receive buffer holds incoming data while they are being processed. The send buffer holds data until they are ready for

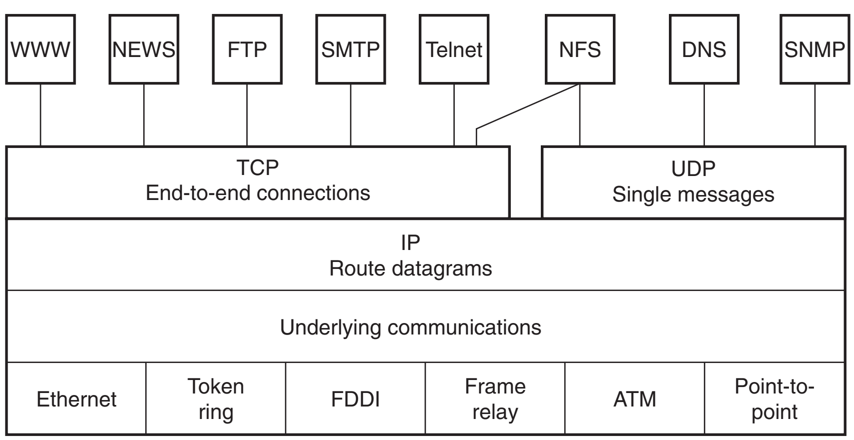


Figure 15.4 The TCP/IP suite. (*After Feit, 1997.*)

transmission. It also holds copies of data already sent until it receives an acknowledgment that the original has been received correctly. The *receive window* is the amount of receive buffer space available at any given time. This changes as the received data are processed and removed from the buffer. The receive TCP layer sends an acknowledgment (ACK) signal to the send TCP layer when it has cleared data from its buffer, and the ACK signal also provides an update on the current size of the receive window.

The send TCP layer keeps track of the amount of data in transit and therefore unacknowledged. It can calculate the amount of receive buffer space remaining, allowing for the data in transit. This remaining buffer space represents the amount of data that can still be sent and is termed the *send window.* The send TCP layer also sets a *timeout period,* and failure to receive an ACK signal within this period results in a duplicate packet being sent. On terrestrial networks, the probability of bit error (see Chap. 10) is extremely low, and *congestion* is the most likely reason for loss of ACK signals. Because a network carries traffic from many sources, traffic congestion can occur. The IP layer of the TCP/IP discards packets when congestion occurs, and hence the corresponding ACK signals from the TCP layer do not get sent. Rather than continually resending packets, the send station reduces its rate of transmission, this being known as *congestion control.* A *congestion window* is applied, which starts at a size of one segment for a new connection. The window is doubled in size for each ACK received until it reaches a maximum value determined by the number of failed ACKs experienced. For normal operation, the congestion window grows in size to equal the receive window. The congestion window increases slowly at first, but as each doubling takes effect, the size increases exponentially. This controlling mechanism is known as *slow start.* If congestion sets in, this will be evidenced by an increase in the failure to receive ACKs, and the send TCP will revert to the slow start.

15.4 Satellite Links and TCP

Although satellite links have formed part of the Internet from its beginning, the rapid expansion of the Internet and the need to introduce congestion control have highlighted certain performance limitations imposed by the satellite links. Before discussing these, it should be pointed out that the increasing demand for Internet services may well be met best with satellite direct-to-home links, and many companies are actively engaged in setting up just such systems (see Sec. 15.6).

In the ideal case, the virtual link between TCP layers should not be affected by the physical link, and certainly the Transmission Control Protocol (TCP) is so well established that it would be undesirable (some would say unacceptable) to modify it to accommodate peculiarities of the physical link. The factors that can adversely affect TCP performance over satellite links are as follows:

Bit error rate (BER). Satellite links have a higher bit error rate (BER; see Chap. 10) than the terrestrial links forming the Internet. Typically, the satellite link BER without error control coding is around 10^{-6}, whereas a level of 10^{-8} or lower is needed for successful TCP transfer (Chotikapong and Sun, 2000). The comparatively low BER on terrestrial links means that most packet losses are the result of congestion, and the TCP send layer is programmed to act on this assumption. When packets are lost as a result of high BER, therefore, as they might on satellite links, the TCP layer assumes that congestion is at fault and automatically invokes the congestion control measures. This slows the throughput.

Round-trip time (RTT). The round-trip time (RTT) of interest here is the time interval that elapses between sending a TCP segment and receiving its ACK. With geostationary (GEO) satellites, the round-trip propagation path is ground station to satellite to ground station and back again. The range from ground station to the satellite (see Chap. 3) is on the order of 40,000 km, and therefore, the propagation path for the round trip is $4 \times 40{,}000 = 160{,}000$ km. The propagation delay is therefore $160{,}000/3 \times 10^{8} = 0.532$ s. This is just the space propagation delay. The total round-trip time must take into account the propagation delays on the terrestrial circuits and the delays resulting from signal processing. For order of magnitude calculations, an RTT value of 0.55 s would be appropriate. The send TCP layer must wait this length of time to receive the ACKs, and of course, it cannot send new segments until the ACKs are received, which is going to slow the throughput. The send TCP timeout period is also based on the RTT, and this will be unduly lengthened. Also, with interactive applications such as Telnet, this delay is highly undesirable.

Bandwidth-delay product (BDP). The RTT is also used in determining an important factor known as the *bandwidth delay product* (BDP). The delay part of this refers to the RTT, since a sender has to wait this amount of time for the ACK before sending more data. The bandwidth refers to the channel bandwidth. As shown in Chaps. 10 and 12, bandwidth and bit rate are directly related. In network terminology, the bandwidth is usually specified in bytes per second (or multiples of this), where it is understood that 1 byte is equal to 8 bits. For example, a satellite bandwidth of 36 MHz carrying a BPSK signal could handle a bit rate given by Eq. (14.30) as 30 megabits per second (Mb/s). This is equivalent to 3.75×10^6 bytes per second (B/s) or about 3662 kilobytes per second (kB/s). If the sender transmits at this rate, the largest packet it can send within the RTT of 0.55 s is $3662 \times 0.55 = 2014$ kilobytes approximately. This is the BDP for the two-way satellite channel. The channel is sometimes referred to as a pipeline, and one that has a high BDP, as a long fat pipe. Now the receive TCP layer uses a 16-bit word to notify the send TCP layer of the size of the receive window it is going to use. Allowing 1 byte for certain overheads, the biggest segment size that can be declared for the receive window is $2^{16} - 1 = 65{,}535$ bytes, or approximately 64 kilobytes. (Recall that 1 kilobyte is equal to 1024 bytes.) This falls well short of the 2014 kilobytes set by the BDP for the channel, and thus the channel is very underutilized.

Variable round-trip time. Where lower earth orbiting satellites are used such as those in low earth orbits (LEOs) and medium earth orbits (MEOs), the propagation delays will be much less than that for the GEO. The slant range to LEOs is typically on the order of a few thousand kilometers at most, and for MEOs, a few tens of thousand kilometers. The problem with these orbits is not so much the absolute value of delay as the variability. Because these satellites are not geostationary, the slant range varies, and for continuous communications there is the need for intersatellite links, which also adds to the delay and the variability. For example, for LEOs, the delay can vary from a few to about 80 ms. Whether or not this will have an impact on TCP performance is currently an open question (RFC-2488).

15.5 Enhancing TCP Over Satellite Channels Using Standard Mechanisms (RFC-2488)

In keeping with the objective that, where possible, the TCP itself should not be modified to accommodate satellite links, the Request for Comments 2488 (RFC-2488) describes in detail several ways in which the performance over satellite links can be improved. These are summarized in Table 15.1. The first two mechanisms listed do not require any changes to the TCP. The others do require extensions to the TCP. As always, any extensions to the TCP must maintain compatibility with networks that do not employ the extensions. Brief descriptions of the mechanisms are included, but the reader is referred to the Requests for Comments (RFCs) for full details (see Sec. 15.6).

MTU stands for *maximum transmission unit,* and *Path MTU-Discovery* is a method that allows the sender to find the largest packet and hence largest TCP segment size that can be sent without fragmentation. The congestion window is incremented in segments; hence

TABLE 15.1 Summary of Objectives in RFC-2488

Mechanism	Use	RFC-2488 Section	Where Applied
Path MTU-Discovery	Recommended	3.1	Sender
FEC	Recommended	3.2	Link
	TCP Congestion Control		
Slow start	Required	4.1.1	Sender
Congestion avoidance	Required	4.1.1	Sender
Fast retransmit	Recommended	4.1.2	Sender
Fast recovery	Recommended	4.1.2	Sender
	TCP Large Windows		
Window Scaling	Recommended	4.2	Sender & Receiver
PAWS	Recommended	4.2	Sender & Receiver
RTTM	Recommended	4.2	Sender & Receiver
TCP SACKS	Recommended	4.4	Sender & Receiver

larger segments allow the congestion window to increment faster in terms of number of bytes carried. There is a delay involved in implementing Path MTU-Discovery, and of course, there is the added complexity. Overall, however, it improves the performance of TCP over satellite links.

Forward error correction (FEC). Lost packets, whether from transmission errors or congestion, are assumed by the TCP to happen as a result of congestion, which means that congestion control is implemented, with its resulting reduction in throughput. Although there is ongoing research into ways of identifying the mechanisms for packet loss, the problem still remains. Application of FEC (as described in Chap. 11) therefore should be used where possible.

Slow start and congestion avoidance. These strategies have already been described in Sec. 15.3, along with the problems introduced by long RTTs. Slow start and congestion avoidance control the number of segments transmitted, but not the size of the segments. Using Path MTU-Discovery as described earlier can increase the size, and hence the data throughput is improved.

Fast retransmit and fast recovery. From the nature of the ACKs received, the fast retransmit algorithm enables the sender to identify and resend a lost segment before its timeout expires. Since the data flow is not interrupted by timeouts, the sender can infer that congestion is not a problem, and the fast recovery algorithm prevents the congestion window from reverting to slow start. The fast retransmit algorithm can only respond to one lost segment per send window. If there is more than one, the others trigger the slow start mechanism.

TCP large windows. As shown in Sec. 15.4 in connection with the *bandwidth delay product,* the receive window size is limited by the address field to 64 kilobytes maximum. By introducing a *window scale extension* into the TCP header, the address field can be effectively increased to 32 bits. Allowing for certain overheads, the maximum window size that can be declared is $2^{30} = 1$ gigabyte (again keeping in mind that 1 gigabyte $= 1024^3$ bytes). The window size and hence the scale factor can be set locally by the receive TCP layer. Note, however, that the TCP extension has to be implemented at the sender and the receiver.

The two mechanisms *PAWS,* which stands for *protection against wrapped sequence,* and *RTTM,* which stands for *round-trip time measurement,* are extensions that should be used with large windows. Maintaining steady traffic flow and avoiding congestion require a current knowledge of the RTT, which can be difficult to obtain with large windows. By including a *time stamp* in the TCP header, the RTT can be measured. Another problem that arises with large windows is that the numbering of old sequences can overlap with new, a condition known as *wrap-around.* The protection against wrapped sequences is an algorithm that also makes use of the time stamp. These algorithms are described fully in RFC-1323.

SACK stands for *selective acknowledgment* and is a strategy that enables the receiver to inform the sender of all segments received successfully. The sender then need only resend the missing segments. The strategy should be used where multiple segments may be lost during transmission, such as, for example, in a satellite link, since clearly, retransmission of duplicate segments over long delay paths would seriously reduce the throughput. Full details of SACK will be found in RFC-2018.

15.6 Requests for Comments

The rapid growth of the Internet resulted in large part from the free and open access to documentation provided by network researchers. The ideas and proposals of researchers are circulated in memos called *Requests for Comments* (RFCs). They can be accessed on the World Wide Web at a number of sites, for example, *http://www.rfc-editor.org/*. Below is a summary of some of the RFCs that relate specifically to satellite links and that have been referred to in Sec. 15.5

- *RFC-2760, Ongoing TCP Research Related to Satellites, February 2000. Abstract:* This document outlines possible TCP enhancements that may allow TCP to better utilize the available bandwidth provided by networks containing satellite links. The algorithms and mechanisms outlined have not been judged to be mature enough to be recommended by the IETF. The goal of this document is to educate researchers as to the current work and progress being done in TCP research related to satellite networks.

- *RFC-2488, Enhancing TCP Over Satellite Channels Using Standard Mechanisms, January 1999. Abstract:* The Transmission Control Protocol (TCP) provides reliable delivery of data across any network path, including network paths containing satellite channels. While TCP works over satellite channels, there are several IETF standardized mechanisms that enable TCP to more effectively utilize the available capacity of the network path. This document outlines some of these TCP mitigations. At this time, all mitigations discussed in this document are IETF standards track mechanisms (or are compliant with IETF standards).
- *RFC-2018, TCP Selective Acknowledgment Options, October 1996. Abstract:* TCP may experience poor performance when multiple packets are lost from one window of data. With the limited information available from cumulative acknowledgments, a TCP sender can only learn about a single lost packet per round-trip time. An aggressive sender could choose to retransmit packets early, but such retransmitted segments may have already been received successfully. A selective acknowledgment (SACK) mechanism, combined with a selective repeat retransmission policy, can help to overcome these limitations. The receiving TCP sends back SACK packets to the sender informing the sender of data that have been received. The sender can then retransmit only the missing data segments. This memo proposes an implementation of SACK and discusses its performance and related issues.
- *RFC-1323, TCP Extensions for High Performance, May 1992. Abstract:* This memo presents a set of TCP extensions to improve performance over large bandwidth*delay product paths and to provide reliable operation over very high-speed paths. It defines new TCP options for scaled windows and timestamps, which are designed to provide compatible interworking with TCPs that do not implement the extensions. The timestamps are used for two distinct mechanisms: RTTM (round-trip time measurement) and PAWS (protect against wrapped sequences). Selective acknowledgments are not included in this memo. This memo combines and supersedes RFC-1072 and RFC-1185, adding additional clarification and more detailed specification. Appendix C of RFC-1323 summarizes the changes from the earlier RFCs.
- *RFC-1072, TCP Extensions for Long Delay Paths, October 1988. Status of this memo:* This memo proposes a set of extensions to the TCP to provide efficient operation over a path with a high band-

width*delay product. These extensions are not proposed as an Internet standard at this time. Instead, they are intended as a basis for further experimentation and research on TCP performance. Distribution of this memo is unlimited.

15.7 Split TCP Connections

The Transmission Control Protocol (TCP) provides end-to-end connection. By this is meant that the TCP layers at the sender and receiver are connected through a virtual link (see Sec. 15.3) so that such matters as congestion control, regulation of data flow, etc. can be carried out without intervention of intermediate stages. It is to preserve this end-to-end connection that many of the extensions to TCP described in the preceding section have been introduced.

If, however, it is assumed that the end-to-end connectivity can be split, new possibilities are opened up for the introduction of satellite links as part of the overall Internet. Figure 15.5 shows one possible arrangement (Ghani and Dixit, 1999). Breaking the network in this way is termed *spoofing*. This refers to the fact that the TCP source thinks it is connected to the TCP destination, whereas the interworking unit (IWU) performs a protocol conversion. In Fig. 15.5, TCP Reno refers to the TCP with extensions: slow start, congestion avoidance, fast retransmit, fast recovery, support for large windows, and delayed ACKs. At the IWU the data are transferred from the TCP Reno to the

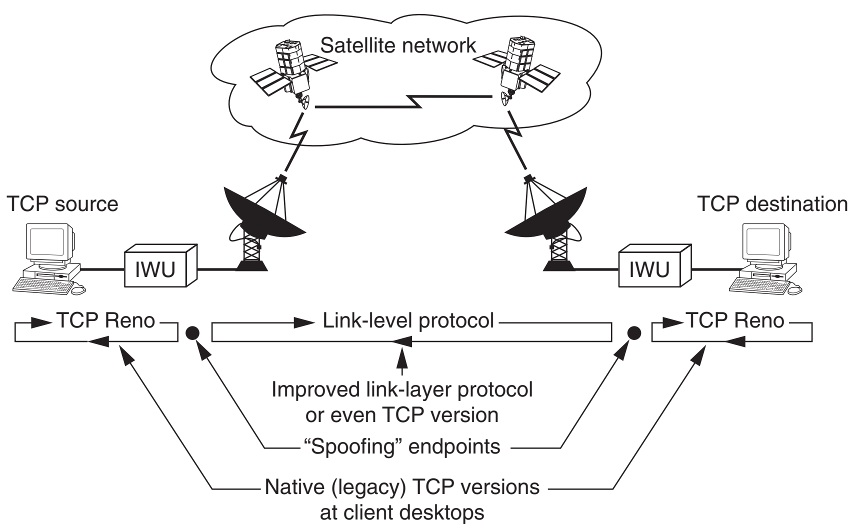


Figure 15.5 TCP/IP satellite link spoofing configuration. (*From Ghani and Dixit, 1999. © 1999 IEEE.*)

data-link protocol. As shown in the figure, any one of a number of link layer protocols may be used. At the destination end, the IWU performs the conversion back to TCP Reno.

One approach developed at Roke Manor Research, Ltd. (West and McCann, 2000) illustrates some of the possibilities and problems associated with splitting. The two key issues are setup and teardown (West and McCann, 2000). To illustrate the process, consider a connection being set up between host *A* and host *B*. In setting up the connection, host *A* sends a synchronizing segment, labeled SYN in the TCP/IP scheme, which specifies certain protocols to be followed. Host *B* responds with its own SYN, which contains its protocol requirements, also an ACK signal that carries the number to be used by host *A* for the first data byte it sends. Host *A* then responds with an ACK signal that carries the number to be used by host *B* for the first data byte it sends. The three signals, SYN, SYN/ACK, and ACK, constitute what is known as a *three-way handshake*.

A three-way handshake is also used to close (teardown). Suppose host *B* wishes to close. It sends a final segment (FIN). Host *A* acknowledges the FIN with an ACK and follows this with its own final segment (FIN). Host *B* acknowledges the FIN with an ACK. On connections with long RTTs, it will be seen that these three-way handshakes will be very time-consuming.

Figure 15.6 shows the system developed at Roke Manor. The *enhancers* perform the same function as the IWUs in Fig. 15.5 in that they terminate the Internet connections and do not require any modifications to the TCP/IP. A propriety protocol is used over the satellite link. Figure 15.7 illustrates a situation where both setup and teardown are spoofed. In this illustration, host *B* refuses the connection, but host *A* receives the RESET signal too late. The spoofed FIN ACK tells host *A* that the data transfer was successful.

Figure 15.8 shows a more appropriate strategy. In this case, the FIN sent by host *A* is not spoofed. Since host *B* has refused the connection, host *A* receives no FIN, ACK back from host *B*. Therefore, host *A* can

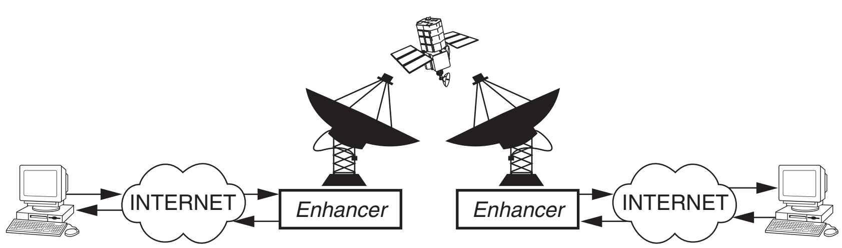


Figure 15.6 Physical architecture incorporating enhancer technology. (*From West and McCann, 2000.*)

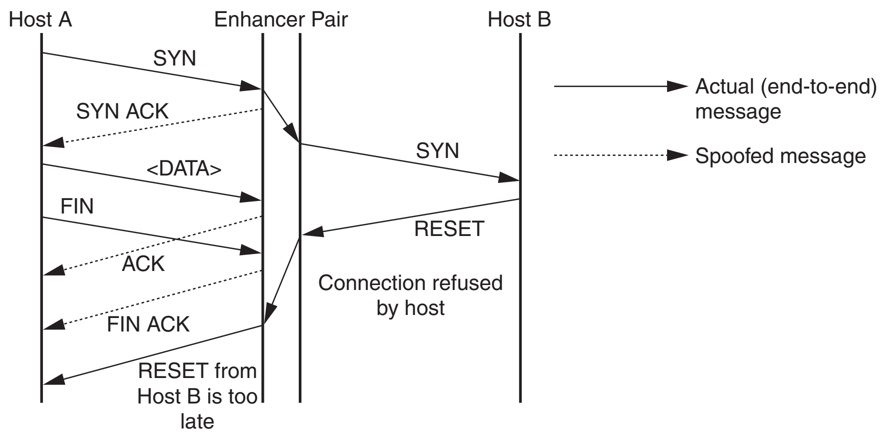


Figure 15.7 Connection establishment and closing. (*From West and McCann, 2000.*)

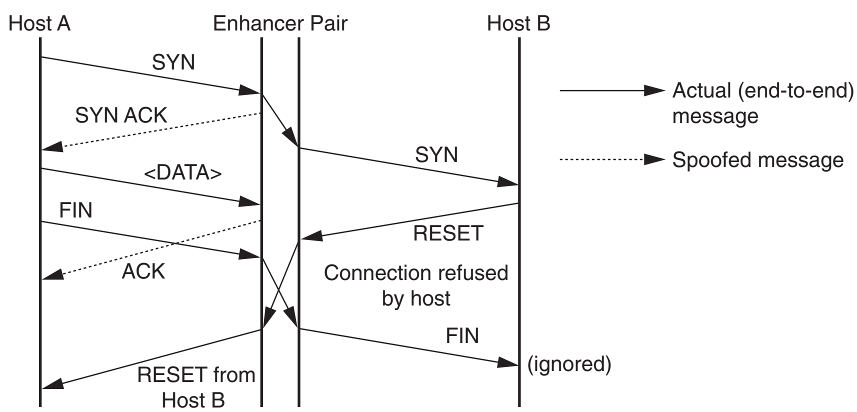


Figure 15.8 Improved connection close. (*From West and McCann, 2000.*)

infer from this that there was an error. While this is not as good as a regular TCP/IP connection, which would have reported a failure to connect on the first SYN, the system does adjust to the error and removes the spoofing on the setup on a second try.

15.8 Asymmetric Channels

The term *asymmetry* applies in two senses to an Internet connection. It can refer to the data flow, which is often asymmetric in nature. A short request being sent for a Web page and the returned Web page may be a much larger document. Also, the acknowledgment packets sent on the return or reverse link are generally shorter than the TCP segments sent on the forward link. Values of 1500 bytes for data segments on the forward link and 40 bytes for ACKs on the reverse link are given in RFC-2760.

Asymmetry is also used to describe the physical capacities of the links. For small earth stations (e.g., VSATs), transmit power and antenna size (in effect, the EIRP) limit the uplink data rate, which therefore may be much less than the downlink data rate. Such asymmetry can result in ACK congestion. Again, using some values given in RFC-2760, for a 1.5 Mb/s data link a reverse link of less than 20 kb/s can result in ACK congestion. The levels of asymmetry that lead to ACK congestion are readily encountered in VSAT networks that share the uplink through multiple access.

In some situations, the reverse link may be completed through a terrestrial circuit, as shown in Fig. 15.9 (Ghani and Dixit, 1999). Here, the TCP source is connected to the satellite uplink through an IWU as before. The downlink signal feeds the small residential receiver, which is a receive-only earth station. An IWU on the receive side converts the data to the TCP format and sends them on to the destination. The ACK packets from the TCP destination are returned to the TCP source through a terrestrial network. As pointed out in RFC-2760, the reverse link capacity is limited not only by its bandwidth but also by queue lengths at routers, which again can result in ACK congestion. Some of the proposed methods of handling asymmetry problems and ongoing research are described in Ghani and Dixit (1999).

DirecPC (*http://www.direcpc.com/*), a product and service of Hughes Network Systems, takes advantage of asymmetric channels to provide satellite transfer of Internet data to the individual consumer and homeowner. The system as shown in Fig. 15.10 can be used for all TCP/IP-based software applications such as email, Telnet, and the World Wide Web (WWW). A customer URL (uniform, or universal resource locator) request goes out through a modem to the Internet service provider (ISP) at normal modem speeds (typically in the range 28.6 to 56 kb/s). Before sending the URL, the DirecPC software installed in the customer's PC attaches what is termed a *tunneling code* to the URL. This instructs the ISP to forward the request to the DirecPC network operations center (NOC) instead of to the Internet server. At the NOC, the tunneling code is stripped away, and the request is forwarded to the Internet site through high-speed lines. The requested content is returned to the NOC, which then downloads it to the customer through the satellite link, at a bit rate of up to 400 kb/s. The advantage of the system, therefore, is that the large data files are returned over a high-speed link rather than back through the low-speed modem.

Hughes 601 spacecraft are used, and the downlink is in the Ka band (20 to 30 GHz). The company offers a variety of system configurations,

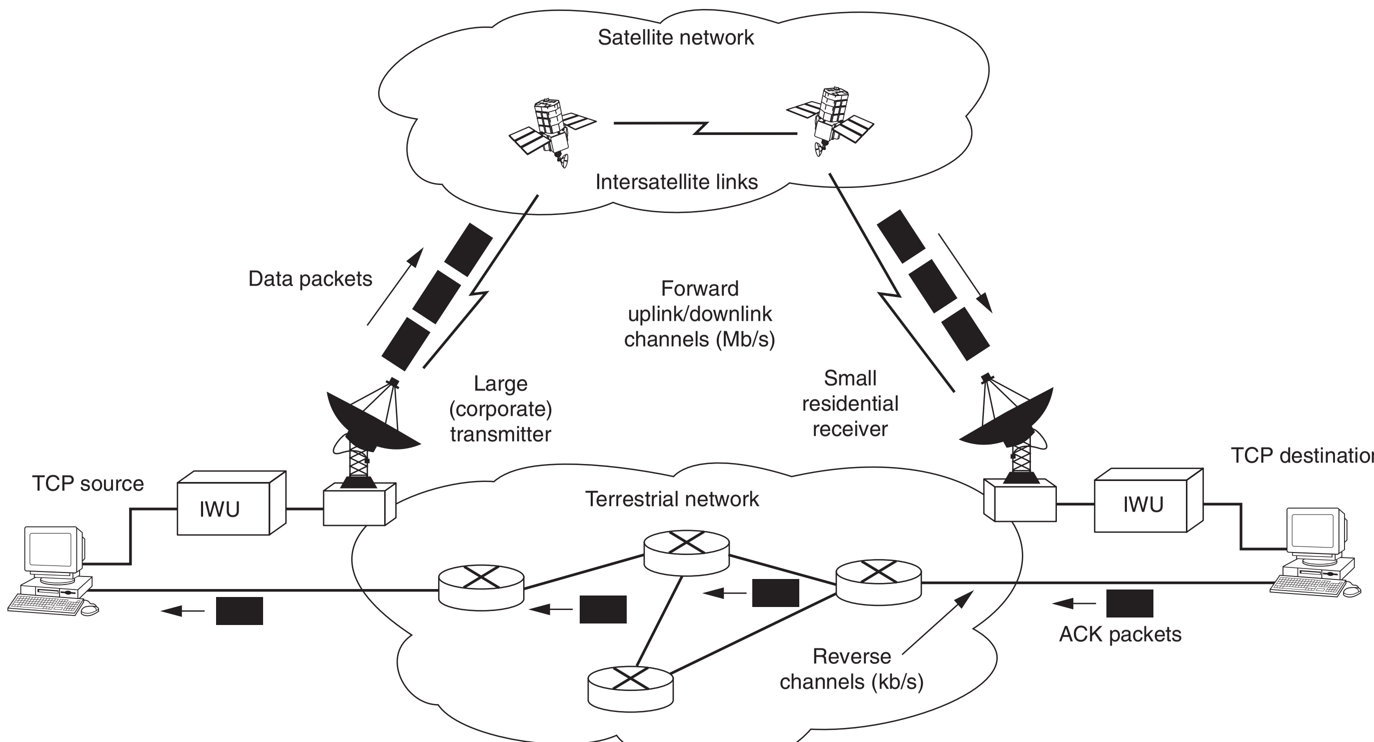


Figure 15.9 Asymmetric reverse ACK channel configuration. (*From Ghani and Dixit, 1999. © 1999 IEEE.*)

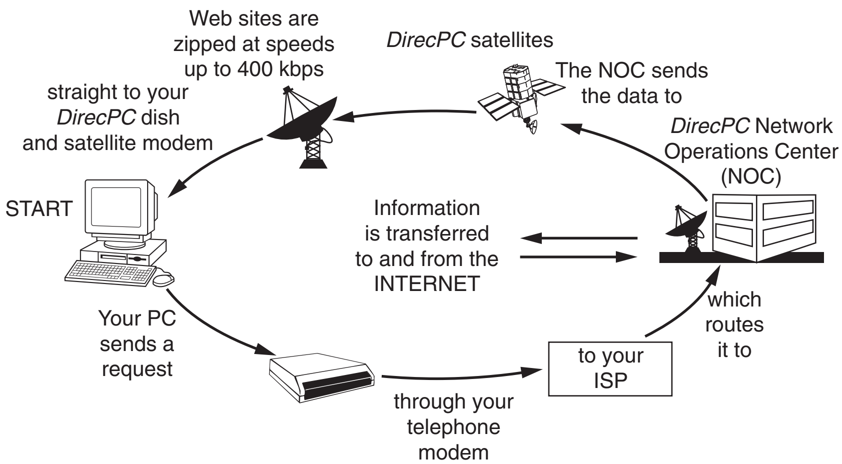


Figure 15.10 DirecPC turbo Webcast. (*From* http://direcpc.com/consumer/what/services.html.)

ranging from Internet-only connections to a combined DirecPC and DirecTV system.

15.9 Proposed Systems

Most of the currently employed satellites operate in what is called a "bent pipe" mode; that is, they relay the data from one host to another without any onboard processing. Also, many of the problems with using geostationary satellites for Internet traffic arise because of the long propagation delay and the resulting high delay-bandwidth product. In some of the newer satellite systems, use is made of low and medium earth orbiting satellites to cut down on the propagation delay time. Also, onboard signal processing is used in many instances that may result in the TCP/IP protocol being exchanged for propriety protocols over the satellite links. The satellite part of the network may carry the Internet Protocol (IP) over what is known as Asynchronous Transfer Mode (ATM). With ATM, the data are partitioned into fixed-length packets, called *cells,* of length 53 bytes total. Five bytes form the header for the cell, and the remaining 48 bytes are the payload. However, with ATM, the "payload" also carries additional overhead bytes, specifically bytes for what is called the *ATM adaptation layer.* The *asynchronous* part of the ATM refers to the fact that cells are transmitted asynchronously; that is, they do not need to be assigned regular time slots. The cell size of 53 bytes is a compromise between what is suitable for time-sensitive traffic such as voice and video and what is suitable for non-time-sensitive data. For high-bit-rate traffic, many cells can be transmitted in sequence, and traffic bursts also can be handled in this way. This is sometimes referred to as *bandwidth on demand.* Note that synchronous digital signals may be carried in the cells, even though the system is referred to as asynchronous. A good introduction to ATM networking is provided in Goralski (1995), and TCP/IP over satellite ATM networks is described in Goyal et al. (1999).

The Ka band, which covers from 27 to 40 GHz, is used in many (but not all) of these newer systems. Wider bandwidths are available for carriers in the Ka band compared with those in the Ku band. A survey of some of these broadband systems will be found in Farserotu and Prasad (2000). A summary is provided in the following paragraphs.

Astrolink. The Astrolink system is scheduled to be operational in 2003. It will begin service with four geostationary satellites, and this number will be expanded to nine as the need arises. The nine satellites are to be distributed among the five orbital positions: 97°W for the Americas; 21.5°W for Europe, the Americas, and Africa; 130°E for East Asia and Australia; 2°E for Europe, Africa, and West Asia; and 175°E for Oceania. Uplink frequencies are in the ranges 28.35 to 28.8 GHz and 29.25 to 30.0 GHz, and downlink frequencies are in the range 19.7 to 20.2 GHz (Farserotu and Prasad, 2000). Each satellite operates 44 spot beams, and onboard switching will be provided. An extensive ground segment is envisioned, having four regional network control centers for the day-to-day running of the network and a master network control center that will supervise and control the overall system. In addition to the 44 spot beams, there will be 12 point-of-presence beams that provide large-scale interconnection with terrestrial facilities. The Astrolink satellite network will be ATM-based and will be compatible with terrestrial technologies, including the Internet Protocol (IP). The ATM-based technology enables the satellites to provide the bandwidth-on-demand service where the customer is only charged for the bandwidth used. Customers will be able to transmit at speeds up to 20 Mb/s and to receive or download at speeds up to 220 Mb/s. Further details are available at www.astrolink.com.

Loral Cyberstar. Scheduled for operation in 2001, the Cyberstar satellite segment consists of three geostationary satellites operating in the Ka band (Farserotu and Prasad, 2000). A number of gateway teleports are stationed around the world, which provide access to and from the satellite segment. In addition, the company operates point-of-presence sites in North America, which are part of the high-speed link between Internet and the satellite. Small WWW requests from a host to the Internet are transported over a terrestrial link. Cyberstar provides IP *multicast* services. Multicast service is where information is sent to a specified group of users; for example, it might be used for conferencing. Cyberstar makes use of *digital video broadcast* (DVB) to send IP packets. Briefly, DVB is a worldwide standard used for the transmission of digital TV via satellites (DVB-S), cable (DVB-C), and terrestrial (DVB-T). DVB supports data transmission as well as video and can carry multiple program streams (Liebowitz, 2000). In the Cyberstar system, the IP data packets are "repackaged" into an MPEG-2 format, which then form the IP/DVB stream. (The MPEG-2 format is described in Chap. 16.) In general, a DVB signal is received on a *integrated receiver decoder* (IRD), which is available either as a set-top box or on a PC card. The Cyberstar system uses a set-top unit that supports data rates up to 4 Mb/s, with an 8-Mb/s IRD to be introduced. For simplex operation, the receiver only is needed, along with the antenna and LNB feed. For most locations, a small antenna, less than 1 m in diameter, can be used. Duplex operation requires in addition the installation of a standard SCPC transmitter for the return link. Basically, this is a VSAT installation. Further details can be obtained at www.cyberstar.com.

SkyBridge. Scheduled for operation in 2001, Skybridge will use 80 low earth orbiting (LEO) satellites in a circular orbit at an altitude of 1469 km (913 miles) and inclination of 53°. The control ground segment (CGS) includes the satellite control center and the telemetry and command ground stations. The CGS controls all aspects of the satellite constellation, including satellite positioning and orientation. Satellite positions are determined by the global positioning satellite system (GPS; see Sec. 17.5). Carrier frequencies are in the Ku band, uplinks being in the range 12.75 to 14.5 GHz and downlinks in the range 10.7 to 12.75 GHz. Right-hand circular (RHC) and left-hand circular (LHC) polarizations are used to maximize frequency utilization. Each satellite operates with a total of 18 spot beams that carry the uplink and downlink signals. No onboard processing is used, the satellites being in the "bent pipe" category. Approximately 200 gateway stations are planned worldwide, which interface with the terrestrial network, and user terminals are linked to the gateways through the satellites. The user terminal consists of an indoor unit and an outdoor unit, which, of course, includes the antenna. Two levels of user terminals are envisioned. One is a residential terminal, where the outdoor unit is small enough to be installed on a residential rooftop. The residential terminal can handle bit rates of up to 20 Mb/s for the forward link and 2 Mb/s for the return link. The other is a larger terminal suited to professional use, including distribution to several users. This professional terminal can handle traffic of up to 100 Mb/s on the forward link and 10 Mb/s on the return link. The equipment is larger than that used with the residential terminal, with the outdoor unit being designed for installation on larger buildings. Great care has been taken in system design to avoid interference with geostationary satellite systems that utilize Ku-band frequencies; also, power densities should be low enough to prevent interference with terrestrial Ku-band systems. Skybridge supports a wide range of applications including multimedia applications over broadband Internet. The low propagation delay with the LEOs makes the system usable for real-time applications. Further details can be obtained at www.skybridgesatellite.com.

Spaceway. Spaceway should begin operating in 2002 with an initial North American constellation of two geostationary satellites. These will by Hughes-built HS 702 satellites, the network being part of the Hughes Network Systems. These satellites will have onboard processing, packet switching, and spot-beam technology. The assigned North American slots are at 101°W and 99°W. International geostationary slots have been assigned at 49°W, 25°E, 54°E, 101°E, 111°E, and 164°E for future expansion (details from www.hns.com/spaceway/spaceway.htm). According to Farserotu and Prasad (2000), the Spaceway system eventually will have 16 satellites in the geostationary earth orbit (GEO) and 20 in medium earth orbits (MEO), the latter being at an altitude of 10352 km. Frequencies will be in the Ka band, and uplink rates will range from 16 kb/s to 6 Mb/s. A number of network protocols are provided, including IP/ATM, and users will be able to access the system through 66-cm (26-in) antennas. A comparison of data rates as given in the Sapceway Web site is reproduced in Table 15.2.

iSky. Formerly KaSTAR Satellite Communications, iSky plans to start its Internet broadband service in late 2001 marketed to homes and small businesses in the United States, Canada, and Latin America. Geostationary satellites will be used, the first, iSky-1 being located at 109.2°W. A large number of spot beams will be used rather than a single continental U.S. (CONUS) beam. Telesat, a Canadian company, has an agreement with iSky to offer Ka-band multimedia capacity through their Anik F2 satellite, scheduled for launch in 2002, at an orbital position of

TABLE 15.2 Comparison of Satellite and Terrestrial bit rates (From *www.hns.com/spaceway/spaceway.htm*)

Application	Information Content	Phone Line at 28.8 kb/s	Spaceway at 384 kb/s	Spaceway at 1.5 Mb/s
Digitized photo	1 megabit	34 s	2.6 s	0.7 s
CAD/CAM	2 megabits	70 s	5.2 s	1.4 s
X-ray	12 megabits	7 min	31.3 s	7.8 s
Digital library	*Washington Post,* Sunday edition	9 min	41.6 s	10.4 s

111.1°W, and in 2002, iSky will launch it iSky-2 satellite, which will cover Latin America. The satellites initially deployed will use the "bent pipe" technology, but the company envisions advanced onboard signal processing for future satellites. The user package includes a 66-cm (26-in) dish and a modem that provides a two-way satellite connection. An initial downlink speed of 1.5 Mb/s will be offered, and it is envisaged that this will be increased to 40 Mb/s. The uplink speed is 0.5 Mb/s. iSky and EchoStar are working in partnership to develop a single dish and single box for use with EchoStar's DISH Network satellite TV and iSky's high-speed data service. iSky's Web site is at www.iSKY.net.

Teledesic. This system should be operational by 2002. It has the trademark name Internet in the Sky, with the stated objective of meeting terrestrial network standards and protocols rather than changing them. To this end, it plans to deploy a total of 288 satellites in low earth orbits (LEOs) and ground terminals that interface between the satellite network, end users, and terrestrial networks. Translation occurs at these terminals between the standard protocols of the terrestrial systems and the internal protocol of the Teledesic satellite network. Most users will employ uplink speeds of up to 2 Mb/s and downlink speeds of up to 64 Mb/s, although broadband terminals also will be available that support 64 Mb/s both ways. Ka-band carriers are used, the uplink being in the range 28.6 to 29.1 GHz and the downlink 18.8 to 19.3 GHz. The satellite constellation ensures that the elevation angle (satellite angle above the horizon) does not fall below 40.25° for any user. This limits the line-of-sight path length through possible rain cells and helps to avoid possible obstacles, allowing the company to claim an availability figure of 99.9 percent or greater. The satellites are at an altitude of 1375 km, and intersatellite links (ISLs) operating at 60 GHz are employed, along with full onboard processing (OBP) and onboard switching (OBS) (Farserotu and Prasad, 2000). Further details can be found at www.teledesic.com.

It should be noted that although this chapter is on Internet applications, these proposed systems carry other types of network traffic such as Asynchronous Transmission Mode (ATM already explained), Integrated Digital Systems Network (ISDN), and Frame Relay. Details of these protocols will be found in Russell (2000). Also, a brief description of these proposed systems and some European systems will be found in Williamson (1999).

15.10 Problems

15.1. Write brief notes on the *Internet* and the *World Wide Web* showing the relationship between them.

15.2. Describe the layered architecture that applies to networks, and indicate which of these layers contain (*a*) the IP and (*b*) the TCP. Explain what is meant by *protocol.*

15.3. Describe how *connectionless* and *connection-oriented* layers differ.

15.4. Describe what is meant by a *packet* in the different layers of a network.

15.5. Binary transmission occurs at a rate of 5000 bytes per second. What is this in bits per second?

15.6. Binary transmission occurs at a rate of 25 kilobytes per second. What is this in kilobits per second?

15.7. Binary transmission occurs at a rate of 30 megabytes per second. What is this in megabits per second?

15.8. What is meant by a *virtual link* in a network?

15.9. Define and explain what is meant by (*a*) *receive window,* (*b*) *send window,* and (*c*) *timeout* with reference to the Internet.

15.10. Explain what is meant by *congestion* and *slow start* in relation to Internet traffic.

15.11. In the early days of the Internet, Internet traffic could be sent over satellite links without any particular difficulty. State briefly the changes that have taken place that introduce difficulties and that require special attention.

15.12. Explain what is meant by *round-trip time* (RTT). Given that the uplink range to a satellite is 38,000 km and the downlink range is 40,000 km, calculate the RTT.

15.13. The overall bandwidth for a satellite link is 36 MHz, and the filtering is raised-cosine with a rolloff factor of 0.2. BPSK modulation is used. The range to the satellite is 39,000 km in both directions. Calculate the size of the largest packet that could be in transit if the bandwidth-delay product (BDP) was the limiting factor. Assume that the RTT is equal to the space propagation delay.

15.14. A low earth orbit (LEO) satellite is used for Internet transmissions. The orbit can be assumed circular at an altitude of 800 km. Assuming that the minimum usable angle of elevation of the satellite is 10° (the angle above the horizon), calculate the approximate values of maximum and minimum RTTs. A spherical earth of uniform mass and radius of 6371 km may be assumed.

15.15. Repeat Prob. 15.14 for a circular medium earth orbit (MEO) at altitude 20,000 km.

15.16. Briefly discuss the enhancements covered in RFC-2488 and how these might improve the performance of Internet traffic over satellite links.

15.17. By accessing the RFC site on the World Wide Web, find the latest version of RFC-2760. Comment on this.

15.18. Explain what is meant by *split TCP connections* and why these might be considered undesirable for Internet use.

15.19. Explain what is meant by *asymmetric channels.* Describe how asymmetric channels may be incorporated in Internet connections via satellites.

15.20. By accessing the WWW addresses given in the text, provide an update on the systems covered in Sec. 15.9.