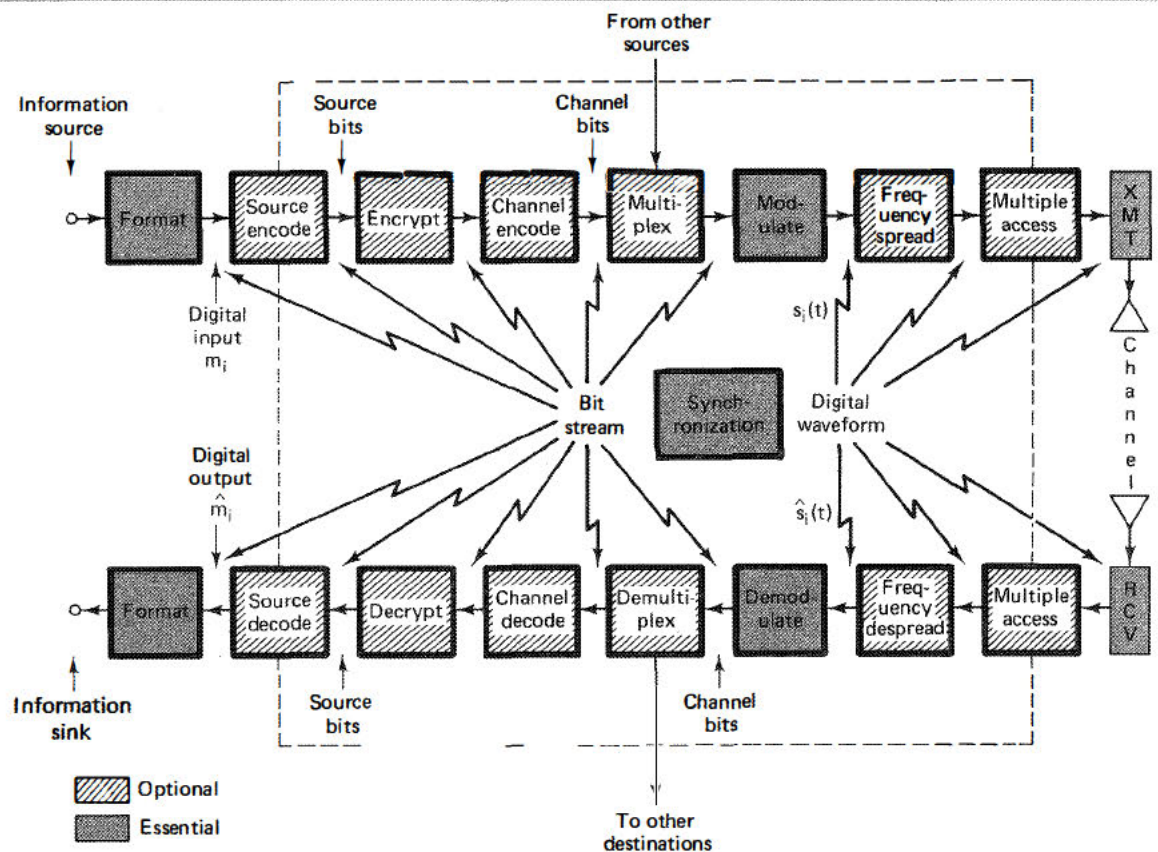


BERNARD SKLAR

DIGITAL COMMUNICATIONS

Fundamentals and Applications



DIGITAL COMMUNICATIONS

Fundamentals and Applications

BERNARD SKLAR

*The Aerospace Corporation, El Segundo, California
and
University of California, Los Angeles*



P T R Prentice Hall
Englewood Cliffs, New Jersey 07632

Library of Congress Cataloging-in-Publication Data

SKLAR, BERNARD (date)
Digital communications.

Bibliography: p.
Includes index.

1. Digital communications. I. Title.

TK5103.7.S55 1988 621.38'0413 87-1316
ISBN 0-13-211939-0

Editorial/production supervision and
interior design: Reynold Rieger
Cover design: Wanda Lubelska Design
Manufacturing buyers: Gordon Osbourne and Paula Benevento



© 1988 by P T R Prentice Hall
Prentice-Hall, Inc.
A Paramount Communications Company
Englewood Cliffs, New Jersey 07632

All rights reserved. No part of this book may be
reproduced, in any form or by any means,
without permission in writing from the publisher.

Printed in the United States of America

20 19 18 17 16 15 14 13 12 11

ISBN 0-13-211939-0

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

*To my mother, Ruth Sklar,
the memory of my father, Julius Sklar,
my wife, Gwen, and our children,
Debra, Sharon, and Dean*

between cycle slips is exponentially distributed; that is, the distribution function of the time between cycle slips, T , is given by

$$p(T) = 1 - \exp\left(-\frac{T}{T_m}\right)$$

Given this distribution function, find the mean time between cycle slips and the variance about this mean as functions of T_m . If the mean time between cycle slips is 1 day, what is the probability of cycle slips less than 1 hour apart? More than 3 days apart?

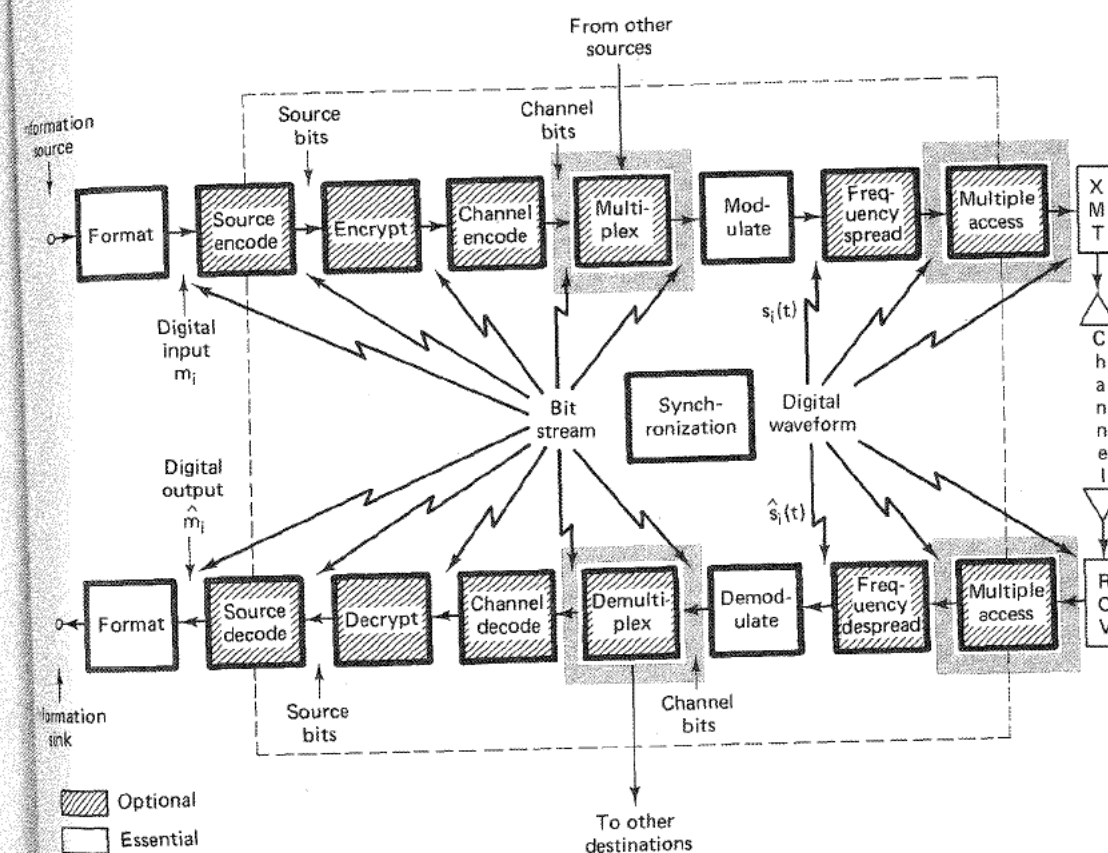
- 8.9. Consider a second-order phase-locked loop with a low-pass loop filter

$$F(\omega) = \frac{\omega_1}{j\omega + \omega_1}$$

During aided acquisition it is desired that this loop be scanned throughout a 1000-radian uncertainty region in 1 s. If the relationship between loop gain and filter constant is $K_0 = 2\omega_1$, what is the required relationship between loop gain and the single-sided additive white Gaussian noise power spectral density, N_0 ? Determine the largest value of N_0 that can be accommodated.

- 8.10. Consider the operation of an open-loop symbol synchronizer whose band pass filter (BPF) has a bandwidth of $0.1/T$ hertz, where T is the symbol period. For a bit energy-to-noise power spectral density ratio (E_b/N_0) of 10 dB, determine the magnitude of the approximate mean and variance of the fractional tracking error, and compute an upper bound on the probability that the tracking error exceeds three times its approximate fractional mean. (Hint: Consider the Chebyshev inequality [4].)
- 8.11. A communication system is used to transmit commands to a payload at a data rate of 100 bits/s. Each command is preceded by an N -bit header that identifies it in the data stream. Assuming that except possibly for the header, the bits appear to be random [$P(1) = P(0) = \frac{1}{2}$], what is the minimum-length header that would provide an expected frequency of false alarms of one per year? For a channel bit error probability of 10^{-5} , what is the probability of missing this header? What is the miss probability if the channel error probability is 2×10^{-2} ? If the system is redesigned to accept the header with up to two errors, what is the minimum required length for an expected false alarm rate of one per year? What is the miss probability with this new system and a channel error probability of 2×10^{-2} ?
- 8.12. A deep-space probe is moving away from the earth at a nominal velocity of 15,000 m/s with a velocity uncertainty of ± 3 m/s. The probe frequency reference is specified to have a drift rate of no more than 10^{-9} Hz/Hz/day. The nominal downlink transmission frequency is 8 GHz. After a 1-month (30-day) silence the probe begins a scheduled transmission toward an earth terminal. The earth terminal contains a cesium standard. What center frequency and frequency search bandwidth should be used by the ground station? Assuming that the range to the probe was accurately known at the beginning of the month, and that the uncertainty in the probe's time and frequency references were zero [$\Delta t(0) = 0$, $\Delta \omega(0) = 0$], what is the uncertainty in the time of arrival of the downlink transmission?
- 8.13. A communications link operates at a nominal center frequency of 10 GHz for a single brief period once a day. The receiver operates with a second-order PLL that has a pull-in range of ± 1 kHz. Assuming that the loop acquires using self-acquisition and that both the transmitter and receiver use the same type of frequency reference, what type of frequency reference must this be?

Multiplexing and Multiple Access



The terms “multiplexing” and “multiple access” refer to the sharing of a fixed communications resource (CR). There is a subtle difference between multiplexing and multiple access. With *multiplexing*, users’ requirements or plans for CR sharing are fixed, or at most, slowly changing. The resource allocation is assigned a priori, and the sharing is usually a process that takes place within the confines of a *local site* (e.g., a circuit board). *Multiple access*, however, usually involves the *remote sharing* of a resource, such as a satellite. With a dynamically changing multiple access scheme, a system controller must become aware of each user’s CR needs; the amount of time required for this information transfer constitutes an overhead and sets an upper limit on the efficiency of the utilization of the CR.

9.1 ALLOCATION OF THE COMMUNICATIONS RESOURCE

There are three basic ways to increase the throughput (total data rate) of a communications resource (CR). The first way is either to increase the transmitter’s effective isotropic radiated power (EIRP) or to reduce system losses so that the received E_b/N_0 is increased. The second way is to provide more channel bandwidth. The third approach is to make the allocation of the CR more efficient. This third approach is the domain of communications multiple access. The problem, in the context of a satellite transponder, is to efficiently allocate portions of the transponder’s fixed CR to a large number of users who seek to communicate digital information to each other at a variety of bit rates and duty cycles. The

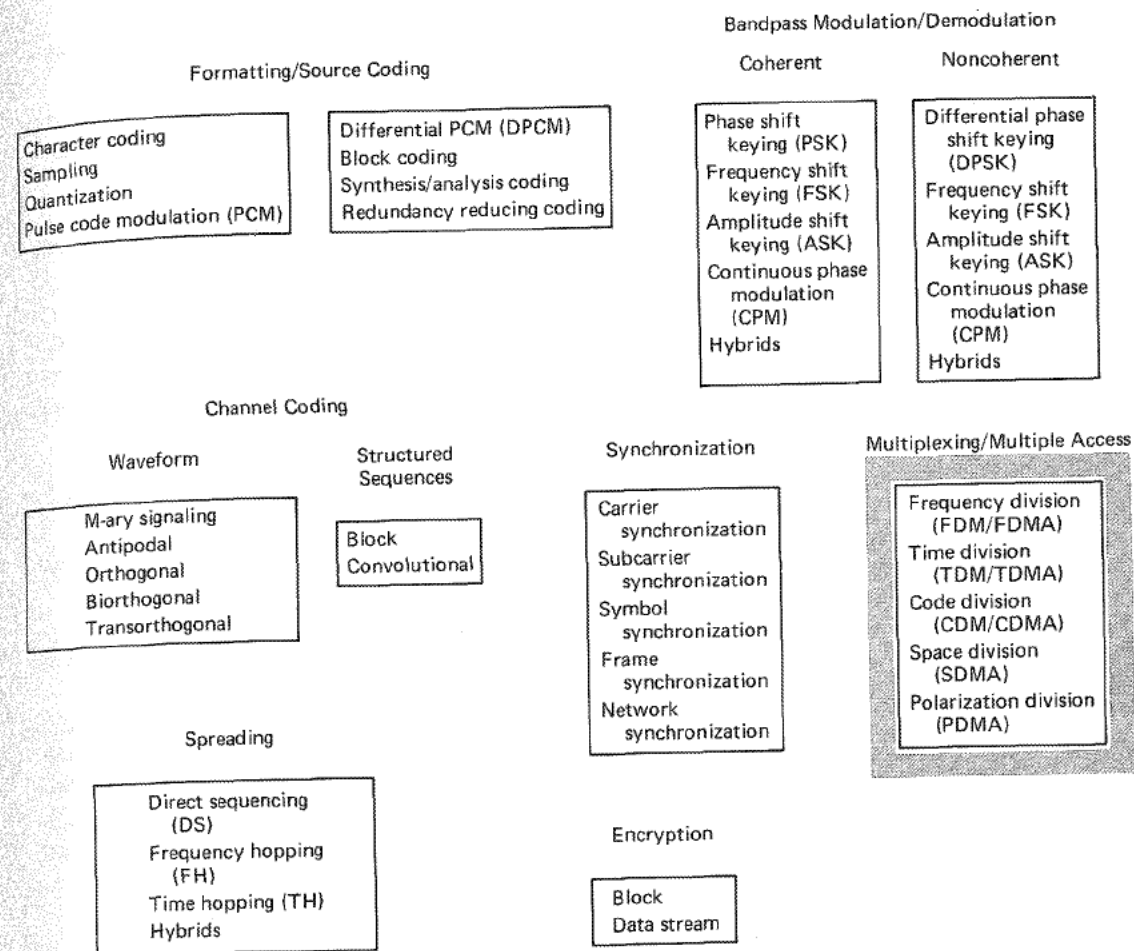


Figure 9.1 Basic digital communication transformations.

basic ways of distributing the communications resource, listed under the heading “multiplexing/multiple access” in Figure 9.1, are:

1. *Frequency division (FD)*. Specified subbands of frequency are allocated.
2. *Time division (TD)*. Periodically recurring time slots are identified. With some systems, users are provided a fixed assignment in time. With others, users may access the resource at random times.
3. *Code division (CD)*. Specified members of a set of orthogonal or nearly orthogonal spread spectrum codes (each using the full channel bandwidth) are allocated.
4. *Space division (SD) or multiple beam frequency reuse*. Spot beam antennas

are used to separate radio signals by pointing in different directions. It allows for reuse of the same frequency band.

5. *Polarization division (PD) or dual polarization frequency reuse.* Orthogonal polarizations are used to separate signals, allowing for reuse of the same frequency band.

The key to *all* multiplexing and multiple access schemes is that various signals share a CR without creating unmanageable interference to each other in the detection process. The allowable limit of such interference is that signals on one CR channel should not significantly increase the probability of error in another channel. Orthogonal signals on separate channels will avoid interference between users. Signal waveforms $x_i(t)$, where $i = 1, 2, \dots$, are defined to be orthogonal if they can be described in the time domain by

$$\int_{-\infty}^{\infty} x_i(t)x_j(t) dt = \begin{cases} K & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \quad (9.1)$$

where K is a nonzero constant. Similarly, the signals are orthogonal if they can be described in the frequency domain by

$$\int_{-\infty}^{\infty} X_i(f)X_j(f) df = \begin{cases} K & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \quad (9.2)$$

where the functions $X_i(f)$ are the Fourier transforms of the signal waveforms $x_i(t)$. Channelization characterized by orthogonal waveforms, as shown in Equation (9.1), is called time-division multiplexing or time-division multiple access (TDM/TDMA), and that characterized by orthogonal spectra, as shown in Equation (9.2), is called frequency-division multiplexing or frequency-division multiple access (FDM/FDMA).

9.1.1 Frequency-Division Multiplexing/Multiple Access

9.1.1.1 Frequency-Division Multiplex Telephony

In the early days of telephony, a separate pair of wires was needed for each telephone trunk circuit (trunk circuits interconnect intercity switching centers). As illustrated in Figure 9.2, the skies of all the major cities in the world grew dark with overhead wires as the demand for telephone service grew. A major development in the early 1900s, frequency-division multiplex (FDM) telephony, made it possible to transmit several telephone signals simultaneously on a single wire, and thereby transformed the methods of telephone transmission.

The communications resource (CR) is illustrated in Figure 9.3 as the frequency-time plane. The channelized spectrum shown here is an example of FDM or FDMA. The assignment of a signal or user to a frequency band is *long term* or *permanent*; the CR can simultaneously contain several spectrally separate signals. The first frequency band contains signals that operate between frequencies f_0 and f_1 , the second between frequencies f_2 and f_3 , and so on. The spectral regions between assignments, called *guard bands*, act as buffer zones to reduce

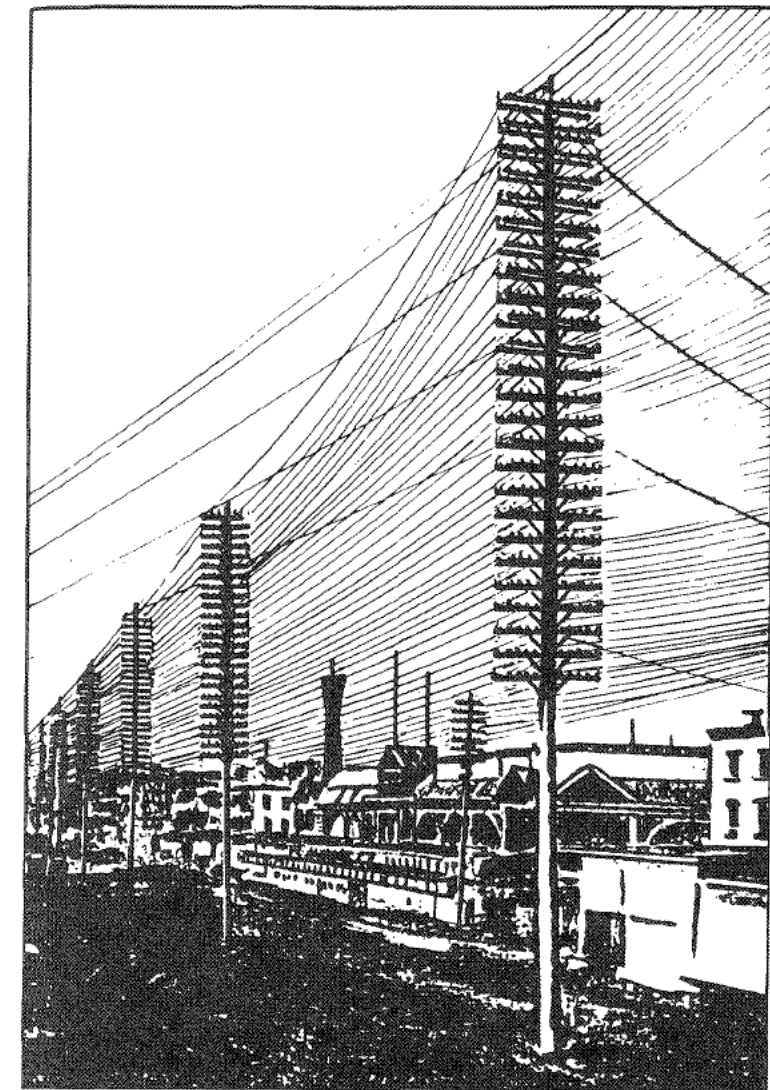


Figure 9.2 In the early days of telephony a pair of wires was needed for each trunk circuit.

interference between adjacent frequency channels. We might ask: How does one transform a baseband signal so that it occupies a higher frequency band? The answer is, by *heterodyning* or *mixing*, also called *modulating* the signal with a fixed frequency from a sine-wave oscillator.

If two input signals to a mixer are sinusoids with frequencies f_A and f_B , the mixing or multiplication will yield new sum and difference frequencies at f_{A+B} and f_{A-B} . The trigonometric identity

$$\cos A \cos B = \frac{1}{2}[\cos(A + B) + \cos(A - B)] \quad (9.3)$$

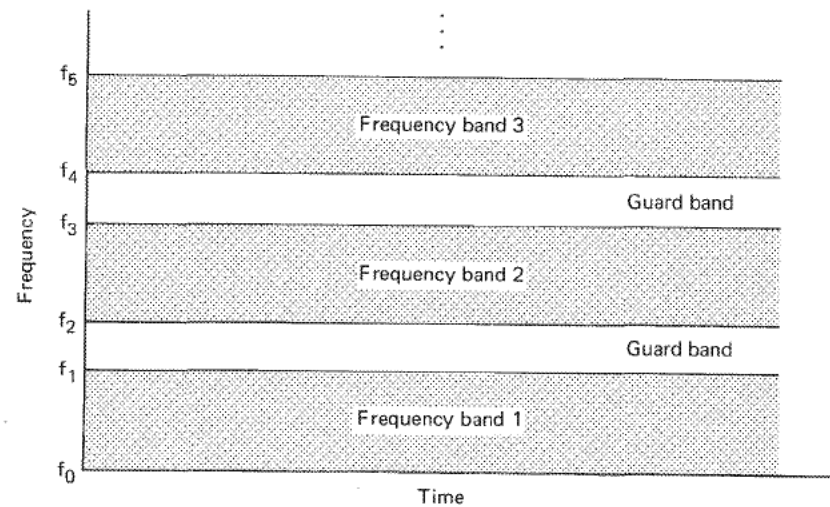


Figure 9.3 Frequency-division multiplexing.

describes the effect of the mixer. Figure 9.4a illustrates the mixing of a typical voice-grade telephone signal, $x(t)$ (baseband frequency range is 300 to 3400 Hz) with a sinusoid from a 20-kHz oscillator. The baseband two-sided magnitude spectrum, $|X(f)|$, is shown in Figure 9.4a. Can the mixer be a linear device? *No*. The output signal of a linear device will only consist of the *same* component frequencies as the input signal, differing only in amplitude and/or phase.

Figure 9.4b illustrates the one-sided magnitude spectrum, $|X(f - f_0)|$, at the mixer output. As a result of the mixing described by Equation (9.3), the output spectrum is a frequency-upshifted version of the baseband spectrum, centered at the oscillator frequency of 20 kHz. This spectrum is called a *double-sideband (DSB) spectrum* because the information appears in two different bands of the positive frequency domain. Figure 9.4c shows the lower sideband (LSB), whose frequency range is 16,600 to 19,700 Hz, the result of filtering the DSB spectrum. This sideband is sometimes referred to as the *inverted sideband* because the order of low-to-high frequency components is the reverse of that of the baseband components. Filtering can similarly be used to separate the upper sideband (USB), whose frequency range is 20,300 to 23,400 Hz, as shown in Figure 9.4d. This sideband is sometimes referred to as the *erect sideband* because the order of the low-to-high frequency components corresponds to that of the baseband components. Each sideband of the DSB spectrum contains the same information. Thus, only one sideband, either the USB or the LSB, is needed in order to retrieve the original baseband data.

A simple FDM example with three translated voice channels is seen in Figure 9.5. In channel 1, the 300- to 3400-Hz voice signal is mixed with a 20-kHz oscillator. In channels 2 and 3, a similar type of voice signal is mixed with a 16-kHz and 12-kHz oscillator, respectively. Only the lower sidebands are retained; the

result of the mixing and filtering (to remove the upper sidebands) yields the frequency-shifted voice channels shown in Figure 9.5. The composite output waveform is just the sum of the three signals, having a total bandwidth in the range 8.6 to 19.7 kHz.

Figure 9.6 illustrates the two lowest levels of the FDM multiplex hierarchy for telephone channels. The first level consists of a *group* of 12 channels modulated onto subcarriers shown in the range 60 to 108 kHz. The second level is made up of five groups (60 channels) called a *supergroup* modulated onto the subcarriers shown in the range 312 to 552 kHz. The multiplexed channels are now treated as a composite signal that can be transmitted over cables or can be further modulated onto a carrier wave for radio transmission.

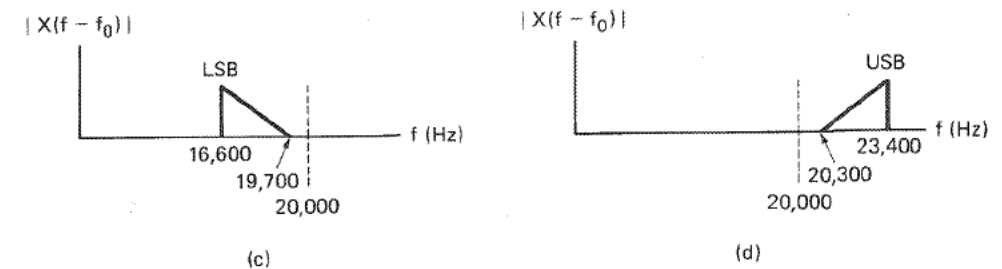
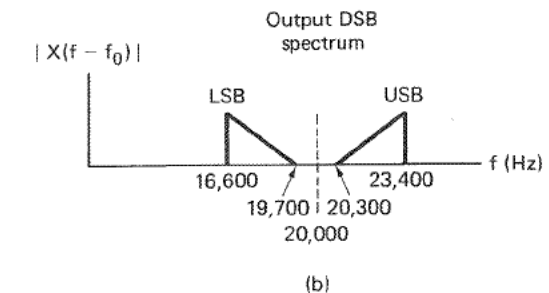
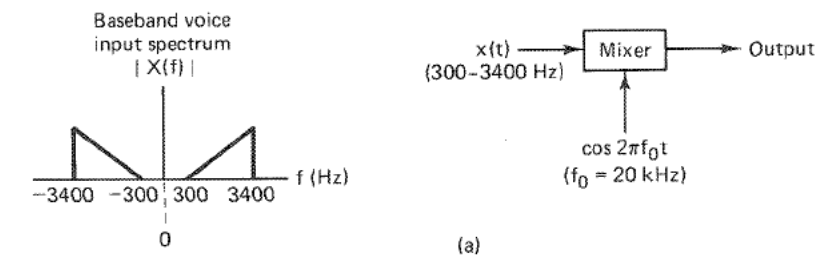


Figure 9.4 Heterodyning (mixing). (a) Mixing operation. (b) Mixer output spectrum. (c) Lower sideband. (d) Upper sideband.

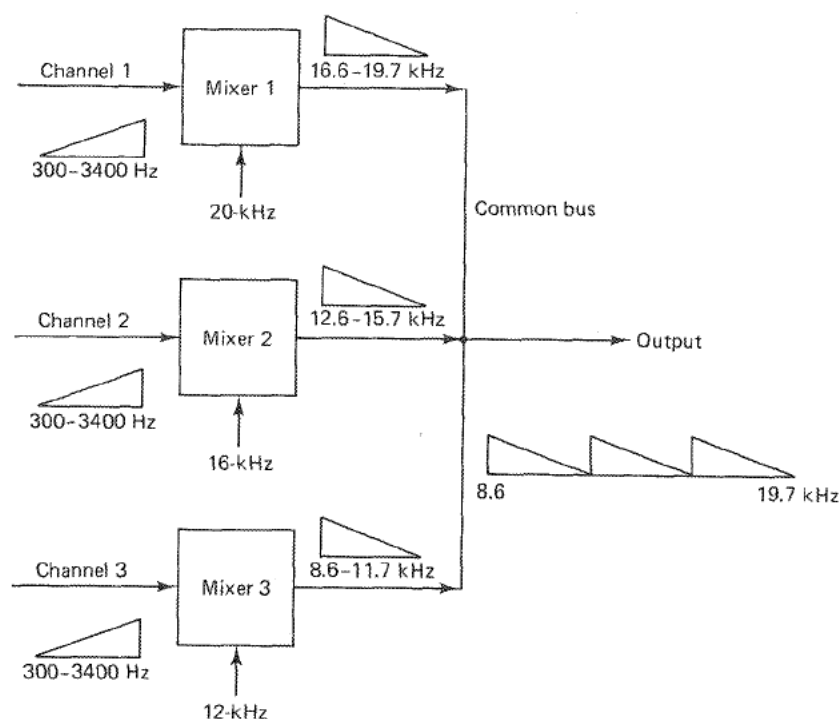


Figure 9.5 Simple FDM example. Three frequency-shifted voice channels.

9.1.1.2 Frequency-Division Multiple Access of Satellite Systems

Most of the free world's communication satellites are positioned in a *geostationary* or *geosynchronous* orbit. This means that the satellite is in a circular orbit, in the same plane as the earth's equatorial plane, and at such an altitude (approximately 19,330 nautical miles) that the orbital period is identical with the earth's rotational period. Since such satellites appear stationary when viewed from the earth, three of them spaced 120° apart can provide worldwide coverage (except for the polar regions). Most communication satellite systems are made up of non-regenerative repeaters or transponders. *Nonregenerative* means that the uplink (earth-to-satellite) transmissions are simply amplified, frequency shifted, and retransmitted on the downlink (satellite-to-earth) without any demodulation/re-modulation or signal processing (see Section 4.7.1). The most popular frequency band for commercial satellite communications, called *C-band*, uses a 6-GHz carrier for the uplink and a 4-GHz carrier for the downlink. For C-band satellite systems, *each satellite* is permitted, by international agreement, to use a 500-MHz-wide spectral assignment. Typically, each satellite has 12 transponders with a bandwidth of 36 MHz each. The most common 36-MHz transponders operate in an FDM/FM/FDMA (frequency-division multiplex, frequency-modulated, fre-

quency-division multiple access) multidestination mode. Let us consider each component of this name:

1. **FDM.** Signals such as telephone signals, each one having a single-sideband 4-kHz spectrum (including guard bands) are FDM'd to form a multichannel composite signal.
2. **FM.** The composite signal is frequency-modulated (FM) onto a carrier and transmitted to the satellite.
3. **FDMA.** Subdivisions of the 36-MHz transponder bandwidth may be assigned to different users. Each user receives a specific bandwidth allocation whereby he or she can access the transponder.

Thus, composite FDM channels are FM modulated and transmitted to the satellite within the bandwidth allocation of an FDMA plan. The major advantage of FDMA (compared to TDMA) is its simplicity. The FDMA channels require no

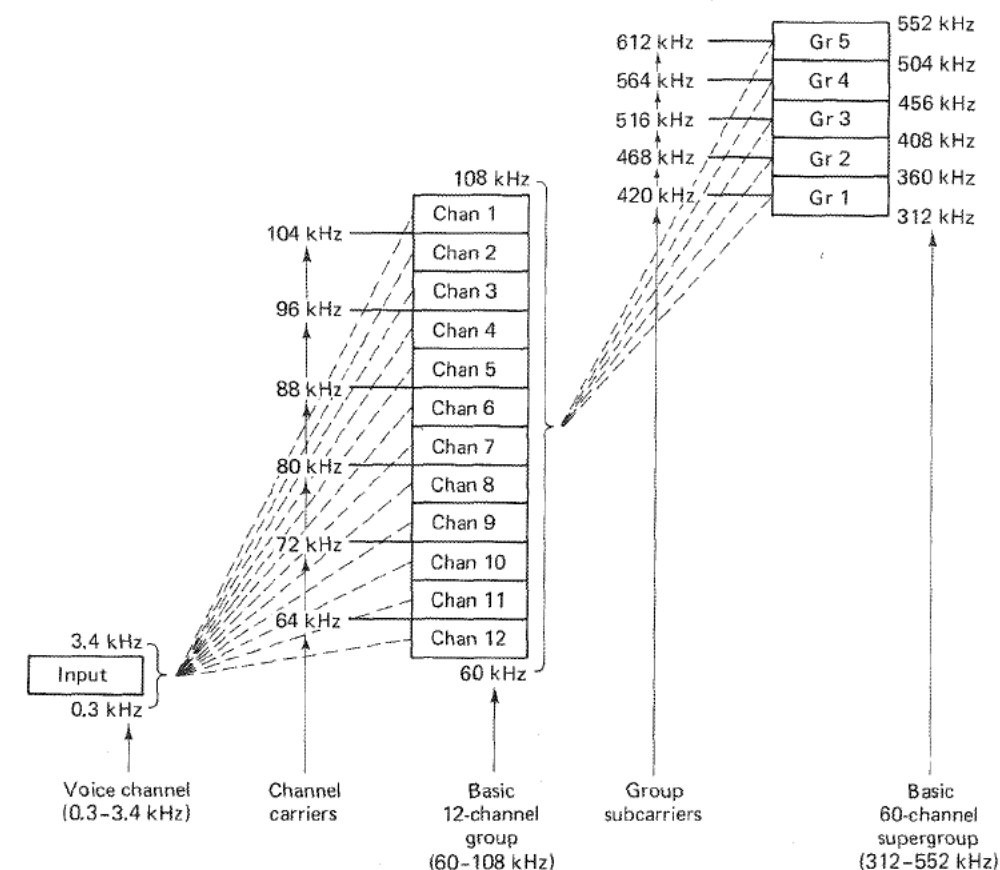


Figure 9.6 Modulation plan of a typical frequency-division multiplex system.

synchronization or central timing; each channel is almost independent of all other channels. Later we discuss some advantages of TDMA compared to FDMA.

9.1.2 Time-Division Multiplexing/Multiple Access

In Figure 9.3, sharing of the communications resource (CR) is accomplished by allocating frequency bands. In Figure 9.7, the same CR is shared by assigning each of M signals or users the full spectral occupancy of the system for a short duration of time called a *time slot*. The unused time regions between slot assignments, called *guard times*, allow for some time uncertainty between signals in adjacent time slots, and thus act as buffer zones to reduce interference. Figure 9.8 is an illustration of a typical TDMA satellite application. Time is segmented into intervals called frames. Each frame is further partitioned into assignable user time slots. The frame structure repeats, so that a fixed TDMA assignment constitutes one or more slots that periodically appear during each frame time. Each earth station transmits its data in bursts, timed so as to arrive at the satellite coincident with its designated time slot(s). When the bursts are received by the satellite transponder, they are retransmitted on the downlink, together with the bursts from other stations. A receiving station detects and demultiplexes the appropriate bursts and feeds the information to the intended user.

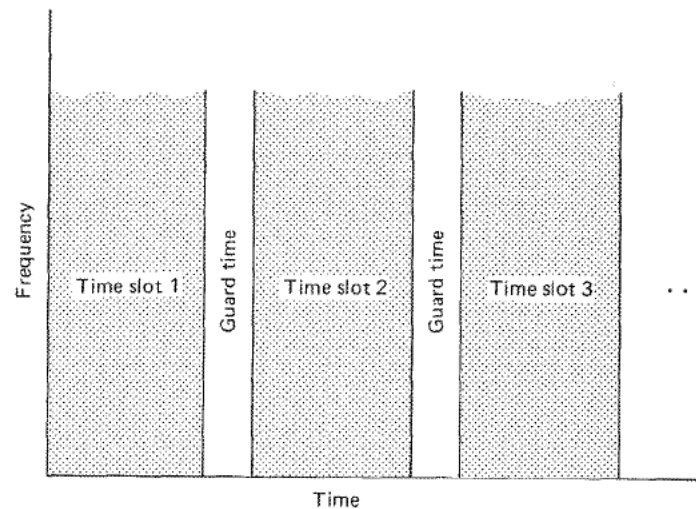


Figure 9.7 Time-division multiplexing.

9.1.2.1 Fixed-Assignment TDM/TDMA

The simplest TDM/TDMA scheme, called *fixed-assignment TDM/TDMA*, is so named because the M time slots that make up each frame are preassigned to signal sources, long term. Figure 9.9 illustrates, in block diagram form, the operation of such a system. The multiplexing operation consists of providing each

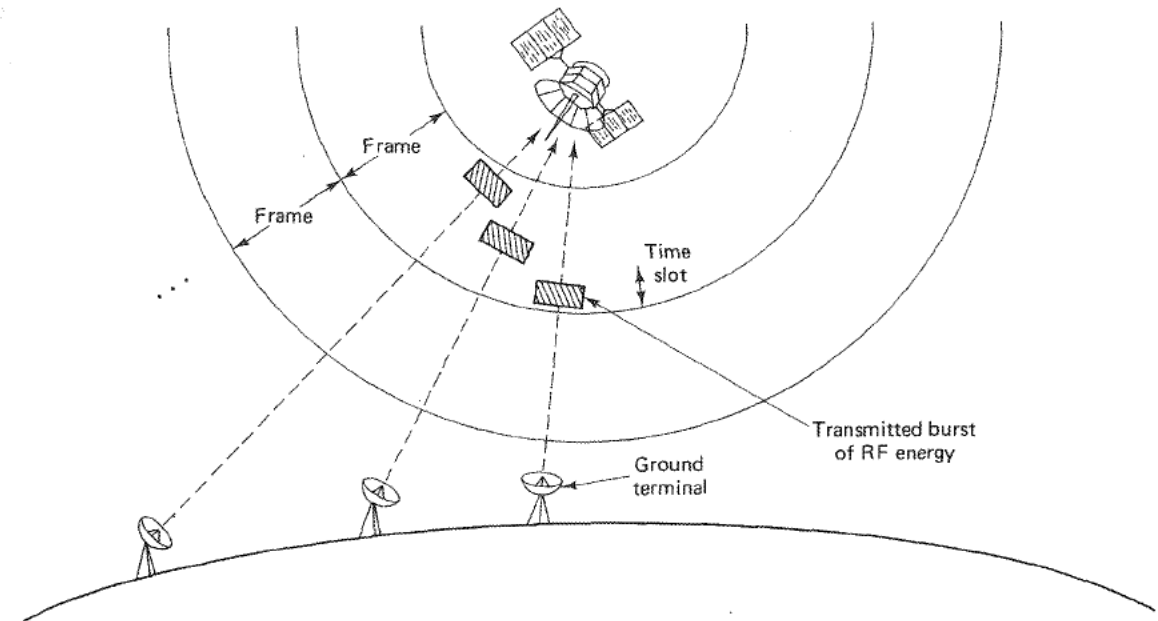


Figure 9.8 Typical TDMA configuration.

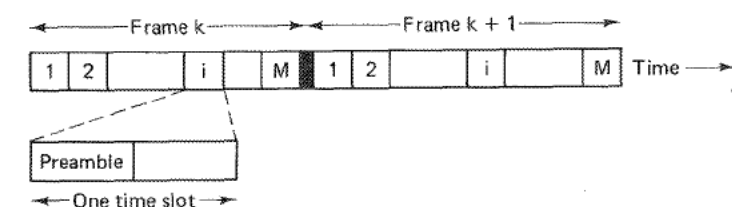
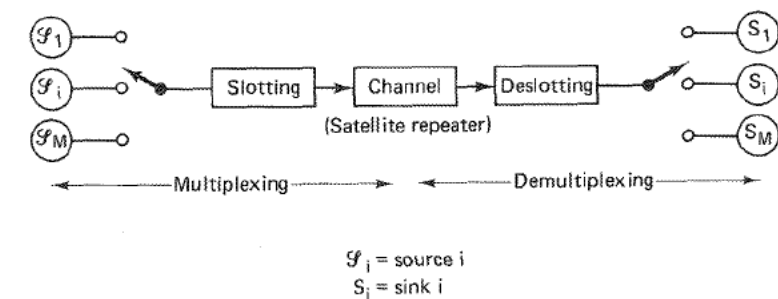


Figure 9.9 Fixed-assignment TDM.

source with an opportunity to occupy one or more slots. The demultiplexing operation consists of deslotting the information and delivering the data to the intended sink. The two commutating switches in Figure 9.9 have to be synchronized so that the message corresponding to source 1, for example, appears on the channel 1 output, and so on. The message itself is generally comprised of a preamble portion and a data portion. The preamble portion usually contains synchronization, addressing, and error control sequences.

A fixed-assignment TDM/TDMA scheme is extremely efficient when the source requirements are predictable, and the traffic is heavy (the time slots are most always filled). However, for bursty or sporadic traffic, the fixed-assignment scheme is wasteful. Consider the simple example shown in Figure 9.10. In this example there are four time slots per frame; each slot is preassigned to users A, B, C, and D, respectively. In Figure 9.10a we see a typical activity profile of the four users. During the first frame time, user C has no data to transmit; during the second frame time, user B has none, and during the third frame time, user A has none. In a fixed-assignment TDMA scheme, all of the slots within a frame are preassigned. If the "owner" of a slot has no data to send during a particular frame, that slot is wasted. The data stream, shown in Figure 9.10b, illustrates the wasted time slots in this example. When source requirements are unpredictable, as in this example, there can be more efficient schemes, involving the dynamic assignment of the slots rather than a fixed assignment. Such schemes are variously known as packet-switched systems, statistical multiplexers, or concentrators; the effect, shown in Figure 9.10c, is to use all the slots in a frame in such a way that capacity is conserved. In later sections we discuss the TDMA systems used in INTELSAT V and VI.

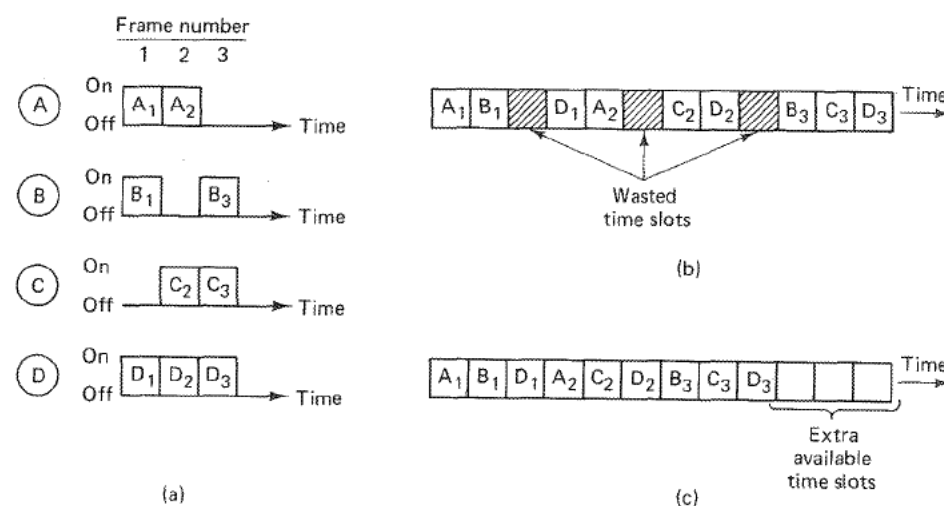


Figure 9.10 Fixed-assignment TDM versus packet switching. (a) Data source activity profiles. (b) Fixed-assignment time-division multiplexing. (c) Time-division packet switching (concentration).

9.1.3 Communications Resource Channelization

In Figure 9.3 we considered that the CR is partitioned into spectral bands, and in Figure 9.7 we viewed the same CR as being partitioned into time slots. Figure 9.11 represents a more general organization of the CR allowing for the assignment of a frequency band for a prescribed period of time. Such a multiple access scheme is referred to as *combined FDMA/TDMA*. For the assignments of frequency bands, let us assume an equal apportionment of the total bandwidth, W , among M user groups or classes, so that M disjoint frequency bands of width W/M hertz are continuously available to their assigned group. Similarly, for the assignment of time slots, the time axis is partitioned into time frames, each of duration T , and the frames are partitioned into N slot times, each of duration T/N . We assume that the users are time synchronized and that the assigned slots are located periodically within the frames. Each user in each frequency band is permitted to transmit during each periodic appearance of the user's assigned slot, and is permitted to use the assigned channel bandwidth for the slot duration. A slot is uniquely determined as the m th slot within the n th frame. Referring to Figure 9.11, we can describe the time of a particular slot (n, m) with reference to time zero as follows:

$$\text{time of slot } (n, m) = nT + \frac{(m-1)T}{N} \leq t \leq nT + \frac{mT}{N}$$

$$n = 0, 1, \dots; m = 1, 2, \dots, N. \quad (9.4)$$

The n th frame time, T , is denoted by the time interval $[nT, (n+1)T]$. As can be seen in Figure 9.11, the domain of the unit signal is the intersection of the time slot (n, m) and the frequency band (j) . Assume that a modulation/coding system is chosen so that the full bandwidth W of the CR can support R bits/s. In any

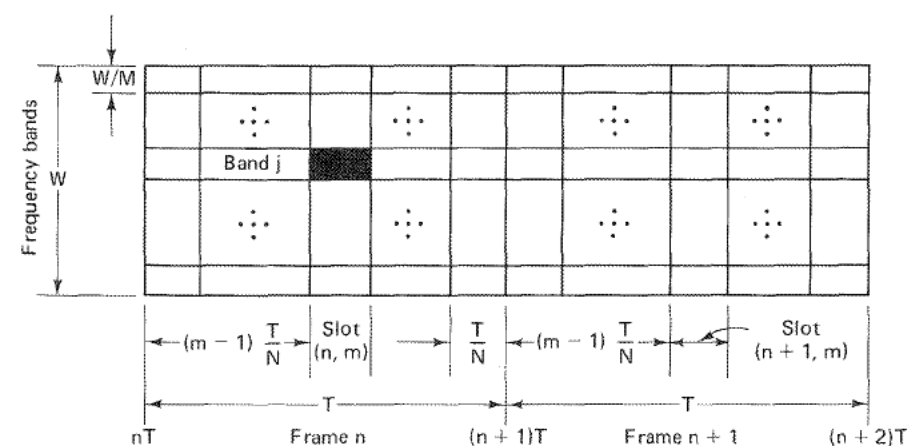


Figure 9.11 Communications resource: time/frequency channelization.

frequency band having a bandwidth of W/M hertz, the associated bit rate will be R/M bits/s. FDMA alone would provide M bands each with a bandwidth of $1/M$ of the full bandwidth of the CR. TDMA alone would provide the full system bandwidth for each of the N slots, where the duration of each slot is $1/N$ of the frame time.

9.1.4 Performance Comparison of FDMA and TDMA

9.1.4.1 Bit Rate Equivalence of FDMA and TDMA

Figure 9.12 highlights the basic differences between an FDMA and TDMA system in a communications resource capable of supporting a total of R bits/s. In Figure 9.12a the system bandwidth is divided into M orthogonal frequency bands. Hence each of the M sources, $\mathcal{S}_m (1 \leq m \leq M)$ can simultaneously transmit at a bit rate of R/M bits/s. In Figure 9.12b the frame is divided into M orthogonal time slots. Hence each of the M sources bursts its transmission at R bits/s, M times faster than the equivalent FDMA user for $(1/M)$ th the time. In both cases, the source $\mathcal{S}_m$ transmits information at an average rate of R/M bits/s.

Let the information generated by each of the sources in Figure 9.12 be organized into b -bit groups, or *packets*. In the case of FDMA, the b -bit packets are transmitted in T seconds over each of the M disjoint channels. Therefore, the

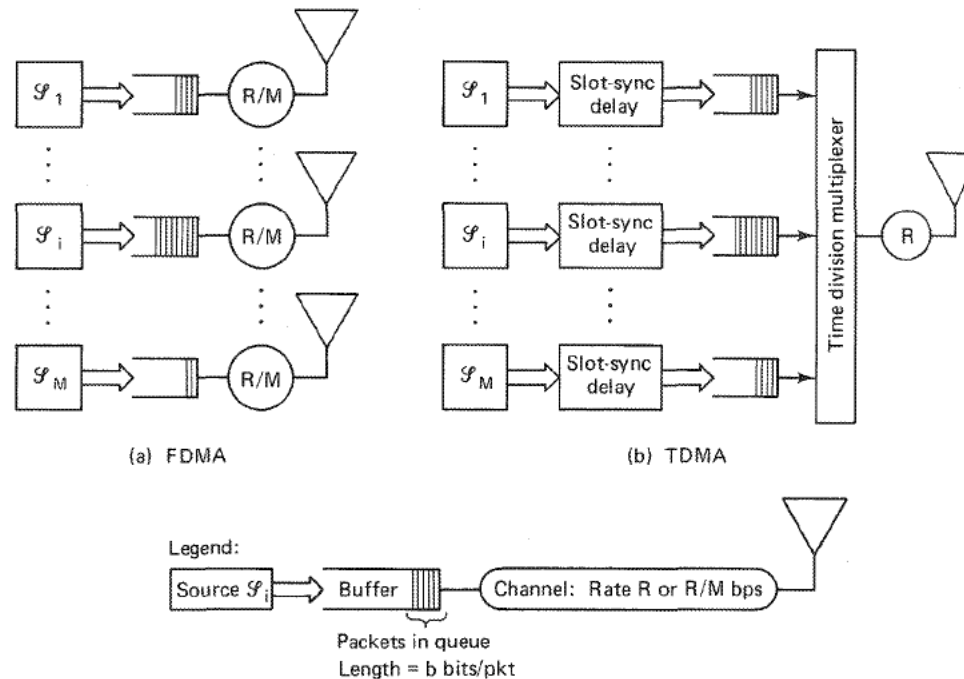


Figure 9.12 (a) FDMA: frequency divided into M orthogonal frequency bands. (b) TDMA: time divided into M orthogonal time slots (one packet per time slot).

total bit rate, R_{FD} , required is

$$R_{FD} = M \frac{b}{T} \quad \text{bits/s} \quad (9.5)$$

In the case of TDMA, the b bits are transmitted in T/M seconds from each source. Therefore, the bit rate, R_{TD} , required is

$$R_{TD} = \frac{b}{T/M} \quad \text{bits/s} \quad (9.6)$$

Since Equations (9.5) and (9.6) yield identical results, we can conclude that

$$R_{FD} = R_{TD} = R = \frac{Mb}{T} \quad \text{bits/s} \quad (9.7)$$

Thus both systems require the same full CR data rate, R bits/s.

9.1.4.2 Message Delays in FDMA and TDMA

From the previous sections it might appear that the duality between FDMA and TDMA will result in equivalent performance. This is not the case when the metric of performance is the average packet *delay*. It can be shown [1, 2] that TDMA is inherently superior to FDMA in the sense that the average packet delay using TDMA is less than the delay using FDMA.

As before, we assume that in the case of FDMA the system bandwidth is divided into M orthogonal frequency bands, and in the case of TDMA the frame is divided into M orthogonal time slots. For the analysis of message delay, the simplest case is that of deterministic data sources. It is assumed that the CR is 100% utilized, so that all frequency bands in the case of FDMA, and all time slots in the case of TDMA, are filled with data packets. For simplicity, it is also assumed that there are *no* overhead costs such as guard bands or guard times.

The message delay, D , can be defined as

$$D = w + \tau \quad (9.8)$$

where w is the average packet waiting time (prior to transmission) and τ is the packet transmission time. In the FDMA case, each packet is sent over a T -second interval, so the packet transmission time for FDMA, τ_{FD} , is simply

$$\tau_{FD} = T \quad (9.9)$$

In the TDMA case, each packet is sent in slots of T/M seconds. We can thus write the TDMA packet transmission time, τ_{TD} , with the use of Equation (9.7), as

$$\tau_{TD} = \frac{T}{M} = \frac{b}{R} \quad (9.10)$$

Since the FDMA channel is continuously available and packets are sent as soon as they are generated, the waiting time, w_{FD} , for FDMA is

$$w_{FD} = 0 \quad (9.11)$$

FDMA and TDMA bit streams are compared in Figure 9.13. For TDMA, Figure 9.13a illustrates that each user's slot begins at a different point in the T -second frame; that is, packet S_{mk} will start at $(m-1)T/M$ seconds ($1 \leq m \leq M$) after the packet generation instant. Therefore, the average waiting time, w_{TD} , that a TDMA packet sustains before transmission begins is

$$w_{TD} = \frac{1}{M} \sum_{m=1}^M (m-1) \frac{T}{M} = \frac{T}{M^2} \sum_{n=0}^{M-1} n = \frac{T}{M^2} \frac{(M-1)(M)}{2} = \frac{T}{2} \left(1 - \frac{1}{M}\right) \quad (9.12)$$

The maximum waiting time before transmission of a packet is $(M-1)T/M$ seconds, and on the average a packet will wait $\frac{1}{2}(M-1)(T/M) = (T/2)(1 - 1/M)$ seconds, as given by Equation (9.12).

To compare the average delay times, D_{FD} and D_{TD} , for FDMA and TDMA, respectively, we combine Equations (9.9) and (9.11) into Equation (9.8), and sim-

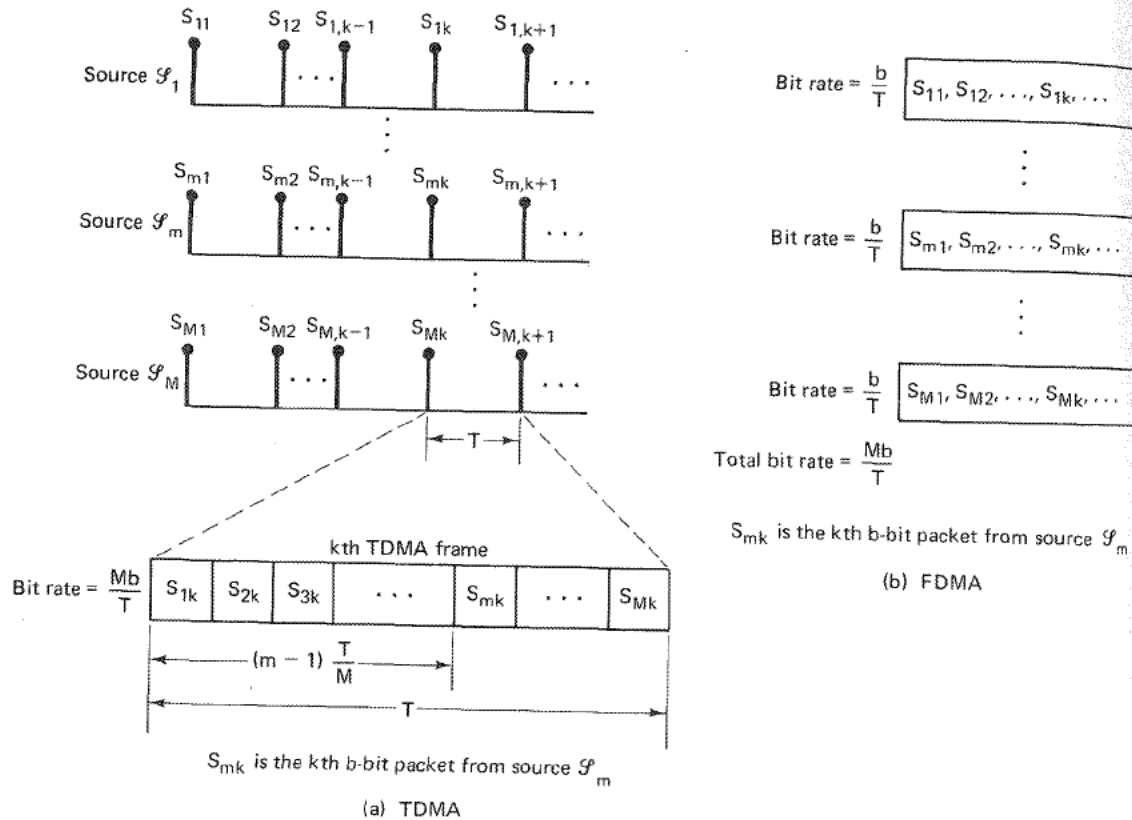


Figure 9.13 (a) TDMA and (b) FDMA channelization.

ilarly combine Equations (9.10) and (9.12) into Equation (9.8), yielding

$$D_{FD} = T \quad (9.13)$$

$$D_{TD} = \frac{T}{2} \left(1 - \frac{1}{M}\right) + \frac{T}{M} = D_{FD} - \frac{T}{2} \left(1 - \frac{1}{M}\right) \quad (9.14)$$

Using Equation (9.7), Equation (9.14) can be written as

$$D_{TD} = D_{FD} - \frac{b}{2R} (M-1) \quad (9.15)$$

The result indicates that TDMA is inherently superior to FDMA, from a message delay point of view. Although Equation (9.15) assumed that the data source is deterministic, the smaller average message delays for TDMA schemes hold up for any independent message arrival process [1, 2].

9.1.5 Code-Division Multiple Access

In Figure 9.3 the CR plane was illustrated as being shared by slicing it horizontally to form FDMA frequency bands, and in Figure 9.7 the same CR plane was illustrated as being shared by slicing it vertically to form TDMA time slots. These two techniques are the most common choices for multiple access applications. Figure 9.14 illustrates the CR being partitioned by the use of a hybrid combination of FDMA and TDMA known as *code-division multiple access* (CDMA). CDMA

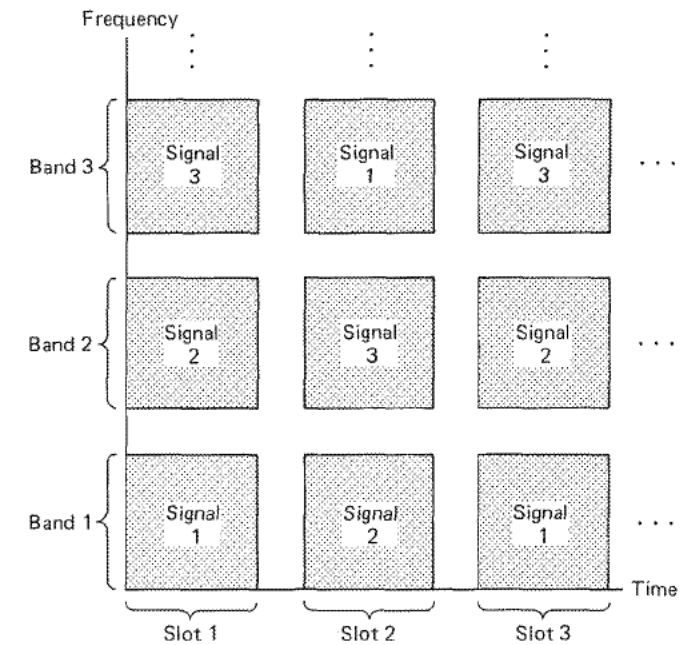


Figure 9.14 Code-division multiplexing.

is an application of spread-spectrum (SS) techniques. Spread-spectrum techniques can be classified into two major categories: *direct-sequence SS* and *frequency hopping SS*. We introduce frequency hopping CDMA (FH-CDMA) in this chapter, and we treat direct-sequence CDMA together with the overall subject of spread-spectrum techniques in Chapter 10.

It is easiest to visualize *frequency hopping* CDMA, illustrated in Figure 9.14, as the short-term assignment of a frequency band to various signal sources. At each successive time slot, whose duration is usually brief, the frequency band assignments are reordered. In Figure 9.14, during time slot 1, signal 1 occupies band 1, signal 2 occupies band 2, and signal 3 occupies band 3. During time slot 2, signal 1 hops to band 3, signal 2 hops to band 1, and signal 3 hops to band 2, and so on. The CR can thus be fully utilized, but the participants, having their frequency bands reassigned at each time slot, appear to be playing “musical chairs.” Each user employs a pseudonoise (PN) code, orthogonal (or nearly orthogonal) to all the other user codes, that dictates the frequency hopping band assignments. Details of PN code sequences are treated in Section 10.2. Figure 9.14 is an oversimplified view of the way the CR is shared in frequency hopping CDMA, since the symmetry implies that each frequency hopping signal is in time synchronism with each of the other signals. This is *not the case*. In fact, one of the attractions of CDMA compared to TDMA is that there is no need for synchronization among user groups (only between a transmitter and a receiver within a group).

The block diagram in Figure 9.15 illustrates the frequency hopping modulation process. At each frequency hop time the PN generator feeds a code sequence to a device called a *frequency hopper*. The frequency hopper synthesizes one of the allowable hop frequencies. Assume that the data modulation has an M -ary frequency shift keying (MFSK) format. The essential difference between a conventional MFSK system and a frequency hopping (FH) MFSK system is that in the conventional system, a data symbol modulates a carrier wave that is *fixed* in frequency, but in the hopping system, the data symbol modulates a carrier

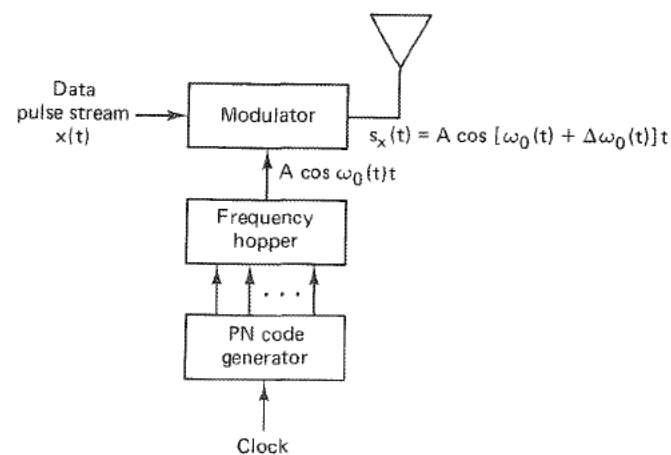


Figure 9.15 CDMA frequency hopping modulation process.

wave that *hops* across the total CR bandwidth. The FH modulation in Figure 9.15 can be thought of as a two-step process—data modulation and frequency hopping modulation—even though it can be implemented in a single step, where the modulator produces a transmission tone based on the simultaneous dictates of the PN code and the data. Frequency hopping systems are covered in detail in Section 10.4.

One might ask: Don't the FDMA and TDMA options provide sufficient multiple access flexibility? FDMA and TDMA methods can surely be relied on to apportion the communications resource equitably. Of what use is this hybrid technique? CDMA offers some unique advantages, as follows:

1. **Privacy.** When the code for a particular user group is only distributed among authorized users, the CDMA process provides communications privacy, since the transmissions cannot easily be intercepted by unauthorized users without the code.
2. **Fading channels.** If a particular portion of the spectrum is characterized by fading, signals in that frequency range are attenuated. In an FDMA scheme, a user who was unfortunate enough to be assigned to the fading position of the spectrum might experience highly degraded communications for as long as the fading persists. However, in a FH-CDMA scheme, only during the time a user hops into the affected portion of the spectrum will the user experience degradation. Therefore, with CDMA, such degradation is shared among all the users.
3. **Jam resistance.** During a given CDMA hop, the signal bandwidth is identical to the bandwidth of conventional MFSK, which is typically equal to the minimum bandwidth necessary to transmit the MFSK symbol. However, over a duration of many time slots, the system will hop over a frequency band which is much wider than the data bandwidth. We refer to this utilization of bandwidth as spread spectrum. In Chapter 10 we develop, in detail, the resistance to jamming that spread spectrum affords a user.
4. **Flexibility.** The most important advantage of CDMA schemes, compared to TDMA, is that there need be no precise time coordination among the various simultaneous transmitters. The orthogonality between user transmissions on different codes is not affected by transmission-time variations. This will become clear upon closer examination of the autocorrelation and cross-correlation properties of the codes, considered in Chapter 10.

9.1.6 Space-Division and Polarization-Division Multiple Access

Figure 9.16a depicts the INTELSAT IVA application of space-division multiple access (SDMA), also called *multiple-beam frequency reuse*. INTELSAT IVA used a dual-beam receive antenna feeding two receivers to allow simultaneous access of the satellite from two different regions of the earth. The frequency band

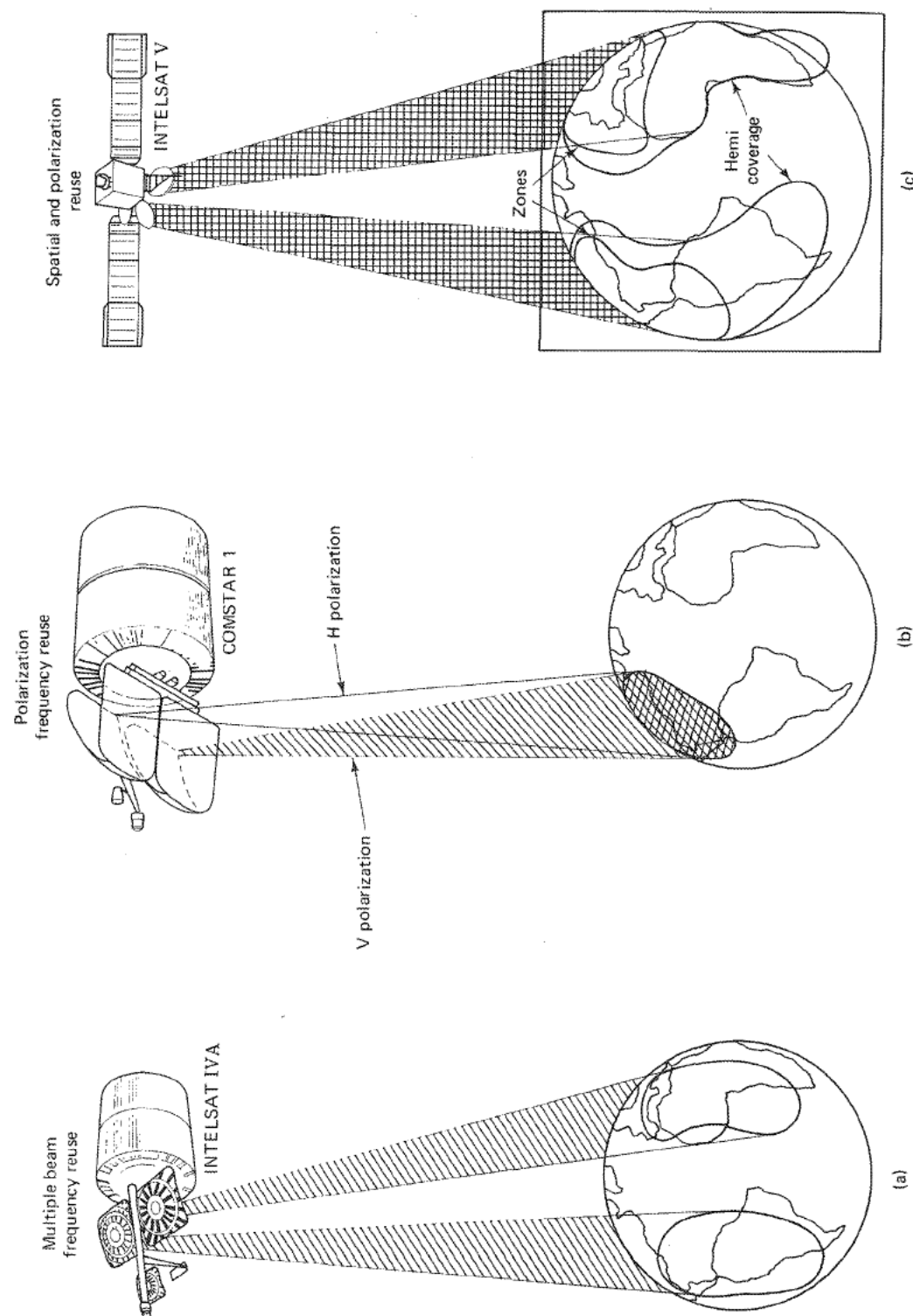


Figure 9.16 SDMA and PDMA. (a) INTELSAT IVA. (b) COMSTAR 1. (c) INTELSAT V (Atlantic coverage).

allocated to each receive beam was identical because the uplink signals were spatially separated. In such cases, the frequency band is said to be *reused*.

Figure 9.16b depicts an application of polarization-division multiple access (PDMA), also called *dual-polarization frequency reuse*, from COMSTAR 1. Here separate antennas are used, each with different polarization and followed by separate receivers, allowing simultaneous access of the satellite from the same region of the earth. Each corresponding earth station antenna needs to be polarized in the same way as its counterpart in the satellite. (This is generally accomplished by providing each participating earth station with an antenna that has dual polarization.) The frequency band allocated to each antenna beam can be identical because the uplink signals are orthogonal in polarization. As with SDMA, the frequency band in PDMA is said to be reused. Figure 9.16c depicts an application of the simultaneous use of SDMA and PDMA in INTELSAT V. There are two separate hemispheric coverages, west and east. There are also two smaller zone beams; each zone beam overlaps a portion of one of the hemispheric beams and is separated from it by orthogonal polarization. Thus there is a fourfold reuse of the spectrum.

9.2 MULTIPLE ACCESS COMMUNICATIONS SYSTEM AND ARCHITECTURE

A *multiple access protocol* or *multiple access algorithm* (MAA) is that rule by which a user knows how to use time, frequency, and code functions to communicate through a satellite to other users. A multiple access system is a com-

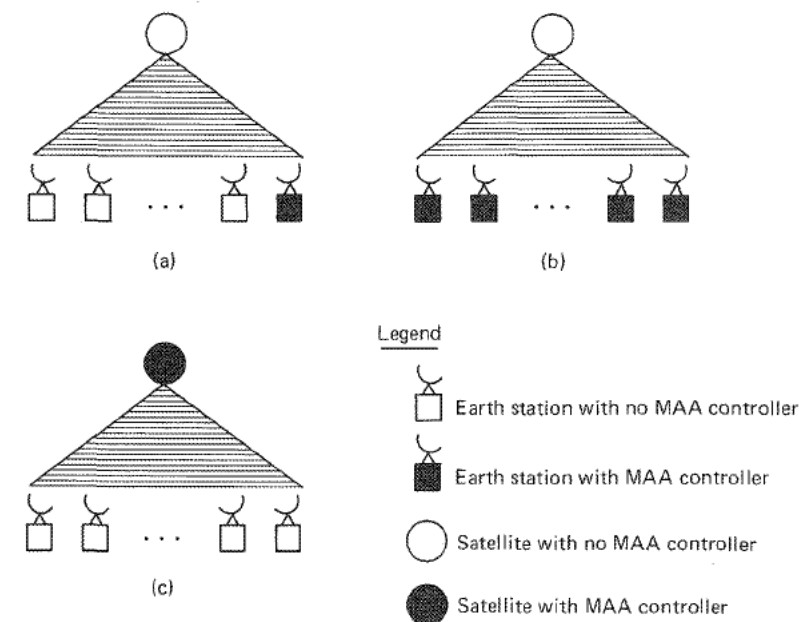


Figure 9.17 Satellite multiple access architecture. (a) Single earth station control. (b) Distributed earth station control. (c) Satellite control.

bination of hardware and software that supports the MAA. The general goal of a multiple access system is to provide communications service in a timely, orderly, and efficient way.

Figure 9.17 illustrates some basic choices for the architecture of a satellite multiple access system. The legend indicates the symbols used for an earth station with and without an MAA controller, and a satellite with and without an MAA controller. Figure 9.17a illustrates the case where one earth station is designated as the master, or the controller. This earth station possesses an MAA computer and responds to the service requests of all other users. Notice that a user's request entails a transmission through the satellite and back down to the controller. The controller's response entails another transmission through the satellite; hence there are two up- and downlink transmissions required for each service assignment. Figure 9.17b illustrates the case where the MAA control is distributed among all the earth stations; there is no single controller. Each earth station uses the same algorithm and they each have identical knowledge regarding access requests and assignments; therefore, only one round trip is required for each service assignment. Figure 9.17c illustrates the case where the MAA controller is in the satellite. A service request goes from user to satellite, and the response from the satellite can follow immediately; therefore, only one round trip is required for each service assignment.

9.2.1 Multiple Access Information Flow

Figure 9.18 is a flow diagram describing the basic flow of information between the multiple access algorithm (MAA) or controller and an earth station; the numbers below correspond to those on the figure. Recall from the preceding section that the control may be lodged in the satellite, in a master station, or distributed among all the earth stations.

1. *Channelization*. This term refers to the most general allocation information [e.g., channels 1 to N may be allocated for the Army and channels $(N + 1)$

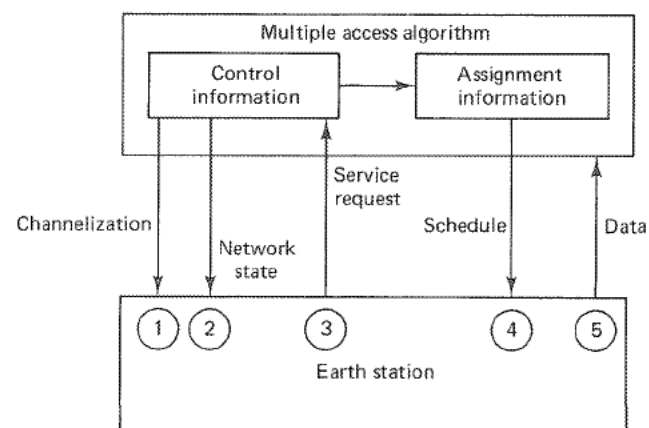


Figure 9.18 Multiple access information flow.

to M for the Navy]. This information seldom changes, and may be distributed to the earth stations by the use of a newsletter rather than via the communication system.

2. *Network state (NS)*. This term refers to the state of the CR. A station is advised regarding the availability of the communications resource and where in the resource (e.g., time, frequency, code position) to transmit its service request(s).
3. *Service request*. Then the station makes its request(s) for service (e.g., allocation for m message slots).
4. Upon receipt of the service request(s), the controller sends the station a schedule regarding where and when to position its data in the CR.
5. The station transmits its data according to its assigned schedule.

9.2.2 Demand-Assignment Multiple Access

Multiple access schemes are termed *fixed assignment* when a station has periodic access to the channel independent of its actual need. By comparison, dynamic assignment schemes, sometimes called *demand-assignment multiple access* (DAMA), give the station access to the channel only when it requests access. If the traffic from a station tends to be burst-like or intermittent, DAMA procedures can be much more efficient than fixed-assignment procedures. A DAMA scheme capitalizes on the fact that actual demand *rarely* equals the peak demand. If a system's capacity is equal to the total peak demand and if the traffic is bursty, the system will be underutilized most of the time. However, by using buffers and DAMA, a system with reduced average capacity can handle bursty traffic, at the cost of some queueing delay. Figure 9.19 summarizes the difference between a fixed system, whose capacity is equal to the sum of the user requirements, and a dynamic system, whose capacity is equal to the average of the user requirements.

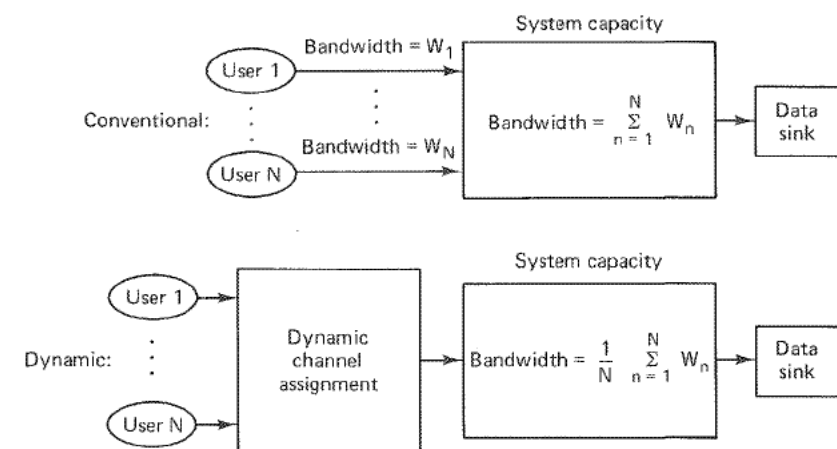


Figure 9.19 Bandwidth reduction for systems using dynamic channel assignment.

9.3 ACCESS ALGORITHMS

9.3.1 ALOHA

In 1971, the University of Hawaii began operation of its ALOHA system. A communication satellite was used to interconnect the several university computers by use of a random access protocol [3-7]. The system concept was extremely simple, consisting of the following modes:

1. *Transmission mode.* Users transmit at any time they desire, encoding their transmissions with an error detection code.
2. *Listening mode.* After a message transmission, a user listens for an acknowledgment (ACK) from the receiver. Transmissions from different users will sometimes overlap in time, causing reception errors in the data in each of the contending messages. We say that the messages have *collided*. In such cases, the errors are detected, and the users receive a negative acknowledgment (NAK).
3. *Retransmission mode.* When a NAK is received, the messages are simply retransmitted. Of course, if the colliding users were to retransmit immediately, they would collide again. Therefore, the users retransmit after a random delay.
4. *Timeout mode.* If, after a transmission, the user does not receive either an ACK or NAK within a specified time, the user retransmits the message.

9.3.1.1 Message Arrival Statistics

Assume that the total system demand requires an average message or packet arrival rate of λ successful or accepted messages per second. Because of the presence of collisions, some of the messages will be unsuccessful or rejected. Therefore, we define the total traffic arrival rate, λ_t , as the acceptance rate, λ , plus the rejection rate, λ_r , as follows:

$$\lambda_t = \lambda + \lambda_r \quad (9.16)$$

Let us denote the length of each message or packet as b bits. Then we can define the average amount of successful traffic or *throughput*, ρ , on the channel in units of bits per second, as

$$\rho = b\lambda \quad (9.17)$$

We can also define the *total traffic*, G , on the channel in units of bits per second, as

$$G = b\lambda_t \quad (9.18)$$

With the channel capacity (maximum bit rate) designated as R bits per second, let us further define a *normalized throughput*, ρ , and a *normalized total traffic*, G , as

$$\rho = \frac{b\lambda}{R} \quad (9.19)$$

$$G = \frac{b\lambda_t}{R} \quad (9.20)$$

Normalized throughput, ρ , expresses throughput as a fraction ($0 \leq \rho \leq 1$) of channel capacity. Normalized total traffic, G , expresses total traffic as a fraction ($0 \leq G \leq \infty$) of the channel capacity. Notice that G can take on values greater than unity.

We can also define the transmission time of each packet as follows:

$$\tau = \frac{b}{R} \quad \text{seconds/packet} \quad (9.21)$$

By substituting equation (9.21) into Equations (9.19) and (9.20), we can write

$$\rho = \lambda\tau \quad (9.22)$$

$$G = \lambda_t\tau \quad (9.23)$$

A user can successfully transmit a message as long as no other user began one within the previous τ seconds or starts one within the next τ seconds. If another user began a message within the previous τ seconds, its tail end will collide with the current message. If another user begins a message within the next τ seconds, it will collide with the tail end of the current message. Thus a space of 2τ seconds is needed for each message.

The message arrival statistics for unrelated users of a communication system is often modeled as a Poisson process. The probability of having K new messages arrive during a time interval of τ seconds is given by the Poisson distribution [8] to be

$$P(K) = \frac{(\lambda\tau)^K e^{-\lambda\tau}}{K!} \quad K \geq 0 \quad (9.24)$$

where λ is the average message arrival rate. Because the users transmit without regard for each other in the ALOHA system, this expression is useful for calculating the probability that exactly $K = 0$ other messages are transmitted during a time interval 2τ . This is the probability, P_s , that a user's message transmission was successful (experienced no collisions). To compute P_s , assuming that all traffic is Poisson, we use λ_t and 2τ in Equation (9.24). Thus

$$P_s = P(K = 0) = \frac{(2\tau\lambda_t)^0 e^{-2\tau\lambda_t}}{0!} = e^{-2\tau\lambda_t} \quad (9.25)$$

In Equation (9.16) we defined total traffic arrival rate λ_t , in terms of the successful portion, λ , and the repetition or unsuccessful portion, λ_r ; then, by definition, the probability of a successful packet can be expressed as

$$P_s = \frac{\lambda}{\lambda_t} \quad (9.26)$$

By combining Equations (9.25) and (9.26), we have

$$\lambda = \lambda_t e^{-2\tau\lambda_t} \quad (9.27)$$

By combining Equation (9.27) with Equations (9.22) and (9.23), we can write

$$\rho = Ge^{-2G} \quad (9.28)$$

Equation (9.28) relates the normalized throughput, ρ , to the normalized total traffic, G , on the channel for the ALOHA system. A plot of this relationship labeled "pure ALOHA" is shown in Figure 9.20. As G increases, ρ increases until a point is reached where further traffic increases create a large enough collision rate to cause a reduction in the throughput. The maximum ρ , equal to $1/2e = 0.18$, occurs at a value of $G = 0.5$. Therefore, for a pure ALOHA channel, only 18% of the CR can be utilized. Simplicity of control is achieved at the expense of channel capacity [7, 9].

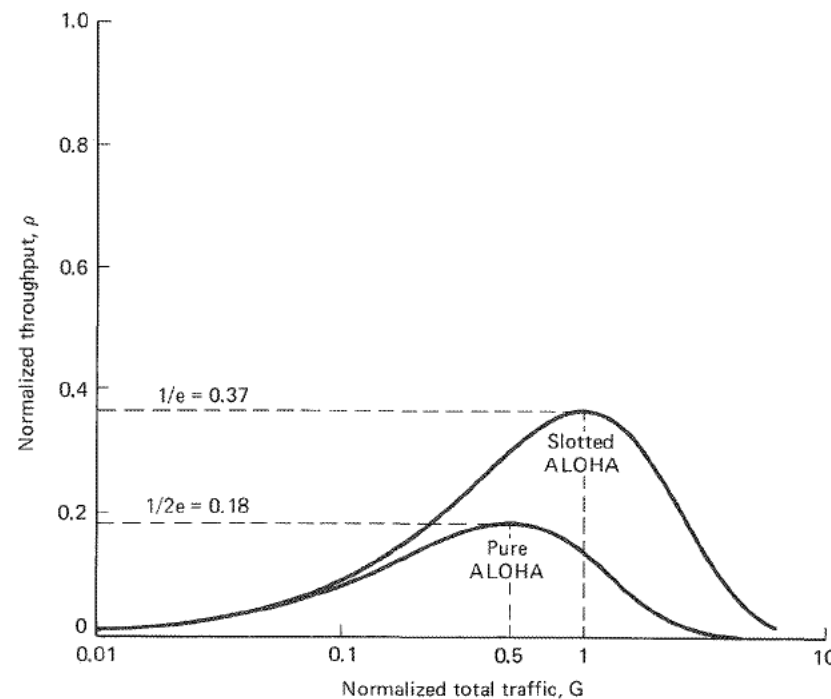


Figure 9.20 Throughput in ALOHA channels (successful transmissions versus total transmissions).

9.3.2 Slotted ALOHA

The pure ALOHA scheme can be improved by requiring a small amount of coordination among the stations. The slotted ALOHA (S-ALOHA) is such a system. A sequence of synchronization pulses is broadcast to all stations. As with pure

ALOHA, packet lengths are constant. Messages are required to be sent in the slot time between synchronization pulses, and can be started only at the *beginning* of a time slot. This simple change reduces the rate of collisions by half, since only messages transmitted in the same slot can interfere with one another. It can be shown [9, 10] that for S-ALOHA, the reduction in the *collision window* from 2τ to τ results in the following relationship between normalized throughput, ρ , and normalized total traffic, G .

$$\rho = Ge^{-G} \quad (9.29)$$

The plot of Equation (9.29) is shown in Figure 9.20 labeled "slotted ALOHA." Here the maximum value of ρ is $1/e = 0.37$, or an improvement of two times the pure ALOHA protocol.

The retransmission mode described for the pure ALOHA system was modified for S-ALOHA so that if a negative acknowledgment (NAK) occurs, the user retransmits after a *random* delay of an integer number of slot times. Figure 9.21 illustrates the S-ALOHA operation. A packet of data bits is shown transmitted by user k followed by the satellite acknowledgment (ACK). Also shown are users m and n simultaneously transmitting packets, which results in a collision; a NAK is returned. Each using station employs a random-number generator to select its retransmission time. The figure illustrates an example of the m and n retransmission at their respective randomly selected times. Of course, there is some probability that users m and n will recollide. However, in that case, they simply repeat the retransmission, using another random delay.

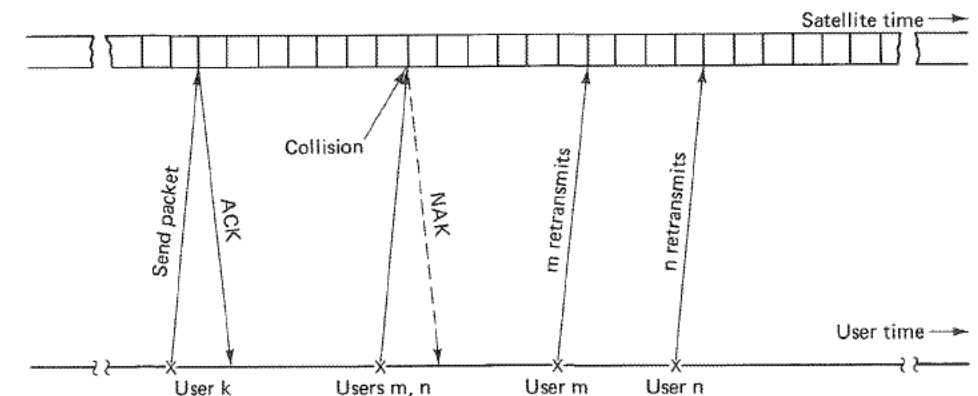


Figure 9.21 Random access scheme: slotted ALOHA operation.

Example 9.1 Poisson Process

Assuming that packet transmissions and retransmission can both be described as a Poisson process, calculate the *probability* that a data packet transmission in an S-ALOHA system will experience a collision with *one other user*. Assume that the total traffic rate $\lambda_t = 10$ packets/s and the packet duration $\tau = 10$ ms.

Solution

$$\begin{aligned}
 P(K = 1) &= \frac{(\tau\lambda_t)^K e^{-\tau\lambda_t}}{K!} \bigg|_{K=1} \\
 &= (10 \times 0.01)^1 e^{-0.1} = 0.1e^{-0.1} \\
 &= 0.09
 \end{aligned}$$

9.3.3 Reservation-ALOHA

A significant improvement was made to the ALOHA system with the introduction of the reservation-ALOHA (R-ALOHA) [11] scheme. The R-ALOHA system has two basic modes: an unreserved mode and a reserved mode; each is described below.

Unreserved Mode (Quiescent State)

1. A time frame is established and divided into a number of small reservation subslots.
2. Users use these small subslots to reserve message slots.
3. After requesting a reservation, the user listens for an acknowledgment and a slot assignment.

Reserved Mode

1. The time frame is divided into $M + 1$ slots whenever a reservation is made.
2. The first M slots are used for message transmissions.
3. The last slot is subdivided into subslots to be used for reservation/requests.
4. Users send message packets only in their assigned portions of the M slots.

Consider the R-ALOHA example shown in Figure 9.22. In the quiescent state, with no reservations, time is partitioned into short subslots for making reservations. Once a reservation is made, the system is configured so that $M = 5$ message slots followed by $V = 6$ reservation subslots becomes the timing format. The figure illustrates a request and an acknowledgment in progress. In this example the station seeks to reserve three message slots. The reservation acknowledgment advises the using station where to locate its first data packet. Since the control is distributed so that all participants receive the downlink transmissions and are thus aware of the reservations and time format, the acknowledgment need not disclose any more than the location of the first slot. As shown in Figure 9.22, the station sends its second packet in the slot following the first packet. The user further knows that the next slot is comprised of six subslots for reservations, so no packets are transmitted during this time. The third and final packet is sent in the following slot. When there are no reservations taking place, the system reverts back to its quiescent format of subslots only. Since the control is distributed, all the participants are made aware of the quiescent format by receiving appropriate

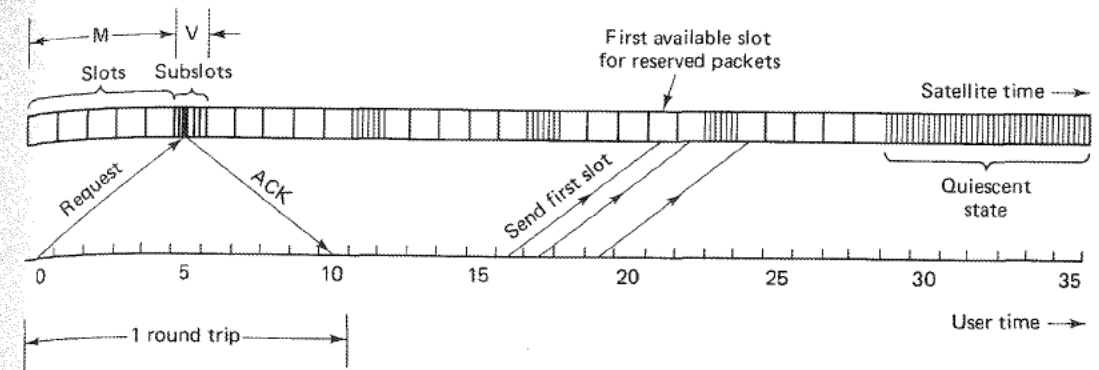


Figure 9.22 Example of reservation ALOHA. Station seeks to reserve three slots ($M = 5$ slots, $V = 6$ subslots).

synchronizing pulses on the downlink. Other interesting reservation schemes are discussed in References [12, 13].

9.3.4 Performance Comparison of S-ALOHA and R-ALOHA

From Chapter 3 the basic quality measure of a digital modulation scheme is its P_B versus E_b/N_0 curve. This measure is particularly useful because E_b/N_0 is a *normalized signal-to-noise ratio*; being normalized, the curves allow us to compare the performance of various modulation schemes. There is a similar performance measure for multiple access schemes. Here we are interested in the average delay versus normalized throughput. What would an *ideal delay-throughput curve* look like? Figure 9.23 illustrates such a curve. For normalized throughput values

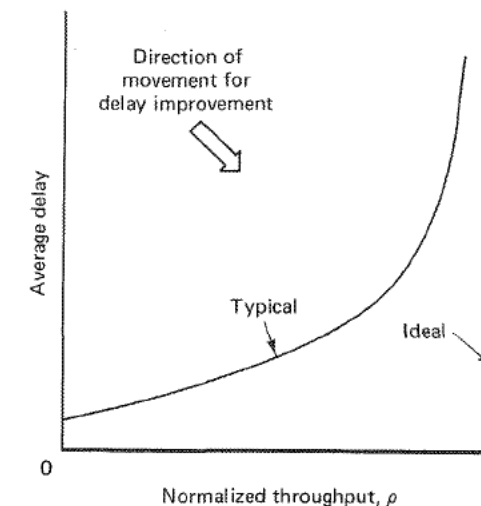


Figure 9.23 Delay-throughput characteristic.

of, $0 \leq \rho < 1$, the delay equals zero until $\rho = 1$; then the delay increases without bound. Figure 9.23 also shows a *typical* delay-throughput curve and the direction in which the curve will move as delay performance improves.

Figure 9.24 compares the delay-throughput performance of S-ALOHA with that of R-ALOHA (formatted with two message slots and six reservation subslots). Knowing the location of the *ideal* curve it is easy to compare the delay performance of these two systems. For a throughput of less than approximately 0.20, the S-ALOHA manifests less average delay than does R-ALOHA. But for values of ρ between 0.20 and 0.67, it is apparent that R-ALOHA is superior, since the average delay is less. Why does the S-ALOHA perform better at low traffic intensity? The S-ALOHA algorithm does not require the overhead of the reservation subslots as does R-ALOHA. Therefore, at low values of ρ , R-ALOHA pays the price of greater delay due to the greater overhead. For $\rho > 0.2$, the collisions and retransmissions inherent in the S-ALOHA system cause it to incur greater delay (unbounded at $\rho = 0.37$), more quickly than the R-ALOHA system. At higher throughput ($0.2 < \rho < 0.67$), the overhead structure of R-ALOHA ensures that its delay degradation grows in a more orderly manner than S-ALOHA. For R-ALOHA, an unbounded delay is not reached until $\rho = 0.67$.

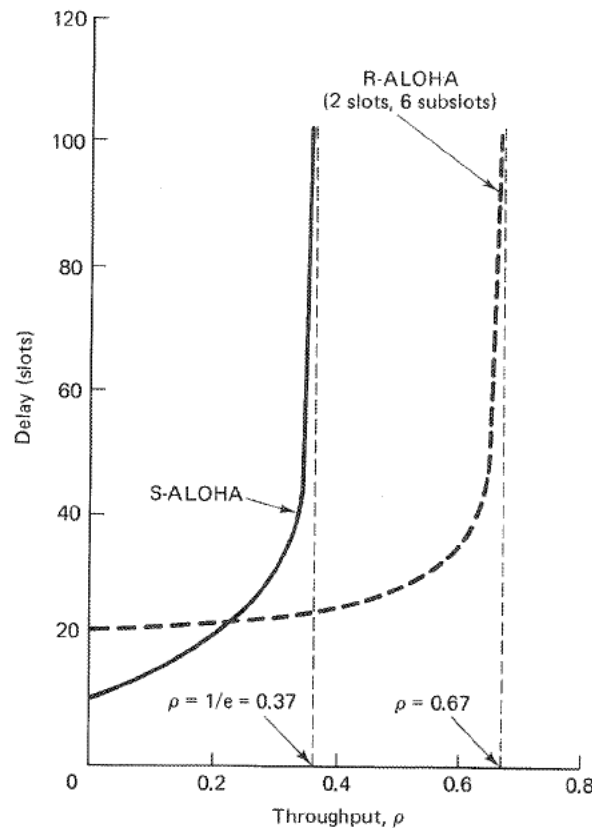


Figure 9.24 Delay-throughput comparison: S-ALOHA versus R-ALOHA on a satellite channel.

Example 9.2 Channel Utilization

- Normalized throughput, ρ , is a measure of channel utilization. It can be found by forming the ratio of the successfully transmitted message traffic, in bits per second to the total message traffic, including rejected messages, in bits per second. Calculate the normalized throughput of a channel that has a maximum data rate $R = 50$ kbits/s and operates with $M = 10$ ground stations, each station transmitting at the average rate of $\lambda = 2$ packets/second. The system format provides for $b = 1350$ bits/packet.
- Which of the three ALOHA schemes discussed—pure, slotted, and reservation—could be successfully used with this channel?

Solution

- Generalizing Equation (9.19) to allow for traffic from multiple stations, we have

$$\begin{aligned}\rho &= \frac{Mb\lambda}{R} \\ &= \frac{10(1350)(2)}{50,000} \\ &= 0.54\end{aligned}$$

- Only the R-ALOHA scheme could be used for this system, since with each of the other schemes, 54% of the resource cannot be utilized.

9.3.5 Polling Techniques

One way to impose order on a system with multiple users having random access requirements is to institute a controller that periodically polls the user population to determine their service requests. If the user population is large (e.g., thousands of terminals) and the traffic is bursty, the time required to poll the population can be an excessive overhead burden. One technique for rapidly polling a user population [4, 14] is called a *binary tree search*. Figure 9.25 illustrates a satellite example of such a tree search to resolve contention among users. In this example, assume that the total user population is eight terminals; let them be identified by the binary numbers 000 to 111 as shown in Figure 9.25. Assume that terminals 001, 100, and 110 are contending for the service of a single channel. The tree search operates by continually partitioning the population until there is just a single branch remaining. The terminal corresponding to that branch is the “winner” and hence the first terminal to access the channel. The operation is repeated and again yields a single terminal that may next use the channel. The algorithm proceeds according to the following steps (see Figure 9.25):

- The satellite requests the transmission of the contending terminals’ first (left-most) bit of their identification (ID) numbers.
- Terminal 001 transmits a zero, and terminals 100 and 110 each transmit a one. The satellite, on the basis of received signal strength, selects one or zero as the bit it “heard.” In this example the satellite chooses binary one

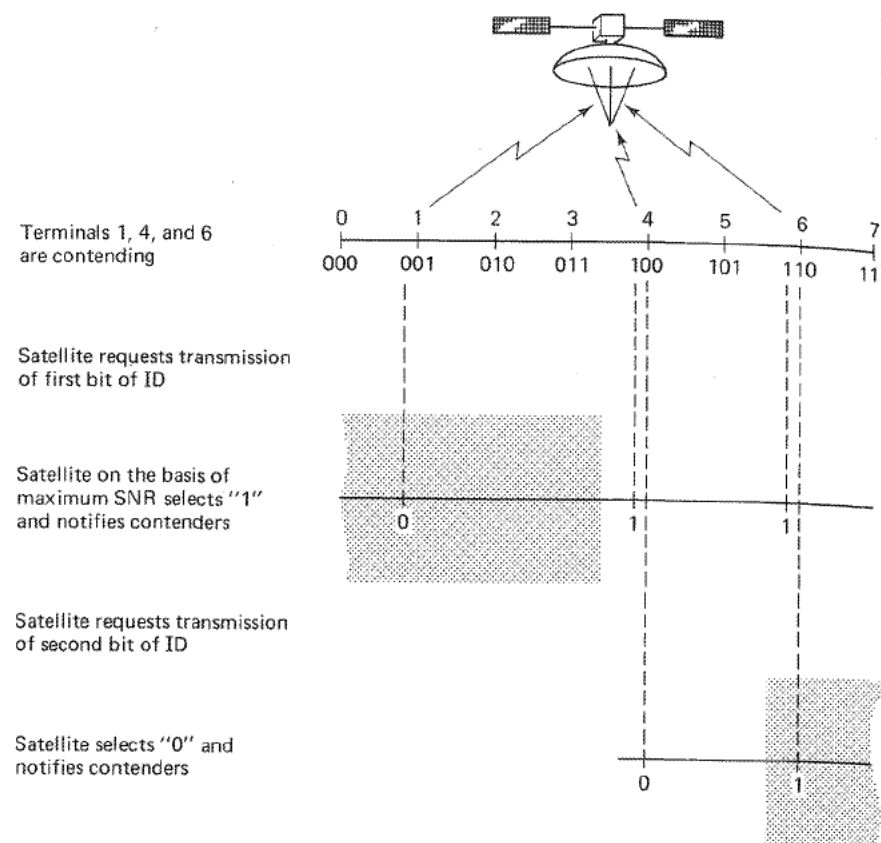


Figure 9.25 Tree search to resolve contention (eight terminal example).

and informs the users accordingly. Half the user population now knows that it has not been selected. The terminals in the "losing" half "bow out" of contention during this pass through the tree. In this example terminal 001 bows out.

3. The satellite requests the transmission of the second identifying bit from the remaining contending terminals.
4. Terminal 100 transmits a zero, and terminal 110 transmits a one.
5. Assume that the satellite selects the zero and notifies the contenders accordingly. Terminal 110 bows out. The process continues until it is clear that terminal 100 is free to access the satellite.
6. When the channel becomes available, steps 1–5 are repeated.

Example 9.3 Comparison between Binary Tree Search and Straight Polling

- (a) A binary tree search requires $n = \log_2 Q$ decisions for each pass through a population of Q terminals. A savings in time is possible with a tree search if the population is large and the average demand for service is small. Calculate the

time needed for the straight polling of a population of 4096 terminals, to provide channel availability to 100 terminals requesting service. Compare the result with the time needed to perform a binary tree search 100 times, over the same population. Assume that the time required to poll one terminal and the time required for one decision of a binary tree search are each equal to 1 s.

- (b) Develop an expression for Q' , the largest number of terminals that results in the same (or less) time expended for binary tree searching as compared to straight polling.
- (c) Compute Q' for part (a).

Solution

- (a) Straight polling of 4096 terminals:

$$T = 4096 \times 1 \text{ s} = 4096 \text{ s}$$

Binary tree search for 100 terminals requires 100 passes through the binary tree:

$$T' = (100 \times \log_2 4096) \times 1 \text{ s} = 1200 \text{ s}$$

- (b) Q' is the maximum number of terminals that will result in $T' \leq T$ in part (a). This will occur when

$$Q'' \log_2 Q \times 1 \text{ s/decision} = Q \times 1 \text{ s/poll}$$

$$Q' = \lfloor Q'' \rfloor = \left\lfloor \frac{Q}{\log_2 Q} \right\rfloor \quad (9.30)$$

where $\lfloor x \rfloor$ is the largest integer no greater than x .

- (c) Q' for part (a)

$$Q' = \left\lfloor \frac{4096}{\log_2 4096} \right\rfloor = 341 \text{ terminals}$$

A binary tree search for 341 terminals entails a search time of 4092 s.

9.4 MULTIPLE ACCESS TECHNIQUES EMPLOYED WITH INTELSAT

The first commercial, geostationary communication satellite (INTELSAT I, or Early Bird) launched in 1965, represented the start of a new telecommunications era. Its 240 voice circuits provided more capacity than the undersea cables laid between the United States and Europe during the previous 10 years [15].

Early Bird featured a hard-limiting nonlinear transponder using FDMA. When several signals having different carrier frequencies simultaneously occupy a nonlinear device, the result is the production of intermodulation products which are signals at all combinations of sum and difference frequencies [16–18]. The energy apportioned to these intermodulation or IM products represents a *loss* in the useful signal energy. In addition, if these IM products appear within the bandwidth of other signals, the effect is that of added *noise* for the other signals.

The nonlinear transponder in Early Bird allowed for only two earth stations (one in the United States and one in Europe) to simultaneously access the satellite.

Figure 9.26 illustrates this satellite's operation between the United States and Europe. Three European earth stations were interconnected via a terrestrial network. Each month a different European station accessed the satellite and distributed the traffic to the other two stations.

9.4.1 Preassigned FDM/FM/FDMA or MCPC Operation

INTELSAT II and III improved multiple access capability by operating their travelling-wave tube amplifiers (TWTAs) in the linear region instead of the hard-limiting region. This kept the IM products at an acceptable level, allowing more than two simultaneous accesses. (The price paid was a reduction in power amplifier efficiency.) Thus many FM carriers from various earth stations could simultaneously access these satellites. The operation is designated preassigned multdestination FDM/FM/FDMA or simply FDM/FM, or multichannel per carrier (MCPC), and is illustrated in Figure 9.27. Long-distance calls originating in country A enter the telephone exchange and are multiplexed into a supergroup (five groups of 12 voice circuits each). Country A transmits the supergroup on a single FM carrier at frequency f_A . Each group within the supergroup has been preassigned to an earth station in country A for telephone traffic destined to countries

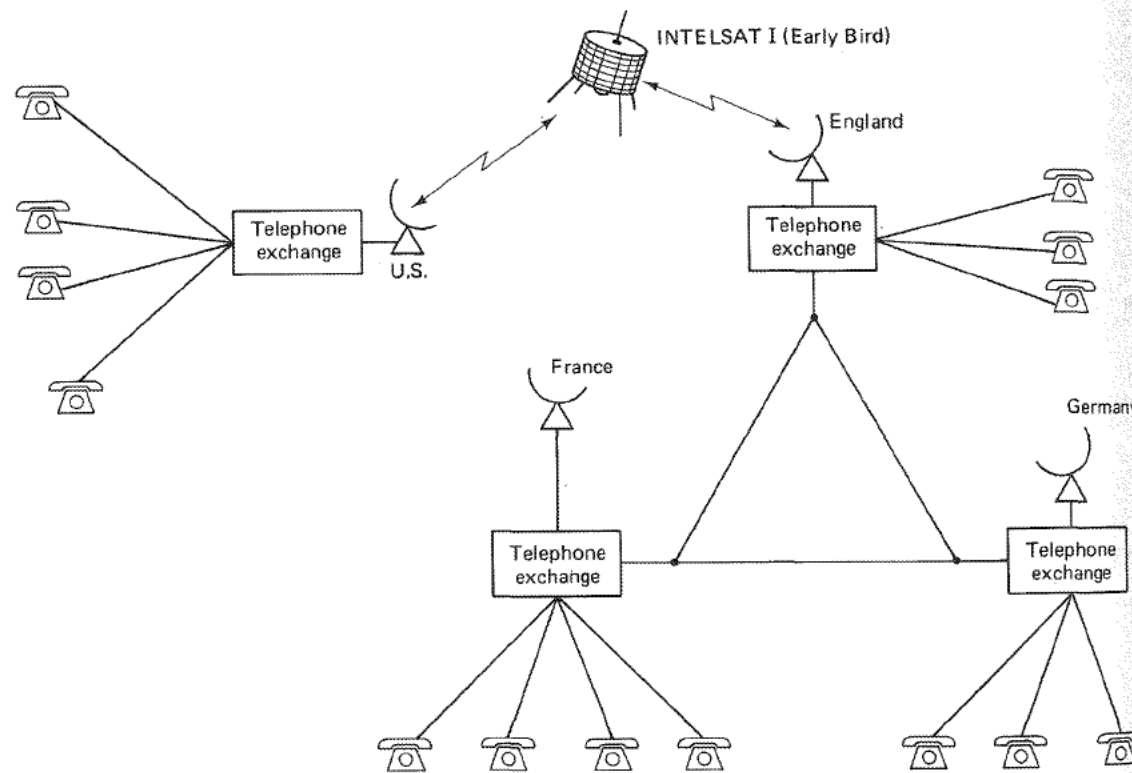


Figure 9.26 Early satellite operation.

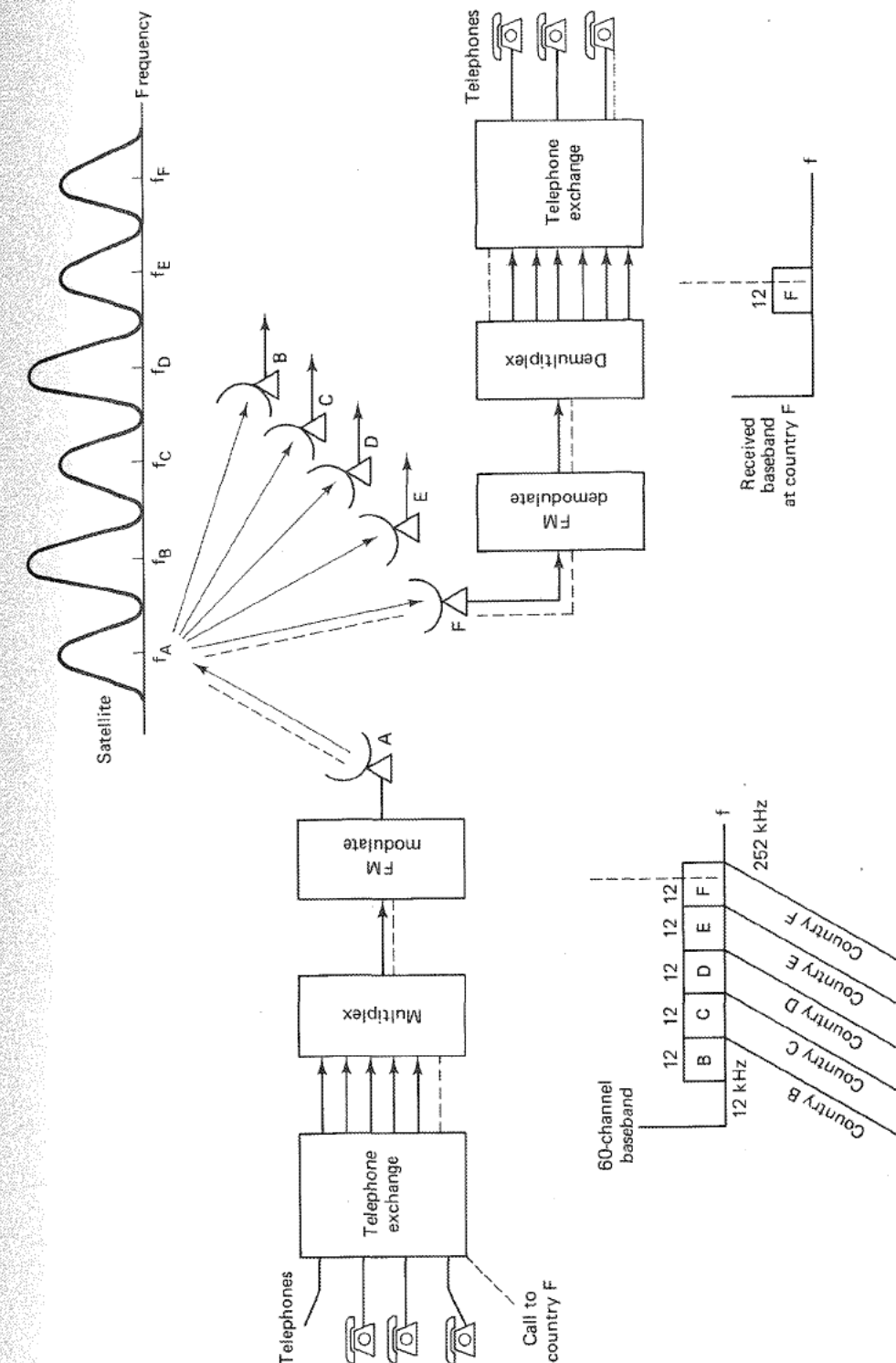


Figure 9.27 Preassigned multdestination FDM/FM carriers. (Reprinted with permission from J. G. Puente and A. M. Werth, "Demand-Assigned Service for the INTELSAT Global Network," *IEEE Spectrum*, Jan. 1971, © 1971 IEEE.)

B through F. These countries each receive the signal on frequency f_A . The received signal is demodulated and demultiplexed at the destination country, selecting only those 12 channels preassigned to it.

9.4.2 MCPC Modes of Accessing an INTELSAT Satellite

INTELSAT has standardized the ways in which each 36-MHz transponder may be shared by specifying the occupied RF bandwidth and the number of 4-kHz channels per user. Some of these standard channels are shown in Table 9.1. Notice that the capacity of the transponder (last column in Table 9.1) drops as the number of carriers increases. The reasons are as follows:

1. Guard bands are needed between carrier bands; the more carriers there are, the more guard bands are needed. Hence capacity is reduced.
2. Multiple carriers in the nonlinear TWTA cause intermodulation (IM) products. If the TWTA is backed off into the linear region to reduce interference, the TWTA can provide less overall power. The channel becomes power limited and can service fewer carriers.

TABLE 9.1 Standard INTELSAT MCPC Accessing Modes

Number of carriers per transponder	Carrier bandwidth	Number of 4-kHz channels per carrier	Number of 4-kHz channels per transponder
1	36 MHz	900	900
4	3 at 10 MHz 1 at 5 MHz	132 60	456
7	5 MHz	60	420
14	2.5 MHz	24	336

Table 9.1 indicates that a single carrier provides the most efficient use of the transponder. Why doesn't INTELSAT always operate its transponders in this mode? The answer is that not all earth stations have enough traffic to justify the assignment of an entire 36-MHz transponder. The other modes are needed so that various combinations of stations having less traffic will be able to share a transponder.

9.4.2.1 Bandwidth-Limited versus Power-Limited Conditions

In the preceding section it was stated that the backed-off transponder cannot support as many channels as the fully saturated transponder. It is useful to examine the two extreme transponder conditions, bandwidth limited and power limited, in the context of a satellite transponder. In Figure 9.28 we assume a 36-MHz transponder with a maximum power output of 20 W. Figure 9.28a illustrates an MCPC mode of operation whereby four carriers share the 36-MHz bandwidth. Assume that each carrier requires 4 W. The total output power is 16 W (less than the maximum capability of the amplifier); therefore, there is still power to spare.

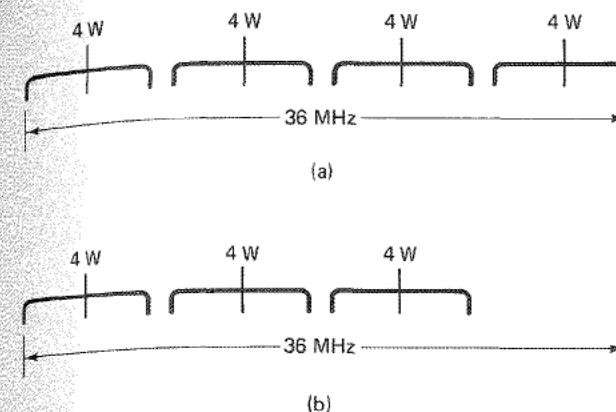


Figure 9.28 Bandwidth-limited versus power-limited configurations. (a) Bandwidth-limited example. (b) Power-limited example.

However, should another user want to access the transponder, the total 36-MHz bandwidth has already been allocated to the existing four carriers; there is no additional bandwidth to spare. Figure 9.28a illustrates this bandwidth-limited case.

Suppose that the previous example results in the production of serious IM products at the transponder. Assume that it is necessary to linearize the transponder by operating it at a reduced maximum power output of 12 W. With only a 12-W capability, the transponder can no longer support four users with 4 W each. One of the users must be "thrown off," as illustrated in Figure 9.28b. Therefore, we have bandwidth to support another user, but not sufficient power. Figure 9.28b illustrates this power-limited case.

9.4.3 SPADE Operation

The preassigned MCPC multiple access scheme is very efficient when the traffic is heavy enough so that the channels are most always filled. However, if out of a 12-channel group, only one channel is active, the other 11 cannot be turned off. The FDM/FM transmission is made with or without actual telephone traffic on the channels. Therefore, the long-term preassignment of carriers to stations having light traffic is wasteful. Since there are many light traffic links, a flexible method to service them was needed. Also, an efficient way to handle overflow traffic from medium-capacity preassigned links was needed. Such was the motivation for a novel DAMA scheme known as SPADE, first used with INTELSAT IV. The acronym SPADE stands for "single-channel-per-carrier PCM multiple access demand assignment equipment." The principal features characterizing SPADE operation [15] are:

1. A single voice-grade channel is analog-to-digital (A/D) converted at a bit rate of 64 kbits/s.
2. This baseband digital signal modulates a carrier using quadrature phase shift keying (QPSK). Unlike the MCPC case, there is *only one* voice channel per carrier.

3. The channel spacing is 45 kHz. Within a transponder, there is bandwidth available for 800 channel carriers. Six carrier positions are vacant by design; thus there are 794 usable carriers.
4. The carrier is dynamically assigned, *upon demand*.
5. The dynamic assignment is accomplished over a 160-kHz common signaling channel (CSC) used as an "order-wire" or control circuit. The bit rate on the CSC is 128 kbits/s, and the modulation is binary phase shift keying (BPSK).

Figure 9.29 illustrates the frequency allocations for the CSC and the 800 carriers in the SPADE system. The SPADE operation can best be understood with the aid of Figure 9.30. The CSC operates in a fixed-assignment TDMA broadcast mode; that is, all earth stations monitor the CSC and are aware of the current state of channel assignments. Each earth station has a 1-ms time slot on the CSC (once every 50 ms) for requesting or releasing a channel. When an earth station needs a channel, it "seizes" a free one by requesting a frequency pair at random and transmitting its selection on the CSC. Random selection makes it unlikely that two stations will simultaneously request the same channel unless there are very few remaining. As soon as the channel is allocated, each of the other earth station processors deletes it from its list of available channels. The list is continually kept updated via the CSC. Thus control of the SPADE access scheme is *distributed* among all the participating earth stations.

When the station finishes with the channel, the station indicates the channel's release by transmitting a signal in its time slot on the CSC. Each station receives this signal and designates the released channel as available. If two stations simultaneously seize the same channel, they each get a "busy" indication. They try again, selecting at random from the pool of available channels.

9.4.3.1 Transponder Capacity Utilization with SPADE

Table 9.2 is a continuation of Table 9.1. We see that the transponder bandwidth utilization with SPADE results in a total capacity of 800 voice channels per transponder. Compare Table 9.2 with Table 9.1. In Table 9.1, as the number of carriers increases from 1 to 14, the total number of channels decreases from 900 to 336. Why doesn't the SPADE system in Table 9.2 exhibit less capacity than the 336 channels associated with 14 carriers? The improved utilization comes about as follows. When there is only one voice channel per carrier, the carrier can be switched off when no speech is detected. Even with all channels operating, they can be switched off approximately 60% of the time. The transponders are power limited; power savings means that more channels can be transmitted. Also, SPADE uses digital voice transmission (QPSK); the bandwidth efficiency of the system is commensurate with the single-carrier FDM/FM case.

9.4.3.2 SPADE Efficiency

With MCPC, capacity is preassigned; a station's unused channels cannot be reallocated to other stations. SPADE is a DAMA system where all channels are shared. The channels are allocated to users as needed. An important telephone

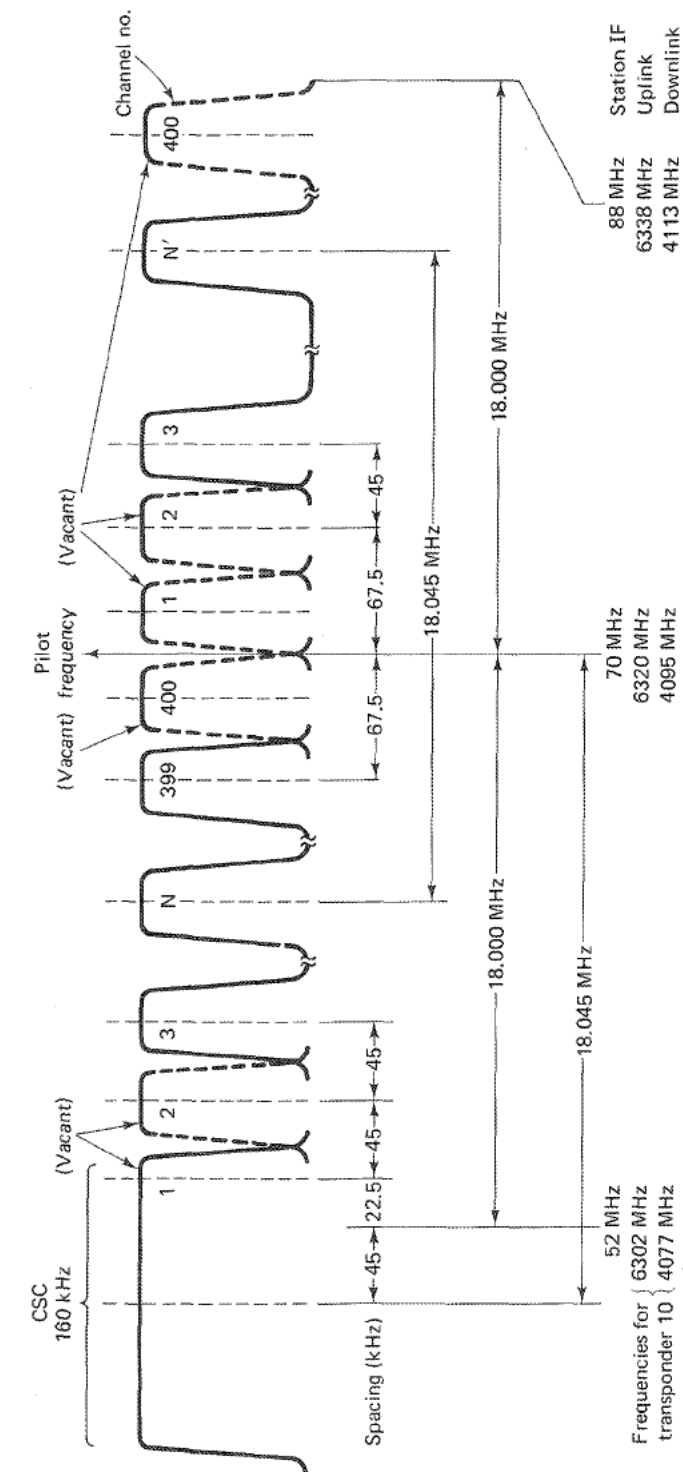


Figure 9.29 SPADE frequency allocations. (Reprinted with permission from J. G. Puente and A. M. Werth, "Demand-Assigned Service for the INTELSAT Global Network," *IEEE Spectrum*, Jan. 1971. © 1971 IEEE.)

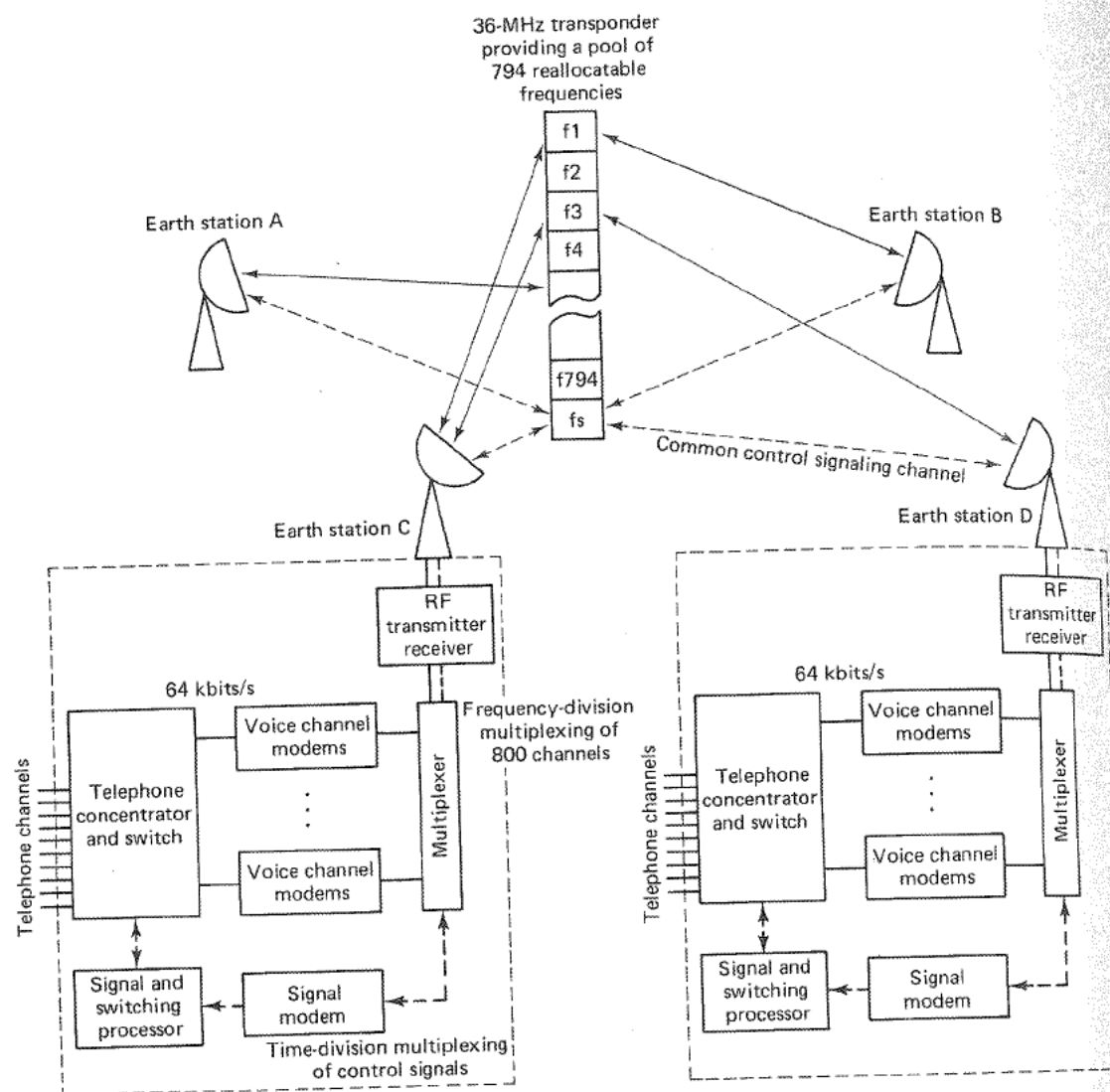


Figure 9.30 SPADE operation. (From James Martin, *Communications Satellite Systems*, © 1978, Fig. 15.2, p. 236. Reprinted by permission of Prentice-Hall, Englewood Cliffs, N.J.)

TABLE 9.2 SPADE Accessing

Number of carriers per transponder	Carrier bandwidth	Number of 4-kHz channels per carrier	Number of 4-kHz channels per transponder
800	45 kHz	1	800

system quality measure, called the probability of blocking, is the probability that a requested circuit is not available. To achieve 1% probability of blocking requires four times as many MCPC channels as SPADE channels. A SPADE transponder with 800 channels is equivalent to 3200 MCPC channels [15].

9.4.3.3 Mixed-Size Earth Station Network Using SPADE

A standard-size INTELSAT earth station has a receiver sensitivity $G/T^{\circ} = 40.7$ dB/K, whereas the smaller size stations have a $G/T^{\circ} = 35$ dB/K. If 125 SPADE channels are destined for small stations, the total transponder capacity of 800 standard channels is reduced to 525 channels. This is the point at which half the available power is used to service the standard stations. The relationship between transponder capacity and channels allocated to small stations is shown in Figure 9.31. An explanation of this relationship can best be seen in Figure 9.32. When the total TWTA power provides service to large stations, Figure 9.32a illustrates that the 36-MHz bandwidth transponder is occupied by approximately 800 carriers each at a power level of x dBW (the bandwidth-limited case). When half the power is required to service small stations, Figure 9.32b illustrates that 400 carriers (half of the original 800) each at a power level of x dBW are reserved for the standard stations. Consider what happens to the remaining 400 carrier positions. From Chapter 4 we know that the error performance of a link is directly related to the product of EIRP and G/T° . For any link, one can trade off these two parameters, thereby maintaining a fixed level of performance. Since the small station has a G/T° of 5.7 dB less than that of the standard station, it is necessary to supply the small station with 5.7 dB more EIRP for equivalent performance. The carrier power is increased by approximately 5.7 dB for each small station, thus the quan-

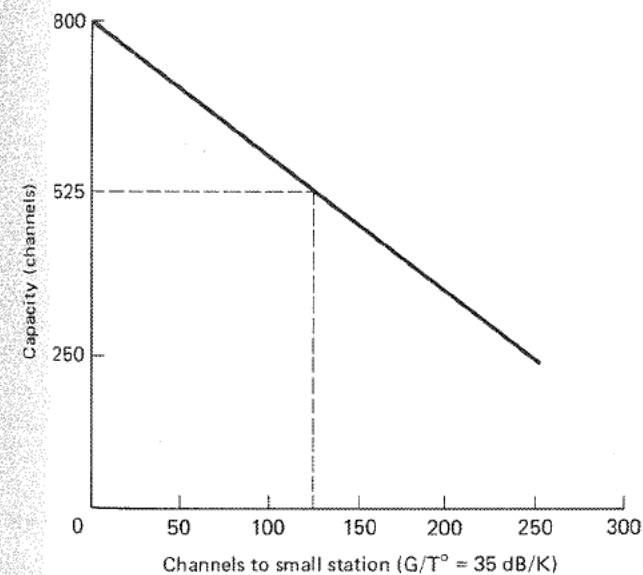


Figure 9.31 SPADE transponder capacity in a mixed-size earth station network.

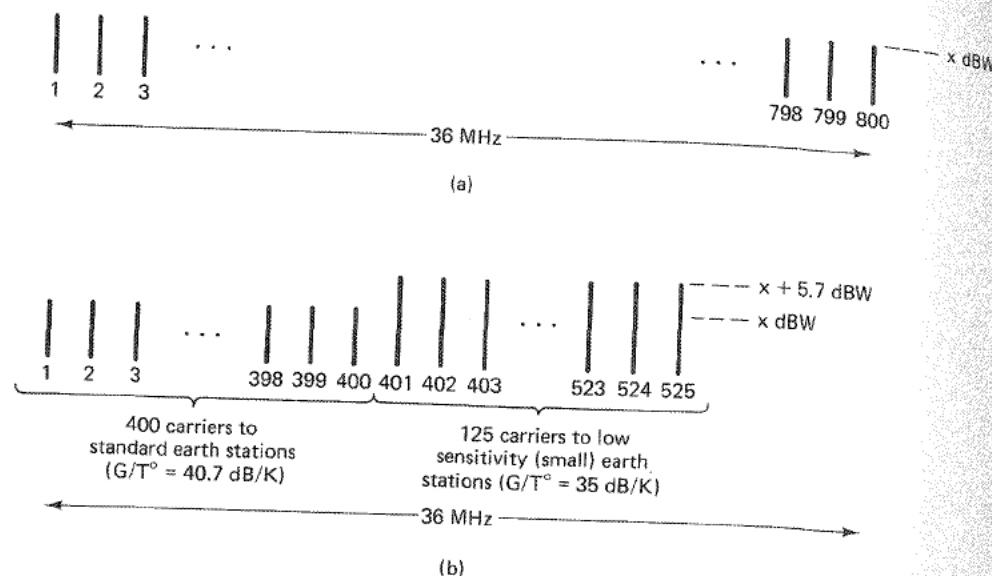


Figure 9.32 Mixed-size earth station network. (a) When the total TWTA power services large stations: bandwidth limited (800 carriers). (b) When half the TWTA power services small stations: power limited (525 carriers).

tity of the remaining carriers serving these small stations is *decreased* by a similar amount. Therefore instead of 400 carriers, 125 (a reduction of 5.1 dB) are used to serve the small stations; the transponder is now power limited.

At the time a channel is assigned to a call, the transmitting station is apprised of the size of the destination station. Recall that these satellites are non-regenerative so that the apportionment of downlink EIRP is established by the transmitting station (see Section 4.7.1). The transmitting station sets its power level according to the needs of the receiving station.

9.4.4 TDMA in INTELSAT

The first generation of multiple access communication systems has been dominated by FDMA systems. The trend, however, is now in favor of TDMA systems, made possible by the availability of precise clocks and high-speed switching elements [19–24]. INTELSAT IV used a 128-kbits/s TDMA scheme for the common signaling channel that controls the SPADE network. Intelsat V introduced a 120-Mbits/s TDMA scheme for multiple-beam international digital service. One disadvantage or cost in implementing a TDMA scheme is the need for providing precise *synchronization* among the participating earth stations and the satellite. FDMA systems, not having such requirements, are less complex from a networking point of view. Comparisons of TDMA versus FDMA operation are summarized as follows:

1. FDMA can cause IM products. This can be avoided by operating the TWTA in its linear region, thereby reducing the available power output.
2. With TDMA, there is only one carrier present at a time in the TWTA. Thus IM distortion cannot occur.
3. TDMA earth station equipment is more sophisticated and hence more costly than FDMA equipment. However, for earth stations providing multiple point-to-point channels, FDMA stations require separate radio-frequency (RF) up-conversion and down-conversion signal processing stages. Thus with FDMA, the amount of equipment grows with the amount of simultaneous connectivity. With TDMA, such growth does not take place since channel selectivity is accomplished in time rather than frequency. Therefore, for a large multiply connected earth station, TDMA can be more cost-effective than FDMA.
4. In multiple-beam systems, each beam may need to communicate with every other beam. TDMA lends itself to conveniently forming connections sequentially as in satellite-switched TDMA (SS/TDMA). INTELSAT VI uses such satellite-switched TDMA (SS/TDMA), described in Section 9.4.5.

An example of the comparative performance of TDMA, FDM/FM, and SPADE is shown for an INTELSAT IV transponder as a plot of channel capacity versus earth station G/T° in Figure 9.33. Figure 9.33a is for an earth coverage antenna, and Figure 9.33b is for a spot-beam antenna. From synchronous altitudes these antennas have half-power beamwidths of 17° and 4.5° , respectively. From these plots it is seen that single-carrier FDM/FM is as efficient as TDMA when the system is operated with standard earth stations ($G/T^\circ = 40.7$ dB/K). For smaller earth stations ($G/T^\circ \leq 31$ dB/K) working through earth-coverage transponders, SPADE is more efficient than TDMA and multicarrier FDM/FM (MCPC); only the four-carrier case is plotted. For earth stations having G/T° in the range 19 to 40.7 dB/K working through a spot-beam transponder, TDMA is superior to SPADE and MCPC. For smaller earth stations having G/T° in the range 6 to 19 dB/K working through a spot-beam transponder, SPADE is superior to TDMA and MCPC. In general, when working through *standard* earth stations it is seen [19] that TDMA is the most efficient multiple access scheme for INTELSAT IV.

9.4.4.1 PCM Multiplex Frame Structures

There are two digital telephony standards for PCM frame structures in operation. The North American standard is called *T-Carrier*; it is built around the 193-bit frame shown in Figure 9.34a. There are 24 channels; each channel contains an 8-bit voice sample. Also, there is one bit per frame with alternating value 1 0 1 0 . . . from frame to frame, used for frame alignment. Since a voice-grade telephone channel has a bandwidth of $W = 4$ kHz (including guard bands), the Nyquist sampling rate for recovering the analog information within 4 kHz is $f_s = 2W = 8000$ samples/s. Therefore, the basic PCM frame, called the *Nyquist*

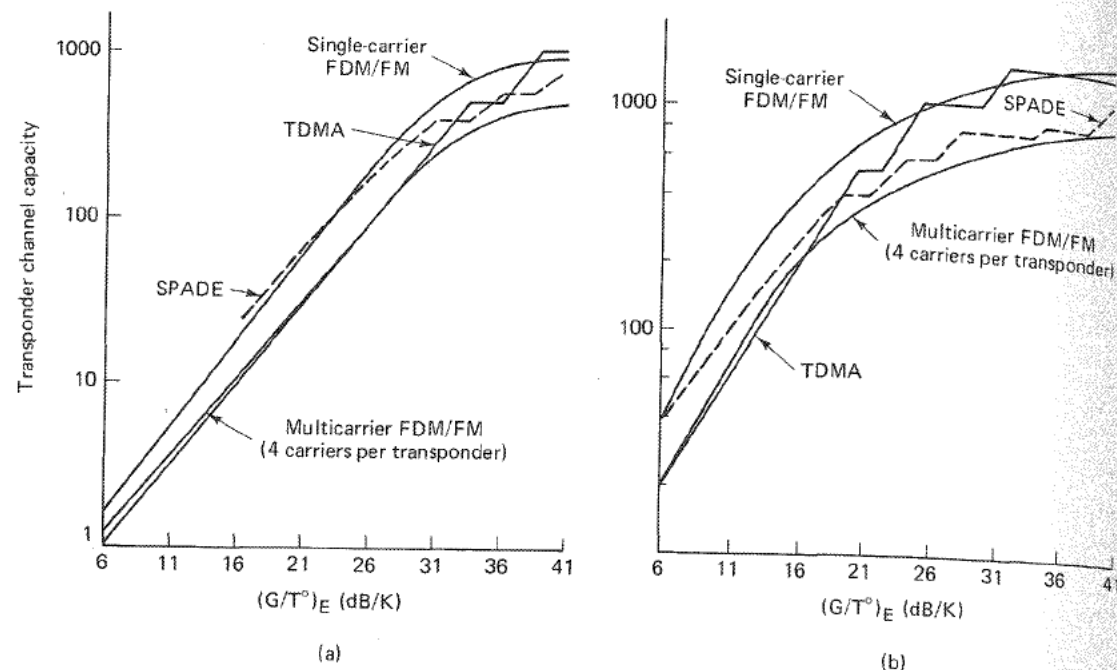


Figure 9.33 Channel capacity versus earth station G/T° for FDMA, TDMA, and SPADE. (a) Global-beam transponder channel capacity as function of $(G/T^\circ)_E$, where $(G/T^\circ)_E$ means earth station G/T° . (b) Spot-beam transponder channel capacity as function of $(G/T^\circ)_E$. [From D. Chakraborty, "INTELSAT IV Satellite System (Voice) Channel Capacity versus Earth Station Performance," *IEEE Trans. Commun. Tech.*, vol. COM19, no. 3, June 1971, pp. 355-362. © 1971 IEEE.]

frame, which contains 24 voice samples from 24 different message sources, has a frame rate of 8000 frames/s (duration of 125 μ s). Thus the basic T-Carrier bit rate is 193 bits/frame $\times$ 8000 frames/s = 1.544 Mbits/s.

The European standard is built around a 256-bit frame shown in Figure 9.34b. There are 30 message channels, each containing an 8-bit voice sample. Also, one 8-bit time slot is used for frame alignment and another 8-bit time slot is used for signaling (addressing) information. The European frame rate is the same as that of the T-Carrier. Therefore, the basic European bit rate is 256 bits/frame $\times$ 8000 frames/s = 2.048 Mbits/s.

9.4.4.2 The High-Rate TDMA Frame for Europe

Sixteen Nyquist frames of the European PCM Multiplex format are shown in Figure 9.35a. Each frame contains an 8-bit sample from each of 30 terrestrial channels, plus 8 bits of framing and 8 bits of signaling information. The TDMA frame duration is

$$16 \text{ Nyquist frames} \times 125 \mu\text{s/Nyquist frame} = 2 \text{ ms}$$

Within this 2-ms frame are contained

$$16 \text{ Nyquist frames} \times 256 \text{ bits/Nyquist frame} = 4096 \text{ bits}$$

The basic idea behind TDMA is that a user's low-rate data stream can share the CR with similar streams from other users by *bursting* the transmission at a much faster rate than the rate at which it is generated. Figure 9.35b illustrates a 2-ms high-rate TDMA frame. The frame begins with a reference burst, RB1, emitted by a reference station. The burst contains information necessary to enable other stations to precisely position their message traffic bursts in the frame. There may be a second burst, RB2, for reliability, followed by a sequence of traffic slots. The traffic slots may be preassigned, or they may be assigned according to a DAMA protocol [20].

The PCM multiplex signal with a bit rate of $R_0 = 2.048$ Mbits/s and a frame duration of $T = 2$ ms is compressed (by a factor of 59) and transmitted using QPSK modulation at a burst rate of $R_T = 120.832$ Mbits/s (symbol rate of 60.416 megasymbols/s). The duration of the traffic data field T_{tr} in the high rate TDMA

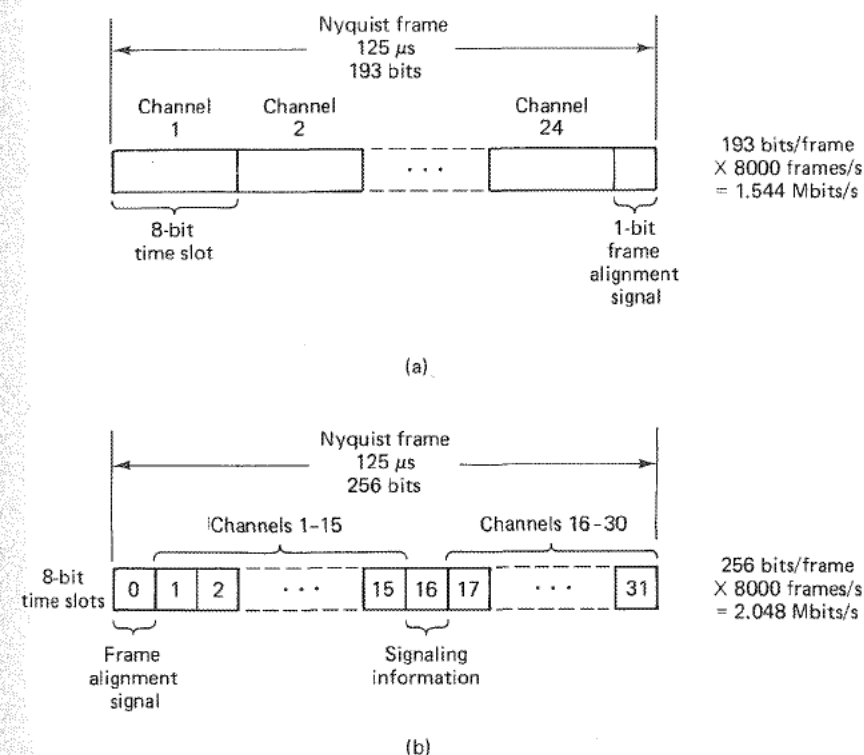


Figure 9.34 PCM multiplex frame structure. (a) Frame structure for T-Carrier (North American) PCM multiplex. (b) Frame structure for the European PCM multiplex.

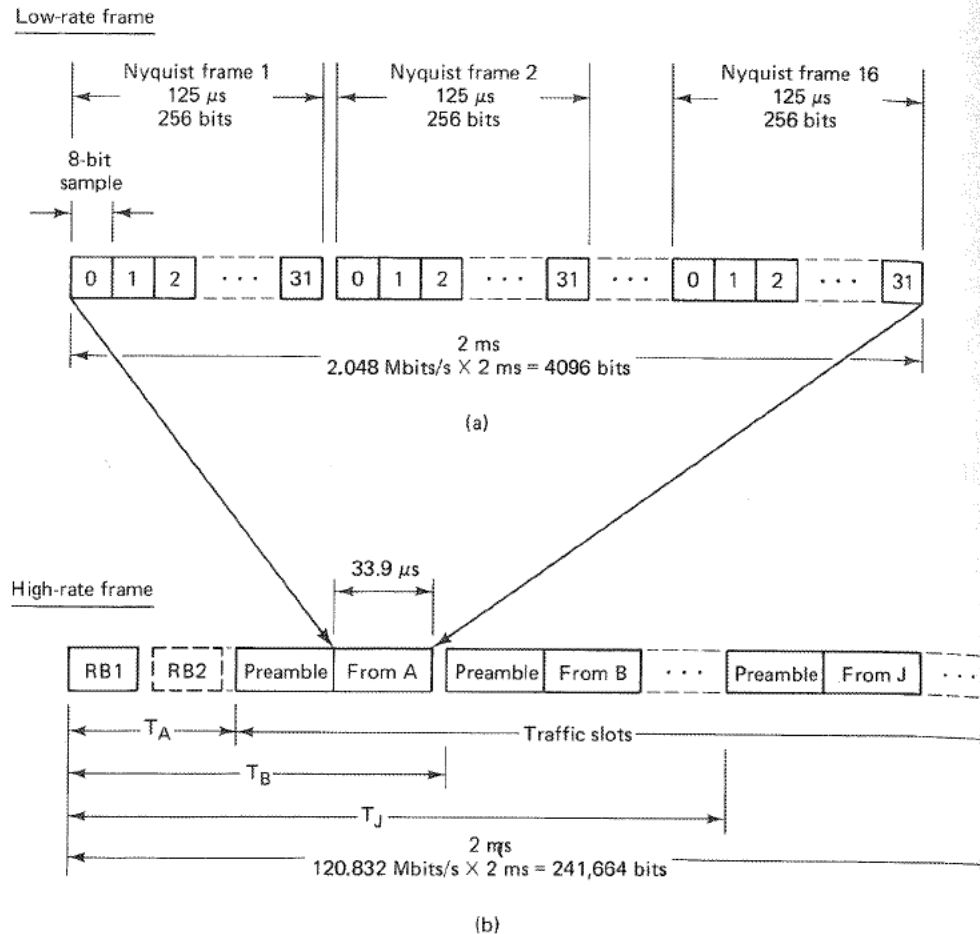


Figure 9.35 INTELSAT digital transmission standards for Europe. (a) Terrestrial PCM multiplex. (b) High-rate frame.

frame is calculated as follows:

$$T_{tr} = \frac{R_0 T}{R_T} \quad (9.31)$$

$$= \frac{2.048 \times 10^6 \times 2 \times 10^{-3}}{120.832 \times 10^6}$$

$$= 33.9 \mu s$$

To obtain the total duration of a traffic burst, the time used for the preamble must be added. If the preamble contains S_P symbols, then assuming QPSK modulation, the total length of the traffic burst measured in number of symbols, S_T , is

$$S_T = \frac{R_0 T}{2} + S_P \quad (9.32)$$

and the burst-time duration is

$$T_T = \frac{2S_T}{R_T} \quad (9.33)$$

If the preamble contains 300 symbols, then

$$S_T = \frac{2.048 \times 10^6 \times 2 \times 10^{-3}}{2} + 300$$

$$= 2348 \text{ symbols}$$

Using this in Equation (9.33), we obtain

$$T_T = \frac{2 \times 2348}{120.832 \times 10^6} = 38.9 \mu s$$

9.4.4.3 The High-Rate TDMA Frame for North America

The INTELSAT TDMA burst (bit) rate of $R_T = 120.832$ Mbits/s was chosen to be compatible with both the European and North American standards. Figure 9.36 is similar to Figure 9.35 except that the PCM multiplex signal is the 24-channel T-Carrier instead of the 30-channel European standard. The essential T-Carrier features that are different from the European standard are listed below and are shown on the figure.

1. Each Nyquist frame is comprised of 24 channels or samples $\times$ 8 bits + 1 frame alignment bit = 193 bits.
2. The 16 Nyquist frames contain $16 \times 193 = 3088$ bits.
3. The T-Carrier data rate is 1.544 Mbits/s.
4. The duration of the traffic data field in the high-rate TDMA frame is calculated from Equation (9.31).

$$T_{tr} = \frac{1.544 \times 10^6 \times 2 \times 10^{-3}}{120.832 \times 10^6}$$

$$= 25.6 \mu s$$

9.4.4.4 INTELSAT TDMA Operation

At the transmitting earth station, the continuous low-rate data stream enters one of a pair of buffers illustrated in Figure 9.37a. When one buffer is filling at the low rate (1.544 Mbits/s or 2.048 Mbits/s), the other is emptying at the burst rate (120.832 Mbits/s). The buffers alternate functions at each TDMA frame. The time of application of the high-rate clock is controlled so that the traffic burst is transmitted in the proper interval to arrive at the satellite in its assigned position in the TDMA frame.

At the receiving station, the received traffic burst is routed to one of a pair of expansion buffers, shown in Figure 9.37b, that have the inverse function of

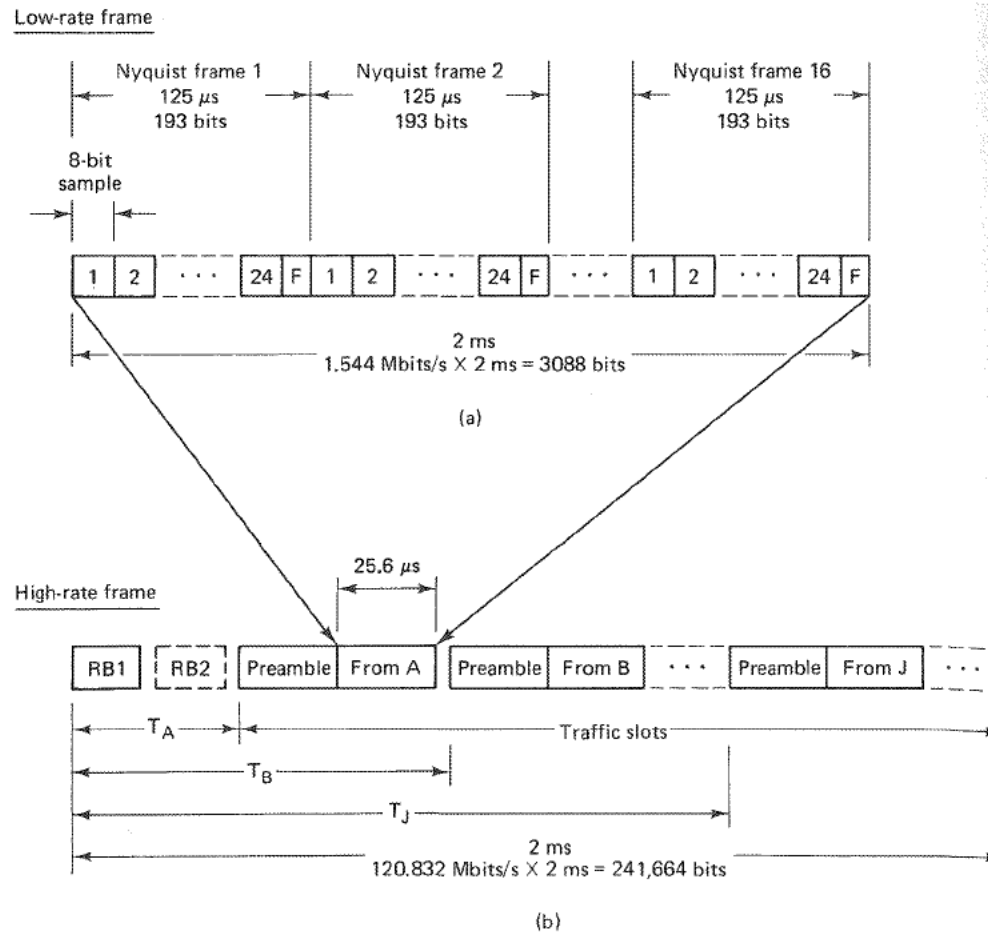


Figure 9.36 INTELSAT digital transmissions standards for T-Carrier. (a) Terrestrial PCM multiplex. (b) High-rate frame.

the compression buffers in Figure 9.37a. When one buffer is filling at the high rate, the other is emptying at the desired output rate.

The most critical aspect of TDMA operation is the precise synchronization needed to assure orthogonality of the time slots [20]. Figure 9.38 illustrates the general idea behind most commercial satellite synchronization schemes. One station is designated as the master or control station. This station transmits periodic bursts of reference timing pulses. User stations also transmit their timing pulses, designated as slave pulses in Figure 9.38. On the downlink, the using station receives the master or reference pulses in addition to its own slave pulses. The time difference between the master and slave pulses corresponds to the timing error. The station adjusts its clock so as to reduce this timing error.

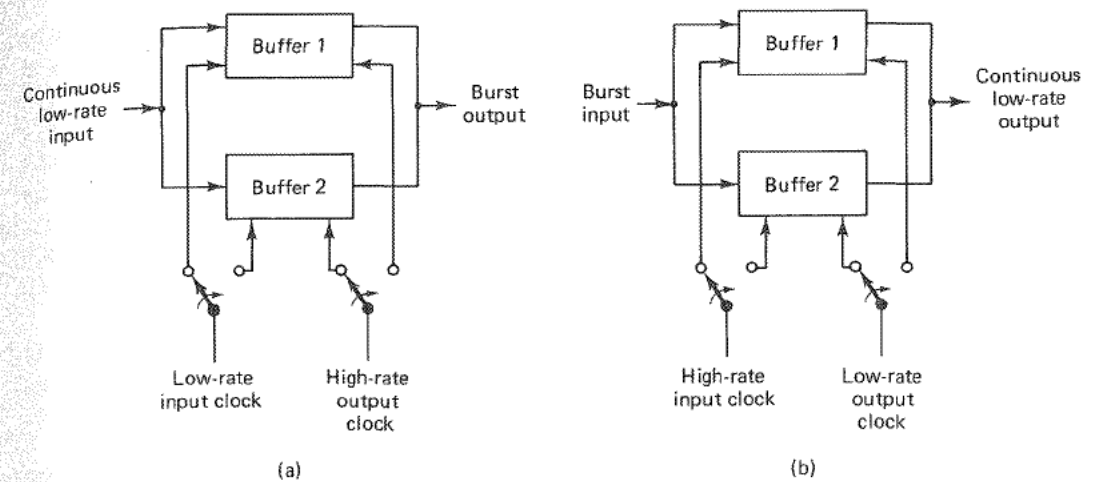


Figure 9.37 Burst compression and expansion buffers. (a) Compression buffers at transmitter. (b) Expansion buffers at receiver.

9.4.5 Satellite-Switched TDMA in INTELSAT

Modern communication satellites often employ several regional antenna beams. For a satellite based over the Atlantic Ocean, separate beams might be aimed at North America, Europe, South America, and Africa. Switches are used to allow the interconnection of stations in one region to communicate with stations in another region. The basic goal of a satellite-switched TDMA (SS/TDMA) scheme is to provide an efficient way of cyclically providing interconnection of TDMA data among various coverage regions.

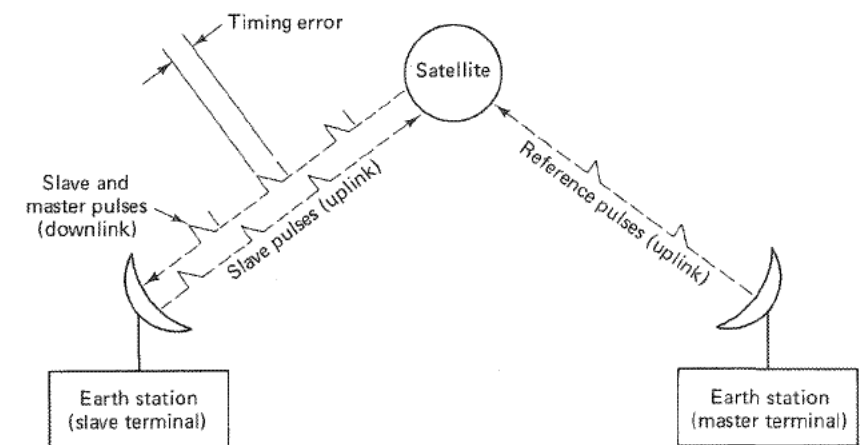


Figure 9.38 TDMA synchronization concept.

The heart of the system consists of a microwave switch matrix, located in the satellite, that is programmed via ground control to change states cyclically in rapid sequence, thus interconnecting distinct uplink beams to distinct downlink beams at each switching time. An earth station in the network communicates with those in other beams by transmitting TDMA bursts in the proper timing positions in the sequence. The pattern of switch states is selected so as to maximize the usable system capacity under the constraints of the traffic demands [21]. For complete interconnectivity between N beams, a total of $N!$ different satellite switch states or *modes* are required. Table 9.3 illustrates the six modes required for the full interconnectivity of a three-beam system.

In mode 1, the satellite receivers in beams A, B, and C are connected to the satellite transmitters for beams A, B, and C, respectively. An earth station in one of these beams can then communicate with other earth stations in the same beam. The beam is said to be *looped back* on itself.

Figure 9.39 illustrates a three-beam (beams A, B, and C) example of a SS/TDMA system. The satellite microwave switch matrix is configured in a *crossbar* design. This design can be thought of as being made up of row and column lines; when one row and one column are energized, contact is made at the intersection. A crossbar design only permits a single row to communicate with a single column at a time. If uplink A_U is connected to downlink B_D , neither A_U nor B_D can be simultaneously connected to any other beam.

In Figure 9.39, three different traffic patterns during time slot intervals T_1 , T_2 , and T_3 , with three different switch states S_1 , S_2 , and S_3 are shown. During interval T_1 , switch state S_1 interconnects the beams in a loop-back fashion which permits the uplink messages in slot T_1 to be delivered to their correct destinations. During time interval T_2 , switch state S_2 interconnects uplink beam A_U to downlink beam B_D , uplink beam B_U to downlink beam C_D , and uplink beam C_U to downlink beam A_D . This connection pattern assures that the uplink messages in slot T_2 are delivered to their correct destinations. During time interval T_3 , switch state S_3 similarly connects uplink transmissions to downlink beams to assure correct delivery of the data.

The traffic patterns and their durations are programmed to optimize the resource capacity and to serve the users as efficiently as possible. The cyclic pattern can be reprogrammed by ground command to meet changing traffic requirements.

9.4.5.1 Traffic Matrix

Figure 9.40 is a matrix describing the communication traffic among N spot-beam coverages. In this figure, t_{ij} is the traffic volume from the i th beam to the j th beam. The subtotal S_i is the total traffic originating from the i th uplink beam, expressed as

$$S_i = \sum_{j=1}^N t_{ij} \quad (9.34)$$

TABLE 9.3 Three-Beam Satellite Switch Modes

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

and R_j is the total traffic received in the j th downlink beam, expressed as

$$R_j = \sum_{i=1}^N t_{ij} \quad (9.35)$$

When the traffic in a SS/TDMA system is controlled by a nonblocking switch (one that allows for the transmission of *all* messages, without any "busy" signals)

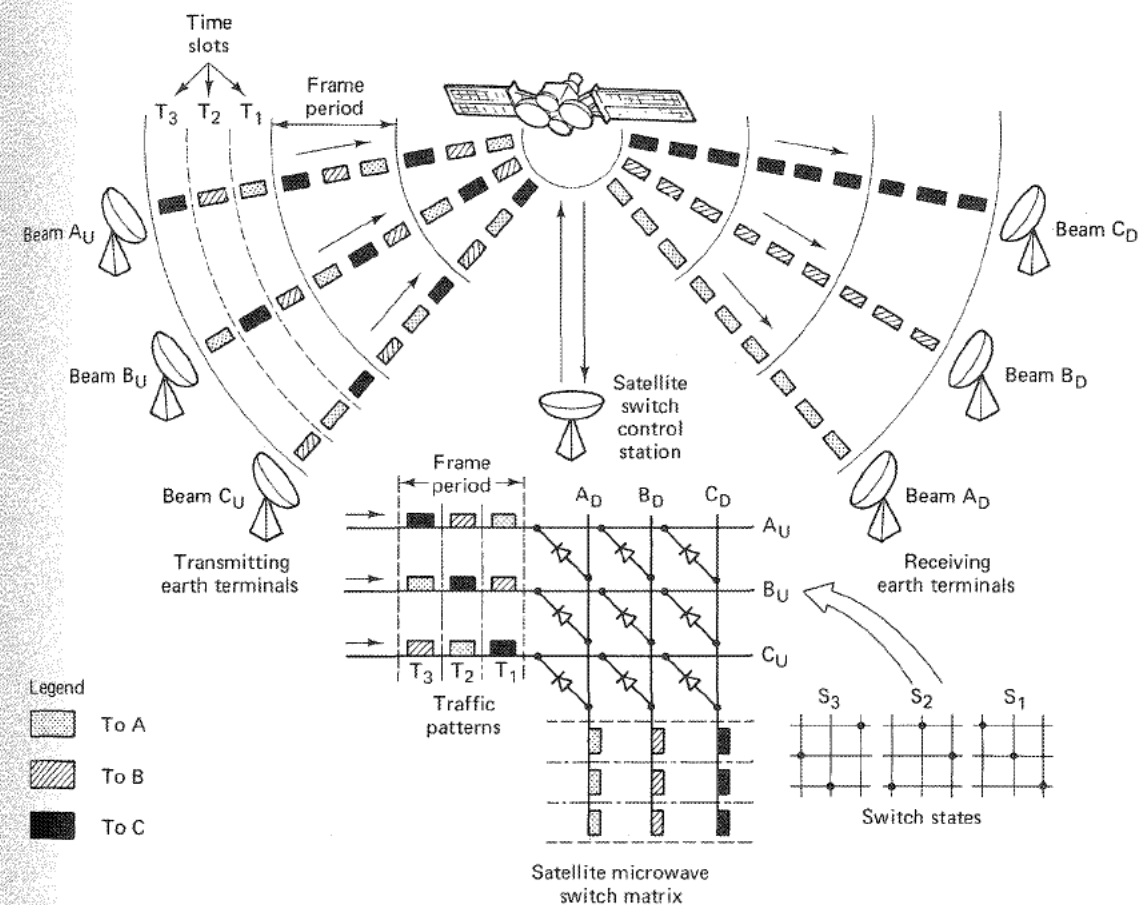


Figure 9.39 Satellite-switched TDMA (SS/TDMA).

		Destination						Originating traffic subtotals
		1	2	...	j	...	N	
Origin	1	t_{11}	t_{12}		t_{1j}		t_{1N}	S_1
	2	t_{21}	t_{22}		t_{2j}		t_{2N}	S_2
	...							
	i	t_{i1}	t_{i2}		t_{ij}		t_{iN}	S_i
	...							
	N	t_{N1}	t_{N2}		t_{Nj}		t_{NN}	S_N
Received traffic subtotals		R_1	R_2		R_j		R_N	Total

Figure 9.40 Traffic matrix.

a k -second time slot will be assigned to each channel in the TDMA frame. For efficient utilization of the CR, the total traffic in Figure 9.40 should be transmitted within a frame time T which should be made as short as possible. The minimum frame time, $T_{\min}$, for providing such nonblocking connectivity can be expressed [22] as follows:

$$T_{\min} = k \max (\{S_i\}, \{R_j\}) \quad (9.36)$$

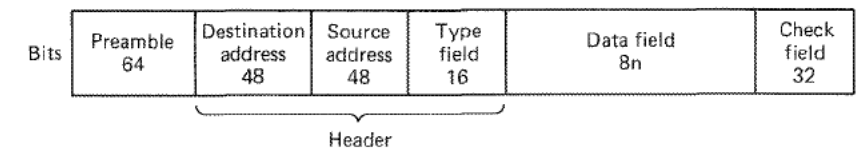
where $\max (\{S_i\}, \{R_j\})$ is the maximum value taken over the set of all $\{S_i\}$ and $\{R_j\}$. Equation (9.36) describes the minimum time to communicate *all* of the traffic in the traffic matrix, for equal bandwidth per channel.

9.5 MULTIPLE ACCESS TECHNIQUES FOR LOCAL AREA NETWORKS

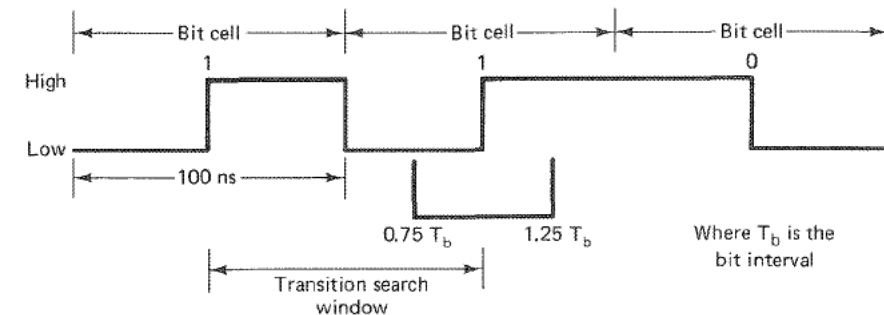
A local area network (LAN) can be used to interconnect computers, terminals, printers, and so on, located within a building or a small set of buildings. While long-haul networks use the public telephone network for economic reasons, LAN designers usually lay their own high-bandwidth cables. Bandwidth is not as scarce as it is in the long-haul cases. Not being forced to optimize bandwidth, a LAN can use simple access algorithms [6, 25–27].

9.5.1 Carrier-Sense Multiple Access Networks

Ethernet is a LAN access scheme developed by the Xerox Corporation. The Ethernet scheme is based on the assumption that each local machine can sense the state of a common broadcast channel before attempting to use it. The technique is known as *carrier-sense multiple access with collision detection* (CSMA/CD). The word “carrier,” here, means *any* electrical activity on the cable. Figure 9.41a



(a)



(b)

Figure 9.41 Ethernet bit field and PCM format. (a) Ethernet specification. (b) Manchester PCM format.

illustrates the bit field format for the Ethernet specification; the details are listed below.

1. The maximum packet size is 1526 bytes, where a byte is 8 bits. The packet breakdown is 8-byte preamble + 14-byte header + 1500-byte data + 4-byte parity.
2. The minimum packet size is 72 bytes, consisting of an 8-byte preamble + 14-byte header + 46-byte data + 4-byte parity.
3. The minimum spacing between packets is 9.6 μ s.
4. The preamble contains a 64-bit synchronization pattern of alternating ones and zeros, ending with two consecutive ones: (1 0 1 0 1 0 ... 1 0 1 0 1 1).
5. The receiving station examines a destination address field in the header to see if it should accept a particular packet. The first bit indicates the type of address (0 = single address, 1 = group address); an entire field of ones means an all-station broadcast.
6. The source address is the unique address of the transmitting machine.
7. The type field determines how the data field is to be interpreted. For example, bits in the type field can be used to describe such things as data encoding, encryption, message priority, and so on.
8. The data field is an integer number of bytes from a minimum of 46 to a maximum of 1500.

9. The parity check field houses the parity bits which are generated by the following generating polynomial (see Section 5.6):

$$X^{32} + X^{26} + X^{23} + X^{22} + X^{14} + X^{12} + X^{11} + X^{10} \\ + X^8 + X^7 + X^5 + X^4 + X^2 + X + 1$$

The Ethernet multiple access algorithm defines the following user action or response:

1. *Defer*. The user must not transmit when the carrier is present or within the minimum packet spacing time.
2. *Transmit*. The user may transmit if not deferring until the end of the packet or until a collision is detected.
3. *Abort*. If a collision is detected, the user must terminate packet transmission and transmit a short jamming signal to ensure that all collision participants are aware of the collision.
4. *Retransmit*. The user must wait a random delay time (similar to the ALOHA system) and then attempt retransmission.
5. *Backoff*. The delay before the n th attempt is a uniformly distributed random number from 0 to $2^n - 1$, for $(0 < n \leq 10)$. For $n > 10$, the interval remains 0 to 1023. The unit of time for the retransmission delay is 512 bits (51.2 μ s).

Figure 9.41b illustrates a 10-Mbits/s data stream with Manchester PCM formatting from the Ethernet specification. Notice that with such formatting, each bit cell or bit position contains a transition. A binary one is characterized by transitioning from a low level to a high level, while a binary zero has the opposite transition. Therefore, the presence of data transitions denotes to all "listeners" that the carrier is present. If a transition is not seen between 0.75 and 1.25 bit times since the last transition, the carrier has been lost, indicating the end of a packet.

9.5.2 Token-Ring Networks

A carrier-sense network consists of a cable onto which all stations are passively connected. A *ring network*, by comparison, consists of a series of point-to-point cables between consecutive stations. The interfaces between the ring and the stations are active rather than passive. Figure 9.42a illustrates a typical unidirectional ring with interface connections to several stations. Figure 9.42b illustrates the state of the interface for the listen mode and the transmit mode. In the *listen mode* the input bits are copied to the output with a delay of one bit time. In the *transmit mode*, the connection is broken so that the station can enter its own data onto the ring. The token is defined as a special bit pattern (e.g., 1 1 1 1 1 1 1) which circulates on the ring whenever all stations are idle. How does the system ensure that message data do not contain a tokenlike sequence? *Bit stuffing* is used to prevent this pattern from occurring in the data. For the

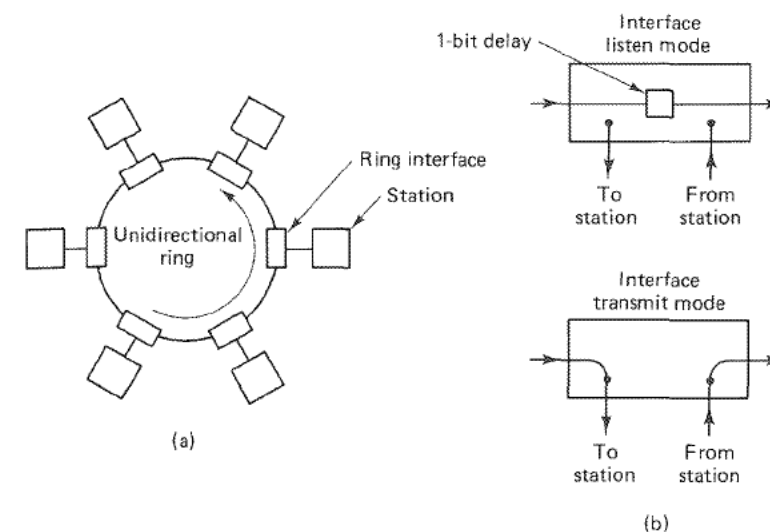


Figure 9.42 Token-ring network. (a) Network. (b) Listen and transmit modes.

8-bit token example shown, a bit-stuffing algorithm would insert a zero into the data stream after each sequence of seven consecutive ones. The data receiver would use a similar algorithm to dispose of the inserted bit following any sequence of seven consecutive ones. The token-ring access scheme works as follows:

1. A station wanting to send a message monitors the token appearing at the interface. When the last bit of the token appears, the station inverts it (e.g., 1 1 1 1 1 1 0). The station then breaks the interface connection and enters its own data onto the ring.
2. As bits come back around the ring, they are removed by the sender. There is no limit on the size of the packets, because the entire packet never appears on the ring at one instant.
3. After transmitting the last bit of its message, the station must regenerate the token. After the last data bit has circled the ring and has been removed, the interface is switched back to the listen mode.
4. Contention is not possible with a token-ring system. During heavy traffic, as soon as a token is regenerated, the next downstream station requiring service will see and remove the token. Thereby, permission to transmit rotates smoothly around the ring. Since there is only one token, there is no contention.

The ring itself must have sufficient delay to enable a complete token to circulate when all stations are idle. A major issue in ring network design is the propagation distance or "length" of a bit. If the data rate is R Mbits/s, a bit is emitted every $(1/R)$ microseconds. Since the propagation rate along a typical coaxial cable is 200 m/ μ s, each bit occupies $200/R$ meters on the ring.

Example 9.4 Minimum Ring Size

If an 8-bit token is to be used on a 5-Mbits/s token-ring network, calculate the minimum propagation distance, d_p , needed for the ring circumference. Assume that the propagation velocity v_p is 200 m/ μ s.

Solution

$$R = 5 \text{ Mbits/s}$$

Time to emit one bit, t_b :

$$t_b = \frac{1}{5 \times 10^6} \text{ s}$$

Time to emit the 8-bit token, t_t :

$$t_t = \frac{8}{5 \times 10^6} \text{ s}$$

Propagation distance for the 8-bit token:

$$\begin{aligned} d_p &= t_t \times v_p \\ &= \frac{8}{5} \mu\text{s} \times 200 \text{ m}/\mu\text{s} \\ &= 320 \text{ m} \end{aligned}$$

9.5.3 Performance Comparison of CSMA/CD and Token-Ring Networks

Figure 9.43 compares the delay-throughput characteristics of a CSMA/CD network with a token-ring network. In each case, the cable length is 2 km, there are 50 stations on the network, the average packet length is 1000 bits, and the header size is 24 bits. Figure 9.43a, the case where the transmission rate is 1 Mbits/s, illustrates that under these assumptions, CSMA/CD and token ring perform almost equally well. In Figure 9.43b, only one parameter has been changed as compared to Figure 9.43a; the transmission rate was increased to 10 Mbits/s. The difference for CSMA/CD is considerable; for normalized throughput, $\rho < 0.22$, CSMA/CD performs better than token ring. However, for $\rho > 0.22$, token ring clearly manifests better delay-throughput characteristics. To understand the reason for the poor CSMA/CD performance in Figure 9.43b, let us review the definition of ρ , described in Equations (9.17) and (9.19) and shown as

$$\rho = \frac{b\lambda}{R} = \frac{\rho'}{R}$$

where $\rho' = b\lambda$ is channel throughput in bits per second and R is the channel capacity (maximum transmission bit rate). As R increases, channel throughput must increase accordingly for a given value of ρ . At higher channel throughput rates a significant portion of the CSMA/CD transmission attempts ends in collision [26].

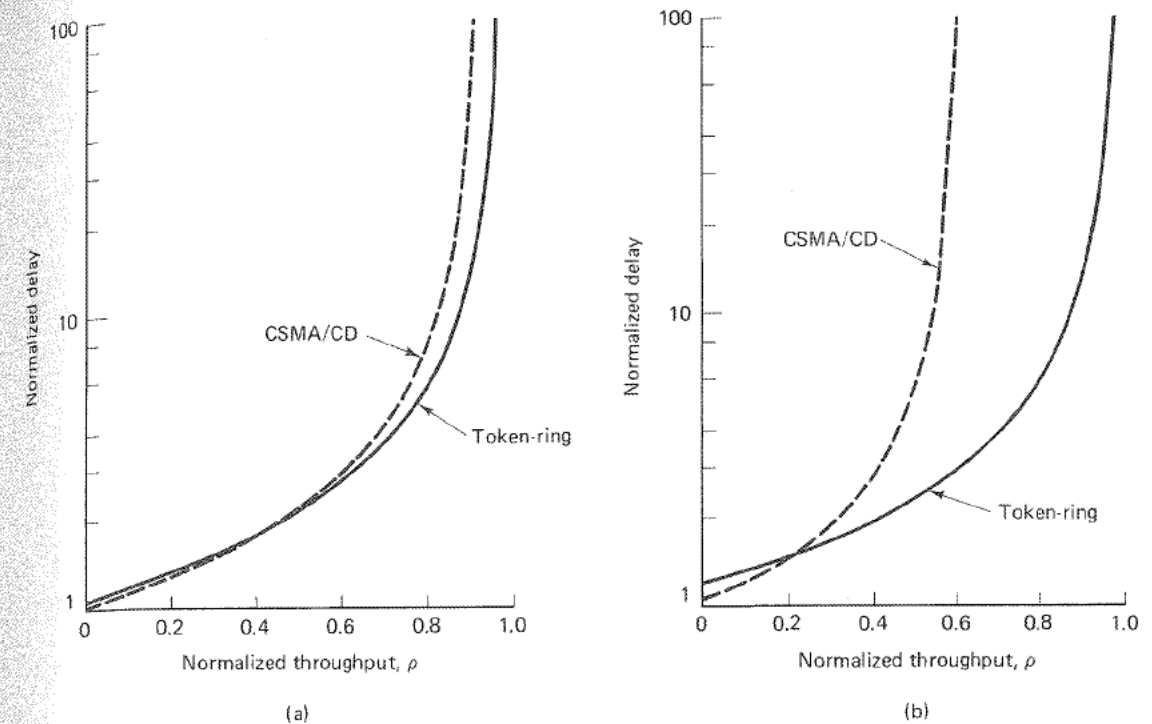


Figure 9.43 Delay versus throughput performance for CSMA/CD and token-ring networks. (a) Transmission rate = 1 Mbits/s. (b) Transmission rate = 10 Mbits/s. (Reprinted with permission from W. Bux, "Local-Area Subnetworks: A Performance Comparison," *IEEE Trans. Commun.*, vol. COM29, no. 10, Oct. 1981, pp. 1465–1473. © 1981 IEEE.)

9.6 CONCLUSION

In this chapter we have outlined the concepts of resource sharing. The classical approaches of FDM/FDMA and TDM/TDMA were discussed in some detail. We also described a hybrid multiple access technique called CDMA, and introduced some of the satellite multiple access techniques that became popular in the 1970s and 1980s, known as multiple-beam frequency reuse and dual-polarization frequency reuse.

We described the demand-assignment (DAMA) techniques in the context of several versions of the ALOHA algorithm, and we considered several of the multiple-access techniques employed with INTELSAT, such as FDM/FM, SPADE, TDMA, and SS/TDMA. Finally, we examined two popular algorithms used for local area networks: carrier-sense multiple access with collision detection (CSMA/CD) and a token-ring network. The goals of the chapter were to introduce an assortment of multiple access techniques rather than attempting a rigorous treatment of any of them.

REFERENCES

1. Rubin, I., "Message Delays in FDMA and TDMA Communication Channels," *IEEE Trans. Commun.*, vol. COM27, no. 5, May 1979, pp. 769-777.
2. Nirenberg, L. M., and Rubin, I., "Multiple Access System Engineering—A Tutorial," *IEEE WESCON/78 Professional Program*, Modern Communication Techniques and Applications, session 21, Los Angeles, Sept. 13, 1978.
3. Abramson, N., "The ALOHA System—Another Alternative for Computer Communications," *Proc. Fall Joint Comput. Conf. AFIPS*, vol. 37, 1970, pp. 281-285.
4. Hayes, J. F., "Local Distribution in Computer Communications," *IEEE Commun. Mag.*, Mar. 1981, pp. 6-14.
5. Schwartz, M., *Computer-Communication Network Design and Analysis*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1977.
6. Tanenbaum, A. S., *Computer Networks*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1981.
7. Abramson, N., "The ALOHA System," in N. Abramson and F. F. Kuo, eds., *Computer Communication Networks*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1973.
8. Kleinrock, L., *Queueing Systems*, Vol. 1, *Theory*, John Wiley & Sons, Inc., New York, 1975.
9. Abramson, N., "Packet Switching with Satellites," *AFIPS Conf. Proc.*, vol. 42, June 1973, pp. 695-702.
10. Rosner, R. D., *Packet Switching*, Lifelong Learning Publications, Wadsworth Publishing Company, Inc., Belmont, Calif., 1982.
11. Crowther, W., Rettberg, R., Walden, D., Ornstein, S., and Heart, F., "A System for Broadcast Communication: Reservation ALOHA," *Proc. Sixth Hawaii Int. Conf. Syst. Sci.*, Jan. 1973, pp. 371-374.
12. Roberts, L., "Dynamic Allocation of Satellite Capacity through Packet Reservation," *AFIPS Conf. Proc.*, vol. 42, June 1973, p. 711.
13. Binder, R., "A Dynamic Packet-Switching System for Satellite Broadcast Channels," *Proc. Int. Conf. Commun.*, June 1975, pp. 41-1-41-5.
14. Capetanakis, J., "Tree Algorithms for Packet Broadcast Channels," *IEEE Trans. Inf. Theory*, vol. IT25, Sept. 1979, pp. 505-515.
15. Puente, J. G., and Werth, A. M., "Demand-Assigned Service for the INTELSAT Global Network," *IEEE Spectrum*, Jan. 1971, pp. 59-69.
16. Jones, J. J., "Hard Limiting of Two Signals in Random Noise," *IEEE Trans. Inf. Theory*, vol. IT9, Jan. 1963, pp. 34-42.
17. Bond, F. E., and Meyer, H. F., "Intermodulation Effects in Limiter Amplifier Repeaters," *IEEE Trans. Commun. Technol.*, vol. COM18, no. 2, Apr. 1970, pp. 127-135.
18. Shimbo, O., "Effects of Intermodulation, AM-PM Conversion, and Additive Noise in Multicarrier TWT Systems," *Proc. IEEE*, vol. 59, Feb. 1971, pp. 230-238.
19. Chakraborty, D., "INTELSAT IV Satellite System (Voice) Channel Capacity versus Earth-Station Performance," *IEEE Trans. Commun. Technol.*, vol. COM19, no. 3, June 1971, 355-362.
20. Campanella, S., and Schaefer, D., "Time Division Multiple Access Systems

(TDMA)," in K. Feher, *Digital Communications, Satellite/Earth Station Engineering*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1983.

21. Scarcella, T., and Abbott, R. V., "Orbital Efficiency Through Satellite Digital Switching," *IEEE Commun. Mag.*, May 1983, pp. 38-46.
22. Muratani, T., "Satellite-Switched Time-Domain Multiple Access," *Proc. IEEE Electron. and Aerosp. Conf. (EASCON)*, 1974, pp. 189-196.
23. Dill, G. D., "TDMA, The State-of-the-Art," *Rec. IEEE Electron. Aerosp. Syst. Conv. (EASCON)*, Sept. 26-28, 1977, pp. 31-5A-31-5I.
24. Jarett, K., "Operational Aspects of Intelsat VI Satellite-Switched TDMA Communication System," *AIAA Tenth Commun. Satell. Syst. Conf.* Mar. 1984, pp. 107-111.
25. Stallings, W., "Local Network Performance," *IEEE Commun. Mag.*, vol. 22, No. 2, Feb. 1984, pp. 27-36.
26. Bux, W., "Local-Area Subnetworks: A Performance Comparison," *IEEE Trans. Commun.*, vol. COM29, no. 10, Oct. 1981, pp. 1465-1473.
27. Dixon, R. C., Strole, N. C., and Markov, J. D., "A Token-Ring Network for Local Data Communications," *IBM Syst. J.*, vol. 22, no. 1-2, 1983, pp. 47-62.

PROBLEMS

- 9.1. Design an FDM signal set consisting of five voice channels, each in the frequency range 300 to 3400 Hz. The multiplexed composite is to be made up of inverted sidebands and is to occupy the spectral region from 30 to 50 kHz.
 - (a) Draw the composite spectrum, indicating individual spectrum and guard band frequency locations.
 - (b) Draw a block diagram showing the heterodyning and filtering details and the required local oscillator values.
- 9.2. A receiver is tuned to receive the lower sideband (LSB) of a radio-frequency (RF) carrier wave with frequency, $f_c = 8$ MHz. The bandwidth of the LSB signal is 100 kHz. The receiver employs a local oscillator (LO) with frequency, f_{LO} , for heterodyning the received signal down to a lower intermediate frequency (IF). Assume that $f_{LO} > f_c$, and that the IF amplifier is centered at 2 MHz. Draw a block diagram of the heterodyning conversion, including the RF filter, the LO, and the IF filter. Indicate the center frequency of each filter and typical spectra of the signals at various points in the diagram.
- 9.3. Equations (9.13) to (9.15) demonstrate that the average message delay time for TDMA is less than that for FDMA. Discuss the practical benefits of such reduced delay in TDMA, as a function of frame time, for a satellite link with a one-way range of 36,000 km. For what values of frame time can there be a significant advantage of TDMA over FDMA?
- 9.4. A group of stations share a 56-kbits/s pure ALOHA channel. Each station outputs a packet on the average of once every 10 s, even if the previous one has not yet been sent (i.e., the stations buffer the packets). Each packet is comprised of 3000 bits. What is the maximum number of stations that can share this channel, assuming that the arrival process is Poisson?
- 9.5. A group of three stations share a 56-kbits/s pure ALOHA channel. The average bit

rate transmitted from each of the three stations is $R_1 = 7.5$ kbits/s, $R_2 = 10$ kbits/s, and $R_3 = 20$ kbits/s. The size of each packet is 100 bits/packet. Find the normalized total traffic on the channel, the normalized throughput, the probability of successful transmission, and the arrival rate of successful packets. Assume that the arrival process is Poisson.

- 9.6. Verify that for a pure ALOHA access scheme, the normalized throughput is bounded by $1/2e$ and that this maximum occurs when the normalized total traffic is equal to 0.5.
- 9.7. (a) Verify that Equation (9.24) is a valid probability density function (pdf) for a discrete random variable.
 (b) Calculate the mean of a discrete random variable having a pdf like the one given in Equation (9.24).
 (c) Show that your result in part (b) is consistent with the claim that λ is the average packet arrival rate.
- 9.8. Consider the pure ALOHA arrival scenario shown in Figure P9.1. The vertical arrows indicate packet arrival times. N_n is the number of arriving packets in the time interval $(T_{n-1}, T_n]$, where $(t_x, t_y]$ indicates the interval $t_x < t \leq t_y$. N_{n+1} is the number of arriving packets in $(T_n, T_{n+1}]$, and τ is the time duration per packet in seconds. The average arrival rate is λ . Assume the arrivals are independent of each other.
 (a) Write an expression for the joint pdf of N_n and N_{n+1} .
 (b) Let T_n define the time at which user A's packet arrives. Express, in terms of the joint pdf of N_n and N_{n+1} , the probability that user A's transmission will be successful.

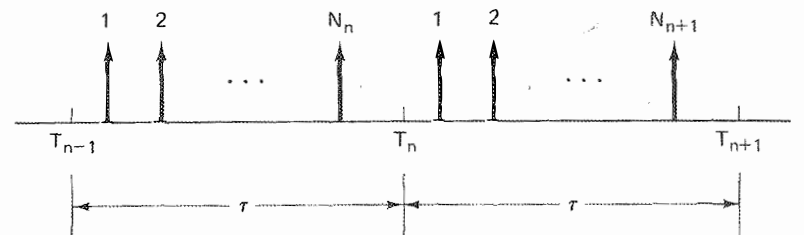


Figure P9.1

- 9.9. Let $N = N_n + N_{n+1}$, where N_n and N_{n+1} are as defined in Problem 9.8. Write an expression for the pdf of N , and give an interpretation for N .
- 9.10. Six thousand stations are competing for the use of a single slotted ALOHA channel. The average station makes 30 requests per hour, where each request is for one slot of 500- μ s duration. Calculate the normalized total traffic on the channel.
- 9.11. Consider the arrival scenario of Figure P9.1; the location of the packet arrival times are permissible as shown under pure ALOHA, but not under slotted ALOHA, where arrivals are permitted only at the discrete times T_i , where $i = 0, 1, \dots$. Assume that the average arrival rate is λ .
 (a) How would Figure P9.1 need to be modified if slotted ALOHA is used? How would the pdfs of N_n and N_{n+1} change?
 (b) If user A's packet arrives at time T_n , what is the probability of successful transmission?
- 9.12. A group of slotted-ALOHA stations generate a total of 120 requests per second,

including both original and retransmissions. Each request is for a 12.5-ms duration slot.

- (a) What is the normalized total traffic on the channel?
 (b) What is the probability of a successful transmission on the first attempt?
 (c) What is the probability of exactly two collisions before a successful transmission?
- 9.13. Measurements of a slotted-ALOHA channel show that 20% of the slots are idle.
 (a) What is the normalized total traffic on the channel?
 (b) What is the normalized throughput?
 (c) Is the channel underloaded or overloaded?
- 9.14. Show that the sum of two Poisson processes, with rates λ_1 and λ_2 , is also a Poisson process, with rate $\lambda = \lambda_1 + \lambda_2$. Generalize your result for the sum of n Poisson processes.
- 9.15. A 10-MHz transponder is occupied by 200 identical carriers, half servicing stations with $G/T = 40$ dB/K, the other half servicing stations with $G/T = 37$ dB/K. All stations have a requirement to operate with a bit error probability of 10^{-5} . The transponder is power limited under this configuration.
 (a) What is the maximum possible bandwidth for each carrier?
 (b) Suppose that each carrier has a bandwidth of 40 kHz, and the transponder is required to service a group of larger ($G/T = 40$ dB/K) stations only. How many stations can the transponder handle? Will the transponder be power or bandwidth limited?
 (c) Repeat part (b) for the case where the transponder is to service a group of small ($G/T = 37$ dB/K) stations only.
- 9.16. A TDMA system operates at 100 Mbits/s with a 2-ms frame time. Assume that all slots are of equal length and that a guard time of 1 μ s is required between slots.
 (a) Compute the efficiency of the communications resource (CR) for the case of 1, 2, 5, 10, 20, 50, and 100 slots per frame.
 (b) Repeat part (a) assuming that a 100-bit preamble is required at the start of each slot. Compute the efficiency of the CR in terms of the desired information transmission.
 (c) Graph the results of parts (a) and (b).
- 9.17. With reference to Equation (9.36):
 (a) Discuss the efficiency of the CR use if all S_i and R_j are equal.
 (b) Discuss the effect of a few S_i or R_j being much larger than the majority. How can the efficiency of the CR be improved?
 (c) When are the distributions of S_i and R_j likely to be similar? Dissimilar?
- 9.18. (a) Consider a token-ring network operating at a transmission rate of 10 Mbits/s over a cable having a propagation velocity of 200 m/ μ s. How many meters of cable is equal to a delay of 1 bit at each ring interface?
 (b) If the token is 10 bits long, and all but three stations are switched off during evening hours, what is the minimum cable length needed for the ring?