THE

Satellite Communication
Applications Handbook

SECOND EDITION · BRUCE R. ELBERT

The Satellite Communication Applications Handbook

Second Edition

The Satellite Communication Applications Handbook

Second Edition

Bruce R. Elbert

To Cathy, my wife and parnter

Contents

Preface

The first edition of *The Satellite Communication Applications Handbook* established an important milestone in industry publications by defining the different application segments and providing up-to-date design and development information. As with any handbook, a sufficient percentage of the material lost its timeliness not long after the start of the new millennium. It was imperative, therefore, to update and expand its content to reflect the changes in application focus and industry structure. We did this in a way to preserve the methodical approach of the first edition while introducing a considerable amount of new technical and application information that has been gained through more recent experience and research. The handbook is intended for anyone interested in satellite communications, whether an active member of the industry or someone considering entry into one of its segments. The book can be read sequentially so as to follow the thread of developing ideas and processes, or it can be used as a reference on any of the specific topics, outlined next. A technical background, while helpful, is not necessary for understanding the principles and the majority of concepts in this book.

Throughout the 1990s, the satellite communication industry experienced tremendous growth, surpassing the expectations of all who have contributed to its success. The gross revenues in 2000 reached $60 billion, big chunks of which were contributed by satellite manufacture, launch, satellite transponder sales and leases, ground equipment supply, and direct-to-home (DTH) TV and very small aperture terminal (VSAT) data networks. This book provides a comprehensive review of the applications that have driven this growth. It discusses the technical and business aspects of the systems and services that operators and users exploit to make money, serve and protect, and even have fun.

The book is organized into four parts, which deal with the most fundamental areas of concern to application developers and users: the technical and business fundamentals, the application of simplex (broadcast) links to multiple users, duplex links that deliver two-way interactive services, and regulatory and business affairs that drive investment and financial performance. The 13 chapters of the book fall nicely into these general categories.

Chapters 1 through 6 follow the first edition rather closely—they have been changed only to account for some of the new features developed over the intervening 7 years. Part I consists of the first three chapters. Chapters 1 and 2 provide the basis for designing any satellite communications application, which includes finding the most appropriate structure for and suppliers of systems and technology. As in the first edition, Chapter 2 takes the reader through the entire process of designing a satellite link with the methodology of the link budget (explained line by line). Issues

for the space segment are covered in Chapter 3 and now include details on both analog (bent-pipe) and digital onboard processing repeaters. The reason we include this here is because of the close tie between the application and the construction of the satellite repeater, particularly if it is of the digital processing variety.

Chapters 4 through 6 (Part II) are presented as in the first edition to review the scope and detail of creating a satellite television application and system. The basics are covered in Chapter 4 from the standpoint of service possibilities: entertainment TV for local TV stations and cable, videoconferencing and business video, and distance learning. Chapter 5 covers the range of digital TV standards such as MPEG 2 and the H series of the International Telecommunication Union (ITU) standards. This provides the base for Chapter 6, which deals with the largest single application segment in our industry—DTH television broadcasting.

New to the handbook (Chapter 7, also in Part II) is the application called Digital Audio Radio Service (DARS), now an established service in the United States thanks to XM Satellite Radio and Sirius Satellite Radio. Borne out of the innovative World-Space system that provides satellite radio programming to Africa, DARS is beginning to have the same strategic impact on terrestrial AM and FM radio as DTH had on cable and over-the-air TV. Part III consists of Chapters 8 through 11 and deals with two-way interactive applications for data and voice. Two chapters, rather than one, are now devoted to the important topic of VSAT networks for provision of two-way interactive data communications. Focusing on Internet-based services (e.g., IP networks), Chapters 8 and 9 cover the enhanced capabilities of satellite-delivered interactive data to homes and businesses. Chapter 8 reviews the uses of star and mesh VSAT networks for various applications, and Chapter 9 provides technical criteria and guidelines for how a VSAT network is sized and optimized.

Chapters 10 through 13 follow the same content flow as Chapters 8 through 11 in the first edition. In Chapter 10, which covers fixed telephony networks, we have added material on the all-important topic of voice over IP (VoIP) over satellites. This adds to the foundation of satellite telephony for providing basic communications in remote locations and for temporary operations. Mobile telephony is covered in Chapter 11, from both geostationary Earth orbit (GEO) and non-GEO perspectives. Most of the Mobile Satellite Service (MSS) providers continue to use GEO satellite platforms to extend service beyond ships to include handheld devices and IP-based satellite modems. The technical and operational issues of providing MSS applications are covered in detail in this chapter.

To conclude the second edition, we provide updated regulatory and business guidance in Chapters 12 and 13, respectively (Part IV). The procedures and issues surrounding how one obtains a satellite orbit slot and Earth station license are covered in Chapter 12. In some ways, the process has been simplified, such as with the 2001 edition of the *Radio Regulations* of the International Telecommunication Union. Issues of gaining access and licenses in specific countries continue to be a challenge, and so we cover this topic to give readers a head start in the process. Finally, the business of satellite communication is described in Chapter 13, where the industry is divided up by the elements of a typical satellite application. This gives developers of new applications a framework for organizing and managing the process of going from the idea to a revenue-generating resource or entire network.

Anyone entering this exciting field at this time has many options to consider and many avenues to follow. Fortunately, there is a great deal of useful information and experience available to anyone who wishes to do the research and explore its many dimensions. The origin of this book comes from the author's journey of more than 30 years as an independent consultant and educator, at Hughes Electronics, COMSAT, Western Union, and the U.S. Army Signal Corps (where one really learns how to communicate). Teachers and other presenters may contact the author by e-mail at bruce@applicationstrategy.com for additional help in using this book as a text for a technical or business course on satellite communication.

PART I

System Considerations

Evolution of Satellite Technology and Applications

Communication satellites, whether in geostationary Earth orbit (GEO) or non-GEO, provide an effective platform to relay radio signals between points on the ground. The users who employ these signals enjoy a broad spectrum of telecommunication services on the ground, at sea, and in the air. In recent years, such systems have become practical to the point where a typical household can have its own satellite dish. That dish can receive a broad range of television programming and provide broadband access to the Internet. These satellite systems compete directly in some markets with the more established broadcasting media, including over-the-air TV and cable TV, and with high-speed Internet access services like digital subscriber line (DSL) and cable modems. In addition, GEO and non-GEO satellites will continue to offer unique benefits for users on the go with such mobile services as two-way voice and data, and digital audio broadcasting. The accelerated installation of undersea fiber optics that accompanied the Internet and telecom boom of the late 1990s put more capacity into service than markets could quickly absorb. Curiously, these new operators claimed that satellites were obsolescent. Quite to the contrary, satellite communication continues to play an increasing role in backbone networks that extend globally. Just how well we employ satellites to compete in markets depends on our ability to identify, develop, and manage the associated networks and applications.

To this end, this book shows how satellite technology can meet a variety of human needs, the ultimate measure of its effectiveness. My first work, *Introduction to Satellite Communication* [1], established the foundation for the technology and its applications. These have progressed significantly since the late 1980s; however, the basic principles remain the same. Satellite communication applications (which we will refer to as simply satellite applications) extend throughout human activity—both occupational and recreational. Many large companies have built their communications foundations on satellite services such as cable TV, direct-to-home broadcasting satellite (DBS), private data networks, information distribution, maritime communications, and remote monitoring. For others, satellites have become a hidden asset by providing a reliable communications infrastructure. Examples abound in their use for disaster relief by the Red Cross and other such organizations, and for instant news coverage from areas of conflict. In the public and military sectors, satellite applications are extremely effective in situations where terrestrial lines and portable radio transceivers are not available or ineffective for a variety of reasons.

We can conclude that there are two basic purposes for creating and operating satellite applications, namely, to make money from selling systems and services (efficient communications) and to meet vital communications needs (essential communications). The composition of satellite communication markets has changed over the years. Initially, the primary use was to extend the worldwide telephony net. In the 1980s, video transmission established itself as the hottest application, with data communications gaining an important second place position. Voice services are no longer the principal application in industrialized countries but retain their value in rural environments and in the international telecommunications field. Special-purpose voice applications like mobile telephone and emergency communications continue to expand. The very fact that high-capacity fiber optic systems exist in many countries and extend to major cities worldwide makes satellite applications that much more important as a supplementary and backup medium. Satellites are enjoying rapid adoption in regions where fixed installations are impractical. For example, ships at sea no longer employ the Morse code because of the success of the Inmarsat system. And people who live in remote areas use satellite dishes rather than large VHF antenna arrays to receive television programming.

Satellite operators, which are the organizations that own and operate satellites, must attract a significant quantity of users to succeed as a business. As illustrated in Figure 1.1, the fixed ground antennas that become aligned with a given satellite or constellation create synergy and establish a "real estate value" for the orbit position. Some of the key success factors include the following:

- The best orbit positions (for GEO) or orbital constellation (for non-GEO);
- The right coverage footprint to reach portions of the ground where users exist or would expect to appear;
- Service in the best frequency bands to correspond to the availability of low-cost user terminal equipment;

Figure 1.1 A neighborhood created by a GEO satellite with many fixed antennas aligned with it.

- Satellite performance in terms of downlink radiated power and uplink receive sensitivity;
- Service from major Earth stations (also called teleports) for access to the terrestrial infrastructure, particularly the Public-Switched Telephone Network (PSTN), the Internet, and the fiber backbone;
- Sufficient funding to get the system started and operating at least through a cash-flow break-even point.

Optimum footprint and technical performance allow a satellite to garner an attractive collection of markets. Importantly, these do not necessarily need to be known with precision when the satellite is launched because new users and applications can start service at any time during the operating lifetime of the satellite (typically 15 years). Anywhere within the footprint, a new application can be introduced quickly once ground antennas are installed. This provides what is called high operating leverage—a factor not usually associated with buried telecom assets such as fiber optic cables and wireless towers.

Ultimately, one can create a *hot bird* that attracts a very large user community of antennas and viewers. Galaxy I, the most successful cable TV hot bird of the 1980s, established the first shopping center in the sky, with anchor tenants like HBO and ESPN and boutiques like Arts & Entertainment Channel (A&E) and The Discovery Channel. Many of the early boutiques have become anchors, and new boutiques, like The Food Network and History International, arrive to establish new market segments. New hot birds develop as well, such as Astra 1 in Europe and AsiaSat 3S in Asia. Users of hot birds pay a premium for access to the ground infrastructure of cable TV and DBS receiving antennas much like tenants in a premium shopping mall pay to be in an outstanding location and in proximity to the most attractive department stores in the city. In the case of cable TV, access is everything because the ground antenna is, in turn, connected to households where cable services are consumed and paid for. DBS delivers direct access to subscribers, bypassing cable systems. For a new satellite operator to get into an established market often requires them to subsidize users by paying some of the switching costs out of expected revenues. From this experience, those who offer satellite services to large user communities know that the three most important words in satellite service marketing are LOCATION, LOCATION, and LOCATION! This refers to the factors previously listed. Stated another way, it is all about connectivity to the right user community.

Satellite operators, who invest in the satellites and make capacity available to their customers, generally prefer that users own their own Earth stations. This is because installing antennas and associated indoor electronics is costly for satellite service providers. Once working, this investment must be maintained and upgraded to meet evolving needs. On the other hand, why would users want to make such a commitment? There are two good reasons for this trend toward ownership of the ground segment by the user: (1) the owner/user has complete control of the network resources, and (2) the cost and complexity of ownership and operation have been greatly reduced because of advances in microcircuitry and computer control. A typical small Earth station is no more complex than a cellular telephone or VCR. As a result of strong competition for new subscribers, DBS and the newer S-DARS have

to subsidize receiver purchases. Larger Earth stations such as TV uplinks and international telephone gateways are certainly not a consumer item, so it is common for several users to share a large facility in the form of a teleport.

User organizations in the public and private sectors that wish to develop their own unique satellite networks have a wide array of tools and technologies at their disposal (which are reviewed in detail in this book). One need not launch and operate satellites as on-orbit capacity may be taken as a service for as long or as short a period as needed. On the other hand, it can be bewildering when one considers the complexity of the various satellite systems that could potentially serve the desired region and community. The associated Earth stations and user terminals must be selected, purchased, installed, and properly integrated with applications and other networks that they access. Happily for the new user, there are effective methodologies that address this complexity and thereby reduce risk and potentially cost. Satellite communications can also reduce entry barriers for many information industry applications. As a first step, a well-constructed business plan based on the use of existing satellites could be attractive to investors. (More on finance can be found in Chapter 11.)

The history of commercial satellite communications includes some fascinating startup services that took advantage of the relatively low cost of entry. The following three examples illustrate the range of possibilities. The Discovery Channel made the substantial commitment to a Galaxy I C-band transponder and thereby gained access to the most lucrative cable TV market in North America. Another startup, Equatorial Communications, pioneered very small aperture terminal (VSAT) networks to deliver financial data to investors. Their first receive-only product was a roaring success, and in 1985 the company became the darling of venture capitalists. Unfortunately, they broke their sword trying to move into the much more complicated two-way data communication market. Their technology failed to gain acceptance, and the company disappeared through a series of mergers. SpeedCast was founded in Hong Kong in 2000 to allow content providers and information services to overcome the limited broadband infrastructure in the Asia-Pacific region. Utilizing existing C-band capacity on AsiaSat 3C, SpeedCast built the needed hub in Hong Kong at the terminus of broadband capacity on a trans-Pacific fiber optic cable.

Several U.S. corporations attempted to introduce DTH satellite broadcasting at a time when cable TV was still establishing itself. The first entrants experienced great difficulties with limited capacity of existing low- and medium-power Ku-band satellites, hampering the capacity of the networks and the affordability of the home receiving equipment. Europe and Japan had problems of their own in finding the handle on viable DTH systems, choosing first to launch high-power Ku-band satellites with only a few operating channels. It was not until BSkyB and NHK were able to bring attractive programming to the public exclusively on their respective satellites that consumers moved in the millions of numbers.

In the United States, the only viable form of DTH to emerge in the 1980s was through the backyard C-band satellite dish that could pull in existing cable TV programming from hot birds like Galaxy I and Satcom 3R. In the 1980s there were already millions of C-band receive dishes in North America. This clearly demonstrated the principle that people would vote with their money for a wide range of attractive programming, gaining access to services that were either not available or priced out of reach. Early adopters of the dishes purchased these somewhat

expensive systems because the signals were not scrambled at the time. A similar story can be told for Asia on the basis of Star TV, which continues to provide advertiser-supported C-band satellite television to the broad Asian market. HBO and other cable networks in the United States changed the equation markedly when they scrambled their programming using the Videocipher 2 system, resulting in a halt to the expansion of backyard dishes. This market settled back into the doldrums for several years. In today's world, C-band home dishes are rare in the United States and Europe but have a significant following in tropical regions that effectively employ this band.

In 1994, Hughes Electronics introduced its DIRECTV service through three high-power satellites colocated at 101° EL (all receivable by a single Ku-band home dish). With more than 150 digitally compressed TV channels, DIRECTV demonstrated that DTH could be both a consumer product and a viable alternative to cable. As an important footnote, DIRECTV shared one of the satellites with another company called USSB; however, the latter was subsequently bought out to aggregate all programming under one trademark. An older competing service, PrimeStar, was first introduced by TCI and other cable operators as a means to serve users who were beyond the reach of their cable systems. DIRECTV moved to acquire this competitor, resulting in a quantum increase of subscribers. A single competitor remained in the form of EchoStar with their DISH Network. DIRECTV was first to be acquired by DISH, but as a result of U.S. government objections, the acquirer would be News Corp.

Satellite communication applications can establish a solid business for companies that know how to work out the details to satisfy customer needs. A stellar example is the mobile satellite service business pioneered by Inmarsat. Through a conservatively managed strategy, Inmarsat has driven its service from initially providing ship-to-shore communications to being the main source of emergency and temporary communications on land. Whether we are talking about reporters covering a conflict in southern Asia or the provision of disaster relief in eastern Europe, lightweight Inmarsat terminals fit the need.

1.1   Satellite Network Fundamentals

Every satellite application achieves its effectiveness by building on the strengths of the satellite link. A satellite is capable of performing as a microwave repeater for Earth stations that are located within its coverage area, determined by the altitude of the satellite and the design of its antenna system. The arrangement of three basic orbit configurations is shown in Figure 1.2. A GEO satellite can cover nearly one-third of the Earth's surface, with the exception of the polar regions. This includes more than 99% of the world's population and economic activity.

The low Earth orbit (LEO) and medium Earth orbit (MEO) approaches require more satellites to achieve this level of coverage. Due to the fact that non-GEO satellites move in relation to the surface of the Earth, a full complement of satellites (called a constellation) must be operating to provide continuous, unbroken service. The trade-off here is that the GEO satellites, being more distant, incur a longer path length to Earth stations, while the LEO systems promise short paths not unlike

Figure 1.2   The three most popular orbits for communication satellites are LEO, MEO, and GEO. The respective altitude ranges are 500 to 900 km for LEO, 5,000 to 12,000 km for MEO, and 36,000 km for GEO. Only one orbit per altitude is illustrated, even though there is a requirement for constellations of LEO and MEO satellites to provide continuous service. The standard GEO orbit is perfectly circular and lies in the plane of the equator; other 24-hour orbits are inclined and/or elliptical rather than circular.

those of terrestrial systems. The path length introduces a propagation delay since radio signals travel at the speed of light. This is illustrated in Figure 1.3, which is a plot of orbit period and propagation delay for various altitudes. Depending on the nature of the service, the increased delay of MEO and GEO orbits may impose some degradation on quality or throughput. The extent to which this materially affects the acceptability of the service depends on many factors, such as the degree of interactivity, the delay of other components of the end-to-end system, and the protocols used to coordinate information transfer and error recovery. This is reviewed in detail in Part III of this book, which consists of Chapters 8–11.

Figure 1.3   A graph that plots orbit period in hours versus the mean altitude of the orbit in kilometers. One-way (single-hop) propagation delay is indicated at the top in milliseconds.

Three LEO systems have begun service since the publication of the first edition of this handbook: Orbcomm, Iridium, and Globalstar. Orbcomm was designed for two-way messaging service, while Iridium and Globalstar were designed for mobile telephony. Early advertising for Iridium suggested that with one of their handheld phones, you could be reached anywhere in the world. This would be the case only if you remained out of doors with a clear view of the sky from horizon to horizon. Globalstar had a slightly less ambitious claim that its service was cheaper than that of Iridium. While these systems could deliver services, all have resulted in financial failures for their investors. The non-GEO system that has yet to begin operation at the time of this writing is ICO Communications (ICO originally stood for interme-diate circular orbit, but that was subsequently dropped when they spun the com-pany off) and its MEO constellation. The developers of this system explain that their strategy does not rely on service to handheld telephones and such instruments, but rather is a means to provide near-broadband service to small terminals.

Except for Orbcomm, which is in VHF band, all of the satellites just discussed have microwave repeaters that operate over an assigned segment of the 1- to 80-GHz frequency range. As microwaves, the signals transmitted between the satellite and Earth stations propagate along line-of-sight paths and experience free-space loss that increases as the square of the distance. The spectrum allocations are given in the following approximate ranges, as practiced in the satellite industry:

- L-band: 1.5 to 1.65 GHz:
- S-band: 2.4 to 2.8 GHz;
- C-band: 3.4 to 7.0 GHz;
- X-band: 7.9 to 9.0 GHz;
- Ku-band: 10.7 to 15.0 GHz;
- Ka-band: 18.0 to 31.0 GHz;
- Q-band: 40 to 50 GHz;
- V-band: 60 to 80 GHz.

Actual assignments to satellites and Earth stations are further restricted in order to permit different services (and the associated user community) to share this valu-able resource. In addition to microwaves, laser systems continue to be under evalua-tion. Rather than being simple repeaters, laser links require modulated coherent light sources and demodulating receivers that include mutually tracking telescopes. An example of such a device is shown in Figure 1.4. So far, commercial laser links are not in use, but there is interest in them principally to allow direct connections between satellites—called intersatellite links or cross links.

Applications are delivered through a network architecture that falls into one of three categories: point-to-point (mesh), point-to-multipoint (broadcast), and mul-tipoint interactive (VSAT). Mesh-type networks mirror the telephone network. They allow Earth stations to communicate directly with each other on a one-to-one basis. To make this possible, each Earth station in the network must have sufficient transmit and receive performance to exchange information with its least effective partner. Generally, all such Earth stations have similar antennas and transmitter systems, so their network is completely balanced. Links between pairs of stations

Figure 1.4   Illustration of a laser intersatellite link by the Artemis satellite. (Courtesy of ESTECH.)

can be operated on a full-time basis for the transfer of broadband information like TV or multiplexed voice and data. Alternatively, links can be established only when needed to transfer information, either by user scheduling (reservation system) or on demand (demand-assignment system).

A broadcast of information by the satellite is more efficient than terrestrial arrangements using copper wires, fiber optic cables, or multiple wireless stations. By taking advantage of the broadcast capability of a GEO satellite, the point-to-multipoint network supports the distribution of information from a source (the hub/uplink Earth station) to a potentially very large number of users of that information (the remote Earth stations, also called receive-only terminals). Any application that uses this basic feature will usually find that a GEO satellite is its most effective delivery vehicle to reach a national audience.

Many applications employ two-way links, which may or may not use the broadcast feature. The application of the VSAT to interactive data communication applications has proven successful in many lines of business and more recently to the public. As will be covered in Chapter 8, a hub and spoke network using VSATs can be compared to almost any terrestrial wide-area network topology that is designed to accomplish the same result. This is because the satellite provides the common point of connection for the network, eliminating the requirement for a separate physical link between the hub and each remote point. Other interactive applications can employ point-to-point links to mimic the telephone network, although this tends to be favored for rural and mobile services. The incoming generation of satellite and ground equipment, which involves very low-cost VSATs, is reducing barriers to mass market satellite networks.

The degree to which satellite communications is superior to terrestrial alternatives depends on many interrelated factors. Experience has shown that the following features tend to give satellite communication an advantage in appropriate applications:

- Wide area coverage of a country, region, or continent;
- Wide bandwidth available throughout;
- Independent of terrestrial infrastructure;
- Rapid installation of ground network;

- Low cost per added site;
- Uniform service characteristics;
- Total service from a single provider;
- Mobile/wireless communication, independent of location.

While satellite communications will probably never overtake terrestrial tele-communications on a major scale, these strengths can produce very effective niches in the marketplace. Once the satellite operator has placed the satellite into service, a network can easily be installed and managed by a single organization. This is possible on a national or regional basis (including global using at least three GEO satellites). The frequency allocations at C-, Ku-, and Ka-bands offer effective bandwidths of 1 GHz or more per satellite, facilitating a range of broadband services that are not constrained by local infrastructure considerations. Satellites that employ L- and S- bands constrain bandwidth to less than 100 MHz but may propagate signals that bend around obstacles and penetrate nonmetallic structures. Regardless of the band, the satellite delivers the same consistent set of services at costs that are potentially lower than those of fixed terrestrial systems. For the long term, the ability to serve mobile stations and provide communications instantly are features that offer strength in a changing world.

Originally, Earth stations were large, expensive, and located in rural areas so as not to interfere with terrestrial microwave systems that operate in the same frequency bands. These massive structures had to use wideband terrestrial links to reach the closest city. Current emphasis is on customer premise Earth stations—simple, reliable, low cost. An example of a modern small VSAT is illustrated in Figure 1.5. Home receiving systems for DTH service are also low in cost and quite inconspicuous. The current generation of low-cost VSATs introduced since 2002 encourage greater use of bidirectional data communications via satellite. As terminals have shrunk in size, satellites have grown in power and sophistication. There are three general classes of satellites used in commercial service, each designed for a particular mission and capital budget. Smaller satellites, capable of launch by the Delta II rocket or dual-launched on the Ariane 4 or 5, provide a basic number of transponders usually in a single frequency band. Satellite operators in the United States, Canada, Indonesia, and China have established themselves in business through this class of satellite. The Measat satellite, illustrated in Figure 1.6, is an example of this class

Figure 1.5   Example of a VSAT for broadband communications. (Courtesy of Gilat Satellite Communications.)

Figure 1.6   The Measat 1 satellite provides services to Malaysia and throughout Southeast Asia.

of vehicle. The introduction of mobile service in the LEO involves satellites of this class as well. Moving up to the middle range of spacecraft, we find designs capable of operating in two frequency bands simultaneously. AsiaSat 3S, shown in Figure 1.7, provides 24 C-band and 24 Ku-band transponders to the Asia-Pacific market. A dual payload of this type increases capacity and decreases the cost per transponder.

Finally, some satellites serve specialized markets such as GEO mobile satellites that connect directly with specially designed handheld phones. An example of one of these satellites, Thuraya, is shown in Figure 1.8 with its 12-m antenna deployed.

Figure 1.7   AsiaSat 3C is a hybrid C/Ka satellite with a total of 48 transponders.

Figure 1.8   Thuraya 1 provides high-power mobile satellite links to handheld terminals. (Courtesy of Boeing Satellite Systems.)

Also, the trend to use the smallest possible DTH home receiving antenna and to cover the largest service area combine to demand the largest possible spacecraft. The total payload power of such satellites reaches 15 kW, which is roughly 12 times that of Measat. At the time of this writing, there are drawing board designs for satellites that can support payload powers of up to 20 kW. An example of this is the 2020 program from Space Systems/Loral.

While most of the money in satellite communications is derived from the broadcast feature, there are service possibilities where remote Earth stations must transmit information back to the hub Earth station (and this is not necessarily by satellite). Examples of such return link applications include:

- Control signals to change the content of the information being broadcast (to achieve narrow casting on a broadcast link);
- Requests for specific information or browsing of documents (to support Internet or intranet services);
- Responsive information to update the record for a particular customer;
- Point-to-point information that one remote user wishes to be routed to another remote user (like e-mail).

Adding the return link to the network tends to increase the cost of the remote Earth station by a significant amount since both a transmitter and controller are required. However, there are many applications that demand a two-way communication feature. The relative amount of information (bandwidth) on the forward and return links can be quantified for the specific application, as suggested in Figure 1.9. Most of the bandwidth on GEO satellites is consumed in the forward direction, as indicated by the area in the lower right for TV broadcast or distribution. There are also uses for transmitting video in both directions, which is indicated in the upper

Figure 1.9 The approximate relationship of bandwidth usage between the forward link (hub transmit) and return link (remote transmit) in satellite applications.

right-hand corner. Cutting the bandwidth back on the forward link but not on the return link supports an application where bulk data is transferred from a remote to a centralized host computer. Reduced bandwidth in both directions expands the quantity of user channels to offer low data rate switched service for fixed and mobile telephone markets.

These general principles lead to a certain set of applications that serve telecommunication users. In the next section, we review the most popular applications in preparation for the detailed evaluations in the remaining chapters.

1.2 Satellite Application Types

Applications in satellite communications have evolved over the years to adapt to competitive markets. Evolutionary development, described in [1], is a natural facet of the technology because satellite communication is extremely versatile. This is important to its extension to new applications yet to be fielded.

1.2.1 Broadcast and Multicast of Digital Content

The first set of applications follow the predominant transmission mode of the GEO satellite—that of point-to-multipoint information distribution. We have chosen to focus exclusively on the broadcast and multicast of content in digital form to a community of users. In the past, signals were transmitted in their original analog form using frequency modulation (FM). While some of this equipment is still in use around the world, it is being phased out. One of the main reasons for this is that signals in digital form can be compressed appreciably without impairing their quality.

A bandwidth compression factor of 10 to 20 is now common, with the primary benefit of reducing transponder occupancy per channel of transmission, thereby increasing useful capacity. Rather than paying, say, $1.5 million per TV channel per year, transponder cost is reduced to $250,000 or less. Therefore, analog ground equipment has become expensive to operate even if its sunk cost is zero.

Once in digital form, information can be managed in a wide variety of manners and forms. The resulting bit stream can be expanded to include different content, addressable to subsets of users or even an individual user. In addition to the current heavy use of satellites to transmit digital TV channels, we see new applications in digital content distribution appearing and developing. These new applications may employ features of the Internet in terms of permitting Web browsing; however, multicast techniques are better suited to the GEO platform than the Internet itself.

1.2.1.1   Entertainment Television (Network, Cable, and Direct Broadcast Satellite)

Commercial TV is the largest segment of the entertainment industry; it also represents the most financially rewarding user group to satellite operators. The four fundamental ways that the satellite transfers TV signals to the ultimate consumer are:

- Point-to-multipoint distribution of TV network programming contribution from the studio to the local broadcast station;
- Point-to-point transmission of specific programming from an event location to the studio (alternatively, from one studio to another studio);
- Point-to-multipoint distribution of cable TV programming from the studio to the local cable TV system;
- Point-to-multipoint distribution of TV network and/or cable TV programming from the studio directly to the subscriber (i.e., DTH).

It may have taken 10 or more years for the leading networks in the United States and Europe to adopt satellites for distribution of their signals, but since 1985, it has been the main stay. Prior to 1985, pioneering efforts in Indonesia and India allowed these countries to introduce nationwide TV distribution via satellite even before the United States had made the conversion from terrestrial microwave. European TV providers pooled their resources through the European Broadcasting Union (EBU) and the EUTELSAT regional satellite system. Very quickly, the leading nations of Asia and Latin America adopted satellite TV delivery, rapidly expanding this popular medium to global levels.

Over-the-Air TV Broadcasting
The first of the four fundamental techniques is now standard for TV broadcasting in the VHF and UHF bands, which use local TV transmitters to cover a city or market. The satellite is used to carry the network signal from a central studio to multiple receive Earth stations, each connected to a local TV transmitter. This has been called TV distribution or TV rebroadcast. When equipped with uplink equipment, the remote Earth station can also transmit a signal back to the central studio to allow the station to originate programming for the entire network. U.S. TV networks like CBS and Fox employ these reverse point-to-point links for on-location

news reports. The remote TV uplink provides a transmission point for local sporting and entertainment events in the same city. This is popular in the United States, for example, to allow baseball and football fans to see their home team play an away-from-home game in a remote city. More recently, TV networks employ fiber optic transmission between studio and broadcast station, and between stadium and studio; but the satellite continues to be the alternate flexible routing system.

Satellite transmissions have gone digital, as discussed previously, but broadcast stations depend heavily on the conventional analog standards: NTSC, PAL, and SECAM. In developed countries, governments are encouraging broadcasters to digitize their signals to open up bandwidth for more TV channels and for use in other radio services such as mobile telephone. In the United States, many local stations provide some quantity of their programming in digital form, offering high-definition television in some cases.

Revenue for local broadcast operations is available from two potential sources: advertisers and public taxes. Pay TV services from cable, satellite, and local microwave transmissions permit greater revenue when TV watchers become monthly subscribers. In some countries, nationally sponsored broadcasters are supported directly through a tax or indirectly by government subsidy. Since its beginnings in the United States, TV provided an excellent medium to influence consumer purchase behavior. In exchange for watching commercials for soap, airlines, and automobiles, the consumer is entertained for nothing. This has produced a large industry in the United States as stations address local advertisers and the networks promote nationwide advertising. The commercial model was also adopted in Latin America.

An alternative approach was taken in many European countries and in Japan, where government-operated networks were the first to appear. In this case, the consumer is taxed on each TV set in operation. These revenues are then used to operate the network and to produce the programming. The BBC in the United Kingdom and NHK in Japan are powerhouses in terms of their programming efforts and broadcast resources. However, with the rapid introduction of truly commercial networks, cable TV, and DTH, these tax-supported networks are experiencing funding difficulties.

Public TV in the United States developed after commercial TV was well established. Originally called Educational TV, this service existed in a fragmented way until a nonprofit organization called the Public Broadcasting Service (PBS) began serving the nation by satellite in 1978. The individual stations are supported by the local communities through various types of donations. Some are attached to universities; others depend on donations from individuals and corporations. PBS itself acquires programming from the member stations and from outside sources like the BBC. Programs are distributed to the members using satellite transponders purchased by the U.S. government. It must therefore compete with other government agencies for Congressional support. PBS programming is arguably of better quality than some of the popular shows on the commercial networks. Even though PBS addresses a relatively narrow segment, its markets are under attack by even more targeted cable TV networks like A&E, The Discovery Channel, The Learning Channel, The History Channel, Home and Garden TV, and the Food Network. All of these competitors built their businesses on satellite delivery to cable systems and DTH subscribers.

The local airwaves provide a reasonably good medium to distribute programming with the added benefit of allowing the local broadcaster to introduce local programs and advertising. Satellite transmission, on the other hand, is not limited by local terrain and thus can be received outside the range of terrestrial transmitters, extending across a nation or region. In extreme cases where terrestrial broadcasting has been destroyed by war or conflict, or has not been constructed due to a lack of economic motivation, satellite TV represents the only effective alternative.

Cable Television
Begun as a way to improve local reception in rural areas, cable TV has established itself as the dominant force in many developed countries. This was facilitated by organizations that used satellite transmission to distribute unique programming formats to cable subscribers. The cable TV network was pioneered by HBO in the 1970s. Other early adopters of satellite delivery include Turner Broadcasting, Warner Communications, and Viacom. By 1980, 40% of urban homes in the United States were using cable to receive the local TV stations (because the cable provided a more reliable signal); at the same time, the first nationwide cable networks were included as a sweetener and additional revenue source. During the 1980s, cable TV became an $8 billion industry and the prototype for this medium in Europe, Latin America, and the developed parts of Asia.

By 2002, about 80 million U.S. households were connected to cable for TV, with about 6 million benefiting from broadband Internet access through two-way cable technology. The vitality of the cable industry actually benefited from the digital DTH revolution, which forced cable systems to digitize and expand services.

Cable TV networks, discussed in Chapter 4, offer programming as a subscriber service to be paid for on a monthly basis or as an almost free service like commercial TV broadcasting. HBO, Showtime, and the Disney Channel are examples of premium (pay) services, while The Discovery Channel, CNN, and MSNBC are examples of commercial channels that receive most of their revenue from advertisers. The leading premium channels in North America and Europe are successful in financial terms, but the business has yet to be broadly accepted in economies with low-income levels.

Cable TV became the first to offer a wide range of programming options that are under the direct control of the service provider. The local cable system operator controls access and can therefore collect subscription fees and service charges from subscribers. If the fees are not paid, the service is terminated. Wireless cable, a contradiction in terms but nevertheless a viable alternative to wired cable, uses portions of the microwave spectrum to broadcast multiple TV channels from local towers. It has proven effective in urban areas in developing economies where the density of paying subscribers is relatively high, such as Mexico City and Jakarta, Indonesia. Just as in the case of DTH, wireless cable depends on some form of conditional access control that allows the operator to electronically disconnect a nonpaying user. Theft of signals, called piracy, is a common threat to the economic viability of wired and wireless cable (as it is to DTH, discussed next).

Direct-to-Home Broadcasting Satellite

The last step in the evolution of the satellite TV network is DTH. After a number of ill-fated ventures during the early 1980s by USCI, COMSAT, CBS, and others, DTH has established its niche in the broadcasting and cable spheres. BSkyB in the United Kingdom, NHK in Japan, DIRECTV and EchoStar in the United States, Sky Latin America, and STAR TV in Asia are now established businesses, with other broadcasters following suit. Through its wide-area broadcast capability, a GEO satellite is uniquely situated to deliver the same signal throughout a country or region at an attractive cost per user. The particular economics of this delivery depend on the following factors.

- *The size of the receiving antennas:* Smaller antennas are easier to install and maintain and are cheaper to purchase in the first place. They are also less noticeable (something that is desirable in some cultures).
- *The design of the equipment:* This is simple to install and operate (this author's Digital Satellite System (DSS) installation, needed to receive DIRECTV, took only 2 hours—that is, 105 minutes to run the cables and 15 minutes to install and point the dish).
- *Several users can share the same antenna:* This is sensible if the antenna is relatively expensive, say, in excess of $1,000; otherwise, each user can afford his or her own. A separate receiver is needed for each independent TV watcher (the same now applies to digital cable service).
- *The number of transponders that can be accessed through each antenna (typically 32):* Due to the high power required as well as concerns for single-point failure, DTH operators place more than one satellite in the same orbit position in order to achieve the desired total transponder count. The more channels that are available at the same slot, the more programming choices that the user will have.
- *The number of TV channels that can be carried by each transponder (typically 10):* Capacity is multiplied through digital compression and statistical multiplexing techniques discussed in Chapter 6.
- *Inclusion of local TV channels in the United States:* This simplifies home installation and meets a government mandate that satellites "must carry" these channels to all potential markets.

The ideal satellite video network delivers its programming to the smallest practical antenna on the ground, has a large number of channels available (200 or more), and permits some means for users to interact with the source of programming. A simple connection to the PSTN allows services to be ordered directly by the subscriber; alternatively, a broadband connection is offered either over the satellite or through wireline or wireless access.

1.2.1.2   Content Delivery Networks

A content delivery network (CDN) is a point-to-multipoint satellite network that uses the broadcast feature to inject multimedia content (particularly Web pages and specific content files such as software updates and films) into remote servers and

other types of caching appliances. The basic structure of a CDN is illustrated in Figure 1.10. The remote cache could be a dedicated server connected to the local infrastructure of the Internet. This greatly reduces the delay associated with accessing and downloading the particular content. Another style of CDN is to put the content directly into the PC hard drive; for this to work, the PC must have a direct electrical connection to the remote CDN terminal.

The first CDNs appeared during the Internet boom of 1999–2000; many have not survived the shakeout. However, some organizations are using and developing CDNs as a structure to propagate content to remote locations to bypass the cost and congestion of the terrestrial Internet. The ground equipment and software to create a CDN may be blended with that used for digital TV, as will be discussed in Chapter 5. The fact that the content appears to be local to the user enhances the interactive nature of the service. Thus, the central content store does not directly process requests from users.

1.2.1.3   Satellite Delivered Digital Audio Radio Service

We conclude the discussion of point-to-multipoint applications with an introduction to digital audio broadcasting (DARS). By focusing on sound programming without a visual element, S-DARS addresses itself to networks where (1) spectral bandwidth is limited, (2) users are mobile in their cars and boats, and/or (3) isolated from major sources of radio and other mass media. While DARS is a term generally reserved for terrestrial digital radio, the version we are interested in is satellite delivered digital audio radio service (S-DARS). The first to introduce S-DARS principally as a solution to (3) was WorldSpace, a startup company with the vision of delivering multichannel radio programming to the underdeveloped regions of Africa and Asia. Subsequently, the FCC auctioned off L-band spectrum for S-DARS for the U.S. market. XM Satellite Radio and Sirius Satellite Radio implemented 100 digital audio radio services that are comparable to FM broadcasting. Both companies launched S-band satellites in 2001 and initiated service on a commercial basis in 2002. Through a package of subscription radio channels as well as conventional advertiser supported formats, XM and Sirius serve subscribers in their cars and homes.

Figure 1.10   Structure of a content delivery network with reliable file transfer. (Courtesy of Scopus.)

Satellite construction and launch was hardly a challenge for S-DARS; however, producing the appropriate receiving terminal proved to be more time consuming than the original business plans considered. Examples of the types of units offered for S-DARS service are shown in Figure 1.11. As in any satellite communications service, a line-of-site path is usually required; thus, the antenna must be in plain view of the geostationary orbit. Vehicular installations are best; however, obstructions like tall buildings, trees, tunnels, and overpasses may block the signal. This are countered through three techniques: receiver storage of several seconds of channel stream, allowing for catch-up when a blocked receiver again "sees" the satellite; use of two or more satellites to increase the probability of a line-of-sight path; and rebroadcast of the satellite signal into concrete canyons and inside tunnels through the use of land-based "gap filler" relays.

1.2.2 Voice and Telephony Networks

Voice communications are fundamentally based on the interaction between two people. It was recognized very early in the development of satellite networks that the one-way propagation delay of one-quarter second imposed by the GEO tends to degrade the quality of interactive voice communications, at least for some percentage of the population. However, voice communications represent a significant satellite application due to the other advantages of the medium. For example, many developing countries and lightly inhabited regions of developed countries continue to use satellite links in rural telephony and as an integral part of the voice network infrastructure. Furthermore, an area where satellite links are essential for voice communications is the mobile field. These developments are treated in detail in Chapters 10 and 11.

The PSTN within and between countries is primarily based on the requirements of voice communications, representing something in the range of 50% to 60% of all interactive traffic. The remainder consists of facsimile (fax) transmissions, low- and medium-speed data (both for private networks and access to public network services such as the Internet), and various systems for monitoring and controlling remote facilities. Direct access to the Internet via a dial-up modem will be a supporting factor for the PSTN in coming years. The principal benefit of the PSTN is that it is truly

Figure 1.11   Sanyo WorldSpace receiver.

universal. If you can do your business within the limits of 3,000 Hz of bandwidth and can tolerate the time needed to establish a connection through its dial-up facility, the PSTN is your best bet.

Propagation delay became an issue when competitively priced digital fiber optic networks were introduced in the 1990s. Prior to 1985 in the United States, AT&T, MCI, and others were using a significant amount of analog telephone channels both on terrestrial and satellite links. An aggressive competitor in the form of U.S. Sprint invested in an all-digital network that employed fiber optic transmission. Sprint expanded their network without microwave or satellite links and introduced an all-digital service at a time when competition in long distance was heading up. Their advertising claimed that calls over their network were so quiet "you can hear a pin drop." This strategy was so successful that both MCI and AT&T quickly shifted their calls to fiber, resulting in rapid turn-down of both satellite voice channels and analog microwave systems.

A similar story is told in Europe, Latin America, and Asia, albeit at a slower pace in most countries due to the persistence of local monopolies. In time, fiber links and digital voice switching have become the standard of the PSTN.

The economics of satellite voice communications are substantially different from that of the fiber-based PSTN, even given the use of digital technology with both approaches. With low-cost VSAT technology and high-powered satellites at Ku- and Ka-bands, satellite voice is the cheapest and quickest way to reach remote areas where terrestrial facilities are not available. It will be more attractive to install a VSAT than to extend a fiber optic cable over a distance greater than a few hundred meters. A critical variable in this case is the cost of the VSAT, which dropped from the $10,000 level in 1995 to as low as $1,500 in 2003. Fiber, however, is not the only terrestrial technology that can address the voice communication needs of subscribers. Fixed wireless systems have been installed in developing countries to rapidly turn up telephone services on the local loop. Low-cost cordless phones or simple radio terminals are placed in homes or offices, providing access to the PSTN through a central base station. The base stations are concentrating points for traffic and can be connected to the PSTN by fiber or even satellite links. The cost of the base station and network control is kept low by not incorporating the automatic hand-off feature of cellular mobile radio. Instead, user terminals of different types make the connection through the closest base station, which remains in the same operating mode throughout the call. The ability of the wireless local loop to support Internet access at 56 Kbps depends on the degree of compression used to provide sufficient channel capacity.

High-speed Internet access has been introduced on wireline local loops through the class of technologies known as DSL. Using the basic approach of frequency division multiplex (FDM), DSL adds the baseband bandwidth needed to allow bidirectional transfer speeds of 100 Kbps to as much as 1 Mbps over copper twisted-pair. In the absence of copper, traditional fixed wireless local loop networks cannot support DSL-like services. More recently, some service providers have begun to offer wireless Internet access using the IEEE 802.11b standard (also called Wi-Fi). The advantage of this approach is that the spectrum is unlicensed in the United States and most other countries and therefore freely available (although potentially crowded); furthermore, many individuals already carry Wi-Fi modems within their

laptops. Likewise, to add high-speed access to satellite telephony amounts to providing the appropriate bandwidth over the same or even another VSAT. The notion that bandwidth is free certainly does not apply to wireless systems, whether speaking of the local or satellite varieties.

New classes of public network services may appear in coming years under the general category of Broadband Integrated Services Digital Networks (B-ISDN). The underlying technology is asynchronous transfer mode (ATM), a flexible high-speed packet-switched architecture that integrates all forms of communications [2]. ATM services can be delivered through fiber optic bandwidths and advanced digital switching systems. ATM includes the following capabilities:

- High-speed data on demand (384 Kbps to 155 Mbps, and greater);
- Multichannel voice;
- Video teleconferencing and video telephone;
- Video services on demand;
- High-resolution color images;
- Integrated voice/data/video for enhanced Internet services.

Due to the high cost of upgrading the terrestrial telephone plant for ATM services, many of these services will not appear in many places for some time. However, they represent the capability of the coming generation of public networks being implemented around the globe. Even in the absence of a public B-ISDN, the ATM approach has been applied within private LANs and campus networks, interconnection between LANs to form a WAN, and within the backbone of the Internet itself by Tier 1 Internet service providers (ISPs) like UUNET and Genuity.

Fiber optic networks are attractive for intra- and intercity public networks and can offer broadband point-to-point transmission that is low in cost per user. The economics of long-haul fiber dictate that the operator must aggregate large volumes of telephone calls, private leased lines, and other bulk uses of bandwidth in order to make the investment pay. In 2002, the financial failure of several new fiber carriers illustrates the dilemma they face. Yet this is easier to do with a satellite because it provides a common traffic concentration point in the sky. The bandwidth is used more effectively (a principle of traffic engineering), and therefore the network can carry more telephone conversations and generate more revenue regardless of where the demand arises.

Satellite networks are very expandable because all points are independent and local terrain does not influence performance. Consider the example of the largest German bank, Deutsche Bank, which needed to offer banking services in the new states of the former East Germany. The telecom infrastructure in East Germany in 1990, while the best in the Soviet Block, was very backward by Western European standards. Deutsche Bank installed medium-sized Earth stations at new bank locations and was then able to offer banking services that were identical to those of their existing branches in the West. In a more recent example, the devastation caused in Afghanistan during years of repression and the ensuing conflict destroyed any semblance of telecommunications infrastructure. Existing GEO satellites in the region provide the bandwidth, and low-cost satellite Earth stations are introduced at cities, towns, and villages to restore reliable communications for the entire country.

Another excellent example is the early VSAT network deployed by Wal-Mart department stores, reviewed later in this chapter.

1.2.3   Data Communications and the Internet

Satellite networks are able to meet a wide variety of data communication needs of businesses, government agencies, and nongovernmental organizations (NGOs), which include charities and religious groups. The wide-area coverage feature combined with the ability to deliver relatively wide bandwidths with a consistent level of service make satellite links attractive in the developing world as well as in the geographically larger developed countries and regions. Furthermore, the point-to-multipoint feature renders GEO satellites superior to the terrestrial Internet for the distribution of IP-based multimedia content such as Web pages and movies.

The data that is contained within the satellite transmission can take many forms over a wide range of digital capacities. The standard 36-MHz transponder, familiar to users of C- and Ku-bands worldwide, can transfer up to 80 Mbps, which is suitable for wideband applications and multimedia. Most applications do not need this type of bandwidth; therefore, a somewhat diminished transponder capacity is often divided up among users who employ a multiple access system of some type. In fact, the multiple access techniques used on the satellite mirror the approaches used in wireless networks, local area networks (LANs), and wide area networks (WANs) over terrestrial links. As in any multiple access scheme, the usable capacity decreases as the number of independent users increases. Satellite data networks employing VSATs offer an alternative to terrestrial networks composed of fiber optics and microwave radio. There is even a synergy between VSATs and the various forms of terrestrial networks, as both can multiply the effectiveness of their counterpart.

This is the principle of *complementarity*, where the relative advantages of satellite networks are combined with those of fiber optics and fixed and mobile wireless to address the widest range of needs. Some of the must successful users of VSATs are familiar names in North American consumer markets. Wal-Mart, the largest U.S. department store chain, which has opened outlets in Europe, Latin America, and China, was an early adopter of the technology and has pushed its competitors to use VSATs in much the same manner that they pioneered the megastore. With their Earth station hub at the Arkansas headquarters, Wal-Mart centralizes its credit authorization and inventory management functions. Chevron Oil likewise was first among the gasoline retailers to install VSATs at all of their company-owned filling stations to speed customer service at the pump, prevent fraud, and gain a better system-wide understanding of purchasing trends.

While telephone networks like the PSTN are standardized, data networks cover an almost infinite range of needs, requirements, and implementations. In business, information technology (IT) functions are often an element of competitive strategy [3]. In other words, a company that can use information and communications more effectively than its competitors could enjoy a stronger position in the ultimate market for its products or services. This is a subset of the general principle of *competitive strategy* as outlined by Michael E. Porter in his seminal book of the same title [4]. A data communication system is really only one element of an architecture that is intended to perform business automation functions. However, defining the right

architecture and putting the associated hardware, software, and network pieces in place is often a tall order. As cited previously, major corporations have found that a properly designed and managed satellite data network can overcome some of the most challenging difficulties in this area.

A given IT application using modern client/server computing networks or broadband multimedia will demand efficient transfer of data among various locations and users. Satellite communication introduces a relatively large propagation delay, but this is only one factor in the overall response time. There are many contributors to response time: the input data rate (in bits per second), the propagation delay through the link, the processing and queuing delay in data communication equipment, and any contention for a terrestrial data line or computer processing element. This is shown in Figure 1.12. Each of these contributors must be considered carefully when comparing terrestrial and satellite links.

Propagation delay from a satellite link could reduce the throughput if there is significant interaction between the two ends of the link. The worst case condition occurs where the end devices need to exchange control information to establish a connection or confirm receipt. Modern data communication standards, like Transmission Control Protocol/Internet Protocol (TCP/IP) and Systems Network Architecture/Synchronous Data Link Control (SNA/SDLC), guard against the loss of throughput by only requesting retransmission of blocks of data that have errors detected by the receiver. One the other hand, Frame Relay and ATM, two very popular link-layer protocols in heavy use for WANs, do not guarantee reliable data delivery in this manner. To optimize throughput, there is a need to test and tune the circuit for the delay, type of information, error rate, and resulting protocol and application performance. Still, suppliers of VSAT network hardware and software include protocol acceleration to compensate for the added delay of a GEO satellite connection.

End-to-end data transfer delay on a GEO satellite link can be made to be lower than on a terrestrial telephone circuit. Consider a telephone circuit with an input rate of 48 Kbps; it will take 0.5 second (500 ms) to pass a block of 3,000 bytes (or 24,000 bits). This is calculated by dividing the number of bits by the data rate in bits per second (24,000/48,000). The terrestrial PSTN will add a little to this, perhaps 50 ms, yielding a total of about 550 ms. We all experience (and can measure) this type of delay when connected to the Internet through a dial-up modem. The 260-ms propagation delay for a GEO satellite link increases the total delay to about 760 ms, about 50% greater than the landline service. Now if we were to expand the satellite

Figure 1.12   The total end-to-end latency for data transfer results from several components: access lines, equipment processing, uplink and downlink propagation, and data processing in servers.

link to 2 Mbps, the direct transfer time for the 3,000 bytes would be 24,000/2,000,000, or 12 ms. The total delay of 270 + 12 = 282 ms of the satellite link is now substantially less than the 550 ms for the terrestrial connection. As long as the satellite network can deliver superior transfer speed (e.g., bandwidth) to the terrestrial alternative, it will have an advantage.

One can have a requirement that favors the terrestrial link. Suppose the PSTN is all fiber and has excellent transmission properties (which is the case in developed regions and modern cities). Now, it is possible to input data at 512 Kbps. The total terrestrial delay is only about 100 ms, which is significantly shorter than for the satellite link. However, we must actually obtain 512 Kbps throughout the PSTN, something that is not particularly consistent. If you establish an international connection to many countries in the world, you would find that 512 Kbps is not feasible on an end-to-end basis. The satellite systems of the world, however, can produce links of sufficient quality to permit data transfer at 2 Mbps or higher; whether this is attractive on a financial basis depends on the local availability of consumer-grade DSL or cable modem services.

Satellite links maintain an important position as part of the backbone of the Internet. These are particularly valuable for hooking ISPs in developing countries and second cities to major nodes in the United States, Europe, Japan, and other popular access points. The links provide bidirectional data transfer at rates between about 256 Kbps and 155 Mbps, depending on the expected demand and the financial ability of the particular ISP. Under the basic rules of connection, a lower tier ISP will need to pay the Tier 1 ISP for this type of access; often, the Tier 1 ISP will provide the backhaul satellite circuit as part of the service. Private companies also use such point-to-point links as part of an internal WAN to bypass points of congestion in public networks and to allow medium and high data rate services like video teleconferencing. In general, point-to-point links are established for an extended period using fixed Earth stations and dedicated GEO satellite bandwidth. Demand-assigned services such as those provided using VSAT networks may be employed for connection-oriented applications and when temporary service is required.

1.2.4   Mobile and Personal Communications

The world has experienced an explosion in the demand for wireless telephone and data communications, typically of a mobile nature. The basis for this is the technology of cellular radio systems, which connect vehicular and handheld mobile phones to the public network as if they had access by wire [5]. Availability of cellular radio at a reasonable cost has allowed a much larger subscriber base to develop than was even possible during earlier generations of mobile phone technology. The allure of terrestrial cellular led to the development and operation of two non-GEO MSS systems: Iridium and Globalstar. While both systems began operation, neither succeeded in the marketplace. The Inmarsat GEO satellite system grew in a much more gradual and sustained manner. As a result, Inmarsat successfully offers a range of low and medium speed digital services to user terminals on ships, aircraft, and vehicles, as well as many operated by individuals.

While the satellite industry has been working to compete with conventional cellular telephone, the telephone and mobile radio business has been working to

produce a more capable wireless service. With digital cellular and Personal Communications Network (PCN) and Personal Communications Service (PCS) having become the mainstay, existing and new operators have begun to pursue the third generation (3G) cellular market. As readers are aware, this was an expensive endeavor by the operators, some of which teetered on the edge of bankruptcy due to the high cost of acquiring the new spectrum at auctions from governments. There are two standards vying for the 3G market: IMT-2000, being pursued by the GSM camp, and CDMA2000, the banner for Qualcomm's CDMA system that follows the IS-95 standard. The pure form of these 3G systems may never appear; however, each has variants that address the need for greater channel capacity, improved voice quality, and the introduction of higher throughput for data. That the cellular networks will grow in capability and service performance is not in doubt. Thuraya and Inmarsat 4 represent the satellite response to 3G, assuming an interesting future in mobile communications.

References

[1]   Elbert, B. R., *Introduction to Satellite Communication*, Norwood, MA: Artech House, 1987.

[2]   Elbert, B. R., and B. Martyna, *Client/Server Computing—Architecture, Applications, and Distributed Systems Management*, Norwood, MA: Artech House, 1994.

[3]   Elbert, B. R., *Networking Strategies for Information Technology*, Norwood, MA: Artech House, 1992.

[4]   Porter, M. E., *Competitive Strategy*, New York: The Free Press, 1980.

[5]   Macario, R. C. V. (on behalf of the Institution of Electrical Engineers), *Personal and Mobile Radio Systems*, London, England: Peter Peregrinus, 1991.

Satellite Links, Multiple Access Methods, and Frequency Bands

Satellite links employ microwave frequencies above 1 GHz—the upper end is limited to about 30 GHz for currently active uses. Frequencies above 30 GHz, or equivalently, wavelengths smaller than 1 cm (including optical wavelengths), are the subject of research and development and are discussed briefly at the end of this chapter. The microwave engineering process is no different from the practice developed during and immediately after World War II, when the application of this medium was accelerated for radar and communications. While the principles remain the same, many innovations in digital processing, microelectronics, software, and array antennas allow more options for new applications. In this chapter, we briefly review the basics of the satellite link and relate it as much as possible to the needs of the application. Multiple access systems, including frequency division multiple access (FDMA), time division multiple access (TDMA), and code division multiple access (CDMA), are also discussed and their strengths and weaknesses identified. Once this review is completed, we consider the popular frequency bands used in commercial satellite communication (i.e., L, S, C, X, Ku, and Ka) along with higher frequencies (Q- and V-bands), as well as space-based optical communications.

The information that follows introduces the engineering side of developing satellite communication applications. Readers who have a technical background should have little difficulty following along with the engineering-related discussion and format of the link budget. This is approached as a review and is not a substitute for a more detailed engineering evaluation of the specific transmission systems, losses, and signal impairments that would be experienced in a particular design (in particular, see [1–3]). Nontechnical readers may wish to skim the text concerning microwave link engineering and instead focus on the final sections that consider multiple access methods and frequency band selection.

2.1   Design of the Satellite Link

The satellite link is probably the most basic in microwave communications since a line-of-sight path typically exists between the Earth and space. This means that an imaginary line extending between the transmitting or receiving Earth station and the satellite antenna passes only through the atmosphere and not ground obstacles (the impact of obstacles is considered in our discussion of mobile services in Chapter 11). Such a link is governed by free-space propagation with only limited variation

with respect to time due to various constituents of the atmosphere. Free-space attenuation is determined by the inverse square law, which states that the power received is inversely proportional to the square of the distance. The same law applies to the amount of light that reaches our eyes from a distant point source such as an automobile headlight or star. There are, however, a number of additional effects that produce a significant amount of degradation and time variation. These include rain, terrain effects such as absorption by trees and walls, and some less-obvious impairment produced by unstable conditions of the air and ionosphere.

It is the job of the communication engineer to identify all of the significant contributions to performance and make sure that they are properly taken into account. The required factors include the performance of the satellite itself, the configuration and performance of the uplink and downlink Earth stations, and the impact of the propagation medium in the frequency band of interest. Also important is the efficient transfer of user information across the relevant interfaces at the Earth stations, involving such issues as the precise nature of this information, data protocol, timing, and the telecommunications interface standards that apply to the service. A proper engineering methodology guarantees that the application will go into operation as planned, meeting its objectives for quality and reliability.

The RF carrier in any microwave communications link begins at the transmitting electronics and propagates from the transmitting antenna through the medium of free space and absorptive atmosphere to the receiving antenna, where it is recovered by the receiving electronics. Like your automobile FM radio or any other wireless transmission, the carrier is modulated by a baseband signal that transfers information for the particular application. The first step in designing the microwave link is to identify the overall requirements and the critical components that determine performance. For this purpose, we use the basic arrangement of the link shown in Figure 2.1. This example shows a large hub type of Earth station in the uplink and a small VSAT in the downlink; the satellite is represented by a simple frequency-translating type of repeater (e.g., a bent pipe). Most geostationary satellites employ bent-pipe repeaters since these allow the widest range of services and communication techniques. Bidirectional (duplex) communication occurs with a separate

Figure 2.1   Critical elements of the satellite link.

transmission from each Earth station. Due to the analog nature of the radio frequency link, each element contributes a gain or loss to the link and may add noise and interference as well.

The result in the overall performance is presented in terms of the ratio of carrier power to noise (the carrier-to-noise ratio, C/N) and, ultimately, information quality (bit error rate, video impairment, or audio fidelity). Done properly, this analysis can predict if the link will work with satisfactory quality based on the specifications of the ground and space components. Any uncertainty can be covered by providing an appropriate amount of link margin, which is over and above the C/N needed to deal with propagation effects and nonlinearity in the Earth stations and satellite repeater.

2.1.1   Meaning and Use of the Decibel

Satellite application engineers are more comfortable working with sums and differences than with multiplication and division (which is how the link actually relates to the real world). The decibel (dB) has been settled upon as the most convenient metric because, once converted into decibels, complex factors can be added and subtracted on paper with a calculator (or even in your head) because any number that is either a ratio of two powers or a power expressed in watts can be converted to decibels. This employs the base-10 logarithm (or common logarithm):

$$\text{Power ratio in decibels} = 10 \log (P_2/P_1) \tag{2.1}$$

where P_2 is the output of the device or link, and P_1 is its input. Table 2.1 provides this conversion for integer values of the ratio of P_2 to P_1. The first entry for the number 1.0 represents the ratio of two equal power values. The output power does not change for a ratio of 1, and the corresponding decibel value is 0 dB. If the power increases from this point by 12%, the power ratio is 1.12, which corresponds to a 0.5-dB increase. All decibel values are relative to the starting point; so if we increase the power again by 12% we must add another 0.5 dB, giving a total increase of 1 dB. According to Table 2.1, a 1-dB increase corresponds to a change of 26% or, equivalently, multiplication of the original value by 1.26. This is also equal to 1.12 multiplied by 1.12. The decibel column has a single decimal place to correspond to reasonable therefore measurement accuracies.

Table 2.1 can be used to represent changes that go in the opposite direction—that is, where the ratio of the output power to the input is less than one. For example, if the output is 1 and the input is 10, then the ratio is 0.10. Using the formula, the decibel value is −10 dB. When we invert the relationship between the power values, we simply put a minus sign in front of the decibel value. In Table 2.1, dividing by 2 (instead of multiplying by 2) produces a −3-dB change (instead of a +3-dB change).

An experienced satellite application engineer maintains several values of this table in his or her head and uses basic arithmetic to quickly analyze the main factors in a particular link. We call this "decibel artistry," the rules for which are as follows.

Table 2.1 Conversion Between Power Ratios, Percentage Changes, and Decibels

Power Ratio	% Change	Decibels
1.0	0	0
1.12	12	0.5
1.26	26	1.0
1.41	41	1.5
1.59	58	2.0
2.00	100	3.0
2.51	151	4.0
3.16	216	5.0
3.98	298	6.0
5.01	401	7.0
6.31	531	8.0
7.94	694	9.0
8.91	791	9.5
10.0	900	10.0
20.0	—	13.0
100.0	—	20.0
500.0	—	27.0
1,000.0	—	30.0

1. $10 \log(x) = X$ expressed in decibels (note the use of common logarithms). The true ratio may use a lower-case symbol while the corresponding decibel variable is upper case.
2. Power ratios are expressed in decibels.
3. Voltages must first be squared to convert to power; that is, $10 \log(V_2/V_1)^2$ or, equivalently, $20 \log(V_2/V_1)$, a formula often seen in the radio and sound world.
4. Increases are positive when expressed in decibels; decreases are negative.
5. $10 \log(1/x) = -10 \log(x) = -X$.
6. By definition, the numerical ratio 0 expressed in decibels is $-\infty$.
7. For each multiplication by 10, add 10 dB; for each division by 10, subtract 10 dB.
8. Power relative to 1W is expressed as dBW; that is, 1W is identically equal to 0 dBW.
9. Power relative to 1 mW (0.001W) is expressed in dBm; that is, 1 mW is identically equal to 0 dBm; also, 0 dBW = 30 dBm, and −30 dBW = 0 dBm.

    A proficient decibel artist can do a rough mental link analysis by remembering these rules along with a few values from Table 2.1. For example, if the power at the transmitter is doubled, then it has increased by 3 dB. If the power is cut in half, then it decreases by the same 3 dB (rule 5). The net result is no change at all; that is, +3, −3, nets 0. A 10-dB increase simply means that the power has increased by a factor

of 10 (rule 7). If we multiply by 100, then we have introduced 20 dB, which is two increases of 10 dB (rule 7 again).

The following is a practical example that demonstrates the power of decibel artistry. Say that a satellite radiates a power of 100W toward us on the Earth. If we want the effective power to be 200W instead, then we are looking for a 3-dB increase. This is found by taking the ratio 200/100 = 2, which is converted to 3 dB in Table 2.1. If we need the effect of 2,000W, then according to rule 7, we must add another 10 dB, giving a total increase from 100W of 13 dB. When we discuss the link budget, it will be shown that the signal that reaches the geostationary satellite from the Earth at a frequency of 6 GHz is approximately $1/10^{20}$ of the transmitted power. The corresponding reduction in signal strength (called the path loss) is by a total of 200 dB. A minus sign should be in front of this value; however, in link analyses, we show losses as positive numbers under the assumption that they will be subtracted from the transmitted power.

In summary, the essential value of decibels is that they help to express what is important in the microwave link over a wide range of very weak to very strong signals. You can easily adjust powers and gains up and down using decibel values obtained from a scientific calculator; alternatively, you can memorize the first 5 to 10 values of Table 2.1 and do the calculations in your head. It has been the experience of this author that it is adequate to work in tenths of decibels (0.1 dB), as this represents the smallest increment that can be measured in practice. The overall link, to be discussed in the next section, tends to be forgiving as long as you identify and quantify the more significant elements. The most accurate and effective approach is to use a software tool that incorporates the correct formulas and presents the results in a clear and consistent manner. We will address such methodologies later in this chapter.

2.1.2   Link Budgets and Their Interpretation

The link between the satellite and Earth station is governed by the basic microwave radio link equation:

$$p_r = \frac{p_t g_t g_r c^2}{(4\pi)^2 R^2 f^2} \tag{2.2}$$

where p_r is the power received by the receiving antenna; p_t is the power applied to the transmitting antenna; g_t is the gain of the transmitting antenna, as a true ratio; g_r is the gain of the receiving antenna, as a true ratio; c is the speed of light (i.e., approximately 300×10^6 m/s); R is the range (path length) in meters; and f is the frequency in hertz.

Almost all link calculations are performed after converting from products and ratios to decibels. This uses the popular unit of decibels, which is discussed in detail in the previous section. The same formula, when converted into decibels, has the form of a power balance [the decibel equivalent of each variable is shown as a capital letter rather than the lower-case letter of (2.2)]:

$$P_r = P_t + G_t + G_r - 20\log(f \cdot R) + 147.6 \tag{2.3}$$

The received power in this formula is measured in decibel relative to 1W, which is stated as dBW. The last two terms represent the free-space path loss (A_0) between the Earth station and the satellite. Being a loss, the sum of $20\log(f \cdot R) - 147.6$ is a positive number that is subtracted from the decibel of the power and gain that precede it in (2.3).

If we assume that the frequency is 1 GHz and that the distance is simply the altitude of a GEO satellite (e.g., 35,778 km), then the path loss equals 183.5 dB; that is,

$$P_r = P_t + G_t + G_r - 183.5 \tag{2.4}$$

for f = 1 GHz and R = 35,788 km.

We can correct the path loss for other frequencies and path lengths using the formula:

$$A_0 = 183.5 + 20\log(f) + 20\log(R/35788) \tag{2.5}$$

where A_0 is the free-space path loss in decibels, f is the frequency in gigahertz, and D is the path length in kilometers. The term on the right can be expressed in terms of the elevation angle from the Earth station toward the satellite, as shown in Figure 2.2 and given in the equation:

$$R = 42643.7\sqrt{1 - 0.295577 \times (\cos\phi\cos\delta)} \tag{2.6}$$

where ϕ is the latitude and δ is the longitude of the Earth station minus that of the satellite (e.g., the relative longitude).

Substituting for R in (2.5), we obtain the correction term in decibels to account for the actual path length. This is referred to as the slant range adjustment and is plotted in Figure 2.3 as a function of the elevation angle, θ.

The link power balance relationship in (2.3) considers only the free-space loss and ignores the effects of the different layers of the Earth's atmosphere. The following listing identifies the dominant contributors that introduce additional path loss, which can vary with time. Some are due to the air and water content of the troposphere, while others result from charged particles in the ionosphere. A general quantitative review of ionospheric effects is provided in Table 2.2 [4]. This includes effects of Faraday rotation, time delay, refraction, and dispersion. It is clear from the

Figure 2.2 Definition of the slant range distance, R, between the Earth station and the GEO satellite. The Earth station elevation angle, θ, is with respect to the local horizon.

Table 2.2 Estimated Maximum Ionospheric Effects in the United States for One-Way Paths at an Elevation Angle of About 30°

Effect	100 MHz	300 MHz	1 GHz	3 GHz	10 GHz
Faraday rotation*	30 rotations	3.3 rotations	108°	12°	1.1°
Excess time delay	25 ms	2.8 ms	0.25 ms	28 ns	2.5 ns
Absorption (polar)	5 dB	1.1 dB	0.05 dB	0.006 dB	0.0005 dB
Absorption (mid Lat)	<1 dB	0.1 dB	<0.01 dB	<0.001 dB	<0.0001 dB
Dispersion	0.4 ps/Hz	0.015 ps/Hs	0.0004 ps/Hz	0.000015 ps/Hz	0.0000004 ps/Hz

*Rotation of angle of linear polarization.

data that ionospheric effects are not significant at frequencies of 10 GHz and above, but must be considered at L-, S-, and C-bands (L being the worst).

Derived from ITU-R Reports 565-3, 263-4, 263-5, and 263-6.

- Tropospheric (gaseous atmosphere) effects:
  - Absorption by air and water vapor (noncondensed): This is nearly constant for higher elevation angles, adding only a few tenths of decibels to the path loss. It generally can be ignored at frequencies below 15 GHz.
  - Refractive bending and scintillation (rapid fluctuations of carrier power) at low elevation angles: Earth stations that must point within 10° of the horizon to view the satellite are subject to wider variations in received or transmitted signal and therefore require more link margin. Tropospheric scintillation is time varying signal attenuation (and enhancement) caused by combining of the direct path with the refracted path signal in the receiving antenna.
  - Rain attenuation: This important factor increases with frequency and rain rate. Additional fade margin is required for Ku- and Ka-band links, based on the statistics of local rainfall. This will require careful study for services that demand high availability, as suggested in Figures 2.4 and 2.5. A standardized rain attenuation predictor, called the Dissanayake, Allnut, Haidara (DAH) model is available for this purpose [1]. Rain also introduces scintillation due to scattering of electromagnetic waves by raindrops, and in a later section we will see that the raindrops also radiate thermal noise—a factor that is easily modeled. In addition, rain beading on antenna surfaces scatters and in very heavy rains can puddle on feeds, temporarily providing high losses not accounted for in the DAH and thermal noise models.
- Ionospheric effects:
  - Faraday rotation of linear polarization (first line of Table 2.2): This is most pronounced at L- and S-bands, with significant impact at C-band during the peak of sunspot activity. It is not a significant factor at Ku- and Ka-bands.
  - Ionosphere scintillation (third and fourth lines of Table 2.2): This is most pronounced in the equatorial regions of the world (particularly along the geomagnetic equator). Like Faraday rotation, this source of fading decreases with increasing frequency, making it a factor for L-, S-, and C-band links.

Rain climatic zones for ITU
Regions 1 and 3; rainfall
intensity at 0.01%

Rain zone	Rainfall intensity, mm/hr
A	8
B	12
C	15
D	19
E	22
F	28
G	30
K	42
M	63
N	95
P	145

Figure 2.4   Rain climactic zones for ITU Regions 1 and 3; rainfall intensity as 0.01%.

At frequencies above C-band (i.e., above 7 GHz), rain introduces a substantial amount of loss that must be taken into account in the link design. Regions with intense thunderstorm activity, particularly in the tropics, tend to complicate link design at Ku-band and above, and offer some challenge at C-band as well. As discussed in the next section, proper engineering practice includes the provision of

Figure 2.5   The rain attenuation, in decibels, for C- and Ku-bands, as related to the rain climactic zones in Figure 2.4.

several decibels of margin above the minimum required. Margin represents extra power in the link that cushions against fades and equipment properties not known with sufficient accuracy (e.g., to within 0.1 dB). In mobile communications, it has become practice to include margin for short-term terrain blockage, as when a vehicle travels under trees or past a tall building. If the mobile link is stable and not experiencing a fade, then the margin shows up as better signal quality.

As readers are aware, rainfall is not a predictable phenomenon from year to year. On average, the statistics follow some patterns that have aided in the design of links in rainy regions, like those indicated in Figure 2.4. This information, based on the DAH rain model developed by Intelsat and adopted by the ITU, provides an indication of how the typical rainfall of a region must be considered in link design. Properly engineered satellite links at the higher frequencies may be as dependable as one operating in the popular C-band [1]. In Figure 2.5, rain attenuation is plotted for the worst rain regions (N and P) at C- and Ku-bands. Readers should keep in mind that this information applies for a general case in the Asia-Pacific region, based on the DAH model. A particular design may, under some situations, involve a thorough study of local rainfall statistics for a system operating a Ku- and Ka-band. This is advised where rain attenuation could be substantially different from a general case in the DAH model.

The satellite application engineer can select a desired value of availability such as 99.9% (along the x-axis of Figure 2.5), which determines the required amount of link margin to counter the most extreme rain attenuation condition (along the y-axis). The link should experience outage for 0.1% of the time during the rainiest month. This equates to 72 minutes during that month or a maximum of 8.8 hours for the year. It is clear from the example in Figure 2.5 that more rain margin must be provided at Ku-band than at C-band for the same link availability. In the case of Ka-band, the corresponding margin is of the order of three times, in decibels, that of Ku-band. The link budget provides the vehicle for finding the best combination of power and gain to achieve this result.

2.2 Link Budget Example

Satellite application engineers need to assess and allocate performance for each source of gain and loss. The link budget is the most effective means since it can address and display all of the components of the power balance equation, expressed in decibels. In the past, each engineer was free to create a personalized methodology and format for their own link budgets. This worked adequately as long as the same person continued to do the work. Problems arose, however, when link budgets were exchanged between engineers, as formats and assumptions can vary. Our approach is to provide a basic understanding of the link budget process; then, we suggest using a standardized link budget software tool that performs all of the relevant calculations and presents the results in a clear and complete manner.

We will now evaluate a specific example using a simplified link budget containing the primary contributors. This will provide the reader with a typical format and some guidelines for a practical approach. Separate uplink and downlink budgets are provided; our evaluation of the total end-to-end link presumes the use of a bent-pipe

repeater. This is one that transfers both carrier and noise from the uplink to the downlink, with only a frequency translation and amplification. The three constituents are often shown in a single table, but dividing them should make the development of the process clearer for readers. The detailed engineering comes into play with the development of each entry of the table. Several of the entries are calculated using straightforward mathematical equations; others must be obtained through actual measurements or at least estimates thereof. This particular example is for a C-band digital video link at 40 Mbps, which is capable of transmitting 8 to 12 TV channels using the Motion Picture Experts Group 2 (MPEG 2) standard. Digital TV standards are covered in detail in Chapter 5.

2.2.1   Downlink Budget

Table 2.3 presents the downlink budget in a manner that identifies the characteristics of the satellite transmitter and antenna, the path, the receiving antenna, and the expected performance of the Earth station receiver. The latter contains the elements that select the desired radio signal (i.e., the carrier) and demodulates the useful information (i.e., the digital baseband containing the MPEG 2 "transport" bit stream). Once converted back to baseband, the transmission can be applied to other processes, such as demultiplexing, decryption, and digital-to-analog conversion (D/A conversion).

Figure 2.6 provides the horizontal downlink coverage of Telstar V, a typical C-band satellite that serves the United States. Each contour shows a constant level of saturated effective isotropic radiated power (EIRP) (the value at saturation of the transponder power amplifier). Assuming the receiving Earth station is in Los Angeles, it is possible to interpolate between the contours and estimate a value of 35.5 dBW. A wideband carrier for TV transmission could consume all of this power.

Table 2.3   Link Budget Analysis for the Downlink (3.95 GHz, C-Band)

Item	Link Parameter	Value	Unit	Computation
1	Transmit power (10W)	10.0	dBW	Assumption
2	Transmit waveguide losses	1.5	dB	Assumption
3	Transmit antenna gain	27.0	dBi	U.S. Continental coverage
4	Satellite EIRP (toward LS)	35.5	dBW	1–2+3
5	Free-space loss	196.0	dB	(2.4)
6	Atmospheric absorption (clean air)	0.1	dB	Typical
7	Receive antenna gain(3.2m)	40.2	dBi	
8	Receive waveguide loss	0.5	dB	
9	Received carrier power	–121.7	dBW	4-5-6+7-8
10	System noise temperature (140K)	21.5	dBK	
11	Earth station G/T	18.2	dB/K	7–8–10
12	Boltzmann's constant	–228.6	dBW/Hz/K	
13	Bandwidth (25 MHz)	74.0	dB Hz	
14	Noise power	–133.1	dBW	10+12+13
15	Carrier-to-noise ratio	11.4	dB	9–14

Figure 2.3   Additional path loss due to slant range, versus ground elevation angle.

Figure 2.6 The downlink coverage footprint of the Telstar V satellite, located at 97° WL. The contours are indicated wuth the saturated EIRP in decibels referred to 1W (0 dBW).

Alternatively, the power could be shared using one of the multiple access methods described in Section 2.3. We review the details of this particular link budget for those readers who wish to gain an understanding of how this type of analysis is performed. This material may be skipped by the nontechnical reader.

The following parameters relate to the significant elements in the link (Figure 2.1) and the power balance equation, all expressed in decibels. Most are typically under the control of the satellite engineer:

- Transmit power (P_t);
- Antenna gain at the peak (G_t) and beamwidth at the −3-dB point (θ_{3dB});
- Feeder waveguide losses (L_t);
- EIRP in the direction of the Earth station (Figure 2.5);
- Receiver noise temperature (T_0);
- Noise figure (NF).

System noise temperature (T_{sys}) is the sum of T_0 and the noise contribution of the receive antenna (T_a).

The overall Earth station figure of merit is defined as the ratio of receive gain to system noise temperature expressed in decibels per Kelvin—for example, $G/T = G_r -$ 10 log (T_{sys}). This combines two factors in the receiving Earth station (or satellite), providing a standard specification at the system level. The same can be said of EIRP for the transmit case. Reception is improved if either the gain is increased or the noise temperature is decreased; hence the use of a ratio.

Each of the link parameters relates to a specific piece of hardware or some property of the microwave path between space and ground. A good way to develop the link budget is to prepare it with a spreadsheet program like Microsoft Excel or Lotus 1-2-3. This permits the designer to include the various formulas directly in the budget, thus avoiding the problem of external calculation or the potential for arithmetic error (which still exists if the formulas are wrong or one adds losses instead of subtracting them). Commercial link budget software, such as SatMaster Pro from Arrowe Technical Services, does the same job but in a standardized fashion.

The following comments and clarifications relate to each item in Table 2.3, with additional references made to the rules of decibel artistry in the previous section.

1. The transponder onboard the satellite has a power output of 10W, equivalently 10 dBW.
2. The microwave transmission line between the satellite power amplifier output and the spacecraft antenna absorbs about 40% of the output, converting it into heat. This total loss of 1.5 dB (Table 2.1, line 4) includes some absorption in microwave filters used to combine carriers and other microwave components that are part of the waveguide assembly. It is the practice to represent this type of loss (and most others to be discussed) as a positive number and then to simply subtract it in the link budget.
3. The satellite is engineered to cover a particular area of the Earth, called the coverage area or footprint (see Figure 2.6). The area of the footprint is primarily what determines the gain of the antenna, there being an inverse

relationship. According to antenna theory, the product of gain and illuminated area is a constant. For example, a doubling of the area reduces the gain at the edge by a factor of two (e.g., 3 dB). Gain is expressed in decibels relative to an isotropic antenna (e.g., an antenna that transmits equally in all directions about a sphere), in the units of dBi. The isotropic antenna (with its gain of 0 dBi) is a physical impossibility but is nevertheless used as a standard for comparison. A value of 27 dBi at the edge of the footprint results for an area roughly the size of the United States, Brazil, China, or Indonesia. For example, a gain of 27 dBi means that the power radiated in the direction of interest is 500 times (see Table 2.1) that from an isotropic antenna that is fed with the same input power.

4. This EIRP specifies the maximum radiated power per transponder in the direction of a specific location on the Earth. When we look at the downlink footprint of a satellite (Figure 2.6), it is really a gain contour plot with a conversion factor that depends on the spacecraft repeater and losses. We see contours of constant EIRP, labeled in dBW, computed by adding the input power in dBW (after internal transmission losses) to the antenna gain. Smaller concentric contours indicate higher EIRP levels associated with the peak of the beam, like height contours on a topographical map. The maximum value, indicated by X's at just one or two locations in the overall footprint, would favor these locations on the ground. It is a common practice to select a contour with good EIRP performance across the expected coverage region. This gain data is measured by the spacecraft manufacturer in a special indoor facility called an anechoic chamber or outdoors on a far-field antenna range. Subsequent to launch, the contours are verified in orbit to be sure that nothing has changed since ground testing.

5. Free-space loss is the primary loss in the satellite link, amounting to 183 to 213 dB for frequencies between 1 and 30 GHz for a GEO satellite. In days prior to the use of PC software, it was convenient to calculate the free-space loss as the sum of the loss for the link from the subsatellite point (e.g., the shortest possibly path length) and the small incremental loss to account for the real path length (e.g., the slant range), provided in Figure 2.3. To use this figure, we need to know the elevation angle from the Earth station toward the satellite. This is determined from the latitude and longitude of the Earth station and the longitude of the satellite. For Los Angeles and Telstar V, the elevation angle is approximately 30°. The worst-case value of slant range loss adjustment is 1.4 dB for a 0° elevation angle toward the satellite. Pointing an Earth station antenna at the local horizon would produce a link with substantial fading from tropospheric scintillation and ducting and is not recommended.

6. At C-band, the elements of clear air absorb a small amount of microwave energy as the wave passes through the lower atmosphere. This absorption loss increases as the elevation angle to the satellite decreases; that is, the more air there is to go through, the greater this loss. This loss also increases significantly with moisture content, particularly rain, although this is treated separately in the link budget. For stations mounted on high-flying aircraft, this absorption is effectively 0 dB. The value of 0.1 dB in the table is typical

for an elevation angle greater than 30° under the condition of normal humidity.

7. In this example, the receiving antenna has a diameter of 3.2m (10 ft). The following formula, based on the physical properties of a circular aperture antenna, provides a good estimate:

$$G = 10\log\left(110\eta f^{2} D^{2}\right)$$

where f is in gigahertz and D is the diameter in meters. The aperture efficiency, η, indicates how effectively the circular aperture is in transforming the received electromagnetic energy into an electrical signal at the output of the antenna feed. A typical value of 0.65 (expressed as 65%) is used in this example. Again, we state the gain in terms of dBi to indicate that we are comparing this antenna to the isotropic model.

8. Waveguide or cable loss between the antenna feed and low-noise amplifier (LNA) or low-noise block converter (LNB) reduces the received signal and increases link noise by nearly the same proportion. We have included 0.5 dB of loss for this effect. The cable that connects the LNB to the receiver does not directly impact the performance because it is after the relatively high gain (usually greater than 50 dB) provided by amplifier stages following the low noise preamplifier stage.

9. Received carrier power is calculated directly by the power balance method. This computed value of –121.7 dBW includes all of the gains and losses in the link. It is an absolute measure in terms of power as received by an isotropic antenna; however, we cannot tell at this point if the signal strength is sufficient for good reception. This will have to wait until we consider the uplink and the threshold performance of the demodulator.

10. The noise that exists in all receiving systems is the main cause of degradation. The system noise temperature includes contributions from the LNB, antenna, and transmission line. The LNB is rated in terms of its noise temperature, typically in the range of 20 Kelvin (20K) to 75K at C-band. (The Kelvin scale starts at absolute zero, which is where electron kinetic energy is zero, and hence noise is nonexistent.) The antenna itself collects background noise from space and the local terrain, typically adding about 40K at C-band. We have assumed a combined system noise temperature of 140K, allowing 60K for the LNB, 45K for the antenna, and 35K for the feeder line. The noise temperature introduced by the loss of the feeder line is calculated from:

$$T_{l} = \left(\frac{l - 1}{l}\right)290$$

where T_{l} is the noise temperature contribution of the line, and l is the line transmission factor ($l \geq 1$) calculated from the decibel line loss, L:

$$l = 10^{L/10}$$

11. Earth station G/T is the difference in decibels between the net antenna gain and the system noise temperature converted to decibels; that is,

$$G/T = G - L - 10\log\left(T_{sys}\right)$$

where L is the line loss between the antenna and the LNB, and T_{sys} is the receiving system noise temperature.

12.–14. The noise power (in watts) that reaches the receiver is equal to the product kTB, where k is Boltzmann's constant (1.38062×10^{-12} W/Hz/K), T is the equivalent noise temperature, and B is the noise bandwidth of the carrier in hertz. In the link budget, we use the decibel equivalents of these factors and hence can use addition instead of multiplication. The noise bandwidth for the digital carrier in this carrier is assumed to be 25 MHz for a quaternary phase shift keying (QPSK) signal, which corresponds to a symbol rate of approximately 20 Msps. The system information rate is calculated by multiplying the symbol rate by 2 (because QPSK transfers 2 bits per symbol) and by the coding rate for the particular type of forward error correction (FEC) employed. The standard way to represent a coding rate is a ratio of input (uncoded) to output (coded) bit rates. For the digital video (DVB-S) standard (discussed in Chapter 5), the FEC coding rate is itself the product of the rate for convolutional coding (such as 0.75) and block coding (typically the ratio of 188/204). This arrangement delivers an information rate of approximately 27.6 Mbps. The resulting noise power for the 25-MHz carrier bandwidth is –133.1 dBW.

15. The difference in decibels between the received carrier power and the noise power is the carrier-to-noise ratio. As mentioned in item 9, we cannot determine at this point if 11.4 dB is adequate for the overall link.

The downlink budget is now complete and may be put aside for the moment as we proceed to the uplink.

2.2.2   Uplink Budget

We next perform nearly the same calculation for the uplink, providing an estimate of the carrier-to-noise ratio as measured at the output of the spacecraft antenna system. The uplink budget is also used to determine the EIRP of the transmitting Earth station (see notes for line 20, Table 2.4). To do this properly, we must know a particular specification of the satellite ahead of time; this is the receive saturation flux density (SFD), also called the flux density to saturate (FTS). Saturation in this case refers to the maximum output of the satellite transponder power amplifier, which in turn produces the saturated EIRP in the downlink. For typical C- and Ku-band satellites with national and regional footprints, the SFD is in the range of –80 to –95 dBW/m2 (the more negative the number, the less power is required to achieve saturation). The process we use to estimate uplink power is to solve the link budget equation for the EIRP needed to produce the SFD that is specified for the particular satellite. Otherwise, the calculation of uplink C/N is the same as for the downlink.

Table 2.4   Link Budget Analysis for the Uplink (6.175 GHz, C-band)

Item	Link Parameter	Value	Units	Computation
16	Transmit power (850W)	29.3	dBW	
17	Transmit waveguide losses	2.0	dB	
18	Transmit antenna gain (7m)	50.6	dBi	
19	Uplink EIRP from Boston	77.9	dBW	16 – 17 + 18
20	Spreading loss	162.2	dB(m^2)	
21	Atmospheric attenuation	0.1	dB	
22	Flux density at the spacecraft	–84.4	dBW/m^2	19 – 20 – 21
23	Free-space loss	200.4	dB	
24	Receive antenna gain	26.3	dBi	
25	Receive waveguide loss	0.5	dB	
26	System noise temperature (450K)	26.5	dB(K)	
27	Spacecraft G/T	–0.7	dB/K	24 – 25 – 26
28	Received C/T	–122.9	dBW/K	19 – 23 – 21 + 27
29	Boltmann's constant	–228.6	dBW/Hz/K	
30	Bandwidth (25 MHz)	74.0	dB Hz	
31	Carrier-to-noise ratio	31.7	dB	28 – 29 – 30

Table 2.4 presents the link budget from the transmitting Earth station to the satellite, reviewed in this section. Since this link budget is very similar to that of the downlink, we occasionally refer to the previous explanations for items 1 through 15.

16. The Earth station high-power amplifier (HPA) provides sufficient power to operate the satellite transponder at saturation. Here, 850W was derived to provide sufficient uplink EIRP to achieve an SFD value at the satellite of approximately –85 dBW/m^2.

17. An allocation of 2 dB is made to account for the loss between the HPA and the Earth station antenna feed. This is consistent with 40m of flexible waveguide at a loss of 0.05 dB per meter.

18. A 7-m Earth station antenna diameter provides 50.6 dBi of C-band gain, according to the formula given in item 7.

19. Uplink EIRP must be sufficient to saturate the satellite transponder. We determine its value from the saturation flux density requirement of the satellite in item 22.

20. The spreading loss allows us to convert from Earth station EIRP to the corresponding value of flux density at the face of the satellite receive antenna. It is calculated as $10 \log(4\pi R_0^2)$, where R_0^2 is the slant range. The units are in dB(m^2).

21. The atmospheric loss at 6 GHz is slightly greater than at 4 GHz, but both values are typically small enough to be ignored.

22. It is customary in commercial satellite communications to specify the uplink driving signal to the transponder in terms of the flux density. The SFD, in particular, is that which causes the transponder to transmit the maximum EIRP in the downlink. Once you know the SFD for the satellite, you can

compute the required EIRP for the Earth station through the reverse of the calculation:

Uplink EIRP = Spreading Loss + Atmospheric Loss − SFD

Figure 2.7 provides the SFD for Telstar V in the form of a coverage footprint. This is based on the antenna gain and takes account of the repeater gain between the antenna and the transponder power amplifier.

23. At this point, we revert to a direct calculation of carrier power received at the satellite. We use the free-space loss in lieu of the spreading loss for certain types of uplink calculations. Free-space loss is calculated according to the method in item 5 and (2.5). Alternatively, we could have used the spreading loss and added a term that is the area of an isotropic antenna at this frequency [i.e., $10\log(\lambda^2/4\pi)$.

The next set of calculations are used to predict the receive G/T of the satellite, based on typical spacecraft antenna and receiver characteristics. In reality, the G/T is specified for the particular satellite. This example provides the approach used to assess G/T in a satellite design and is shown for illustrative purposes only.

24. The spacecraft antenna is designed to cover a particular geographic area, here assumed to be the United States. The typical design provides a minimum value of 27 dBi, although 26.3 dBi is shown.

25. An allocation of 0.5 dB is made for the loss between the spacecraft antenna and the receiver front end (which employs an LNA and downconverter).

26. The typical C-band satellite has a system noise temperature of 450K (equivalently, 26.5 dBK), which includes 270K for the antenna temperature (microwave "brightness" of the Earth in noise terms), 50K for the waveguide line, and 130K for the receiver itself. Because the uplink receiver is already exposed to over 300K of background noise, it has become the practice to design the receiver for excellent linearity, flatness of frequency response, and reliability.

27. The third key satellite performance parameter is the G/T, or receiving system figure of merit. As stated in the explanation of item 10, G/T is the difference between the net antenna (including waveguide loss) and the system noise temperature expressed in decibels. The G/T and SFD differ by a fixed constant since both are dependent on the gain of the spacecraft antenna. From the specific example in Figure 2.6, $G/T = -(SFD + 85.1)$ dB/K. A formula of this type would be specified for each satellite design.

28. The value of C/T received by the satellite is calculated from the power balance as

$$C/T = EIRP - A_0 - A_{at} + G/T$$

where A_0 is the free-space loss and A_{at} is the atmospheric loss.

29.–31. These values are considered in the same manner as item 12. The value of uplink C/N presented in this item (i.e., 31.7 dB) is substantially higher

Satellite position: 97.0° W   Peak: 3.8 dB/K

Figure 2.7 The uplink coverage footprint of the Telstar V satellite, located at 97° WL. The contours are indicated with the SFDM in the direction of the Earth station.

than the downlink value in item 12 (i.e., 11.4 dB). Under this condition, the downlink will dominate the overall link performance, as discussed in the next section.

The repeater in this design is a simple bent pipe that does not alter or recover data from the transmission from the uplink. The noise on the uplink (e.g., N in the denominator of C/N) will be transferred directly to the downlink and added to the downlink noise computed in Section 2.2.1. The process for doing this is reviewed in the next section. In a baseband processing type of repeater, the uplink carrier is demodulated within the satellite and only the bits themselves are transferred to the downlink. In such case, the uplink noise only produces bit errors (and possibly frame errors, depending on the modulation and multiple access scheme) that transfer over the remodulated carrier. This is a complex process and can only be assessed for the particular transmission system design in a digital processing satellite.

2.2.3   Overall Link

The last step in link budgeting for a bent-pipe repeater is to combine the two link performances and compare the result against a minimum requirement—also called the threshold. Table 2.5 presents a detailed evaluation of the overall link under the conditions of line-of-sight propagation in clear sky. We have included an allocation for interference coming from sources such as a cross-polarized transponder and adjacent satellites. This type of entry is necessary because all operating satellite networks are exposed to one or more sources of interference. The bottom line represents the margin that is available to counter rain attenuation and any other losses that were not included in the link budgets. Alternatively, rain margin can be allocated separately to the uplink and downlink, with the combined availability value being the arithmetic product of the two as a decimal value (e.g., if the uplink and

Table 2.5   Combining the Uplink and the Downlink to Estimate Overall Link Performance

Item	Link Parameter	Value	Units	Computation
32	Uplink C/N (31.7 dB)	1,479.1	Ratio	31
33	N_u/C	0.000676	Ratio	
34	Downlink C/N (11.4 dB)	13.8	Ratio	15
35	N_d/C	0.0724	Ratio	
36	Total thermal noise (N_{th}/C)	0.0731	Ratio	33 + 35
37	Total thermal C/N_{th}	13.7	Ratio	
38	Total thermal C/N_{th}	11.4	dB	
39	Interference C/I (18.0 dB)	63.1	Ratio	Assumption
40	I/C	0.015848	Ratio	
41	Total noise ($N_{th} + I$)/C	0.0889	Ratio	36 + 40
42	Total $C/(N_{th} + I)$	11.2	Ratio	
43	Total $C/(N_{th} + I)$	10.5	dB	
44	Required C/N	8.0	dB	Equipment

downlink were each 99.9%, then the combined availability is $0.999 \times 0.999 = 0.998$ or 99.8%). We have included itemized remarks as for the previous examples.

32. The uplink C/N (line 12) is converted to a true value (not decibels) using the transformation, $x = 10^{(X/10)}$.

33. This is simply the inverse, which provides the uplink noise in a normalized form.

34. and 35. See comments for items 32 and 33.

36. This is an important step in the overall evaluation. The normalized uplink and downlink noise terms (items 34 and 35) are added together. Section 2.2.4 provides the procedure for this calculation. We see that the downlink noise is much larger than the uplink noise (because the downlink C/N is 20.3 dB lower). In fact, the uplink only contributes about 1% of the total.

37. The inverse of item 31 is the total C/N_{th}, in normalized units (not decibels). The subscript, th, indicates that the noise comes from thermal sources in the Earth station and satellite repeater.

38. This is the combined C/N_{th} in decibels.

39. An estimated C/I value of 18 dB is shown on this line. As reviewed in Section 2.2.4, we would perform several interference calculations based on the likely sources of interference into the system. Cross-polarized interference will come from transmissions to and from the satellite on the same frequency but in the opposite polarization. For most common dual-polarized satellites, this would be from 25 to 35 dB below a saturated carrier. Adjacent satellite interference would be calculated based on the performance of the adjacent satellite, the orbit separation, and the sidelobe characteristics of the Earth station. While a detailed discussion is beyond the scope of this book, the latter can be estimated using the following formula (as prescribed by the ITU) [5]. For an offset angle, θ, for the main beam greater than $100D/\lambda$ (corresponding to a peak gain greater than about 48 dBi), the sidelobe gain at θ can be estimated from:

$$G(\theta) = 29 - 25\log\theta \ \text{ in dBi}$$

The angle θ is measured from the Earth station location and is the angle between the line of sight to the desired satellite and that to the interfering satellite. Due to the local geometry, θ is slightly larger than the orbit spacing (hence, the estimate of $G(\theta)$ is conservatively high for the interference effect). This formula is a worst case and there are more detailed specifications that may be appropriate. If the interfering satellite has the same EIRP in the direction of our Earth station, then the maximum value of C/I is simply the peak gain of the antenna minus the sidelobe gain, $G(\theta)$, given by this formula. Any difference in EIRP between the interfering and desired satellites will either increase (interfering satellite is weaker) or decrease (interfering satellite is stronger) the resultant C/I. The particular case just considered is for downlink interference; there is a corresponding uplink interference case to be evaluated separately. Further details on the

consideration of interference within the frequency coordination process are discussed in Chapter 12. At C-band there might also be terrestrial interference since this band is often shared with microwave links.

40. See comments for items 32 and 33.

41.–43. With all of the noise and interference sources included, we now have the combined *C/N* for the uplink and downlink using the same formula as in item 36. This is the value that we expect to measure at the input to the receiving Earth station demodulator.

44. The required value of *C/N* is specified for the receiver digital demodulator. This characteristic is determined from the demodulator design and can be verified in the laboratory. Digital receivers usually include FEC and therefore can operate at lower values of *C/N* than analog (FM) demodulators. The actual operating *C/N* is set by a maximum allowable error rate for the type of service provided. This cannot be specified in general as it depends heavily on the type of coding and the tolerance of the end-user device to errors and other signal degradations. Digital demodulator threshold is typically specified not by *C/N* but rather using the ratio of the energy per bit to noise density (E_b/N_0), commonly called the "Eb-No" in the trade. The numerator is calculated by dividing the carrier power, *C*, by information bit rate, R_b. The noise density, N_0, is obtained by dividing the noise power, *N*, by the bandwidth of the carrier, *B*. Consequently, the conversion is simply:

$$\frac{E_b}{N_0} = \frac{C}{N} \cdot \frac{B}{R_b}$$

Some specific guidelines for threshold in digital video are provided in Chapter 5.

45. The link margin is simply the difference between the total *C/N* and the required value. It provides a cushion against variations in the link that the budget does not include directly. Obviously, if we included every possible degrading factor, there would be no need for any excess or "system" margin—which is a goal in link design. External forces that can pull the link down from the value on this item include rain attenuation (the biggest single factor), atmospheric fading due to ducting, ionospheric scintillation, antenna misalignment, and satellite motion. As stated previously, all of these factors can be (and often are) included as individual entries. Some system or excess margin may be appropriate to cover risk that hardware performance may fall short, that installation could be imperfect, and that operating procedures may not be fully adequate to maintain service quality under all conditions.

2.2.4 Additional Sources of Noise and Interference

Since most satellites will have neighbors in orbit that operate on the same frequencies, we need to include the contribution of orbital interference. Chapter 12 contains a more detailed discussion of this subject. In a typical link, orbital interference could add 30% or more to the total thermal noise, although the precise amount is

determined by the number of satellites, their spacing, and the types of signals in operation on both the desired and interfering satellites. Another source of noise is intermodulation distortion produced in the transponder and the Earth station HPA. The contribution here could be as much as 100% of the total thermal noise (i.e., it could be equal). C-band links also include the interference caused by terrestrial microwave stations that are within range of the receiving Earth station. The obvious advantage of using a band not shared with such microwave stations is that this particular interference source is not present.

We can extend the budget shown in Table 2.5 to include the additional sources of link noise. In general, we can apply the following formula for total link C/N (thermal noise, distortion, and interference):

$$\frac{C}{N_{total}} = \left[\frac{N_{up}}{C} + \frac{N_{dwn}}{C} + \frac{IM}{C} + \frac{I_{xpol}}{C} + \frac{I_{asi}}{C} + \frac{I_{ti}}{C} \right]^{-1}$$

This rather imposing formula simply states that you can combine all of the contributions together by first converting each C/N to a true ratio, invert each to show its relative noise contribution, add up the noise contributions in this form, and then invert the sum. The result is the total C/N_{tot} as a true ratio. As a last step, convert the total C/N to a decibel value and compare it to the requirement, as in items 43 to 45.

Link budget analysis is probably the most important engineering discipline in designing a satellite application. We recommend that you take the time to get a basic understanding so that you may at least be able to ask the right questions. A low cost yet very effective software tool for this is SatMaster Pro, offered by Arrowe Technical Services of the United Kingdom (http://www.arrowe.com). The product runs on MS Windows computers and is attractively priced. A free trial version (which lacks features needed to save and print files) can be obtained from the Web site as well as the full product. SatMaster Pro can be used for single link analysis for a pair of Earth stations using a variety of popular modulation schemes; the Multi-link (MLink) version permits computation of a satellite serving a network of hundreds of sites in a single pass.

2.3 Multiple Access Systems

Applications employ multiple-access systems to allow two or more Earth stations to simultaneously share the resources of the same transponder or frequency channel. These include the three familiar methods: FDMA, TDMA, and CDMA. Another multiple access system called space division multiple access (SDMA) has been suggested in the past. In practice, SDMA is not really a multiple access method but rather a technique to reuse frequency spectrum through multiple spot beams on the satellite. Because every satellite provides some form of frequency reuse (cross-polarization being included), SDMA is an inherent feature in all applications. TDMA and FDMA require a degree of coordination among users: FDMA users cannot transmit on the same frequency and TDMA users can transmit on the same frequency but not at the same time. Capacity in either case can be calculated based on the total bandwidth and power available within the transponder or slice of a

transponder. CDMA is unique in that multiple users transmit on the same frequency at the same time (and in the same beam or polarization). As will be discussed, this is allowed because the transmissions use a different code either in terms of high-speed spreading sequence or frequency hopping sequence. The capacity of a CDMA network is not unlimited, however, because at some point the channel becomes overloaded by self-interference from the multiple users who occupy it. Furthermore, power level control is critical because a given CDMA carrier that is elevated in power will raise the noise level for all others carriers by a like amount.

Multiple access is always required in networks that involve two-way communications among multiple Earth stations. The selection of the particular method depends heavily on the specific communication requirements, the types of Earth stations employed, and the experience base of the provider of the technology. All three methods are now used for digital communications because this is the basis of a majority of satellite networks. The digital form of a signal is easier to transmit and is less susceptible to the degrading effects of the noise, distortion from amplifiers and filters, and interference. Once in digital form, the information can be compressed to reduce the bit rate, and FEC is usually provided to reduce the required carrier power even further. The specific details of multiple access, modulation, and coding are often preselected as part of the application system and the equipment available on a commercial off-the-shelf (COTS) basis. The only significant analog application at this time is the transmission of cable TV and broadcast TV. These networks are undergoing a slow conversion to digital as well, which may in fact be complete within a few years of this edition's publication.

2.3.1 Frequency Division Multiple Access

Nearly every terrestrial or satellite radio communications system employs some form of FDMA to divide up the available spectrum. The areas where it has the strongest hold are in single channel per carrier (SCPC), intermediate data rate (IDR) links, voice telephone systems, VSAT data networks, and some video networking schemes. Any of these networks can operate alongside other networks within the same transponder. Users need only acquire the amount of bandwidth and power that they require to provide the needed connectivity and throughput. Also, equipment operation is simplified since no coordination is needed other than assuring that each Earth station remains on its assigned frequency and that power levels are properly regulated. However, intermodulation distortion (IMD) present with multiple carriers in the same amplifier must be assessed and managed as well.

As discussed in Chapter 13, the satellite operator divides up the power and bandwidth of the transponder and sells off the capacity in attractively priced segments. Users pay for only the amount that they need. If the requirements increase, additional FDMA channels can be purchased. The IMD that FDMA produces within a transponder must be accounted for in the link budget; otherwise, service quality and capacity will degrade rapidly as users attempt to compensate by increasing uplink power further. The big advantage, however, is that each Earth station has its own independent frequency on which to operate. A bandwidth segment can be assigned to a particular network of users, who subdivide the spectrum further based on individual needs. Another feature, discussed in Chapter 10, is to assign carrier

frequencies when they are needed to satisfy a traffic requirement. This is the general class of demand assigned networks, also called demand-assigned multiple access (DAMA). In general, DAMA can be applied to all three multiple access schemes previously described; however, the term is most often associated with FDMA.

2.3.2   Time Division Multiple Access and ALOHA

TDMA is a truly digital technology, requiring that all information be converted into bit streams or data packets before transmission to the satellite. (An analog form of TDMA is technically feasible but never reached the market due to the rapid acceptance of the digital form.) Contrary to most other communication technologies, TDMA started out as a high-speed system for large Earth stations. Systems that provided a total throughput of 60 to 250 Mbps were developed and fielded over the past 25 years. However, it is the low-rate TDMA systems, operating at less than 10 Mbps, which provide the foundation of most VSAT networks. As the cost and size of digital electronics came down, it became practical to build a TDMA Earth station into a compact package. Lower speed means that less power and bandwidth need to be acquired (e.g., a fraction of a transponder will suffice) with the following benefits:

- The full cost of a transponder can be avoided.
- The uplink power from small terminals is reduced, saving on the cost of transmitters.
- The network capacity and quantity of equipment can grow incrementally, as demand grows.

Considerable information on the capabilities and design of VSAT networks is provided in Chapter 9.

TDMA signals are restricted to assigned time slots and therefore must be transmitted in bursts. This is illustrated in Figure 9.4 for a hypothetical TDMA time frame of 45 ms. The time frame is periodic, allowing stations to transfer a continuous stream of information on average. Reference timing for start-of-frame is needed to synchronize the network and provide control and coordination information. This can be provided either as an initial burst transmitted by a reference Earth station, or on a continuous basis from a central hub (discussed in detail in Chapter 9). The Earth station equipment takes one or more continuous streams of data, stores them in a buffer memory, and then transfers the output toward the satellite in a burst at a higher compression speed. At the receiving Earth station, bursts from Earth stations are received in sequence, selected for recovery if addressed for this station, and then spread back out in time in an output expansion buffer. It is vital that all bursts be synchronized to prevent overlap at the satellite; this is accomplished either with the synchronization burst (as shown) or externally using a separate carrier. Individual time slots may be preassigned to particular stations or provided as a reservation, with both actions under control by a master station. For traffic that requires consistent or constant timing (e.g., voice and TV), the time slots repeat at a constant rate.

Computer data and other forms of packetized information can use dynamic assignment of bursts in a scheme much like a DAMA network. There is an adaptation for data, called ALOHA, that uses burst transmission but eliminates the

assignment function of a master control. ALOHA is a powerful technique for low-cost data networks that need minimum response time. Throughput must be less than 20% if the bursts come from stations that are completely uncoordinated because there is the potential for time overlap (called a collision). This is illustrated in Figure 9.5 for slotted ALOHA, which is a variant that requires that bursts start and end within the timing intervals. The most common implementation of ALOHA employs a hub station that receives all of these bursts and provides a positive acknowledgement to the sender if the particular burst is good. If the sending station does not receive acknowledgment within a set "time window," the packet is re-sent after a randomly selected period is added to prevent another collision. This combined process of the window plus added random wait introduces time delay, but only in the case of a collision. Throughput greater than 20% brings a high percentage of collisions and resulting retransmissions, introducing delay that is unacceptable to the application. In Chapter 9, we review the performance of ALOHA and compare it to TDMA on a performance basis.

An optimally and fully loaded TDMA network can achieve 90% throughput, the only reductions required for guard time between bursts and other burst overhead for synchronization and network management. The corresponding time delay is approximately equal to one-half of the frame time, which is proportional to the number of stations sharing the same channel. This is because each station must wait its turn to use the shared channel. ALOHA, on the other hand, allows stations to transmit immediately upon need. Time delay is minimum, except when you consider the effect of collisions and the resulting retransmission times.

The standard digital modulation used is phase shift keying (PSK), with the most popular form being quaternary PSK (abbreviated QPSK). The advantage of QPSK is that it doubles the number of bits per second that are carried within a given amount of bandwidth. QPSK modems are now integrated into cellular phones, Inmarsat terminals, and VSATs; the receive portion is also a standard feature of every DBS receiver. Modulator and demodulator design are important to link operation and performance. Variants such as minimum shift keying (MSK) and Gaussian MSK (GMSK) have appeared on the market to better utilize the low power of solid-state transmitters. In FDMA, the modem operates more or less continuously and threshold performance can be optimized. Modems for TDMA must operate in the burst mode, meaning that the demodulator must acquire and reacquire the signal rapidly to capture data from different Earth stations operating on the same frequency.

The most important parameter in digital transmission is the bit error rate (BER). Common requirements in satellite communications are digitized telephone (voice) at 10^{-4}, medium and intermediate data rate data transmission at 10^{-7}, and digital video at 10^{-8}. The BER can be reduced (e.g., going from 10^{-4} to 10^{-8}) by using FEC within the modem, a feature that takes advantage of custom VLSI and DSP chips. These codes automatically correct errors in the received data, yielding a significant improvement in the error rate, which is delivered to the end user. The trade-off with FEC is an increase in data rate (to include the extra FEC bits) in exchange for a decrease in the error rate by at least three orders of magnitude. The precise improvement depends on the percentage of extra bits and the FEC coding and decoding algorithms. Performance has been further improved by performing two encodings of the data through a process called concatenation (see the discussion of the DVB standard in Chapter 5).

More recently, concatenation has been extended through turbo codes, which increase the effectiveness of FEC at least another order of magnitude for the same E_b/N_0 and bandwidth. Alternatively, the turbo code principle reduces the required E_b/N_0 by 1 to 2 dB, which directly helps the link budget bottom line.

Data communication systems use protocols to control transfer of information. These are arranged in a hierarchy of layers that define how communicating nodes and end computing devices control data transfer and verify that no data are corrupted [6]. The standard layers, according to the Open Systems Interconnection (OSI) model (starting from the lowest) are: (1) the physical layer, (2) the link layer, (3) the network layer, (4) the transport layer, (5) the session layer, (6) the presentation layer, and (7) the application layer. Each layer provides services, defined by the protocol, to the layer immediately below it. For satellite applications, the physical layer refers to the actual Earth station equipment that multiplexes and modulates/demodulates the information; thus, this layer includes all of the baseband and RF transmission equipment from originating Earth station, through the satellite, and to the destination receiving Earth station. This is certainly the case for a bent-pipe satellite repeater, although digital processing satellites can incorporate features of the link and network layers to achieve various switching and multiplexing functions and enhance throughput. The link layer defines the protocol structure of every block of data and how requests for retransmission are processed. Detection of errors at the link and network layers is afforded by parity check bits and cyclic redundancy check (CRC) computations. For data, error is also addressed by the end-to-end devices that employ automatic retransmission (character or block oriented). Block-oriented protocols like the high-level datalink control (HDLC), which use the "look back N" scheme, are the most effective for satellite links, with significant propagation delay. Fortunately, these tolerant link layer protocols have become standard in all data communication applications. The transport layer, like TCP of the Internet Protocol, provides these functions as well. Higher layers are outside of the network and are associated with the software applications that require communication services.

TDMA is a good fit for all forms of digital communications and should be considered as one option during the design of a satellite application. The complexity of maintaining synchronization and control has been overcome through miniaturization of the electronics and by way of improvements in network management systems. With the rapid introduction of TDMA in terrestrial radio networks like the GSM standard, we will see greater economies of scale and corresponding price reductions in satellite TDMA equipment.

2.3.3   Code Division Multiple Access

CDMA, also called spread spectrum communication, differs from FDMA and TDMA because it allows users to literally transmit on top of each other. This feature has allowed CDMA to gain attention in commercial satellite communication. It was originally developed for use in military satellite communication where its inherent antijam and security features are highly desirable. CDMA was adopted in cellular mobile telephone as an interference-tolerant communication technology that increases capacity above analog systems. Some of these claims are well founded;

however, it has not been proven that CDMA is universally superior as this depends on the specific requirements. For example, an effective CDMA system requires contiguous bandwidth equal to at least the spread bandwidth. Two forms of CDMA are applied in practice: (1) direct sequence spread spectrum (DSSS) and (2) frequency hopping spread spectrum (FHSS). FHSS has been used by the OmniTracs and Eutel-Tracs mobile messaging systems for more than 10 years now, and only recently has it been applied in the consumer's commercial world in the form of the Bluetooth wireless LAN standard. However, most CDMA applications over commercial satellites employ DSSS (as do the cellular networks developed by Qualcomm).

A simplified block diagram of a basic DSSS link is provided in Figure 2.8. The basic principle of operation is that an input data stream of R_b bps at A is mixed with a pseudorandom scrambling bit sequence at B with a rate n times R_b. The value of n is generally in the range of 10 to 1,000, which has the effect of multiplying the bandwidth of the output at C by the same factor. The higher the value of n, the greater the spread bandwidth and the greater the protection from interference and jamming. A modulator converts the baseband version of the signal to an RF carrier at D using standard PSK or QPSK. The output is the spread spectrum signal that can be subjected to link noise at E. Interference from other CDMA carriers as well as other signals could also be introduced. Often, the bandwidth of an interfering carrier is much less than that of the spread spectrum signal, which translates into suppression of the interference at the CDMA receiver.

At F, the combination of spread spectrum signal, noise, and interference is applied to the receiving PSK demodulator to recover the baseband spread spectrum signal. A mixer is used to multiply the baseband by the same pseudorandom sequence that was used in the transmitter. However, the difference here is that the signal baseband at G has been modulated by the original data. This difference between the received signal code pattern and that of the raw chip sequence is what allows the receiver to recover the original data at I. For this to occur, the receiver's pseudorandom sequence must precisely synchronize itself with the incoming signal.

Figure 2.8 A DSSS communication channel for the application of CDMA over satellite link. The receiver at the bottom synchronizes to the desired spread spectrum carrier using the technique of correlation detection. Despreading of the synchronized received carrier also causes nonsynchronized and interfering carriers to be spread further, making them appear like very wideband noise in the narrowband detector.

This is accomplished by the technique of correlation detection wherein the locally generated spreading sequence is slid in time past the incoming signal until they can be exactly superimposed on each other. Once synchronized, the output at I contains the original bit pattern along with the noise and interference acquired along the satellite transmission path. The bit timing circuitry can recover the original pulses and square shaping is restored. The output data stream at J contains the original data plus occasional errors (inverted bits) that result from noise and interference that remains after the despreading process. However, there is a significant reduction in the effect of interference-induced errors because the despreading process does just the opposite to any interference carrier; namely, it spreads it out over a bandwidth related to n times R_b. The jamming margin by which the receiver rejects narrowband interference is approximately $10\log n$.

As an introduction to nontechnical readers, consider the following summary of the features of spread spectrum technology (whether DSSS or FHSS):

- Simplified multiple access: no requirement for coordination among users;
- Selective addressing capability if each station has a unique chip code sequence—provides authentication: alternatively, a common code may still perform the CDMA function adequately since the probability of stations happening to be in synch is approximately $1/n$;
- Low-power spectral density: bandwidth is spread by the code over a bandwidth, which is n times that of the original data; this reduces the radiated power spectral density in inverse proportion;
- Relative security from eavesdroppers: the low spread power and relatively fast direct sequence modulation by the pseudorandom code make detection difficult;
- Interference rejection: the spread-spectrum receiver treats the other DSSS signals as thermal noise and suppresses narrowband interference.

Selective addressing means that each transmission is automatically identified as to its source by the specific code that spreads the signal in the first place. Further addressability comes from a unique word that is attached to a data header, identifying the source or recipient. Through this and the remaining features, spread spectrum permits many stations to operate on the same frequency channel because only one of the signals will be detected by any given receiver. However, the undesired spread spectrum signals will appear as noise that will still degrade the total C/N performance. The consequence of this particular feature is that it is extremely difficult to accurately determine the maximum number of signals to simultaneously share the same channel.

CDMA has special features that make it advantageous under certain conditions. For example, in a multibeam satellite system, CDMA permits the same frequencies to be used in adjacent beams. Unlike FDMA or TDMA, however, the CDMA signals from these beams will add to the total noise budget. If these conditions of appropriateness are not satisfied, then CDMA may not excel in capability and, in fact, can bring with it a penalty relative to FDMA or TDMA. The difficult part is determining the real situation well ahead of an expensive implementation project. This is because the loading of CDMA transmissions on top of each other may follow

an unpredictable pattern. Individual carriers must be controlled in power since the carriers by their nature look to the system like noise. A given carrier that is, say, 6 dB above the proper level will introduce four times the expected interference (the original plus the effect of three more). On average, the network could experience an excessive loading of self-interference.

The detailed design of the spread-spectrum receiver is critical to proper CDMA operation because it incorporates all of the features of a good PSK modem plus the ability to acquire the spread signal. A significant part of the challenge is that the signal power density, measured in watts per hertz, is already less than the noise power density in the receiver. Stated another way, the RF C/N is actually less than one (i.e., negative in terms of decibels) because the information is below the noise level.

A typical CDMA receiver must carry out the following functions in order to acquire the signal, maintain synchronization, and reliably recover the data:

- Synchronization with the incoming code through the technique of correlation detection;
- Despreading of the carrier;
- Tracking the spreading signal to maintain synchronization;
- Demodulation of the basic data stream;
- Timing and bit detection;
- Forward error correction to reduce the effective error rate;

The first three functions are needed to extract the signal from the clutter of noise and other signals. The processes of demodulation, bit timing and detection, and FEC are standard for a digital receiver, regardless of the multiple access method.

The bottom line in multiple access is that there is no single system that provides a universal answer. FDMA, TDMA, and CDMA will each continue to have a place in building the applications of the future. They can all be applied to digital communications and satellite links. When a specific application is contemplated, our recommendation is to perform the comparison to make the most intelligent selection.

2.4 Frequency Band Trade-Offs

Satellite communication is a form of radio or wireless communication and therefore must compete with other existing and potential uses of the radio spectrum. During the initial 10 years of development of these applications, there appeared to be more or less ample bandwidth, limited only by what was physically or economically justified by the rather small and low powered satellites of the time. In later years, as satellites grew in capability, the allocation of spectrum has become a domestic and international battlefield as service providers fight among themselves, joined by their respective governments when the battle extends across borders. So, we must consider all of the factors when selecting a band for a particular application.

The most attractive portion of the radio spectrum for satellite communication lies between 1 and 30 GHz. The relationship of frequency, bandwidth, and application are shown in Figure 2.9. The scale along the x-axis is logarithmic in order to show all of the satellite bands; however, observe that the bandwidth available for

▮	L-band (1–2 GHz)						Mobile Satellite Service (MSS), UHF TV, terrestrial microwave and studio television links, cellular phone
	▮ S-band (2–4 GHz)						MSS, Digital Audio Radio Service (DARS) NASA and deep space research
		▮ C-band (4–8 GHz)					Fixed Satellite Service (FSS), fixed service terrestrial microwave
			▮ X-band (8–12.5 GHz)				FSS military communication, DARS feeder links, fixed service terrestrial, Earth observation satellites
	Ku-band (12.5–18 GHz) ▮						FSS, Broadcast Satellite Service (BSS) fixed service terrestrial microwave
	K-band (18–26.5 GHz) ▮						BSS, FSS, fixed service terrestrial microwave, local multichannel distribution service (LMDS)
	Ka-band (26.5–40 GHz) ▮						FSS, fixed service terrestrial microwave, LMDS, Intersatellite links (ISL), satellite imaging

```
1      2     3   4  5  6   8  10          20    30   40
```

Frequency, GHz

Figure 2.9   The general arrangement of the frequency spectrum that is applied to satellite communications and other radiocommunication services. Indicated are the short-hand letter designations along with an explanation of typical applications. Note the logarithmic scale and that frequency ranges are the general ranges and do not correspond exactly to the ITU frequency allocations and allotments.

applications increases in real terms as one moves toward the right (i.e., frequencies above 3 GHz). Also, the precise amount of spectrum that is available for services in a given region or country is usually less than Figure 2.9 indicates. Please refer to the latest edition of the ITU Radio Regulations available for purchase at the ITU Web site (http://www.itu.int) or relevant domestic allocations in the country of interest. Chapter 12 provides a review of the regulatory process.

The use of letters probably dates back to World War II as a form of shorthand and simple code for developers of early microwave hardware. Two band designation systems are in use: adjectival (meaning the bands are identified by the following adjectives) and letter (which are codes to distinguish bands commonly used in space communications and radar).

Adjectival band designations, frequency in gigahertz:

Very high frequency (VHF): 0.03–0.3;
Ultra high frequency (UHF): 0.3–3;
Super high frequency (SHF): 3–30;
Extremely high frequency (EHF): 30–300.

Letter band designations, frequency in gigahertz (differs slightly from Chapter 1):
L: 1.0–2.0;
S: 2.0–4.0;
C: 4.0–8.0;
X: 8–12;
Ku: 12–18;

Ka: 18–40;
Q: 40–60;
V: 60–75;
W: 75–110.

Today, the letter designations continue to be the popular buzzwords that identify band segments that have commercial application in satellite communications. The international regulatory process, maintained by the ITU, does not consider these letters but rather uses band allocations and service descriptors listed next and in the right-hand column of Figure 2.9:

- Fixed Satellite Service (FSS): between Earth stations at given positions, when one or more satellites are used; the given position may be a specified fixed point or any fixed point within specified areas; in some cases this service includes satellite-to-satellite links, which may also be operated in the intersatellite service; the FSS may also include feeder links for other services.
- Mobile Satellite Service (MSS): between mobile Earth stations and one or more space stations (including multiple satellites using intersatellite links). This service may also include feeder links necessary for its operation.
- Broadcasting Satellite Service (BSS): A service in which signals transmitted or retransmitted by space stations are intended for direct reception by the general public. In the BSS, the term "direct reception" shall encompass both individual reception and community reception.
- Intersatellite Link (ISL): A service providing links between artificial satellites.

The general properties of these bands are reviewed in [3]. Suffice it to say, the lower the band in frequency, the better the propagation characteristics. This is countered by the second general principle, which is that the higher the band, the more bandwidth that is available. The MSS is allocated to the L- and S-bands, where propagation is most forgiving. Yet, the bandwidth available between 1 and 2.5 GHz, where MSS applications are authorized, must be shared not only among GEO and non-GEO applications, but with all kinds of mobile radio, fixed wireless, broadcast, and point-to-point services as well. The competition is keen for this spectrum due to its excellent space and terrestrial propagation characteristics. The rollout of wireless services like cellular radiotelephone, PCS, wireless LANs, and 3G may conflict with advancing GEO and non-GEO MSS systems. Generally, government users in North America and Europe, particularly in the military services, have employed selected bands such as S, X, and Ka to isolate themselves from commercial applications. However, this segregation has disappeared as government users discover the features and attractive prices that commercial systems may offer.

On the other hand, wideband services like DTH and broadband data services can be accommodated at frequencies above 3 GHz, where there is more than 10 times the bandwidth available. Add to this the benefit of using directional ground antennas that effectively multiply the unusable number of orbit positions. Some wideband services have begun their migration from the well-established world of C-band to Ku- and Ka-bands. In the following sections we provide some additional comments about the relative merits of these bands. These should be considered as

starting points for evaluating the proper frequency band and are not substitutes for a detailed evaluation of the relative cost and complexity of different approaches. Higher satellite EIRP used at Ku-band allows the use of relatively small Earth station antennas. On the other hand, C-band should maintain its strength for video distribution to cable systems and TV stations, particularly because of the favorable propagation environment, extensive global coverage, and legacy investment in C-band antennas and electronic equipment.

2.4.1   Ultra High Frequency

While the standard definition of UHF is the range of 300 to 3,000 MHz (0.3 to 3 GHz), the custom is to relate this band to any effective satellite communication below about 1 GHz. Frequencies above 1 GHz are considered in the next sections. The fact that the ionosphere provides a high degree of attenuation below about 100 MHz makes this the certain low end of acceptability (the blockage by the ionosphere at 10 MHz goes along with its ability to reflect radio waves, a benefit for ground-to-ground and air-to-ground communications using what is termed sky wave or "skip"). UHF satellites employ circular polarization (CP) to avoid Faraday effect, wherein the ionosphere rotates any linear-polarized wave. The UHF spectrum between 300 MHz and 1 GHz is exceedingly crowded on the ground and in the air because of numerous commercial, government, and other civil applications. Principal among them is television broadcasting in the VHF and UHF bands, FM radio, and cellular radio telephone. However, we cannot forget less obvious uses like vehicular and handheld radios used by police officers, firefighters, amateurs, the military, taxis and other commercial users, and a variety of unlicensed applications in the home.

From a space perspective, the dominant space users are military and space research (e.g., NASA in the United States and ESA in Europe). These are all narrow bandwidth services for voice and low-speed data transfer in the range of a few thousand hertz or, equivalently, a few kilobytes per second. From a military perspective, the first satellite to provide narrowband voice services was Tacsat. This experimental bird proved that a GEO satellite provides an effective tactical communications service to a mobile radio set that could be transported on a person's back, installed in a vehicle, or operated from an aircraft. Subsequently, the U.S. Navy procured the Fleetsat series of satellites from TRW, a very successful program in operational terms. This was followed by Leasat from Hughes, and currently the UHF Follow-On Satellites from the same maker (now Boeing Satellite Systems).

From a commercial perspective, the only VHF project that one can identify is OrbComm, a low data rate LEO satellite constellation developed by Orbital Sciences Corporation. OrbComm provides a near-real-time messaging service to inexpensive handheld devices about the size of a small transistor radio. On the other hand, its more successful use is to provide occasional data transmissions to and from moving vehicles and aircraft. Due to the limited power of the OrbComm satellites (done to minimize complexity and investment cost), voice service is not supported. Like other LEO systems, OrbComm as a business went into bankruptcy; it may continue in another form as the satellites are expected to keep operating for some time.

2.4.2   L-Band

Frequencies between 1 and 2 GHz are usually referred to as L-band, a segment not applied to commercial satellite communication until the late 1970s. Within this 1 GHz of total spectrum, only about 30 MHz of uplink and downlink, each, was initially allocated by the ITU to the MSS. The first to apply L-band was COMSAT with their Marisat satellites. Constructed primarily to solve a vital need for UHF communications by the U.S. Navy, Marisat also carried an L-band transponder for early adoption by the commercial maritime industry. COMSAT took a gamble that MSS would be accepted by commercial vessels, which at that time relied on high-frequency radio and the Morse code. Over the ensuing years, Marisat and its successors from Inmarsat proved that satellite communications, in general, and MSS, in particular, are reliable and effective. By 1993, the last commercial HF station was closed down in favor of satellite links. With the reorganization and privatization of Inmarsat, the critical safety aspects of the original MSS network are being transferred to a different quasigovernmental operating group.

As is familiar to readers, early MSS Earth stations required 1-m dish antennas that had to be pointed toward the satellite. The equipment was quite large, complex, and expensive. Real demand for this spectrum began to appear as portable, land-based terminals were developed and supported by the network. Moving from rack-mounted to suitcase-sized to attaché case and finally handheld terminals, the MSS has reached consumers.

The most convenient L-band ground antennas are small and ideally do not require pointing toward the satellite. We are all familiar with the very simple cellular whip antennas used on cars and handheld mobile phones. Common L-band antennas for use with Inmarsat are not quite so simple because there is a requirement to provide some antenna gain in the direction of the satellite so a coarse pointing is needed. Additional complexity results from a dependence on circular polarization to allow the mobile antenna to be aligned along any axis (and to allow for Faraday rotation). First generation L-band rod or mast antennas are approximately 1m in length and 2 cm in diameter. This is to accommodate the long wire coil (a bifilar helix) that is contained within. The antenna for the handheld phone is more like a fat fountain pen.

While there is effectively no rain attenuation at L-band, the ionosphere does introduce a source of significant link degradation. This is in the form of rapid fading called ionospheric scintillation, which is the result of the RF signal being split into two parts: the direct path and a refracted (or bent) path. At the receiving station, the two signals combine with random phase. Then, the signals may cancel, producing a deep fade. Ionospheric scintillation is most pronounced in equatorial regions and around the equinoxes (March and September). Both ionospheric scintillation and Faraday rotation decrease in frequency increases and are nearly negligible at Ku-band and higher. Transmissions at UHF are potentially more seriously impaired and for that reason, and additional fade margin over and above that at L-band may be required.

From an overall standpoint, L-band represents a regulatory challenge but not a technical one. There are more users and uses for this spectrum than there is spectrum to use. Over time, technology will improve spectrum efficiency. Techniques like digital speech compression and bandwidth efficient modulation may improve the utilization of this very attractive piece of spectrum. The business failure of LEO systems

like Iridium and Globalstar had raised some doubts that L-band spectrum could be increased. One could argue that more lucrative land-based mobile radio services (e.g., cellular and wireless data services) could end up winning over some of the L-band. This will require never-ending vigilance from the satellite community.

2.4.3   S-Band

S-band was adopted early for space communications by NASA and other governmental space research activities around the world. It has an inherently low background noise level and suffers less from ionospheric effects than L-band. DTH systems at S-band were operated in past years for experiments by NASA and as operational services by the Indian Space Research Organization and in Indonesia. More recently, the ITU allocated a segment of S-band for MSS and Digital Audio Radio (DAR) broadcasting. These applications hold the greatest prospect for expanded commercial use on a global basis.

As a result of a spectrum auction, two companies were granted licenses by the FCC and subsequently went into service in 2001–2002. S-band spectrum in the range of 2,320 to 2,345 MHz is shared equally between the current operators, XM Radio and Sirius Satellite Radio. A matching uplink to the operating satellites was assigned in the 7,025- to 7,075-MHz bands. Both operators installed terrestrial repeaters that fill dead spots within urban areas. With an EIRP of nominally 68 dBW, these broadcast satellites can deliver compressed digital audio to vehicular terminals with low gain antennas.

As a higher frequency band than L-band, it will suffer from somewhat greater (although still low) atmospheric loss and less ability to adapt to local terrain. LEO and MEO satellites are probably a good match to S-band since the path loss is inherently less than for GEO satellites. One can always compensate with greater power on the satellite, a technique used very effectively at Ku-band.

2.4.4   C-Band

Once viewed as obsolete, C-band remains the most heavily developed and used piece of the satellite spectrum. During recent World Radiocommunication Conferences, discussed in Chapter 12, the ITU increased the available uplink and downlink bandwidth from the original allocation of 500 to 800 MHz. This spectrum is effectively multiplied by a factor of two with dual polarization and again by 180, assuming 2° spacing between satellites. Further reuse by a factor of between two and five takes advantage of the geographic separation of land coverage areas. The total usable C-band spectrum bandwidth is therefore in the range of 568 GHz to 1.44 THz, which compares well with land-based fiber optic systems. The added benefit of this bandwidth is that it can be delivered across an entire country or ocean region.

Even though this represents a lot of capacity, there are situations in certain regions where additional satellites are not easily accommodated. In North America, there are more than 35 C-band satellites in operation across a 70° orbital arc. This is the environment that led the FCC in 1985 to adopt the then radical (but necessary) policy of 2° spacing. The GEO orbit segments in Western Europe and east Asia are becoming just as crowded as more countries launch satellites. European

governments mandated the use of Ku-band for domestic satellite communications, delaying somewhat the day of reckoning. Asian and African countries favor C-band because of reduced rain attenuation as compared to Ku- and Ka-bands, making C-band slots a vital issue in that region.

C-band is a good compromise between radio propagation characteristics and available bandwidth. Service characteristics are excellent because of the modest amount of fading from rain and ionospheric scintillation. The one drawback is the somewhat large size of Earth station antenna that must be employed. The 2° spacing environment demands antenna diameters greater than 1m, and in fact 2.4m is more the norm. This size is also driven by the relatively low power of the satellite, itself the result of sharing with terrestrial microwave. High-power video carriers must generally be uplinked through antennas of between 7m and 13m; this assures an adequate signal and reduces the radiation into adjacent satellites and terrestrial receivers.

The prospects for C-band are good because of the rapid introduction of digital compression for video transmission. New C-band satellites with higher EIRP, more transponders, and better coverage are giving C-band new life in the wide expanse of developing regions such as Africa, Asia, and the Pacific.

2.4.5   X-Band

Government and military users of satellite communication established their fixed applications at X-band. This is more by practice than international rule, as the ITU frequency allocations only indicate that the 8-GHz portion of the spectrum is designated for the FSS regardless of who operates the satellite. From a practical standpoint, X-band can provide service quality on par with C-band; however, commercial users will find equipment costs to be substantially higher due to the thinner market. Also, military-type Earth stations are inherently expensive due to need for rugged design and secure operation. Some countries have filed for X-band as an expansion band, hoping to exploit it for commercial applications like VSAT networks and DTH services. As discussed previously, S-DARS in the United States employs X-band feeder uplinks. On the other hand, military usage still dominates for many fixed and mobile applications. This segregation helps maintain a degree of security for military users for whom availability of a larger consumer market would not necessarily be considered advantageous. X-band is likewise shared with terrestrial microwave systems, somewhat complicating frequency coordination.

2.4.6   Ku-Band

Ku-band spectrum allocations are somewhat more plentiful than C-band, comprising 750 MHz for FSS and another 800 MHz for the BSS. Again, we can use dual polarization and satellites positions 2° apart. Closer spacings are not feasible because users prefer to install yet smaller antennas, which have the same or wider beamwidth than the correspondingly larger antennas for C-band service. Typically implemented by different satellites covering different regions, Ku regional shaped spot beams with geographic separation allow up to approximately 10X frequency reuse. This has the added benefit of elevating EIRP using modest transmit power; G/T likewise increases due to the use of spot beams. The maximum available Ku-band spectrum could therefore amount to more than 4 THz.

Exploiting the lack of frequency sharing and the application of higher power in space, digital DTH services from DIRECTV and EchoStar in North America ushered in the age of low-cost and user-friendly home satellite TV. The United Kingdom, continental Western Europe, Japan, and a variety of other Asian countries likewise enjoy the benefits of satellite DTH. As a result of these developments, Ku-band has become a household fixture (if not a household word).

The more progressive regulations at Ku-band also favor its use for two-way interactive services like voice and data communication. Low-cost VSAT networks typify this exploitation of the band and the regulations. Being above C-band, the Ku-band VSATs and DTH receivers must anticipate more rain attenuation. A decrease in capacity can be countered by increasing satellite EIRP. Also, improvements on modulation and forward error correction are making terminals smaller and more affordable for a wider range of uses. Thin route applications for telephony and data, discussed in Chapters 8, 9, and 10, benefit from the lack of terrestrial microwave radios, allowing VSATs to be placed in urban and suburban sites.

2.4.7   Ka-Band

Ka-band spectrum is relatively abundant and therefore attractive for services that cannot find room at the lower frequencies. There is 2 GHz of uplink and downlink spectrum available on a worldwide basis (500 MHz of this spectrum has been allocated to non-GEO satellites, particularly Teledesic, and another 500 MHz for fixed wireless access). In addition, the fact that ground antenna beamwidths are between one-half to one-quarter the values that correspond at Ku- and C-bands means that more satellites could conceivably be accommodated. Conversely, with enough downlink EIRP, smaller antennas will still be compatible with 2° spacing. Another facet of Ka-band is that small spot beams can be generated onboard the satellite with achievable antenna apertures. (Practical implementations need multiple reflectors to allow feed spacing and avoid scan loss.) The design of the satellite repeater is somewhat more complex in this band because of the need for cross connection and routing of information between beams. Consequently, there is considerable interest in the use of onboard processing to provide a degree of flexibility in matching satellite resources to network demands.

The Ka-band region of the spectrum is perhaps the last to be exploited for commercial satellite communications. Research organizations in the United States, Western Europe, and Japan have spent significant sums of money on experimental satellites and network application tests.

From a technical standpoint, Ka-band has many challenges, the biggest being the much greater attenuation for a given amount of rainfall (nominally by a factor of three to four, in decibel terms, for the same availability). This can, of course, be overcome by increasing the transmitted power or receiver sensitivity (e.g., antenna diameter) to gain link margin. Some other techniques that could be applied in addition to or in place of these include (1) dynamic power control on the uplink and downlink, (2) reducing the data rate during rainfall, (3) transferring the transmission to a lower frequency such as Ku- or C-bands, and (4) using multiple-site diversity to sidestep heavy rain-cells. Consideration of Ka-band for an application will involve finding the most optimum combination of these techniques.

The popularity of broadband access to the Internet through DSL and cable modems has encouraged several organizations to consider Ka-band as an effective means to reach the individual subscriber. Ultra-small aperture terminals (USATs) capable of providing two-way high-speed data, in the range of 384 Kbps to 20 Mbps, are entirely feasible at Ka-band. Hughes Electronics filed with the FCC in 1993 for a two-satellite system called Spaceway that would support such low-cost terminals. In 1994, they extended this application to include up to an additional 15 satellites to extend the service worldwide. The timetable for Spaceway has been delayed several times since its intended introduction in 1999. Almost at the same time, several strong backers introduced another proposal called Teledesic, which would employ the same Ka band from LEO satellites—up to 840 in number (later reduced to 288, then again to 30). While this sounds amazing, strong support from Craig McCaw, founder of McCaw Cellular (now part of AT&T Wireless), and Bill Gates (cofounder of Microsoft) lent apparent credibility to Teledesic. In 2001, Teledesic purchased an interest in ICO and delayed introduction of the Ka-band LEO system. A further development occurred in 2003 when Craig McCaw bought a controlling interest in L/S-band non-GEO Globalstar system.

While the commercial segment has taken a breather on Ka-band, the same cannot be said of military users. The U.S. Navy installed a Ka-band repeater on some of their UHF Follow-On Satellites to provide a digital broadcast akin to the commercial DTH services at Ku-band. It is known as the Global Broadcast Service (GBS) and provides a broadband delivery system for video and other content to ships and land-based terminals. In 2001, the U.S. Air Force purchased three X- and Ka-band satellites from Boeing Satellite Systems. These will expand the Ka-band capacity by about three on a global basis, in time to support a growth in the quantity and quality of Ka-band military terminals. The armed services, therefore, are providing the proving grounds for extensive use of this piece of the satellite spectrum.

2.4.8   Q- and V-Bands

Frequencies above 30 GHz are still considered to be experimental in nature, and as yet no organization has seen fit to exploit this region. This is because of the yet more intense rain attenuation and even atmospheric absorption that can be experienced on space-ground paths. Q- and V-bands are also a challenge in terms of the active and passive electronics onboard the satellite and within Earth stations. Dimensions are extremely small, amplifier efficiencies are low, and everything is more expensive to build and test. For these reasons, few have ventured into the regime, which is likely to be the story for some time. Perhaps one promising application is for ISLs, also called cross links, to connect GEO and possibly non-GEO satellites to each other. To date, the only commercial application of ISLs is for the Iridium system, and these employ Ka-band.

2.4.9   Laser Communications

Optical wavelengths are useful on the ground for fiber optic systems and for limited use in line-of-sight transmission. Satellite developers have considered and experimented with lasers for ISL applications, since the size of aperture is considerably

smaller than what would be required at microwave. On the other hand, laser links are more complex to use because of the small beamwidths involved. Control of pointing is extremely critical and the laser often must be mounted on its own control platform. In 2002, the European Space Agency demonstrated a laser ISL called SILEX, which was carried by the Artemis spacecraft. The developers of this equipment achieved everything that they intended in this government-funded program.

2.4.10   Summary Comparison of the Spectrum Options

The frequency bands just reviewed have been treated differently in terms of their developmental timelines (C-band first, Ka-band last) and applications (L-band for MSS and Ku band for BSS and DTH). However, the properties of the microwave link that relate to the link budget are the same. Of course, properties of different types of atmospheric losses and other impairments may vary to a significant degree. This requires a careful review of each of the terms in the link budget prior to making any selection or attempting to implement particular applications.

References

[1]   Kadish, J. E., and T. W. R. East, *Satellite Communications Fundamentals*, Norwood, MA: Artech House, 2000.
[2]   Sklar, B., *Digital Communications—Fundamentals and Applications*, 2nd ed., Upper Saddle River, NJ: Prentice Hall, 2001.
[3]   Elbert, B. R., The Satellite Communication Ground Segment and Earth Station Handbook, Norwood, MA: Artech House, 2001.
[4]   Flock, W. L., *Propagation Effects on Satellite Systems at Frequencies Below 10 GHz, Second Edition,* NASA Reference Publication 1108(2), National Aeronautics and Space Administration, 1987.
[5]   ITU-R Recommendation S.580-5.
[6]   Elbert, B. R., *Introduction to Satellite Communication*, 2nd ed., Norwood, MA: Artech House, 1999.

Issues in Space Segment and Satellite Implementation

With the global adoption of the technology, satellite communications has brought with it a number of issues that must be addressed before an application can be implemented. Satellite capacity is only available if the right satellites are placed in service and cover the region of interest. Considering the complexity of a satellite and its supporting network, applications can be expensive to install and manage. If the issues are addressed correctly, however, the economic and functional needs of the application will be satisfied.

A viable satellite communications business is built on a solid technical foundation along the lines discussed in the previous two chapters. In addition to frequency band and bandwidths, such factors as orbit selection, satellite communications payload design, and the network topology have a direct bearing on the attractiveness of service offerings. The satellite operator must make the decision whether to launch a satellite with one frequency band or to combine payloads for multiple-frequency operation (called a hybrid satellite); whether or not to design the payload and network around an onboard digital processor is another question. As the payload becomes more unique, the demands on the market and supporting technologies increase. In addition, the business and operation should consider and properly address all of the issues that this chapter raises. We also review the current state of the art in bus design as it has a bearing on payload power and flexibility.

The remainder of this chapter goes into contingency planning from the perspective of the operator and the user. The reliability and flexibility of satellite applications cannot be assured without thorough analysis and proper implementation. For example, a satellite operator should implement a system with multiple satellites so that no single event can terminate vital service to users. Users, on the other hand, must approach satellite communications with an open mind and open eyes. They might arrange for backup transponder capacity for use in the event of some type of failure. Both parties may also need to obtain insurance to reduce financial loss. The information that follows provides background on some of the more critical areas that often hamper the introduction and smooth operation of effective systems. Readers should also consider how other potential problems not identified here could adversely impact their services and plan accordingly.

sold at a negotiated price. Each deal is different, considering the factors of price (lease or buy), backup provisions, and the term. The retail case comes into play where the satellite serves the public directly, such as in MSS and BSS networks. We consider such business issues in detail in Chapter 13.

To create the space segment, the satellite operator contracts with one of the approximately 12 spacecraft manufacturers in the world for many of the elements needed for implementation. Historically, most operators took responsibility for putting the satellite into operation, including the purchase and insurance of the launch itself. More recently, some contracts have required in-orbit delivery of the satellite, which reduces the technical demand and some of risk on the satellite purchaser. However, satellite buyers still need a competent staff to monitor the construction of the satellites and ground facilities, and to resolve interface and specification issues. This can be accomplished with consultants, the quality of which depends more on the experience of individuals than on the cost or size of the consulting organization. The experienced spacecraft consultants include Telesat Canada, The Aerospace Corporation, and SESG Global. Individuals, such as retirees from spacecraft manufacturers, can provide excellent assistance at much lower cost. However, they can be difficult to find.

The capacity demands of cable TV and DTH systems are pushing us toward operating multiple satellites in and around the same orbit position. Successful satellite TV operators like SES and PanAmSat have been doing this for some time, developing and improving the required orbit determination and control strategies. This considers accurately determining the range of the satellite, since we are talking about separating satellites by tenths of degrees instead of multiple degrees. A few of the smaller operators of domestic satellites like Telenor, Thaicom, and NHK double the capacity of an orbit slot by operating two smaller satellites rather than launching a single satellite with the larger combined payload capacity. On the other hand, employing a larger satellite with double or quadruple the number of transponders will generally significantly reduce incremental costs at some increase in risk.

Implementation of the Earth station network can follow a wide variety of paths. One approach is to purchase the network as a turn-key package from a manufacturer such as ViaSat (Carlsbad, California), Hughes Network Systems (Germantown, Maryland), Alcatel (Paris, France), or NEC (Yokohama, Japan). This gives good assurance that the network will work as a whole since a common technical architecture will probably be followed. There are systems integration specialists in the field, including L3 Communications STS, Globecomm Systems, Inc. (both of Hauppauge, New York), IDB Systems (Dallas, Texas), and ND SatCom (Friedrichshafen, Germany), which manufacture and purchase the elements from a variety of manufacturers and perform all of the installation and integration work, again on a package basis. The application developer may take on a significant portion of implementation responsibility, depending on its technical strengths and resources. Another strategy for the buyer is to form a strategic partnership with one or more suppliers, who collectively take on technical responsibility as well as some of the financial risk in exchange for a share of revenue or a guarantee of future sales. Some of the smaller and very capable satellite communications specialists, such as Shiron Satellite Communications (http://www.shiron.com) and EMS Technologies, Inc. (Norcross, Georgia), can provide a targeted solution.

The operations and maintenance phase of the application falls heavily on the service provider and in many cases the user as well. The service may be delivered and managed through a large hub or gateway Earth station. This facility should be supported by competent technical staff on a 24-hour per day, 7-day per week basis (called 24–7)—either on site or remotely from an NOC. Such a facility might be operated by the integrator or supplier and shared by several users or groups of users. This is a common practice in VSAT networks and cable TV uplinking. Inexpensive user terminals, whether receive-only or transmit and receive, are designed for unattended operation and would be controlled from the hub. The systems integrator can operate portions or the entire network, including maintenance and repair of equipment. A properly written contract or Service Level Agreement (SLA) with a competent supplier often gives functional advantages for the buyer, such as backup services and protection from technical obsolescence. Other risks to be addressed are reviewed at the end of this chapter.

The satellite communications industry keeps evolving as satellite operators discover how to enter the businesses of their users and as users experiment with becoming satellite operators. In the case of the former, Hughes Communications created the DIRECTV service to produce much more revenue than would be possible through the wholesale lease of the required Ku-band satellite capacity. On the other hand, PanAmSat was started up by the former management of the Spanish International Network, which was the leading Spanish language network in the United States. As these companies have discovered, their counterpart's business is quite different in the nature of the respective investment.

A basic issue on the space segment side is the degree to which the satellite design should be tailored to the application. Historically, C- and Ku-band satellites in the FSS are designed for maximum flexibility so that a variety of customer's needs can be met. A typical FSS transponder may support any one of the following: an analog TV channel, 4 to 10 digital TV channels, a single 60-Mbps data signal such as would come from a wideband TDMA network, or an interactive data network of 2,000 VSATs. The satellite operator may have little direct involvement in these applications. Alternatively, they may invest in these facilities to provide value-added services.

The alternative is to design the payload to meet the requirements for a specific type of signal, tailored for one application. The latest generation of BSS satellites provides high levels of RF power to deliver the signal to very small antennas, and may have less uplink than downlink coverage. These satellites are very effective for broadcasting but would be less suitable for two-way communications from VSATs. At the other end of the range are the MSS constellations, where the antennas and transponder electronics are specifically designed to receive the very low power signals coming from portable terminals and handheld phones similar to cellular phones. The downlink transmit power, on the other hand, is much higher per voice channel than is typical of an FSS design. With all of this, the satellite must provide a high degree of frequency reuse since the bandwidth available at L-band is only on the order of 30 MHz. As a consequence of specialization, systems like DIRECTV and Globalstar are inflexible when it comes to adapting to substantially different applications. On the other hand, these tailored satellites deliver the service that is intended and do it with relative efficiency.

One trend is toward more highly tailored satellites, which could accelerate as the transponder itself becomes digital in nature. Digital processing offers many benefits, such as much greater flexibility in channel routing and antenna beam control. The trade-off is that once internals of the repeater are made of silicon, the signal format may need to remain nearly constant throughout the life of the satellite. Looking ahead a decade or more, satellite technology will evolve so that the flexibility of today's fixed bandwidth analog transponder might be available in the digital repeater mode.

The ground segment is also undergoing an evolutionary process, where the first designs were versatile, multipurpose (voice, TV, and data), and expensive. Over time, the Earth station has become more specific to the purpose and lower in price. Digital implementation of analog functions like the modem now permit very sophisticated small Earth stations that can sell for about the price of a PC.

The choice of level of integration of the ground and space segment activities is a strategic decision of the satellite operator and application developer. There are some operators who have launched spacecraft with no specific thought as to how their space segment services will function with the ground segment of prospective users. In contrast, other operators have simulated their hypothetical communication network performance at every step from when the concept is defined to factory tests to in-orbit. These results have been fed back into the design of critical elements of the satellite and ground user equipment. If the integration between these is very tight, such efforts are mandatory. The choice somewhat depends on how unique the spacecraft payload and frequency band are in relation to alternative systems and services. We begin the discussion of satellite system issues with a review of some of the most important trade-offs in satellite design. The two halves of the satellite—the communications payload and the spacecraft bus—are presented in Table 3.1 and reviewed next.

3.2   Communications Payload Configurations

Communications payloads are increasing in capability and power, getting more complex as time passes. Examples include the newer DTH missions, which are designed to maintain maximum EIRP and digital throughput for up to 32 high-power transponders in the downlink. Likewise, state-of-the-art MSS satellites have large, deployable antennas to allow portable terminals and handheld phones to operate directly over the satellite path. Onboard digital processors are likewise included as a means to improve the capacity and flexibility of the network, a capability for multibeam Ka-band satellites.

Some traditional issues are still with us. One of the most basic is the selection of the frequency band, which was addressed in Chapter 2. Suppose that the operator wants to be able to address the widest range of applications and has consequently decided to implement both C- and Ku-band. With today's range of spacecraft designs, one can launch independent C- and Ku-band satellites. This was done in the first generation of the Ku-band SBS and C-band Galaxy systems in the United States. Each satellite design can be optimized for its particular service. The alternative is to build a larger spacecraft that can carry both payloads at the same time. The

Table 3.1 Subdivision of a Typical Communications Satellite into the Subsystems of the Communications Payload and Spacecraft Bus

Major Subsystem	Element	Function
Communications payload	Repeater and antenna system	Provide all communications relay and processing functions
	Microwave repeater	Basic weak signal reception of the uplink and high-power transmission to the downlink
	Onboard processor	May be analog or digital; intended to divide bandwidths and communications traffic for proper routing between the uplink and downlink
	Antenna system	Captures electromagnetic waves on the uplink side and radiates electromagnetic waves on the downlink side
Spacecraft bus	Power, control, and structure of the spacecraft vehicle	The vehicle that powers and protects the payload through all phases of the mission and operation
	Power subsystem	Provides prime power from solar arrays and energy storage using batteries to provide power during eclipse
	Telemetry and command	Allow ground operations to monitor and control the entire satellite (payload and bus)
	Attitude control and propulsion	Manage and control the spacecraft vehicle in orbit and keep it properly aligned with the Earth throughout its lifetime
	Thermal control	Maintain a suitable temperature environment for all subsystems
	Structure	Contain all of the subsystems and protect them during launch and deployment on orbit

first such hybrids were purchased in the early 1980s by INTELSAT (Intelsat V), Telesat Canada (Anik B), and Southern Pacific Railway (Spacenet). The hybrid satellites launched by INTELSAT also permit cross-band operation, with the uplink Earth station at one band (e.g., Ku-band) and the downlink at the other (C-band).

A third overall design issue deals with the coverage footprint, which has a direct bearing on the design of the spacecraft antenna system. The two basic alternatives are to either create a single footprint that covers the selected service area or to divide the coverage up into regions that each gets its own spot beam. These two approaches have significant differences in terms of capacity, operational flexibility, and technical complexity.

3.2.1 Single-Frequency-Band Payload

The single-frequency payload represents the most focused approach to satellite design. The concept was first introduced by Early Bird (Intelsat I) in 1965 and advanced in 1975 by the 24-transponder Satcom spacecraft built by RCA Astrospace (now absorbed into Lockheed Martin). Beyond 2000, single-band payloads have become targeted toward specific applications in TV and mobile communications. The TV marketplace is dominated by cable TV and DTH, where the quantity of available TV channels at the same orbit position becomes important. The

majority of the cable TV satellites for the United States, including some of the Galaxy and AMC series, are single-frequency designs, optimized to the requirements of the cable TV networks. This considers all of the technical, operational, and financial factors in providing service. From a technical prospective, the transponder gain and power is made to match the Earth stations used to uplink and receive the signals. What is more important to this class of customer is that the capacity must be there when needed. Because nearly all U.S. families receive satellite-delivered programming, the cable TV networks put a very high value on the reliability of getting the capacity to orbit and operating once it gets there.

The satellite operators that address cable markets therefore need excellent plans for launch and on-orbit backup (discussed further at the end of this chapter). In brief, experience has shown that the best way to do this is to construct a series of identical satellites and launch them according to a well-orchestrated plan. This must consider how the capacity is sold to the cable programmers as well as the strategy for replacing existing satellites that reach end of life. The single-band satellite can fit well into such a plan. Matters are more complicated when an operator wishes to replace individual C- and Ku-band satellites with a dual-frequency hybrid satellite. The benefit of doing this is reduced investment cost per transponder and simpler operation, but the common timing and orbit slot required might be incompatible.

A second area of the TV market where single-band satellites are preferred is in DTH. Spacecraft for SES-Astra, DIRECTV, and EchoStar/DISH are single-band designs tailored to the specific requirements of their respective DTH networks. This considers the quantity of transponders, the size of the receiving antenna (and therefore the satellite EIRP), the signal format (which determines the transponder bandwidth, channel capacity, and quality), and the coverage area. Taken together, these factors have enormous leverage on the economics and attractiveness of the service, second only to the programming. The delivery of the signal to millions of small dishes demands the highest EIRP that is feasible with a given state of the art. DBS satellites tend to push the limits on power, as opposed to mass (often referred to as weight, but technically mass since the satellite is in a zero-g environment). This leaves little left over for C- or L-band repeater elements, which, if included, could force a compromise of some type. (A possible exception, for smaller markets, is considered in the next section.)

The cost of a high-power DTH satellite is often more than that of, say, a C-band satellite that serves cable TV. This should not be a concern because of the much larger quantity of receiving antennas. In fact, in [1] we demonstrate that to achieve an optimum G/T for a network of 10 million receivers, the satellite EIRP must be approximately 50 dBW. By optimum, we mean that the total cost of the satellite and all of the receiving Earth stations is at a minimum. Lower EIRP demands that dishes must be larger, raising the cost of the ground segment faster than the savings in the satellite. Going in the direction of increasing satellite EIRP is likewise unattractive since the increase in satellite cost outweighs the savings in receive dishes. In other words, the high investment cost of the satellite is ultimately very economical on a cost-per-user basis. Increasing satellite EIRP allows smaller receive dishes; however, a limit at around 55 dBW and 45 cm, respectively, is imposed by the adjacent satellite interference. This happens to be the design point for most major DBS networks, reflecting the preference of consumers for compact antennas.

3.2.2 Multiple-Frequency-Band Hybrid Payloads

Hybrid satellites were first introduced by INTELSAT at C- and Ku-bands with the launch of Intelsat V. A third L-band payload was added to Intelsat V-A for use by Inmarsat. The first domestic hybrid, Anik B, was operated by Telesat Canada in the late 1970s; and two American companies—Sprint Communications and American Satellite Corporation (both since merged into GTE and the satellites subsequently sold to Americom)—were also early adopters. The idea behind the use of the hybrid was to address both the C- and Ku-band marketplaces at a reduced cost per transponder. During the 1990s, satellite operators pursued much larger spacecraft platforms like the 8-kW Lockheed Martin A2100, the Boeing 601-HP, and the Astrium Eurostar.

Even higher powers are provided by the Boeing 702 and Loral 1300S series, which reach 15 to 20 kW of prime power. Illustrations of these spacecraft are shown in Figure 3.2. This class of vehicle can support almost 100 transponders, allowing a full DBS repeater to be combined with the high end of C-band services. A criticism leveled at the 15 kW and greater design is that the operator may be putting too many eggs in one basket. However, the other side of the coin is that these designs simplify operation (only one spacecraft need be operated at the orbit position) and markedly reduce the cost per transponder.

3.2.3 Shaped Versus Spot Beam Antennas

The coverage pattern of the satellite determines the addressable market and the flexibility of extending services. The traditional and most successful approach to

SS/Loral 1300S
19 kW
6200 kg at launch

Astrium-Space
9 kW
3200 kg at launch

LM A 2100 AX
3600 kg at launch

Figure 3.2 Large-capacity GEO spacecraft.

date is the shaped area-coverage beam that serves a country or region of a hemi-sphere. This type of antenna pattern permits one signal to be delivered across the entire footprint from a bent-pipe transponder. While versatile, this approach limits the overall satellite throughput bandwidth as well as the effective spacecraft antenna gain (and hence EIRP) at the boundary. The opposite principle of frequency reuse through multiple spot beams is gaining favor for high EIRP MSS satellites like Thuraya and Inmarsat 4; in addition, systems that employ Ka-band to provide broadband Internet access likewise use the multiple spot beam approach. This section reviews the characteristics and trade-offs between these two means of serving users on the ground.

For a constant transponder output power, the EIRP varies inversely with the beam area. Stated another way, for a given spacecraft antenna configuration, the product of gain (as a ratio) and area is a constant. We can estimate the gain of any particular area of coverage using the following relationship:

$$G \cong 27,000/\phi^2 \tag{3.1}$$

where G is the gain as a ratio, and ϕ is the average diameter of a circular coverage area, measured from GEO in degrees. Measuring coverage in degrees comes about because the full Earth extends across approximately 17° as viewed from GEO, resulting in a minimum gain at beam edge of $27,000 / 17^2 = 93.4$ or 19.7 dBi. This value would be further reduced by the aperture efficiency, which depends on the design of the antenna (horn, reflector, or array). A beam of one-tenth this angular diameter would have one-hundredth the area, but the gain would increase by 100 (or 20 dB) to a total of 39.7 dBi.

Figure 3.3 provides an illustration of how the gain and area are related for two differing coverage areas: the country of Colombia and the continent of South America. The Colombian market would be served with a national beam that is directed exclusively toward this country, delivering high gain and no direct frequency reuse

Figure 3.3   Coverage options: single country (Columbia) versus entire continent (South America).

(other than through cross-polarization). From an orbit position of 65 WL, the picture of the land area shows heavy grid lines that are 1° apart. By counting grid squares, we can estimate the landmass of Colombia to cover approximately 2.8 deg^2. A national coverage antenna would of necessity reach beyond the border and is slightly larger at 4.6 deg^2. In comparison, the landmass and example antenna coverage of South America are approximately 40 and 52 deg^2, respectively. Because the area in square degrees of the South American beam is 10 times that of Colombia's, the gain over the entire continent is a full 10 dB less. One could, of course, maintain the same level of EIRP by increasing downlink transmitter power by 10 dB as well.

An alternative that is shown in Figure 3.4 subdivides the coverage area many times over using small spot beams. Assuming that each beam is 0.4° in diameter, it will take approximately 38 such spots to provide the full national coverage. As the size of the spot is only 0.126 deg^2, the gain increases substantially by approximately 15 dB compared to the area beam. The 28 spot beams are arranged in a 7-beam reuse pattern with one-seventh of the allocated spectrum assigned into each spot. Spots that reuse the same piece of spectrum are separated by two adjacent spots that are noninterfering. This need to isolate spots applies to FDMA and TDMA; CDMA offers the possibility of not subdividing the spectrum but rather allowing interference to overlap in adjacent beams. To use these beams more effectively, the satellite can have an onboard beam-routing scheme.

The general relationship among the coverage area, beam size, and number of beams is indicated in Figures 3.5 and 3.6. The first graph allows us to estimate the directivity of a beam of a given area, measured in square degrees. For the case of Colombia, which extends approximately 2.8 deg^2 as viewed from GEO, the satellite can deliver about 40-dBi gain. A doubling of the area to include, say, Venezuela will reduce gain by 3 dB to about 37 dBi. Extending further to cover all of South America in one 50-deg^2 beam pushes the gain down all the way to 27 dBi. A word of caution about the directivity numbers: these are approximate values that do not include the relevant losses in a real antenna system. Also, we have not evaluated the actual beam shaping that would be provided during the design of the antenna system. These numbers are intended to provide a general feel for the relationships.

Figure 3.4   Comparison of multiple spot beam coverage versus single country shaped beam.

Figure 3.5   An estimate of directivity in dBi versus beam area in square degrees.

The second figure plots the number of beams required to cover either Colombia or all of South America. This is the concept behind Figure 3.4, which indicates that it would take about 38 spots of 0.4° each to fully cover the country. Beam area is plotted along the *x*-axis to be consistent with the previous figure. The information tells us that it can take a very large number of beams to cover a large landmass. On the other hand, the gain of such beams is substantially higher and the potential for frequency reuse much greater when considering a high density of small spot beams. There are two application areas where the multiple-beam approach appears to be

Figure 3.6   Approximate number of beams versus area covered and beam size in square degrees.

the most appropriate: L-band MSS networks to serve handheld phones, and Ka-band FSS networks for advanced broadband communications to inexpensive personal VSATs. As discussed in Chapter 2, L-band spectrum is very limited and we must incorporate as much frequency-reuse as possible. This, coupled with the difficult requirement of serving low-power handheld phones, demands a large reflector antenna with many small spot beams. Service to mobile users would be restricted by the resulting link budget to something in the range of 50 to 150 Kbps. The bandwidth is more ample at Ka-band, so the major concern is with delivering high digital bandwidths (up to 20 Mbps) to an antenna of less than 1m.

The choice among the coverage alternatives depends on the interaction of the technical and business factors that confront the satellite operator (who may also be the application provider). From a pure marketing perspective, the single area coverage approach is the most flexible since you can deliver both individual services and broadcast services as well. The wide beamwidth produces a relatively low antenna gain, and the only frequency reuse is from cross-polarization. Moving toward multiple spot beams can greatly improve the attractiveness of the service to large quantities of simple, inexpensive Earth terminals. The ultimate example is the handheld satellite phone, which demands the greatest quantity of beams and their corresponding high gain. When we move in this direction, we restrict the range of services that can be delivered. Broadcasting of video and other content is impractical because bandwidth must be provided within each and every beam to be served. A way around this might be to use dynamic beam forming to create an area footprint for the transmission in question.

3.2.4 Analog (Bent-Pipe) Repeater Design

The repeater is that portion of the communications payload that transfers communication carriers from the uplink antenna to the downlink antenna of the spacecraft. In established C- and Ku-band satellite systems, the repeater is divided into transponders, each of which can transmit a predefined amount of bandwidth and downlink power. It is common practice to call a repeater a transponder and vice versa, although repeater is the more general term. Transponder, on the other hand, more typically refers to one RF channel of transmission, which can be assigned to one customer or group of customers for a common purpose (transmitting a multiplex of TV channels or providing a VSAT network).

In the following, we review the traditional type of transponder, called the bent pipe, along with newer concepts employing digital onboard processing (OBP). An OBP repeater may provide a more sophisticated system for routing analog channels (and hence can offer greater flexibility for bent-pipe services) or may demodulate the bit streams onboard for efficient routing, multiplexing, or additional processing. As one moves toward increasing levels of complexity, the satellite becomes more and more a part of an overall network of ground stations and is inseparable from it. This tends to increase performance and effectiveness for a specific network implementation but renders the satellite less flexible in terms of its ability to support different traffic types not considered prior to launch. The development time for an OBP repeater will generally take extra months or years as compared to the bent pipe, introducing the risk that the market for the planned application could be missed.

Each transponder of a bent-pipe repeater receives and retransmits a fixed-bandwidth segment to a common service area. There is a simple mathematical relationship between the number of transponders and the total available bandwidth that is provided by the particular spectrum band. Simply stated, the number of transponders equals the total bandwidth divided by the bandwidth per transponder. There will be 10% to 15% guard band due to filtering at the edges of each transponder. The example of a six-transponder design in Figure 3.7 has a single wideband receiver that takes the entire uplink frequency band, typically 500-MHz wide, amplifies and transfers the same 500 MHz to the corresponding downlink band. The bank of input filters, labeled F1 through F6, subdivides the total bandwidth into 72-MHz segments (11.3 MHz less than straight division would indicate), each amplified to a high level by a dedicated power amplifier. The individual outputs of six amplifiers (each on a different frequency) are summed with minimum loss in an output multiplexer composed of six reactively coupled waveguide filters. The resulting spectrum of 500 MHz (less the guardbands) is applied to the transmitting antenna system of the satellite, which typically broadcasts these signals across a common footprint.

The engineering design of the transponder channel is a high art because a multitude of specifications and manufacturing issues must be considered. Parameters in the link budget like receive G/T, transmit EIRP, transponder bandwidth, and intermodulation distortion have a direct impact on users. These should be specified for every application. A multitude of others, like gain flatness, delay distortion, spurious and phase noise, and AM-to-PM conversion, are often of less concern to some applications but potentially vital to others. Wideband digital transmission at 155 Mbps in a 54-MHz transponder is an exception because these distortions can significantly reduce throughput or increase the EIRP requirement for the same throughput. The driver/limiter/amplifier (DLA) in Figure 3.7 provides a degree of control over data transfer by adjusting the input power and possibly correcting some of the nonlinear distortion.

Legend:
F1   Bandpass filter for channel 1
DLA   Driver/limiter/amplifier
A   High-power amplifier
LPF   Lowpass filter

Figure 3.7   A simple bent-pipe satellite repeater with six transponders.

Typical transponder characteristics for the bent-pipe design are listed in Table 3.2. Actual values will vary from design to design, in response to the type of amplifier, the frequency of operation, and design choices for the intended service. Some examples of how repeater parameters can be related to particular signal types are shown in Table 3.3. To do this properly, the designer must fully understand the signals being transferred and the distortions to those signals caused by the various elements of the transponder. With the advent of 1-GHz PCs and signal analysis software, this optimization can be performed in minutes. However, the issue remains about whether it is wise to design the transponder for a particular signal and corresponding application. A useful alternative is to utilize a compromise design to accommodate a variety of expected signal types. This is actually how the first 36-MHz transponders were designed for INTELSAT IV, which was based on transferring one analog FM TV carrier or a multiplexed telephone baseband containing up to 1,600 voice channels. Today, the 36-MHz transponder is the standard for bent-pipe satellites that serve analog and digital applications.

If the signal format does not change during the lifetime of the satellite, the transponder is eligible for optimization. Consider first if you will operate the transponder with only a single carrier over a wide bandwidth or if you will carry multiple carriers in the same transponder. The DTH transponder has characteristics that are determined first by the frequency assignments filed for (see Chapter 7) and second by the type of signal modulation (analog or digital). Advanced repeaters that use digital signal processing offer a great deal of flexibility in routing traffic and permitting inexpensive user terminals to gain access to a range of mobile and fixed services; however, they may not have the greater than 30-dB range level flexibility (dynamic range) of the standard bent-pipe transponder. Alternatively, the processor function can be implemented with analog components that nevertheless have a degree of flexibility. The MSAT satellites operated by AMSC and Telesat Mobile contain

Table 3.2 Typical Transponder Characteristics in a Bent-Pipe Repeater

Characteristic	Typical Value
Gain (saturation flux density)	−96.0 dBW/m^2
Linearity for multiple carriers (C/3IM) at saturation	−10 dB with respect to saturation by two equal carriers
Linearity for multiple carriers (C/3IM) with backoff	−20 dB at 8-dB IBO, (with linearizer on)
Noise power ratio (NPR)	16 dB at 4-dB output backoff
Nonlinear phase shift (AM-to-PM conversion)	<40° from saturation to −20-dB input backoff
Amplitude frequency response (gain flatness)	±−0.25 dB over useful bandwidth
Out-of-band attenuation (input channel separation)	−30 dB in adjacent channel
Cross-polarization isolation (XPOL)	30 dB (linear)
Frequency tolerance and stability	10^{-6} in the translation frequency
Gain (attenuation) control	0 to 18 dB in 2-dB steps
Gain stability	±1.5 dB over lifetime

Table 3.3 Specific Signal Types and the Transponder Characteristics That Can Be Optimized for Improved Performance

FM video	TWT AM/AM and AM/PM; filter group delay characteristics
Wideband digital data	Gain slope, group delay; TWT AM/AM and AM/PM
SCPC	Gain flatness and amplitude nonlinearity
Wideband TDMA	Linearity, gain flatness, and transient response of the power supply
VSAT operations	Gain flatness/linearity and frequency stability
Mobile SCPC	Noise power ratio

surface acoustic wave (SAW) filters that, when combined with commandable down/upconverters and switching, allow the ground network operator to alter the bandwidths and routing of SCPC bandwidth segments. This controls the balance of traffic and permits the operator to isolate sources of interference. Processors of this design have been offered to the market by ComDev of Canada, and one is carried on Anik F2.

The selection of the bent-pipe transponder has implications for users of satellite capacity. Bent pipes are nearly transparent to the user and can be subdivided in power and bandwidth, as discussed previously in this section. Moving toward the more sophisticated designs, the satellite becomes integrated with the network and transparency is lost. One of the first specialized commercial repeaters was carried aboard the Spacenet IIIR satellite. This was the Geostar payload, which introduced an L-band vehicular position determination service in the United States. Geostar contracted with the spacecraft manufacturer to have additional antennas and receivers installed on the satellite and paid Spacenet for the operation and use of their payload. Later, Geostar failed as a business. The special transponder could not be reassigned to some other revenue producing application although it was kept operating. The spectrum for this application was subsequently reallocated by the ITU.

3.2.5   Digital Onboard Processing Repeater

The digital OBP repeater is a significant advancement from the analog versions that merely interconnect frequency channels using microwave filters and mechanical switches. At the core of OBP is digital signal processing (DSP), a computational process reduced to solid-state electronics that converts an information signal from one form into another unique form. Historically, the DSP was programmed on a multipurpose digital computer as a way to save the time and energy of doing the transform mathematically with integral calculus. The most well-known DSP process is the fast Fourier transform (FFT), which is related to both the Fourier transform and Fourier series taught to all electrical engineering students. It takes a signal in the time domain (i.e., a waveform) and converts it into a collection of frequencies (i.e., a frequency spectrum). The inverse FFT does just the opposite, transforming a frequency spectrum into a time waveform.

When in either digital format, we can multiply, filter, and modulate the signals to produce a variety of alternate signal types. In this manner, a digital processor can perform the same functions in software that would have to be done with physical hardware elements like mixers, filters, and modulators. Modern DSP chips and

systems can operate over many megahertz of bandwidth, which is what we need to build an effective digital repeater. To do this, the calculation speed must be in the gigahertz range. More recently, OBP has taken on many other roles where the actual bits on the RF carrier are recovered and reconstructed with minimum error, switched and routed, and remodulated onto other RF carriers in the downlink. This permits the OBP to act as a conventional packet switch and multiplexer, common to what is employed in land-based data communications networks. The specific configuration of the OBP repeater is created for the expected network environment, including the specific telecommunications applications to be provided to end users.

3.2.5.1 Generic Processing Repeater Architecture

A block diagram of a hypothetical digital processing repeater is shown in Figure 3.8. The antenna and wideband receivers perform their traditional analog functions, while filtering and switching occur in the digitized sections of the repeater. This is indicated within the box at the center of the figure. The majority of OBP repeaters are used to transfer traffic between multiple beams on the uplink and downlink, as would be the case in an MSS L-band or broadband Ka-band satellite. Each uplink beam is first low noise amplified and then down-converted to an intermediate frequency (IF) that is suitable for input to the digital onboard processor.

The first function at the input of the processor is to convert the incoming frequency spectrum into a digital data stream. This is accomplished by the analog-to-digital (A/D) converter using pulse code modulation (PCM). For an IF bandwidth of 50 MHz, the A/D converter must sample at a speed greater than 100 MHz and convert each sample thus taken into a specified number of bits. The number of bits, in turn, is determined by such factors as the acceptable signal-to-noise ratio (inversely,
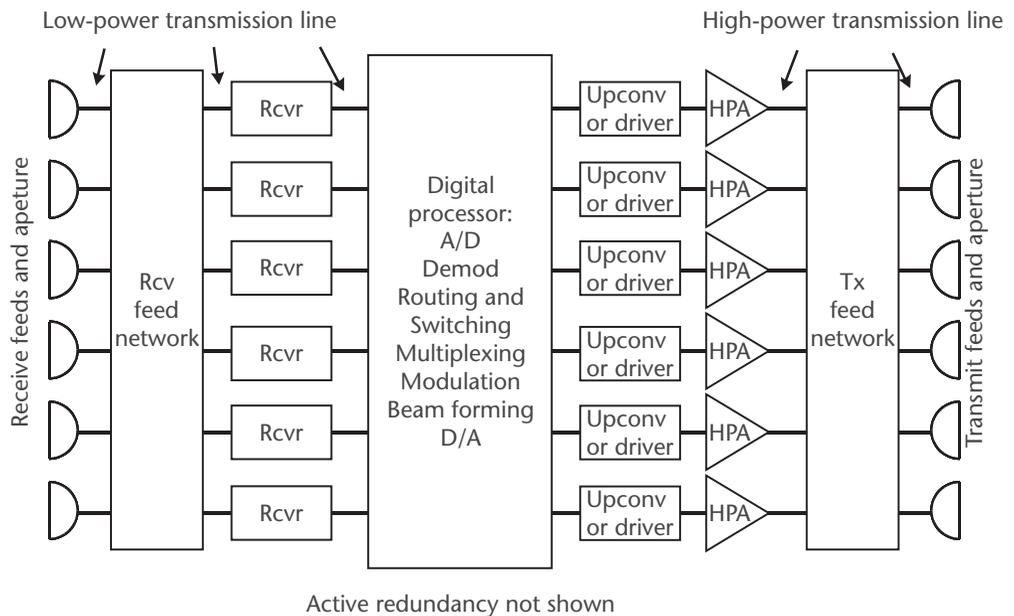

Figure 3.8 Block diagram of a generic digital processing repeater with multiple spot beams, demodulation/remodulation, packet switching, and low-level beam forming.

quantization error) and the dynamic range of input signals. The selection of the number of quantization levels and hence the number of bits per sample determines the amount of degradation to signal quality attributable to the processor. For example, if we assume 150 million samples per second and 10 bits per sample, then the A/D converter must output data at the speed of 1.5 Gbps. Since this is the data per beam, the total data processing capability of the A/D function is 1.5 Gbps times the number of beams. This does not include the processing associated with either the inverse digital-to-analog (D/A) conversion process, nor any of the processing done within the OBP itself. We can see that the processing power of a broadband repeater can be high indeed.

The digitized channels can be routed either as narrow frequency bands or packets. To route frequency bands, the OBP needs to select specific channels, cross connect them to the associated downlink, and reassign the frequencies so as to create a contiguous band. The process for packets requires the additional step of demodulation and potentially forward error correction to deliver the appropriate bit streams to packet switching elements. Other functions that are possible include automatic gain control, phase adjustment (as part of a phased array antenna system), channel multiplexing, linearization, and interference cancellation. After all of the processing is complete, the data is reconverted back into its analog form (i.e., D/A conversion).

While the functions are shown as discrete components, they are actually performed mathematically by the processor and memory chips that implement the desired functionality. This means that the processor is designed for a specific purpose. To convert the processor analog output to the transmit band, each channel is fed to a hardware upconverter that translates from the intermediate frequency range to the RF downlink band. From this point, the signals are amplified in a conventional power amplifier and applied to the appropriate transmit antenna of the satellite.

The onboard digital processor is controlled from the ground to set up the routing instructions and other aspects appropriate to the network. As development continues, it will be possible to transfer the entire uplink/downlink bandwidth through each port on the processor. This only requires greater processor speed, something that one expects to see as the technology is improved over time.

The digital processor repeater was applied commercially in Iridium, which went into service in 1999–2000 incorporating routing, demod-remod, and packet switching functions. Subsequently, ACeS was launched into GEO MSS service with the ability to route narrowband frequency channels between remote handheld terminals and the gateway; likewise, the Thuraya GEO MSS satellite added functionality for dynamic beam forming and direct handheld-to-handheld connectivity. OBP-based repeaters have also been developed for Ka-band satellites that could reach orbit in the mid-2000s. High capacity and sophistication for the processing function translate into the size, mass, and power consumption of the processor itself. In comparison to the typical microprocessor found in a personal computer, the digital repeater processor needs substantially more capacity and tailored functionality. To accomplish this, the OBP employs an architecture that is typically composed of general purpose computer processors, programmable gate arrays, DSP chips, and specialized very large scale integrated (VLSI) circuits and application-specific integrated circuits (AISCs). These tend to run at a much higher speed and therefore

consume more power. Table 3.4 suggests the technical issues and trade-offs required in the design, manufacture and test of the modern OBP repeater [2].

Another consideration is the degree of redundancy that needs to be included. Certain common functions, like clocks, memory, and power supplies, can be made redundant. But the actual channel processing elements would normally be single string. Redundancy must then be provided by including extra strings such that excess capacity may be reallocated in case of a partial failure. The OBP must also be adaptable to ground control and management, a function associated with all land-based digital networks. This must be extended to the payload from the conventional NOC used by telecommunications operators, which in this case, are charged with providing an end-to-end service that meets commercial *quality of service* (QoS) objectives. This will involve expanding the TT&C link for network management and optimization functions and extending to the network provider.

3.2.5.2   Classification of Processing Repeater Designs

Digital onboard processing repeaters are individually designed for a specific mission and therefore are not interchangeable unless produced from the exact same design. We can try to put them into classifications regarding the manner in which the uplink signals are transferred to the downlink, the types of signal processing on board, and whether the carriers are demodulated to recover the bit streams.

A classification matrix of OBP repeaters is presented in Table 3.5, based on missions that have been defined and in some cases launched as of this writing. Within each, there are a wide variety of alternatives, and, in fact, some from one category intersect with others. In the limit, one can imagine an OBP repeater with the

Table 3.4   Considerations in the Design and Production of a Functional Onboard Processing Repeater

- Design and performance considerations
  - Analog to digital quantizing (bits per sample)
  - Number of A/D and D/A operations
  - Fast Fourier transform size
  - Number of points
  - Amplitude granularity
  - Sampling rate
  - Frame overlapping factor
  - Time window
  - Ripple
  - In-band interference
- Physical and electrical considerations
  - Proper interfacing of analog IF circuitry to the A/D function of the processor
  - Design and specification of ASICs and multichip modules (MCMs)
  - Control signals: parallel and serial buses, precise timing
  - Power distribution: voltage and regulation
  - Reliability: 20-year lifetime, radiation shielding, thermal, redundancy, monitoring
  - Gain and phase matching of IF upconversion and downconversion
  - Testing: complex processing and connection, different techniques at different stages
  - Packaging: consumption within allowable budget, integration of thermal control heat pipes, location of mounting hardware

Table 3.5   General Classification of Onboard Digital Processing Repeaters

	RF Switching	IF Routing	Beam Forming	Demod/remod	Packet Routing (with demod/remod)
FDMA		ACeS, ICO	Thuraya	Skyplex mulitplexer (HOT BIRD 5)	Iridium
TDMA	Intelsat 6		Inmarsat 4		Iridium
CDMA		Voice-span		Military	

intelligence and speed to be able to detect any type of uplink signal, recover the bits, and dynamically transfer the resulting data to the most appropriate downlink channel. This is, after all, the role of routers on the Internet and it is possible that a similar type of device could ultimately find its way on board a satellite.

Satellites that provide frequency reuse through multiple spot beams are candidates for OBP because any analog approach is inherently inefficient and inflexible (a possible exception is the Beamlink FDMA routing repeater by Com-Dev). The processor on board Intelsat 6 performed a basic time-division switching function on the full 250-MHz bandwidth of the uplink. TDMA is used on the uplink side to separate carrier bursts according to the desired downlink beam. Switching is done at RF using PIN diode switches, which chop the time frame according to a prestored definition of traffic flow. See our previous work for a more complete description of this somewhat basic approach, which was installed on the ACTS satellite as well [1]. This approach, while efficient in terms of channel capacity (because the downlink amplifier is operated at saturation), had the disadvantage that Earth stations transmit at 100 Mbps or greater. Thus, it was intended for large Earth stations, which were rather expensive, and the entire strategy of RF switching has largely been retired by commercial industry in favor of long-haul fiber.

The U.S. military introduced CDMA onboard a satellite during the 1990s to provide secure and antijam communications. The OBP was one of the first to include A/D conversion as part of the repeater. Subsequently, AT&T Bell Laboratories proposed to combine CDMA with demod/remod as part of the now-defunct Voice-span Ka-band project. The complexity and cost of this approach was beyond what commercial industry could produce at the time in 1997. At the same time, Hughes Space and Communications adapted their military processor experience to the MSS market with the IF routing repeater design. Lacking the complexity and features of CDMA and demod/remod, this approach permits low-power (and therefore low-cost) user terminals to transmit narrowband information to the OBP wherein channels are selected and routed at IF to the appropriate downlink. This is reviewed further in Section 3.2.5.4.

A very useful step that was introduced by Boeing Satellite Systems was digital beam forming on Thuraya. This takes a fixed feed array and through appropriate phase and amplitude adjustment, permits the satellite operator to shape beams to meet traffic requirements after placement into service. Inmarsat selected EADS to produce a similar OBP for their fourth generation satellite. An early demod-remod repeater was put on the HOT BIRD 5 satellite of Eutelsat, thus offering format conversion in space. The concept is that individual uplinks can transmit one video channel per carrier (e.g., SCPC) and the OBP demodulates and multiplexes as many as

six into one TDM stream. Motorola empoyled demod/remod in Iridium by putting a packet switch inside the repeater.

3.2.5.3   Properties of Demodulation/Remodulation OBP

As shown in Figure 3.9, a demod/remod repeater looks nearly identical to the bent pipe, but with a demodulator and modulator added to each channel. The minimum function of this combination is to prevent the direct addition of uplink noise to the downlink noise. Instead, uplink RF noise is transferred to the baseband of the signal where it causes a specific amount of impairment such as increased error rate. The uplink will threshold at a point determined by the demodulator on board the satellite, while the downlink will threshold at a point determined by the demodulator in the receiving Earth station. The only impairment is the additional errors caused by uplink noise, which in many cases is substantially fewer than result in the downlink. Another benefit is that the downlink EIRP will be stable because the carrier that is applied to it is generated in the satellite modulator and not the uplinking Earth station. This same effect is produced by a limiter on the input side of the TWT; however, a limiter is highly nonlinear and cannot be used with multiple carriers.

Some missions might suffice with demod/remod capability alone. For example, we could build a very effective satellite that broadcasts data to millions of receivers where the uplinks come from a variety of locations and sources. The OBP provides the integration of data and proper formatting for distribution. The only variation in downlink received power will be that caused by fading along the path between the satellite and the receiving Earth station. Uplink RF noise will introduce errors in the satellite demodulator, which will be transferred directly to the downlink. For example, if the uplink produces an error rate of 10^{-7} and the downlink produces an error rate of 10^{-6}, then the combined error rate is 1.1×10^{-6}. This condition might correspond to the uplink C/N being only 1 dB greater than the downlink. Figure 3.10 provides an example of how this compares to the bent-pipe repeater, offering up to a 3-dB improvement. The increase in error rate of 10% is almost immeasurable in the recovered data. In comparison, without demod/remod, a 1-dB difference would reduce the total C/N by 2 dB. The error rate for the received data would now be two orders of magnitude less than the downlink by itself, or 10^{-4}.

Once we have recovered the original data in the satellite, it is likely that we would want to do some additional digital processing and switching. Figure 3.8 shows a baseband switch inside the repeater, similar in function to a digital

Figure 3.9   Features of demod/remod repeater; bit errors are transferred from uplink to downlink, not noise.

Figure 3.10   Demod/remod improvement over a bent-pipe repeater.

telephone exchange that switches 64-Kbps channels or a router of the type used by an ISP. With the former, we might set up and connect telephone calls using the switch in the sky. Telephone and switched data can be offered directly to the public, allowing users to pay only for the services that they use. In contrast, bent-pipe transponder capacity cannot easily be sold on a per-minute basis because of the difficulty of managing access. The model of the Internet router allows the satellite to provide a hub in the sky that more or less transparently transfers IP data.

3.2.5.4   FDMA IF Channel Routing Repeater

The IF channel repeater is one OBP approach that has gained some prominence in the MSS field. Perhaps it will be adapted to FSS once adequate processing speed is achieved. For the moment, programs like Thuraya, ICO, ACeS, and Inmarsat 4 make clear that the system works and delivers an effective service comparable to what one can obtain from GSM and GPRS networks on the ground. The basic concept is illustrated in Figure 3.11, which shows the function elements of a multiple-beam L- or S-band satellite. The uplink for each beam (on the upper left side) has 30 MHz of available bandwidth in the MSS allocations, which is shared by a group of individual channels coming from different user terminals. Once digitized, the OBP

selects channels by the process of filtering (as would be done in an analog repeater with physical bandpass filters) using the transfer function of the following type [3]:

$$y\left(mT_y + dDT_x\right) = \frac{W_1^{\left(mT_y + dDT_x\right)f_0}}{M} \sum_{k=0}^{M-1} W_M^{-mk} \cdot X_d\left(k\right) \cdot F\left(k\right)$$

The term on the left side of the equation is the time-domain representation of the selected frequency channel after translation on the downlink channel. On the right, the summation represents the range of the uplink channels that are selected using filter funtions $F(k)$. The ratio in front of the summation takes care of time domain conversion.

Referring back to Figure 3.11, the mixer blocks and vertical arrows represent the frequency translation and cross-connection to the appropriate downlink path on the right side of the OBP. Dynamic beam forming is provided at this point in Thuraya and Inmarsat IV. This is suggested in Figure 3.12 where another digital processing function is introduced after the downlink carriers are compiled [2]. The formulation for doing this is bascially as follows:

$$z\left(mT_y + dDT_x\right) = \frac{1}{M} \sum_{b=0}^{M-1} W_M^{-mb} \cdot \sum_{c=1}^{C} W_1^{dH_c\, D/N} \cdot X_{dc}\left(b + H_c\right) \cdot F_c\left(b + H_c\right)$$

The summation on the far right contains the downlink frequencies, which are basically equal in power and not adjusted for the phased array. The terms X_{dc} provide the desired amplitude adjustment while phase adjustment is indicated in the

Digital onboard processor

Figure 3.11 Block diagram of an FDMA channel routing repeater. In the uplink for each beam (on the upper left side), the available bandwidth is shared by a group of individual channels coming from different user terminals. Once digitized, the OBP selects particular channels, cross-connects and translates them in frequency, and delivers them to the appropriate downlink path on the right side of the OBP. Once converted back to the analog domain, the new aggregate of channels is transmitted in the downlink on the upper right.

Figure 3.12   MSS OBP payload with dynamic beam forming.

exponent W_r. Once converted back to the analog domain by the summation on the immediate right of the equal sign, the new aggregate of channels is transmitted in the downlink on the upper right of Figure 3.11.

3.2.5.5   Demod/Remod and Bit Stream Combining in the Skyplex Multiplexer

The Eutelsat HOT BIRD 5 satellite was one of the first to carry a commercial OBP. Rather than being for narrowband information channels, Skyplex introduced a technique for combining high-speed megabit per second digital channels into an even higher information rate. Designed specifically for use in video distribution service (the mission of HOT BIRD), Skyplex allows multiple uplink sites to individually originate a video channel. The multiplexing of this together into a multichannel MPEG stream is performed in orbit by Skyplex, rather than within a broadcast center Earth station (discussed in Chapter 4). Its functionality is illustrated in Figure 3.11. Additional background on HOT BIRD can be found in Section 6.8.4.

For Skyplex to perform is orbital multiplexing role, it was necessary for Eutelsat and the spacecraft contractor, Alcatel, to alter the standard DVB-S processing system. As illustrated in Figure 5.13, the standard DVB-S scheme employs concatenated forward error correction to produce a very low error rate at the home

Figure 3.13   Demod/remod and bit stream combining as part of the Skyplex multiplexer. Each uplink channel (1 through 6) originates from a different Earth station, using FDMA in this example. However, the channels may alternatively be transmitted by TDMA as the Skyplex OBP includes the necessary burst demodulator.

receiver. A multiplexer on board the satellite would have required the full receive side of this coding scheme in order to reproduce the necessary MPEG 2 stream. The availability of a high power uplink meant that one could rely on just the outer (Reed-Solomon) code and incorporate the inner (convolutional) code on the downlink. This arrangement is illustrated in Figure 3.14.

3.2.6 Repeater Power and Bandwidth

As satellite applications target more toward end users, the demand increases for smaller ground antennas and, as a consequence, higher satellite power. Satellite operators tend to seek a marketing advantage by having greater EIRP in the newest generation of spacecraft. A key parameter for spacecraft design in this environment is the efficiency of conversion from dc (supplied by the solar panels and batteries) to RF (power amplifier output). An overall comparison of the two basic types of power amplifiers is provided in Table 3.6. Traveling-wave tubes tend to have the highest efficiency and are appropriate for broadcasting and digital information distribution. TWTs above 250W have challenged developers because of a lack of adequate on-orbit experience. In comparison, 100W to 200W amplifiers are viewed as dependable, and experience with the generation launched in the 1990s has been very good. Higher power levels are obtained by paralleling pairs of amplifiers. The direction that manufacturers are going now is to integrate a standard TWT with a driver/linearizer that increases gain and cancels a significant amount of nonlinearity. This reduces intermodulation distortion for multiple carriers and/or sideband regrowth for wideband digital signals

Solid-state power amplifiers have become popular for power up to about 50W and may offer longer life because they do not contain a clear-cut wearout mechanism. High-power GaAs FET devices are delicate and must be maintained at a relatively cool temperature over life. SSPAs operate at low voltage and high current and can fail randomly due to design or manufacturing defects (particularly where leads

Figure 3.14   Skyplex repartition of DVB channel to accommodate multiplexing of six individual MPEG video channels onboard a satellite.

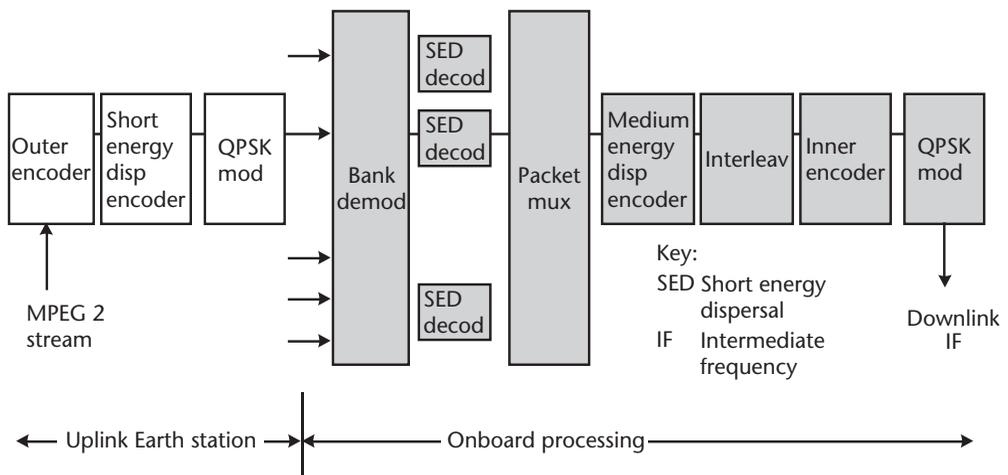are bonded to substrates). A given SSPA uses many FETs arranged in parallel combinations, requiring that all FETs be functioning in order to provide the rated power output. Examples of FET types found in SSPAs include Gallium Arsenide (GaAs), High Electron Mobility Transistors (HEMT), and Indium Phosphide (InP). This same sequence indicates step increases in performance in terms of electron mobility, efficiency, and power output.

TWTAs have maintained their lead over SSPAs because they function as a generator that can be inherently very efficient because energy of the moving element (the electron beam) can be conserved by recycling (via multiple collectors). The state of the art for space TWTs is now four collectors in a direct-radiating case, giving dc-to-RF efficiencies over 70% at Ku-band. TWTs demonstrate a 20-year lifetime through extensive on-orbit experience, even though they have a well-known wearout mechanism in the cathode. The last item in Table 3.6 indicates the maximum operating temperature of the amplifier, which is indicated at 80°C. This corresponds to the case where the TWT transfers most of its dissipated heat through the baseplate. An alternate configuration uses a finned arrangement on the collector, allowing the hot end to radiate directly into space. In this case, at least part of the TWT could run above 100°C.

The high-voltage power supplies of TWTs continue to be a source of concern due to their complexity and potential for high-voltage failure. High-efficiency SSPAs are essentially fast switches with output transformation networks tuned to specific frequency channels to maximize power output and efficiency, resulting in a bandwidth that is substantially less than what can routinely be obtained from TWTs. These differences complicate the choice of the type of amplifier, which cannot be made until the full requirements of the mission are understood and documented.

Table 3.6 also contains a number of secondary parameters that may or may not be significant in a particular application. Power level, bandwidth, gain, and efficiency have direct consequences for the service and spacecraft design. The mass of the amplifier, including its power supply, is important to the size and cost of the

Table 3.6   Comparison of Spacecraft Power Amplifiers Using Typical Values, Without Linearization

Characteristic	TWTA	SSPA
Frequency bands	L through Ka+	L through Ka+
Power output	10W to 200W	5W to 50W (1W at Ka)
Bandwidth	20%	5%
Gain	40 to 60 dB	6 to 20 dB (input driver amplifier can be added)
Efficiency	40% to 65%	25% to 40%
Mass	1.5 to 5.5 kg	0.5 to 2 kg
Linearity:	—	—
C/IM at maximum power	12 dB	16 dB
AM-to-PM conversion	5°/dB	2.5°/dB
Maximum phase shift	40°	20°
Conduction-cooled maximum temperature	80°C	30°C

satellite itself. In some missions, the quantity of amplifiers may indeed be limited by mass. Another consideration is the mass of the thermal control elements needed to remove heat and thereby control amplifier temperature (particularly important for SSPAs and high-power TWTAs). SSPAs and low-power TWTAs may be conduction cooled through mounting surfaces using heat pipe technology. As the power of the tube increases, direct radiation becomes an option.

Linearity specifications are very important in high-speed digital transmission and multiple-carrier applications for VSATs and mobile communications. While this tended to favor the SSPA, current TWTAs can come close through the addition of a linearizer.

With all of the effort that goes into designing and building a powerful and efficient amplifier, there exists the vital importance of minimizing the RF loss between antenna and the amplifier. An output loss of 2 dB results in a diversion of 58% of the amplifier power into resistive heating. This has a direct bearing on the satellite as a whole, because an efficiency improvement means that less total dc power needs to be provided by the bus. Waste power also turns into heat that must be removed from the spacecraft. Another factor due to high power is multiplication breakdown in the output waveguide and antenna feed. A lot of attention has been focused on minimizing this loss and thus maximizing the radiated power. Techniques like shaped reflectors with single-feed horns were introduced with this in mind.

Transponder bandwidth is next in importance to users. For a given application, the signal bandwidth determines the minimum transponder bandwidth required; anything more than that usually cannot be used effectively. Analog video applications generally require bandwidths in the range of 23 to 72 MHz, as summarized in Table 3.7. A VSAT network or other SCPC application could require less bandwidth and hence can share a transponder with other users. The most common approach in spacecraft design is first to determine the minimum acceptable bandwidth for any service that can be anticipated. The designer divides the available bandwidth (typically 2×500 MHz) by the service channel bandwidth to set a number of transponders to be carried (usually in the range of 16 to 48). Considering the required EIRP, coverage, sparing, life, and if it fits an existing spacecraft bus, then one is done and can start working on a detailed spec. If not, one iterates by increasing the transponder bandwidth and looking harder at the available buses.

A more detailed kind of optimization has been performed in recent years using advanced computer analysis and simulation techniques. Using an engineering workstation and sophisticated signal analysis software like the Signal Processing

Table 3.7   Typical Bandwidths Found in C- and Ku-Band Bent-Pipe Repeaters

Frequency Band (Downlink)	ITU-Designated Service	Bandwidth	Derived from
C-band (3.7–4.2 GHz)	FSS, shared with terrestrial fixed service	36 MHz	1 video carrier
		72 MHz	Dual carrier or SCPC
Ku-band (11.7–12.2 GHz)	FSS, not shared with terrestrial fixed service	27 MHz	1 video carrier
		36 MHz	1 video carrier
		54 MHz	Dual carrier or SCPC
Ku-band (12.2–12.7 GHz)	BSS, not shared	27 Mhz	1 video carrier

Workbench (SPW) or PC software like SystemView by Elanix or MathCAD, the satellite engineer can determine which transponder parameters have a significant impact on the overall link and spacecraft design. There are two types of analyses, namely, the analog/digital approach, which looks at the signal as it passes through the linear and nonlinear elements of the repeater; and the discrete time model, which generates simulated traffic (telephone calls or data messages, as appropriate). Either or both might be used for a given system. The critical step in using computer analysis is first to calibrate the model on a real-world system such as an operating satellite link or a laboratory table-top model. Once calibrated, the computer model is useful for testing various equipment arrangements and technical specifications. The approach is very powerful because you can theoretically include everything that can possibly impact performance and thereby adjust relevant characteristics. During the construction of the system or even after it is in operation, the application engineer can reanalyze the links to pull out more capability or troubleshoot problems that arise over time.

3.2.7   Additional Payload Issues

This section discusses a number of additional issues in the design or implementation of the communication payload. While not exhaustive, we provide some ideas and food for thought for specifying a spacecraft to be constructed. These factors may also be considered by users who are in the market to purchase transponder capacity on a long-term basis, or if the spacecraft is still under construction by the manufacturer.

Satellites operating at higher frequencies like Ku- and Ka-band might be fitted with one or more transmitting beacons for reception by communication Earth stations. This provides a reference for determining the amount of rain attenuation being experienced on the link. Another use is as an independent control channel for onboard communication functions such as the digital repeater discussed earlier in this chapter. The command link from the TT&C Earth station must function at all times, which means that the command receiver must be permanently on and physically connected to appropriate antennas. No switches or other interaction with the communication part of the repeater should be allowed. Command encryption might have to be considered for very secure operation, but this also should not interfere with safe operation in the case of an emergency.

Generally, the uplink coverage footprint should be as nearly identical to the downlink as possible. This allows transmitting Earth stations to be located anywhere in the entire area of coverage. However, there are systems like DTH and MSS with only a few ground transmitters (at the broadcast center or gateway) in the fixed uplink part of the spectrum, so consideration may be given to restricting the uplink coverage area. This provides an improvement in spacecraft G/T and SFD, which in turn can improve link quality and availability. Alternatively, smaller uplink antennas can be used, which is a consideration at Ka-band where large antennas are expensive and more complex to operate. Another uplink issue is the appropriate use of uplink power control (UPC) to maintain carrier power at the satellite during heavy rain. UPC has proven effective in Ku-band VSAT hubs and Ka-band video uplinks, where an entire network is dependent on the reception of a strong and

stable broadcast channel. There are concerns for the accuracy and responsiveness of the UPC control loop because any error will translate into a potential for network instability and loss of service. To this end, the UPC should be thoroughly tested prior to operation and maintained in proper working order for as long as it must function. Level control in individual user terminals is also an option, but here it is the service to the single user that is the consideration. This is particularly important in MSS, where mobile terminals can experience deep fading due to multipath and terrain blockage. The reaction time of the UPC will impact service performance and, on an aggregate basis, network capacity for CDMA in particular.

There can be a high degree of design interaction between the antenna and repeater. Top-level design requirements must be allocated to subsystems according to issues of performance, cost, mass, and risk. Components, subsystems, and the entire payload will need to be tested and retested on the ground and in space. A competent analysis and budgeting scheme should define the interfaces and account for the uncertainties on both sides. The approach for sparing the amplifiers and other critical devices needs careful consideration as well, in addition to the overall requirement to isolate and remove any potential single-point-failure mode in the spacecraft or ground segment.

There are other techniques for addressing the greater rain attenuation at Ka-band and higher. A simple type of radome over a ground antenna will greatly reduce the attenuation due to water on the reflector and feed. This could amount to nothing more than a roof that extends above and forward of the reflector, but not within the collimated beam. Site diversity, where two locations are provided to assure that one or the other has a working link to the satellite, provides substantial improvement in service availability. The cost of providing a second site and the interconnecting link can be mitigated if multiple locations are required for other reasons. For example, a major television network in the United States with ground facilities on both coasts of the country can transfer the video feed by fiber to the uplink that is not experiencing rain fade. It will likewise provide redundancy to counter sun outage, equipment failure, and any kind of local disruption or disaster.

TT&C requirements must be considered at the same time as the communication payload is designed and optimized. Most satellites include TT&C frequencies at the upper or lower edge of the communication band. However, this may not be appropriate if existing ground tracking stations cannot employ the same frequencies. In this case, another band, such as C or Ku, may have to be employed and the requisite equipment included on the satellite. The correct number of telemetry points and commands needs to be determined and compared to what the system can provide. In commercial satellites, telemetry is needed to assure reliable operation and to permit troubleshooting in the event of a problem. While it is not usually needed for deep engineering study, adequate telemetry is nevertheless vital to the long-life operation of the entire network.

3.3   Spacecraft Bus Considerations

The spacecraft vehicle, more particularly the bus, must fulfill the requirements of the communication mission, typically lasting 15 years in the case of GEO. Hughes

pioneered the spin stabilization technique for GEO, while European and other U.S. manufacturers pursued the three-axis early on, gaining experience with the nonspinner. Spinners are simpler than three-axis; hence, they tend to be trouble-free. However, three-axis can deliver more power at less overall launch mass and, hence, are preferred for high-power missions (like DTH and MSS). Needing to remain competitive, Hughes adopted the three-axis design and continues to manufacture them as Boeing Satellite Systems. Large classes of three-axis made by Alcatel, Astrium, Boeing, Lockheed Martin, and Loral have the capability of providing 15 kW or more for commercial and government missions. Orbital Sciences, Alenia, and IAI produce small to medium power three-axis satellites, challenging the low-power spinner.

3.3.1   Three-Axis Bus Stability and Control

While three-axis satellites are technically and operationally more complex than earlier spinners, they nevertheless are capable of providing reliable service. The key is to provide sufficient redundancy and protection from failures that take the satellite out of service. Regardless of spacecraft configuration, operations personnel and users have to contend with on-orbit problems that result from flaws in the design, manufacture, or operational procedures.

Control of a three-axis spacecraft relies on the low spin momentum of the vehicle and the complexity that comes from the need to sense and correct for attitude error (which results in mispointing of the footprint or spot beams). There are fundamentally two means for maintaining an inertial reference and degree of control. The first is the "zero-momentum" system that uses three reaction wheels, each along a different orthogonal axis, to allow correction in any of the three directions. Each wheel can spin in either direction, thus affording forward and reverse control. The second, more common approach, is the "momentum bias" system where a wheel provides the net momentum along a virtual spin axis. A pair of such wheels are needed to assure a redundant system. In reality, the former "zero momentum" approach needs some net momentum to provide a basis for antenna pointing.

An emergency in either design occurs from loss of pitch lock, where the body slowly rotates around the north/south direction (perpendicular to the equatorial plane). As long as the wheel is kept within certain speed limits, the spacecraft will tend to remain erect and stable. Problems in three-axis recovery from loss of lock have occurred in the past due to (1) improper use of thrusters (they should have been kept off), (2) inappropriate onboard autonomous protection software that switched the wheel(s) off, (3) lack of sensor (or memory-stored) information so that an autonomously recovering spacecraft does not know where to turn to expect the sun, and (4) operating mistakes made during a problem situation such as one of the above. Depending on where the sun sensors are mounted and how much risk the operator takes, the recovery can take up to 24 hours, during which communication service is almost surely lost. The newer designs are more intelligent and robust, and therefore these problems should become a thing of the past. In any case, it is possible to override such undesirable modes by ground command; using these commands properly is also a key to the safe operation of any spacecraft. The remaining

problem area is the failure of a critical unit or function that cannot be performed otherwise.

Having a stable platform that does not succumb to spin-ups and loss of attitude control is paramount for satellite communication. Users expect their satellite operator to provide dependable service; in fact, they would prefer not to have to worry or even think about issues raised in this section. Selecting the particular bus design and the safeguards at all steps of the manufacture and operation are essential to achieving this result. For this reason, most operators prefer not to be the first in line for a new bus design. Users should gain an understanding of the key issues and review the design and operating history of satellites they intend to use. Users who lease or purchase transponders on a full time basis are typically provided with a monthly report that delineates the satellite health and performance issues.

The responsibility for keeping the spacecraft antenna beam aligned properly with the coverage area falls to the attitude control system (ACS). While a detailed discussion of ACS design and performance is beyond the scope of this book, we want to identify this area as one of potential concern. Spinning satellites can maintain pointing throughout their missions due to the high gyroscopic stiffness produced by rotation of most of the mass of the body. Two directions of control are provided: around the spin axis through adjustment of the despin rate, and north-south using a single motor-actuator. Pointing accuracy can be maintained to about ±0.05° in this manner. However, there is precession of the spin axis over time that causes error along the third axis, called yaw. This is corrected through ground-commanded correction maneuvers using the onboard propulsion system. In noncircular antenna patterns and arrays of multiple beams, yaw error can be maintained as tightly as needed by more frequent corrections.

The situation with regard to three-axis spacecraft is more complex because the vehicle has a potential for tumbling most of the time. The low rotational inertia cannot guarantee beam alignment so the ACS must be in constant control. Interaction of this system with flexible structures like solar panels and antenna reflectors is another consideration. Finally, thruster firings to perform station-keeping and unload momentum wheel inertia cause large attitude transients, which are several times the steady-state pointing error. The overall pointing error thus derived may be of the order of ±0.15°, which can produce footprint EIRP and G/T degradation of 1 dB or more. This will depend on the steepness of the slope of the antenna gain pattern. If this is an issue, then the antennas themselves may require their own pointing mechanisms and means of measuring beam alignment.

3.3.2   Spacecraft Power Constraints

The demand for more downlink RF power has consequences for the design, size, and mass of the spacecraft bus. This is in addition to any impact on the communication hardware, such as power amplifiers and waveguides, which resulted from the use of a high power level. We review in qualitative terms the most significant impacts on the spacecraft bus. Readers who need to quantify the impact should work closely with the spacecraft designer or manufacturer, who is in the best position to evaluate new requirements and their impact on the spacecraft.

3.3.2.1   Power System Limit (Panel Area and Mass)

The most direct impact of increased RF power is the requirement for greater dc power from the electrical power system on board the satellite. The spinner and three-axis designs approach the generation of prime power in a slightly different manner. Generally speaking, flat solar panels on the three-axis can be extended in area, while the cylindrical panels on the spinner cannot exceed the dimensions and volume available within the launch vehicle shroud. Current spinners can generate approximately 2,300W of dc power, of which about 2,000W are available for the payload. The 300W of difference between the two values is needed to power the bus. The comparable numbers for the three-axis are approximately 15,000W and 12,500W for total prime power and available payload power, respectively. Three-axis bus designs that are currently on the drawing board promise up to 20 kW of prime power.

The beginning-of-life power output of the solar panel can be achieved with reasonably good accuracy, although precise on-ground measurements are typically not possible beyond solar cell string level. In orbit, the solar cells in the panel are exposed to the space radiation environment, which consists of free electrons, protons, and UV energy. Most of this charged flux emanates from the sun and follows the 11-year solar cycle. Another source of degradation is surface contamination from outgassing of spacecraft materials that deposit on the panels. Panel designers use projections of the radiation flux that have been compiled by respected organizations like the U.S. National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA). However, no model can predict with certainty what the environment will be for a particular mission. Instead, the approach is to use a reasonably conservative estimate of the environment over the solar cycles in question. In the cycle that peaked around 1990, there were a number of unexpectedly large proton events, the worst of which caused as much as a 1% drop in panel power output. Since panels at launch have around 25% excess power, a 1% drop only becomes an issue near end-of-life.

Early cells were of n-on-p silicon. More efficient cells were introduced in the 1990s, using GaAs semiconductor material. This increased power by up to a factor of two, making 8-kW systems feasible. There was a period of dual junction use and then a quick transition to triple-junction cells, which employ gallium indium phosphide/gallium arsenide/germanium (GaInP/GaAs/Ge). This cell is a sandwich that, aside from the three rectifying solar cell junction, contains two connecting "tunnel" junctions, one between the top GaInP cell and the middle GaAs cell, and the other between the middle GaAs cell and the bottom Ge cell. For correct functioning, all three solar cell junctions need to supply the same current. Complex structures like this convert nearly 30% of incident solar into electrical power.

A last word about solar arrays: the flat arrays on three-satellites are very exposed to the environment and undergo approximately ~200°C temperature swing between the heat of normal sunlit operation and the extreme cold during eclipses. Despite the apparent simplicity of the requirements for a reliable solar wing, achieving long-term mission requirements is not easy. Statistics show that at least 5% of all spacecraft flown have higher than expected solar array power degradation, and, if anything, recent trends with arrays above 10 kW indicate even greater

degradation. Luckily, these problems are usually not sudden-death killers and operational workarounds are possible.

3.3.2.2   Battery System Limit (Volume and Mass)

Assuming that the payload must operate in eclipse as well in sunlight, the power required by the payload would also demand battery capacity. There are two technologies available today for commercial spacecraft, namely, nickel cadmium (NiCd) and nickel hydrogen (NiH_2). NiCd batteries are similar in concept to standard rechargeable NiCd batteries found in home electronic equipment. However, the particular configuration employed in space is designed for much longer life, low mass, and greater discharge capacity. Recently, it has been proposed that satellites employ lithium ion batteries similar to what is used in laptop computers and other portable consumer devices. The other technology, NiH_2, has also been used since the mid-1980s on small spacecraft like the Boeing 376 GEOs and Orbital Orbcomm little LEOs as well as the largest hybrids launched today. NiH_2 can obtain 50% to 100% greater charge capacity for the same mass as NiCd. Up until recently, NiH_2 was only found on the larger satellites but is now available on the smaller spinners as well. Battery technology choice has to do with the spacecraft configuration and depends on the total power required and the capability of the spacecraft design to manage battery temperature and charge. NiH_2, if built properly, is quite robust to charge state extremes, whereas NiCd and lithium are more sensitive.

Some services could employ a satellite with reduced eclipse capacity. For instance, MSS is really a telephone network in space, where subscribers make calls according to their own particular usage patterns. On an aggregate basis, telephone calling follows a predictable pattern of rising in the morning and hitting a peak, called the busy hour, sometime in the early to mid-afternoon. Not having all of the power available at 2 a.m. would be an acceptable compromise, provided that the payload and power system are designed for this kind of variability. This argument has also been used to justify locating BSS satellites to the west of the area service, putting the peak of eclipse after midnight. While this saved battery mass, later operators have chosen to power their broadcasting satellites for the full 24-hour period, so as not to disappoint late-night viewers.

3.3.2.3   Thermal System Capacity

The thermal control subsystem of the spacecraft must remove excess heat so that internal temperatures do not exceed design limits at any time during the life of the satellite. Any increase in input and output power will introduce more heat that the thermal system must reject. With the growth in power demand and packing density of electronics, spacecraft manufacturers have adopted the heat pipe as an effective means to move heat from the source to the external radiating surface. A heat pipe is a long tube containing a substance that can transition between liquid and gaseous states. As with any classical refrigeration cycle, the liquid is converted to gas at the hot end, flowing to the cool end that radiates directly to space. The gas condenses, releases its heat, and returns through the pipe to the hot end. High-power three-axis

satellites have evolved to become cages of heat pipes that move heat from points of high concentration, such as underneath high-power TWTAs and batteries, to external surfaces that can radiate the heat to space. If we generate more heat, then greater radiation surface as well as possibly more heat pipe capacity is needed. This is not a trivial change because it could require extensive rearrangement of the spacecraft structural system and additional mass to provide more thermal control capability. In some cases, physical volume constraints could preclude adding the needed thermal control facilities. One alternative is to design all electronic and mechanical equipment to operate over a wider temperature range. Another alternative is to design radiators that deploy away from the main body, using flexible heat pipes to transfer the heat across the joint that moved. If these steps are not taken correctly, the result will be a hotter spacecraft and a potential reduction in lifetime and performance. For users, there is little that one can do except to ask the right questions of the satellite operator or manufacturer. They should be able to demonstrate by analysis and measurement how their spacecraft will fare over the operating life. Furthermore, large radiator areas will also require significant heaters when the payload is off or only partially in service.

3.3.2.4 Propellant Capacity and Loading

Along with the increase of the spacecraft power and thermal control support, the satellite engineer must consider the propellant load to maintain lifetime. For a given type of propulsion system and propellant, any increase in dry mass of the spacecraft will require the same proportional increase in propellant mass. We typically want to keep the satellite within a north/south and east/west box that is no greater than 0.2° on each side. Most of the station-keeping propellant is required to control the inclination, that is, north/south station-keeping. Obviously, if the need for north/south station-keeping can be eliminated, then substantially less propellant will be required. This can work for MSS missions where users employ broad-beam or tracking antennas but is probably not feasible for broadband applications with fixed dish antennas like DTH and VSAT networks.

It is possible to improve thrust performance without adding propellant by increasing the specific impulse (I_{sp}), which measures thrust per unit mass (in units of seconds) of the propellant. In typical GEO and non-GEO spacecraft designs, I_{sp} is in the range of 170 seconds to 300 seconds. The conventional reaction control system (RCS) formerly employed hydrazine as a single propellant and used a blow-down system with a simple gas pressurant. Moving to a bipropellant system with monomethyl hydrazine and nitrogen tetroxide oxidizer, along with a regulated system to maintain constant pressure during apogee burns, increased station-keeping I_{sp} by up to 50% on the long burns used in north/south station-keeping. Short thrust pulses for momentum dumps or attitude correction bring bipropellant performance way down because the thruster does not have time to reach full operating temperature, but are luckily normally a small fraction of the thrusting required.

Of recent interest are various forms of electric and ion propulsion, including xenon-ion propulsion systems (XIPS) from Boeing, Arc Jets from Olin, and Hall Effect thrusters produced in Russia. These state-of-the-art propulsion technologies

yield values of I_{sp} in excess of 1,000 seconds. The main issue with any of these new concepts (some of which have been around for a decade or more) is that of life expectancy. This is being addressed through life-test programs and, in conjunction with on-orbit experience, will prove the dependability. Additional issues are wide thrust exhaust plumes and thermal inputs to the host spacecraft.

The trend in GEO satellites has been to reach as long a life as possible, extending past 10 years to 20 years or more. If we were to move back to 10 years or even 8 years, it would be possible to reduce total propellant mass. This requires a proper analysis of the mission, which is beyond the scope of this book. However, it is possible to have the satellite operator or manufacturer determine the amount of the required propellant. A shorter lifetime may often have little impact on profitability of the investment because of the effect of discounting the later years of revenue. It might also be possible to increase lifetime by using a different launch vehicle to place the satellite into transfer orbit. In most missions, some of the propulsion system propellant is reserved for orbit corrections and even for the perigee kick function. If the launcher can be depended on to carry out more of these functions (or with improved accuracy), then RCS propellant can be saved for station-keeping. Evaluation of this also requires a thorough analysis by the launch vehicle provider.

If in the final analysis you determine that there must be an increase in propellant load, then there still is the consideration of tank capacity. If the demand is going to be significantly greater than current designs, larger tanks can be installed. A first impact is the difficulty of integrating the larger tanks. A second impact is the structural loading during launch of the larger and heavier (full) tanks. A final consideration is the qualified lifetime impulses of the station-keeping thrusters. Such changes can have a significant impact on the overall design, including its ability to qualify for launch on a particular launch vehicle. Therefore, a thorough evaluation of such a proposed change is justified.

In summary, the required radiated RF power has a significant and possibly major impact on the spacecraft, so it has to be carefully considered. To do this right, one must involve many parties—the satellite operator, the spacecraft manufacturer, and possibly the launch service provider. This is not an impossible task provided it is dealt with in a thorough manner.

3.4   Contingency Planning

Satellite operators and users must engage in contingency planning, which involves making arrangements for backup satellite capacity and succession when operating satellites reach end of life. For operators, this is a matter of maintaining the business in the face of possible launch and on-orbit failures. Users of these satellites share that concern and would probably not use a given satellite system if capacity is not available in the event of a failure. Providing the backup and replacement capacity is costly and if done wrong can lead to a disastrous result for all parties. For all of these reasons, operators and users can participate in the solution to providing continuity of orbital service.

3.4.1 Risks in Satellite Operation

The following subsections identify risks that affect the delivery of space segment service to users. We offer some basic approaches to the resolution of each of these risks. However, this is not a substitute for a detailed plan that is compiled for the unique circumstances of the particular operator and/or user.

3.4.1.1 Launch Failure

The satellite operator and user must make provision for the distinct possibility that a given launch will not be successful. Spacecraft manufacturers can provide a variety of services to compensate for the probability of approximately 10% that the satellite will not reach its specified orbit and provide service. For example, the contract for the satellite might include a provision for a second spacecraft to be ready for backup launch within a specified period after the failure. The contract might even provide for delivery in orbit by a specified date, which implies that the spacecraft manufacturer will have to go through the (expensive) steps that would otherwise fall upon the operator. In the end, however, the operator pays the costs of covering the risk.

It is not unusual for a satellite operator to offer an attractive deal on one or more transponders aboard a satellite that has not been launched. With a significant savings, there is strong motivation to pursue this type of offer. The considerations would be (1) the newness of the design (e.g., it is better to employ a proven design that is essentially a copy of one that is flying successfully), (2) the stage of construction (e.g., in final test with no open issues or already shipped to the launch site), and (3) employing a proven launch vehicle (e.g., one that has a success record of at least 90% and that has not undergone any changes in technology or process). The contract for the purchase should consider the possibility that the satellite may never reach orbit, allowing for a return of deposit and use of alternative capacity. Transponders on a successor or replacement satellite may be attractive provided that the timetable is still of interest for the particular application.

Other steps that an operator can take include having an on-orbit satellite available to maintain the service during the period between the failure and the next launch. This is covered later in this chapter, under the topic of succession strategy. On the user side, some form of contingency plan must be put together. This could involve contracting with another satellite operator to have backup transponders available in the event that the new satellite does not go into operation on time.

3.4.1.2 Loss of On-Orbit Lifetime

Newcomers to satellite communication may have a somewhat negative view of satellite operations, possibly driven by highly visible launch and on-orbit failures along with the business failure of at least two major LEO satellite systems. The actual experience is that most satellites live out their life expectancies and can be counted upon to provide service for a duration of 10 to 15 years. There are exceptions where some kind of catastrophic failure after launch ended the satellite's life prematurely, but the percentage of these is in the low single digits.

An important but often overlooked task of the satellite operator is the proper and efficient maintenance of orbit control. Many GEO satellites enter service using a single TT&C Earth station with one antenna. This has adequate ranging accuracy if the satellite is to be controlled to 0.2° on each side of the station-keeping box. As more satellites are added to the same orbit position, improved accuracy becomes a requirement. Improved ranging methods, which may include a second TT&C station, are then needed to provide range data to enhance the orbit determination process. This allows the software to come up with an accurate orbit more quickly. For non-GEO operators there is also the need to maintain multiple satellites and to coordinate the arrangement of multiple orbits to assure continuous service. Non-GEO systems are different in that many of the satellites are not in view of TT&C stations at any given time.

Even with the excellent experience to date at GEO, the risk of loss is so great that operators and users must have contingency plans. If the risk of reduced lifetime could be anticipated by a few years, then the parties can simply plan on launching the replacement ahead of the originally planned date. Launch failure can be countered by making a first attempt at replacement at least 18 to 24 months prior to the deadline, thus allowing an adequate window to construct and launch a carbon copy. This introduces very little disruption in the normal planning cycle. Planning for an unexpected loss of life, such as that experienced when a satellite abruptly loses a significant fraction of its transponder capacity, means having an extra satellite available in orbit. Rather than sitting idle, this on-orbit spare can be employed for preemptible services that are discontinued when and if the satellite is needed. Preemptible services will produce revenues that help offset the investment and operating cost of the spare.

3.4.1.3   Reduced Technical Capability

Any organization that is engaged in a high-technology activity is exposed to the risk that it will not be able to maintain a sufficient level of technical competence. This depends on the people who work for the company and includes their qualifications and level of training. Historically, companies and government agencies have attempted to build competence through in-house education programs and on-the-job training. There has been a trend in recent years to require that new people come to the company already trained, either because they worked for another organization in the same or a similar line of business or because of their individual educational experiences. This reduces the training burden on companies but increases the risk from poaching—the tendency of companies to lure qualified people away from each other with attractive offers of employment.

In satellite operation and application, loss of technical capability has not been a problem in developed countries, probably because of the outflow of engineers and technicians from the defense and aerospace industries. People are also becoming available as large telecommunications companies downsize to become more competitive and as many Internet and wireless startups fail as businesses. Eventually, this overhang will diminish as more and more people reach retirement age. We will, at that point, depend on the production level of new engineering and other technical

graduates. As a consequence, satellite communication organizations may find it more and more difficult to maintain adequate staff.

A related aspect of this problem is that the technical demand on an organization can increase as new classes of satellite systems and communications technologies are introduced on a large scale. The transfer from analog to digital video along with the popularity of DTH put pressure on the satellite job market, but organizations, universities, and individuals responded to fill the void. Experienced people in the industry also have demonstrated flexibility, possibly because of the range of services available on satellite networks worldwide. For many of us, the fact that the business is getting more complicated makes it that much more interesting and challenging.

3.4.1.4   Loss of Ground Facilities

Ground facilities tend to be less reliable than the satellites that they support. Part of the reason is that they are exposed to many environmental risks, such as flood, Earthquake, fire, wind, theft, and civil unrest. The equipment within an Earth station or control center is designed to perform its function for 5 to 10 years, not 15 to 20. In addition, ground facilities are dependent on external support to keep them running. Some of this can be countered through backup means, such as an uninterruptible power supply (UPS), local water storage or supply, and storage of large quantities of supplies and spare equipment. At some point, however, the ground facility will not be able to fulfill its role either as a control point for the satellite or as a communication node.

Assuming that we have taken appropriate measures to strengthen a particular facility against the expected hazards, the only thing that remains is to provide an independent backup. For a satellite operator, this means having a backup TT&C station and satellite control center. This type of strategy provides a very high degree of confidence that service will be maintained even if the primary site goes out of service. The physical facilities can probably be more easily replaced than the people who operate them. As stated under the section on reduced technical capability, having qualified people available can become a challenge. If we routinely have one operating site to control the satellites, it would be quite a burden to try to maintain a backup site with qualified staff as well. Most of the time, this staff will have little to do and therefore might not be as experienced as those who work at the main operating site. This may be countered by providing routine training and exercises for the staff and by rotating qualified people between the two locations or by assigning complementary prime and backup roles, so that each site nominally works at less than a peak capability, which is only reached in contingencies.

For satellite users and their communication Earth station facilities, the trend has been for the sites to be unattended. Trained staff would normally be located at the network control center and at distributed maintenance facilities. The key here is to have enough staff deployed at different locations so that there is inherent diversity in the operation. A concern is with a single network control point and the possibility that it will be knocked out. The best approach here is to have at least two such facilities, each supporting half of the network. In the event of an outage, the other facility takes over management of the entire network.

In the 1990s, ground facilities were a valued asset and cost a great deal to purchase and develop. This situation has turned around due to many well-equipped Earth stations and teleports coming on the market as their owners seek to reduce capital obligations (or just go bankrupt). In every major city in the developed world, it is possible to either rent space or purchase a working facility on the cheap. Once done, obligations to pay fall to the new owner, so one should do a proper evaluation before making a commitment.

3.4.1.5 Harmful Interference

Any radiocommunication service is potentially a victim of harmful radio frequency interference, which can be either accidental or intentional. A complete discussion of the regulatory aspects of this issue is covered in Chapter 12. In this instance, we are concerned with accidental or intentional disruption of legitimate satellite transmission by another party. By harmful we mean that authorized services are disrupted or rendered unsatisfactory to users. This is different from unacceptable interference, which is a term in frequency coordination to indicate that the calculated interference level is above some detection threshold. The vast majority of harmful interference events are accidental in nature, resulting from an error in operation or an equipment failure of some type. This means that whatever the cause, the interference will be found and corrected as a matter of course because the error or failure produces a direct loss of performance for the unknowing perpetrator.

Intentional interference is rare and often quite notorious. In many countries, particularly in the developed world, intentionally causing harmful radio frequency interference is a crime. This has been an effective deterrent, mainly because the lawful operators want it that way. Satellite communication is particularly vulnerable because any transmitter on the ground that is within the satellite footprint can be a source of harmful interference if it has sufficient EIRP. In area-coverage systems, it is difficult but not impossible to locate the source. Most of our efforts are expended on monitoring all of the transponders in the downlink so that interference can be observed as soon as it appears. This is augmented with good direct telephone communication with users, which are usually the first to notice an interference event.

The key to controlling and eliminating harmful interference is to constantly maintain this type of vigilance over the system. As soon as any interference is detected, the operations staff must move quickly to identify the source and demand correction. In the vast majority of the cases, this is effective in a matter of minutes. The remaining cases take longer to correct, sometimes hours or days if the source cannot be isolated quickly. The approach here is to reduce the impact of the interference by moving users to different frequencies or different transponders. This allows the problem to be studied more carefully without the pressure of having to maintain service.

Intentional interference is a source of anxiety among satellite operators and users alike. There is always the possibility that a radio pirate might either take over an existing legitimate Earth station or build one for the express purpose of causing some kind of abuse (harm or theft). In the rare cases where this has happened (only three that this author can recall), the perpetrator was identified and prosecuted. It

turns out that the type of person who would do this sort of thing has emotional problems. This allows the police and other authorities to track down the individual. In the meantime, people in the industry are given the opportunity to think about how this kind of disruption can be detected more quickly and how to prepare for the next episode. It provides an opportunity for all to increase the level of vigilance. Looking at the international environment, there is no common police force other that the rules of the road provided by the ITU. Therefore, interference problems involving satellites of different countries require some diplomacy in their resolution.

3.4.1.6   Sabotage

Another source of intentional disruption is the physical type, which we call sabotage. Since the satellite is controlled from the ground, it is conceivable that someone might attempt to vandalize an operating TT&C station. Any high-power Earth station used for TV uplinking might also be used to jam the command frequency or even take control, given the proper command encoding equipment. The newer generation of commercial satellites tends to have secure command systems to make a takeover a very remote possibility.

Most Earth stations that are capable of causing sabotage to the satellite are protected with security perimeters. The amount of this type of physical security will depend on the risk. In the United States, it is normal practice to provide security fences, doors, and even guards. Facilities in remote areas might have less physical security, but some minimum amount is still justified. Recall the old adage that most locks are designed to keep an honest person honest.

Satellite control facilities that provide government communications services must be protected to the fullest. This is a special case and is really beyond the normal scope of commercial affairs. However, the thought processes that the government applies can be useful to protect high-value installations. It is always prudent to think about what type of attack might be possible and what could be done to minimize the risk or impact. Some time spent in anticipation of this kind of event is well worth the effort. For example, the only thing that may need to be done is to use the physical security that is already in place but is currently not being used. For example, security doors with TV monitors might be installed but are deactivated as a matter of convenience. If the risk increases for some reason, then all you might need to do is to reinstate the use of these doors and TV monitors. Computer systems are very capable of providing greater security than is used on a routine basis. Tightening security might only be a matter of using and changing the access control mechanisms (including passwords) already provided in the operating systems and data communications network.

3.4.2   Available Insurance Coverage

The policies and procedures described in the previous section deal with the operational impact of risk. There is always the financial impact of loss, for which insurance is an effective preventive measure. We consider some of the more common types of insurance that can be purchased by satellite operators and users.

3.4.2.1 Launch Insurance

A completed but unlaunched satellite stands between an 85% and 95% chance on the average of successfully reaching orbit (GEO, MEO, or LEO) and being capable of a planned start of service. Some launch vehicles and supporting services have achieved the higher end of the range, including Arianespace's Ariane 3 and 4 launch vehicles and McDonnell Douglas's Delta 2 series. Lockheed Martin's Titan and Atlas Centaur have nearly as good a record as the leaders. The launch vehicles available from China Great Wall Industry Corporation of the People's Republic of China are potentially good performers, but the record to date is still advancing from the low end of the scale. And lastly, fully developed Russian launch vehicles like Proton and Zenit are popular in the commercial marketplace.

Planning for GEO systems is on the basis of launching one or two satellites at a time and amounts to betting against "snake eyes" on the toss of dice. As we shall see later in the chapter, building a multiple-satellite GEO system is a step-by-step basis and can proceed more or less in a serial manner. Global LEO systems like Iridium and Globalstar, discussed in Chapter 11, cannot start service with a single satellite but require an initial operating constellation of dozens of working satellites. On a relative basis, one launch by itself does not pose as much of a financial risk as with GEO systems. On the other hand, a serious problem with the design or manufacturing process of the launch vehicle used to create the LEO or MEO constellation can halt implementation and the start of service by between 6 months and 1 year.

The simple fact is that launching satellites is a risky business and demands every possible step to assure the financial and operational viability of the user and the satellite operator. Spacecraft manufacturers may or may not bear part of the risk, depending on the nature of the particular contract. They, too, need to consider how to insure their financial exposure. Launch insurance is generally available to the parties who stand to lose in the event of a failure to reach orbit or maintain service after an initial operating period. The satellite operator can purchase coverage equal to the purchase price of the spacecraft and launch vehicle. This is typically increased to assure that all of the expected cost of a replacement is included. To do this, the number may be adjusted upward for inflation between the contracted price (which is probably 2 to 3 years old) and the time when a negotiation for the new spacecraft and launch would happen. Alternatively, the operator may use option prices that were previously negotiated with the manufacturer and launch vehicle provider. The last item to be included is the cost of the next launch insurance policy.

Operators of LEO and MEO networks take an entirely different stance with respect to launch insurance. Some may prefer to self-insure, meaning that no specific launch insurance with be purchased. By purchasing sufficient extra spacecraft and launch vehicles, they will provide the needed insurance against the expected failure rate of the launch vehicle systems employed. This will not protect them from a systematic problem with a particular system, but the operator could reduce risk by selecting a second source early in the program.

Major users who purchase transponders for the life of the satellite can also obtain launch insurance. In the 1980s, cable TV networks like HBO and Turner Broadcasting purchased such insurance from the same sources as the operators. The issue here is that there could be a very substantial insurance liability placed on a single launch, representing such a large loss as to be uninsurable. The same applies to a

multiple launch where two satellites are insured. The simple answer to this kind of problem is to stay within the limits imposed by the marketplace. If the risk associated with a particular launch exceeds what the insurers will cover, then the insured might get together and divide their risk.

The entities that require the insurance have multiple sources for coverage, depending on the country of origin. Often an insured party will work with its existing underwriter. In the (distant) past, much of the coverage found its way to the largest insurance market in the world for high-risk activities—Lloyds of London. As many readers know, Lloyds is not an insurance company but rather a coordinator. They represent literally thousands of insurers, called *names*. These are companies and even individuals who attempt to make money by betting against disaster. Unfortunately, there have been more disasters in the satellite, shipping, air, and other industries to make the business of being a name rather unattractive. The launch insurance game is now in the hands of a new breed of underwriter who tends to take less risk by charging high premiums in the range of 20% to 30% and further sells much of the risk onward to large reinsurers.

Other insurance coverage is typically bundled in with the purchase price of the satellite and launch vehicle. There is risk of financial and human life loss due to some kind of catastrophe at the launch site. This is a rare, but not unknown, type of loss. The providers of hardware and services typically insure against these losses. Failure to launch on time is typically not a risk borne by providers, except possibly that the spacecraft manufacturer could, under contract, be held to such a claim. This depends on the particular arrangements made ahead of time.

3.4.2.2   On-Orbit Life Insurance

Commencing with the initiation of service, satellite operators usually insure their operating satellites against loss of lifetime. The price of this coverage is proportional to the value of the satellite reduced by the number of years already expended in orbit. A direct analogy is the kind of warranty that automobile tire manufacturers provide, which is reduced by either the years remaining or the consumed tread.

The cost of life insurance has been in the range of 1.5% to 4% per year. Owners of transponders can also purchase life insurance, or, alternatively, it could be provided as part of the transponder purchase agreement (i.e., similar to the tire warranty). Users who rent their satellite capacity have no direct need to insure the remaining life because they simply do not have to pay if the capacity is not available due to a satellite failure. Their situation could be difficult, however, if they have not made other provisions for replacement service.

3.4.2.3   General Liability Coverage

There is a wide variety of other insurance coverage that is valuable to those engaged in the satellite communications field. Some examples include standard workman's compensation insurance, insurance for loss during transportation of equipment, patent liability coverage, insurance to provide replacement of lost facilities or services, and liability insurance to cover the intentional and unintentional actions of employees and management. There is likely to be a need for insurance against

liability for injury or damage that result from a launch failure or the possibility that a satellite may reenter the atmosphere before it reaches its final orbit.

3.4.3 Space Development—Estimating Lead Time

Communication spacecraft used in GEO, MEO, and LEO networks require a considerable time for the design and manufacturing cycles. These last from as long as 6 years for a complex new design with an OBP to as little as 12 months for a very mature design with some existing inventory of parts or subsystems. A typical GEO-class spacecraft of standard design will be contracted to take about 24 months to deliver to the launch site from the time that the manufacturer is authorized to proceed with construction, and will probably take closer to 36 months. The launch service provider also will require lead time to arrange for construction of the launch vehicle and to reserve the launch site. The resulting waiting time to launch could be as long as 30 months once the order is placed. This means that the developer of a new application or system must allow sufficient lead time.

An overall timeline for a typical spacecraft development program is shown in Figure 3.15. This takes the perspective of the satellite operator or developer of an application that is dependent on the availability of a new satellite type. It allows for a precontract period of about 6 months to collect business and technical requirements and to prepare technical specifications. The period could be shortened if the requirements are standard and no new development is required, such as for a "plain vanilla" C-band satellite for video distribution. On the other hand, if we are talking about a new concept for which no precursor exists, the precontract period could last 1 or 2 years.

The satellite operator will normally procure the spacecraft according to a competitive process to provide some confidence that the best performance under reasonable terms have been obtained. This considers the technical, cost, and schedule requirements for the project. Such a procurement may take anywhere from 3 to 6 months or more, depending on the same circumstances mentioned previously. We assume in Figure 3.15 that a decision has been made during the precontract period on the supplier and specifications. The supplier will proceed with the design engineering portion of the program, culminating with a preliminary design review (PDR) some time around the sixth month. The PDR will be evaluated by management from the supplier as well as the satellite operator/buyer and its technical and business
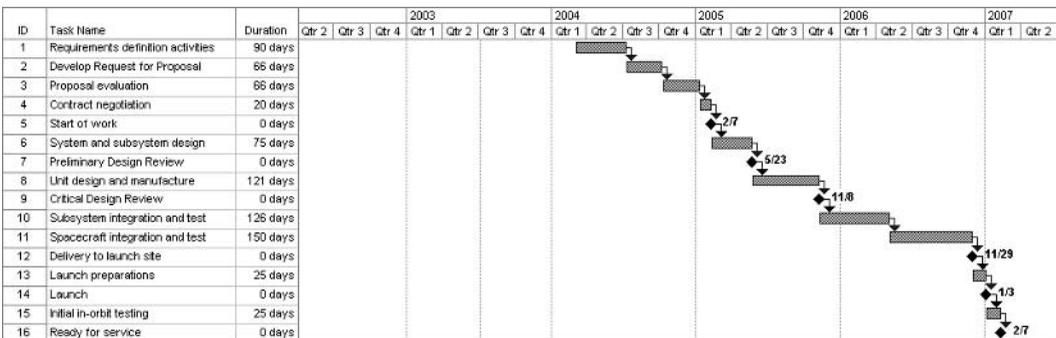

Figure 3.15 A typical spacecraft design and manufacturing program (24-month delivery).

staff. Once completed and progress affirmed, the spacecraft supplier will move into the detailed design and manufacturing phase. At some point, perhaps 12 months from start, a critical design review (CDR) will be held for the same reviewers with the objective of ratifying that the spacecraft program is proceeding correctly.

Units may now be integrated into systems and tested for compliance with the specifications of the satellite. During the integration and test (I&T) phase, the subsystems are installed in the spacecraft and tested both in ambient air and in a chamber that simulates the space environment. The objectives here include:

- Build a spacecraft that is capable of surviving the launch and initial deployments;
- Demonstrate that the satellite will meet its performance specifications throughout its operating lifetime and under all expected on-orbit conditions;
- Demonstrate that the satellite can be operated effectively by radio control through the TT&C system;
- Verify that the spacecraft will withstand the space environment, which includes solar radiation and heating, cooling with battery discharge during eclipses, spacecraft electrostatic charging, and contamination from external and internal sources.

After the spacecraft completes the I&T phase, it is ready to be shipped to the launch site where it is again checked for integrity and operability and placed on top of the launch vehicle to be tested once again. The launch site preparations take between 1 and 2 months, depending on the type of vehicle and the number of spacecraft to be launched at the same time. Also to be arranged are the tracking stations and services that allow the satellite manufacturer and operator to conduct transfer orbit operations and initial on-station testing. Once launched, the satellite is tracked and commanded through the various phases, thoroughly tested and put into service. If this is a new type of network service, then users may need to conduct network checkout testing to demonstrate that everything is in working order and ready for commercial service.

3.4.4   Satellite Backup and Replacement Strategy

Under the assumption that an operator's satellites will work as planned, one must still plan for replacement of the satellites at end of life. This can be a complex and somewhat uncertain process because of (1) the time needed to design and manufacture the replacement satellite (not to mention the time it takes to figure out what kind of satellite to buy), and (2) the operating lifetime of a particular satellite, which is only known within something on the order of a plus or minus 3 months accuracy.

An example of a replacement strategy for a hypothetical satellite system consisting of three orbit positions is shown in Figure 3.16. As this suggests, the best and simplest approach is to start with the current orbital arrangement and build a series of timelines (arrayed from the top to the bottom of the page). The satellite operator in this example starts in 2004 with three operating satellites: F1 and F2, launched in 1994, and F3, launched in 1997. This particular situation might have come about because F1 and F2 were launched within 6 months of each other to provide a
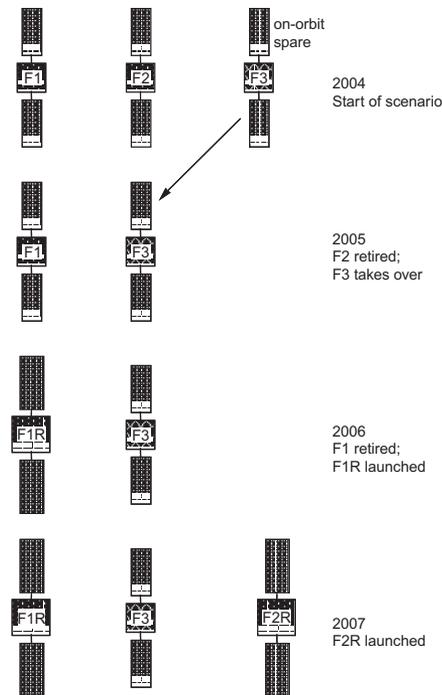
Figure 3.16   A typical satellite replacement scenario.

reliable system of two satellites; since both reached orbit successfully, the third satellite, a launch spare, could be delayed until demand materialized. The operator chose to place F3 into service in 1997 as an on-orbit spare and use it for occasional video and other preemptible services. This provides high confidence that at least two satellites will be available. We assume here that the operating lifetime of each satellite is approximately 12 years.

The satellite operator purchased two replacement spacecraft (F1R and F2R) for delivery and launch in 2005 and 2006. This will ensure continuity of service, provided that both launches are successful and as long as either F1 or F2 exceeds its specified life by at least a year. Figure 3.16 indicates that in 2005, F3 will be taken out of service and drifted over to F2's orbit position. This will allow F3 to take over for F2 when its lifetime runs out. Next, the replacement for F1, called F1R, will be launched in 2006 so that services can be transferred to it in a timely manner. In 2007 F2R will be launched and placed into F3's old orbit position, which will have been vacant for about a year. This scenario provides high confidence that at least two orbit positions will be maintained during the entire transition. If there had been a launch failure, then F3 would have lasted long enough to permit another spacecraft to be built and launched.

Satellites that work but are running out of propellant can be extended in lifetime by switching to inclined-orbit operations. In this mode, a small amount of propellant is reserved to maintain the assigned orbit longitude. Inclination is allowed to build up (at a rate of approximately 0.8° per year), requiring that Earth station

antennas track to satellite during its north-south excursion every day. MSS systems are generally operated in this mode because of the broad beamwidths of user terminal antennas and because antennas on ships and airplanes must track to compensate for relative motion. Also, large antennas in C-, X-, Ku-, and Ka-band networks must track to deal with normal station-keeping and so extending life through inclined-orbit operation could be allowed with little impact. The fact that the satellite is operating beyond its years allows the operator to offer a deep discount; alternatively, it provides a means to hold an important orbit slot until a replacement can be launched.

Obviously, there are many possible replacement scenarios and it makes sense to examine as many as can be imagined. There are financial as well as regulatory considerations. While this is done, users of the system must be kept informed so they understand how the operator is replacing the satellite that makes their respective businesses possible. This is very critical because experience has shown that an operator who practices good replacement planning will tend to have better acceptance in the marketplace. The same goes for the application developer and user.

References

[1]   Elbert, B. R., *Introduction to Satellite Communication*, 2nd ed., Norwood, MA: Artech House, 1999.

[2]   Craig, A. D., and F. A. Petz, "Payload Digital Processor Hardware Demonstration for Future Mobile and Personal Communications Systems," in *Signaling Processing in Telecommunications*, E. Biglieri and M. Luise, (eds.), New York: Springer-Verlag, 1996, p. 271.

[3]   Chiassarini, G., and G. Gallarino, "Frequency Domain Switching: Algorithms, Performances, Implementation Aspects," in *Signaling Processing in Telecommunications*, E. Biglieri and M. Luise, (eds.), New York: Springer-Verlag, 1996, p. 283.

PART II

Broadcast and Multicast Links to Multiple Users

Television Applications and Standards

Television represents approximately 70% of commercial communications satellite use, and it is very suited to the characteristics of the medium. The point-to-multipoint nature of video communications fits the wide-area broadcast feature of the satellite link. The ability of a satellite to serve a particular TV market is simply determined by the coverage area footprint. Thus, the Galaxy satellites are optimized to serve the 50 United States, the JSAT satellites cover Asia, and the SES-Astra satellites are intended for services to the broader European region. The TV signals themselves must be consistent with the technical and content characteristics of the region served, aiming for one or more particular user segments. The important segments include network broadcasting to local over-the-air TV stations, cable TV (CATV) systems, and DTH subscribers who own their own dishes. A complete discussion of the DTH segment is provided in Chapter 6.

Figure 4.1 provides a framework for the discussions in this chapter. Of fundamental importance are the standards used in the creation, organization, and distribution of the programming product. In the analog domain, the same format is used during each stage of preparation and delivery. This imposes tight specifications on the transmission performance of the channel, particularly the video signal-to-noise ratio (S/N) and various impairments that distort the picture. These are covered at the end of this chapter. Digital formats, which are more tolerant of noise and distortion, are described in Chapter 5. During the preparation of the product, the general view is that no impairment should be introduced. In digital terms, this means that the highest data rate possible should be used. Distribution of the signal to the consumer can be with the lowest data rate that is consistent with an adequate perceived quality—the quality generally going down with the data rate due to compression artifacts.

The attractive nature of modern TV programming is a tribute to the technology and skill used to put the product together. The other vital aspect of this medium is the range of possible applications to which the product is put to use. Commercial network broadcast is the broad category that includes the various forms of entertainment television. Once carried out as a local service, network broadcasting is now an international medium. Businesses also exploit the live-action nature of television through private broadcasting and two-way interactive video teleconferencing. In the case of the latter, the medium provides the means to engage in interpersonal communication where the content is also supplied in real time. For the sake of simplicity, we divide the satellite TV industry roughly into entertainment programming and business TV (the economic value of the entertainment sector substantially outweighing the business sector).
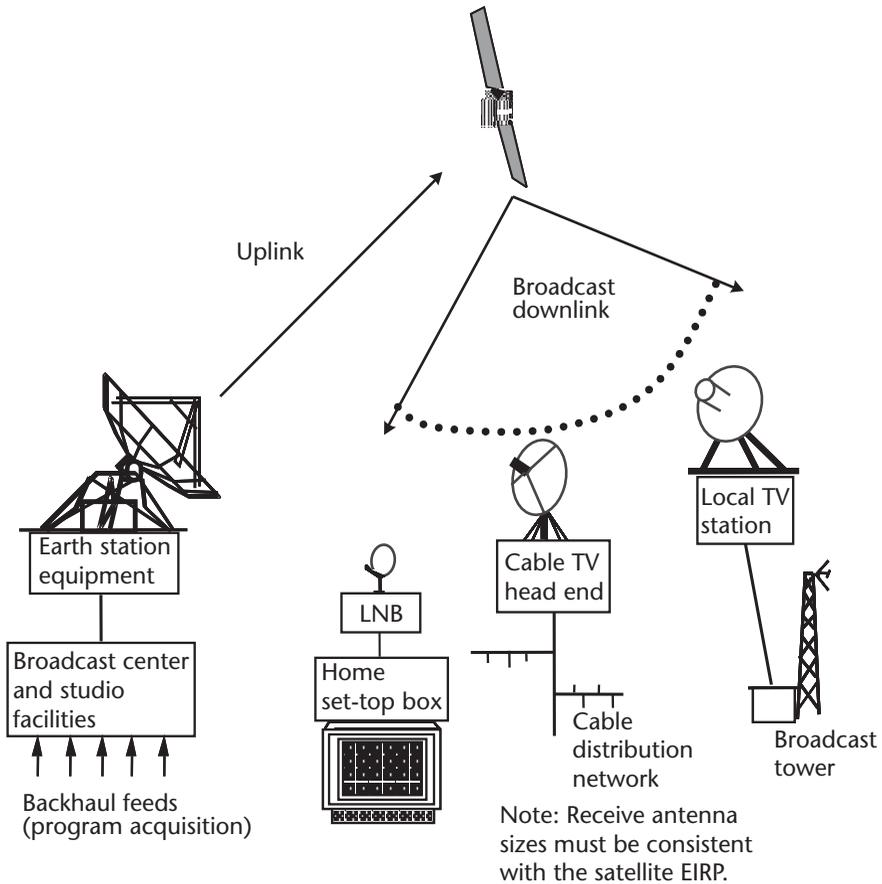
Figure 4.1   Framework for video distribution or direct broadcasting.

The delivery of the product is determined by the equipment used to receive it coupled with the mechanism used to pay for it. With DTH systems, individuals can receive the signal with consumer equipment, providing the most direct connection between the programming supplier and the public. On the other hand, the supplier will often want to restrict distribution to those who have paid for it or to those for whom it is intended for a variety of other reasons. Restrictions on delivery can be based on geography or association with a group. In the following sections, we review the categories of satellite video applications as a means to better define the requirements for each.

4.1   Entertainment Programming

The ultimate consumer of TV programming is the household. In modern times, people have more available time for recreation and are therefore looking for the best value [1]. Beyond a doubt, this is from the home TV set. There are approximately 100 million homes with television sets in the United States. The average time spent

viewing TV in each of these homes is about 8 hours. In advanced economies, the quality of this programming is quite high and everybody can find something that they enjoy. The price of doing this on an hourly basis, after you have purchased the set, is nearly zero. Adding more variety usually means spending more disposable income. For any other form, such as movies on the screen, sporting events, and live theatrical events, the price per hour is much higher. Computer games, like VHS tapes and DVDs, are costly on a per-purchase basis, but can be replayed ad infinitum. But programming is still going to be the best entertainment buy, even when obtaining it through pay services over cable and satellite.

The nature of TV programming is quite subtle, as its value is determined by the size and type of audience that it draws. By far, the most valuable programming is that which is developed for and used in the major commercial TV networks that broadcast their signals through local transmitters and over cable TV systems. The total revenue of the U.S. TV networks is about $60 billion, which now exceeds what newspapers collect from their advertisers. About one-fifth of this revenue goes to the national networks themselves, with the rest divided up among the local TV stations, cable TV networks, and DTH operators. The most attractive type of programming is that which appeals to a broad cross-section of the public divided into certain key age groups. Behind this is the money that comes from manufacturers, retailers, and service providers, who advertise what they sell through the TV medium.

The following discussion suggests what the nature is of this most valuable segment of programming, based on the lineup of TV shows and their relative popularity [2]. Since our first edition, a new classification of programming, called reality television, has appeared (dubbed "Voyeur TV" by *The Economist*). This gained popularity with *Survivor* on CBS. The format is real-life to the extent that "normal" human beings (not actors) are left to fend for themselves in some "hostile" environment such as a desert island or remote jungle. Whether the situations and participants actually represent reality is debatable. At the time of this writing, Fox Television Network led the ratings with a reality show called *American Idol: The Search for a Superstar*. *American Idol* is a hybrid of a talent search (like Ted Mack's *Original Amateur Hour*) and a reality contest show (like *Survivor*) where contestants can vote each other off the program. In the case of *American Idol*, amateurs audition in front of three critics who lay it on the line without mercy. The apparent brutality of the criticism lends to the drama and humor of the spectacle of people being rather foolish. Eventually, after thousands of auditions—shown on screen like the classic *Candid Camera* TV series—one person prevails as the surviving "American Idol." The show has been enormously popular, having approximately 50% audience of all TV watchers during the finals held in May 2003.

Coming to an end after many seasons, is NBC's *Friends*, a situation comedy (sitcom) that stars several of the most popular thirty-something actors and actresses who play out their personal lives within the context of apartment living in New York City. This particular program appeals to the most sought-after age group by advertisers—that is, 18- to 45-year olds. It has been on top at NBC for 3 years. One might ask if its popularity is due to the story line or to the cast—and the answer must be both. Another very good draw is the *Law & Order* franchise, a detective/prosecution hybrid drama that has spun off similar programs under the same

brand. Still a top-five series, *Law & Order* gave NBC a valuable brand and avenue for very substantial revenue through cable syndication.

A category of global appeal is that of sporting event coverage. The baseball leagues in North America and soccer leagues in Europe and Latin America are big draws for TV audiences. Special events, such as the Olympic Games and the World Cup, draw audiences in the billions—something quite incredible when you think about it (only satellites could achieve this end). Recent innovations in sports coverage come from providing coverage of the local team no matter where they play in the nation or world. This requires replication of the functions of origination and the contribution of multiple programming inputs (feeds) to the studio for eventual distribution. The satellite and terrestrial telecommunications facilities required for this in a major country can be substantial.

One could spend a lot of time studying the different programs, their makeup, and source of popularity. The business is very dependent on the creativity of the writers, producers, directors and actors—just like in the movies. Also, what is popular this year can lose appeal with the passage of time. All popular shows have substantial value as reruns on cable TV networks and independent TV stations. *Law & Order*, for example, is very successful as a rerun on the TNT and A&E cable networks at the same time that new shows in the series are appearing on NBC.

The national network TV programs discussed previously are produced and first released in the United States. This is because the United States still represents the best market in the world for English-language programming and because the networks and their program producers are based in the United States. Many of these programs are distributed throughout the world, resulting in substantial additional revenues for their producers. In Europe, the quality of satellite-delivered programming tends to be higher (in commercial value) than what passes through the terrestrial networks. This probably evolved because TV stations were largely installed and operated by governments as a public service and not as a business. The same could be said of TV in Asia and other parts of the world. For this reason, services from Sky of the United Kingdom, RTL of Luxembourg, and Canal Plus of France have become powerhouses and mainstays in the programming field. The programming mix on these systems includes locally developed shows and, more particularly, many of the most popular entertainment shows from the U.S. networks and studios. The Sky services, in particular, are offered by British Sky Broadcasting (BSkyB), formed from the merger of Sky, controlled by News Corp., and British Satellite Broadcasting. The combined company reached 10 million subscribers in 2003 through a combination of cable and DTH access, split roughly 33% and 66%, respectively.

In Asia, the Japanese probably enjoy the most extensive mix of commercial television broadcasts and networks. This is because of the economic status of the country and the great appetite of the Japanese for entertainment and information. There is an extensive infrastructure of microwave and fiber, as well as several TV stations in each major city. Satellite DTH is available from Sky PerfecTV, a service that reaches nearly 10 million households. Much of the programming is locally produced, but there is an interest in American TV shows and movies as well. Hong Kong and Singapore also have excellent local broadcasting services in multiple languages to reflect the diverse population. Some of these stations, particularly TVB in Hong Kong, are extending themselves throughout Asia in response to the rapidly

growing demand for TV. Cable TV has become fairly popular in these cities, but DTH services are generally lacking.

The Star TV services, initiated by Hutchison Wampoa and sold in 1993 to News Corp., are attracting audiences through C-band DTH and cable TV. This supplements and in some cases substitutes for local TV in areas where broadcasting is weak or nonexistent. Star began its operation at the start of the 1990s using the relaunched Westar 6 satellite AsiaSat 1 and a second satellite, AsiaSat 2, which went into service in 1996. The coverage of these satellites extends throughout Asia and all the way to the Middle East, making the programming services available to more than two-thirds of the world's population. Programming is a mix of U.S.- and European-derived entertainment, news, music, and regional movies. Some of the channels are locally produced in Hong Kong and at least one service is originated for distribution in India. That U.S. programming maintains its popularity in Europe and Asia is a source of concern to the national governments that may have a policy of encouraging local content and culture. On the other hand, the satellites that deliver these services, particularly the Astra and Eutelsat series in Europe, are the recognized hot birds where ground antennas remain focused.

It is worth mentioning the great interest in using the Internet as a distribution mechanism for video and audio content. The notion that the Internet could play a substantial role and even supplant commercial and satellite broadcast was premature at the least and no doubt part of the "new economy" boom that led to the "tech wreck" of 2000. The Internet has great strengths in allowing everyone to exchange data and to access stores of information available on Web servers literally anywhere in the world. However, the networks that extend these resources to the end user are typically of limited bandwidth and quality. This makes it impossible to assure delivery of the wide bandwidth needed for commercial programming. We will address the subject of broadband interactive satellite networks later in this book.

While programming can be distributed through land-based systems and even over the Internet, satellites have been the principal carrier of commercial TV material. When satellites first became popular for TV transmission, it was a simple matter to format the signal for distribution. As discussed in Section 4.4, the analog signal format that comes from the camera is essentially the same format used in production and transmission and for viewing by the home receiver. With the introduction of digital TV, new formats are being exploited in certain situations. The typical low-cost home TV receiver is still the same. In fact, television itself was first demonstrated at the 1939 World's Fair, where large numbers of people saw it. Early receivers became available in 1940s for limited broadcast reception. According to Paul Resch, a television industry observer and expert, one of the original receivers would still take the signal off the air today—except, of course, for Channel 1, which has been reassigned to mobile radio. Interestingly, the same principle applies to the first telephones in that you could receive a call over most telephone lines. However, time marches on and digital television sets (like digital cell phones) are popular due to the added features they offer.

Program distribution systems throughout the industry employ all or some of the architecture shown in Figure 4.2. There are many options as to how this architecture can be implemented, but the most economical approach is to combine these elements into a single facility. This is unique to the satellite industry, where one

organization can create, distribute, and sell a product from one location. Network broadcasters like ABC and Fox in the United States, Star TV in Hong Kong, and Tokyo Broadcasting in Japan exploit the efficiency of this arrangement. The basic functions provided are:

- Program origination in a studio;
- Display of prerecorded material, on film, tape, or disk;
- Program contribution by terrestrial links to remote studios and other venues;
- Program contribution by satellite links (typically C-band) from remote venues and other sources;
- Reception of material using electronic news gathering (terrestrial microwave);
- Reception of material using Ku-band satellite news gathering (SNG);
- Recording and playback using tape and disk media;
- Editing of these various inputs into the actual program to be broadcast;
- Relay of programming captured from satellite links;
- Transmission from the studio to a satellite uplink (either C- or Ku-band);
- Transmission from the studio to a local broadcast transmitter and tower;
- Control and switching by computer of input and output video and audio signals to prepare program material for recording and distribution.

The selection and use of combinations of these functions depend on the type of programming and service involved. For example, some programmers produce all of their own material, including sports event coverage, news, and movies, and therefore require the most extensive and reliable facilities. NBC, for example, has redundant broadcast centers in New York City and Burbank, California. A smaller operation like the Disney Channel in Singapore will emphasize the local replay of tape and the retransmission of programming that is received over a Pacific Ocean satellite. The least impressive arrangement that this author has seen was a pair of tape machines connected to a video uplink that represented the complete program distribution system for a startup cable TV subscription movie channel.

One of the most important features of modern broadcast centers is the use of backhaul circuits, many of which are provided over other satellites. As indicated in Figure 4.2, there are several backhaul inputs available at the facility. The satellite capacity that is used for this could be from a pool of available transponders that is maintained by the network. In markets like the United States and Europe, occasional video services from larger satellite operators provide the temporary transponders when needed at a substantially lower cost. The TV network or station reserves the transponder ahead of time and arranges for an uplink for use during the programming event. Backhaul is also available over fiber optic networks. A company called Vyvyx in the United States uses the fiber optic network of Williams Communications to make occasional connections between networks like Fox Television with football stadiums around the country.

The truck in the center of the figure indicates how SNG service is introduced so that the network may cover an event no matter where it is located (compact transportable "flyaway" stations are also used). TV networks and stations own these trucks since they provide a great deal of versatility and flexibility in covering remote
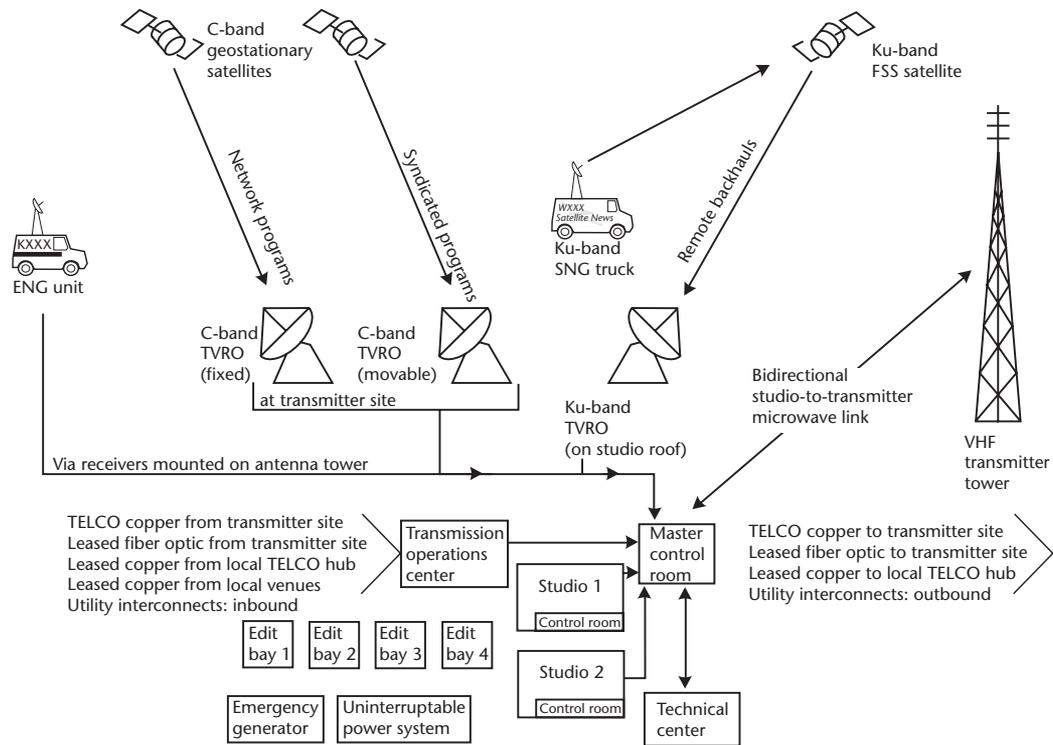
Figure 4.2 Main Broadcast center architecture for program development, acquisition, and distribution to affiliate TV stations. (Courtesy of Paul Resch.)

venues and events. The typical SNG truck includes a 2.4-m antenna and 400-W Ku-band transmitter, along with a limited quantity of production equipment to permit some local editing. The one condition of using SNG and occasional video service, of course, is that the location has access to a satellite that can also be received at the broadcast center. This could be a problem with coverage of very remote places, particularly in other countries. A double-hop arrangement would then have to be made with an international service provider such as Intelsat or PanAmSat.

4.1.1 Network Broadcast

The network broadcast to TV stations originates from the video uplink shown in Figure 4.1 (this particular feature is not shown in Figure 4.2 but could be added to one of the program contribution antennas). A typical network feed can last from a few hours per day, during evening prime time, to a full 24-hour service. The basic characteristic is that the feed contains network-wide programs intended for local rebroadcast by TV stations. U.S. commercial networks have begun transmitting separate regional distribution feeds of the same program but containing targeted advertising and other announcements. In North America, each TV station is separately owned and managed. There are also groups of stations under common ownership, including stations in major cities like New York, Chicago, Los Angeles, and San Francisco that are owned by the national networks themselves.

Network broadcasting in free-market economies has generally been supported by advertisers who pay to have their message put in front of the mass market. Well-known brands like Coca-Cola, Sony, Nestlé, and Texaco got that way through the medium of commercial television. As far as the sponsor is concerned, the most important piece of material to be transmitted is the advertising message. The rest is used as a draw for the audience, and the larger the audience the better. This is not meant as a condemnation, because all of us can enjoy much of the programming that is created with advertising dollars. Rather, we need to understand that the economics of network programming in the main are driven by the need to influence buying behavior (e.g., revenue to advertisers).

National brands can be advertised via the network feed, but much of the programming and advertising can be generated at the local level. In the United States, there has been a government mandate to encourage local control of broadcast TV. This means that the local TV station must include facilities to create as well as distribute program content, including news and commercials. Local sports events are popular, so the station needs to be able to collect sports feeds. This includes the situation when, say, the Los Angeles Dodgers go to New York to play the New York Mets baseball team. The L.A. sports announcers follow the team to New York where they occupy a broadcast booth at the stadium. Their coverage of the away-from-home game is carried by satellite (under special arrangement or as a network feed) back to the hometown, where it can be broadcast locally.

Local stations also receive compensation from the network with which they are affiliated. This source of revenue has declined as the networks have sought to reduce costs in the face of competition from cable TV and DTH. There have been a number of new startup networks in the United States, most notably the now established and highly successful Fox TV network and later UPN, which is backed by Paramount.

The money appears to be there in national advertising, but getting an adequate share of it is becoming more of a challenge.

In Chapter 5, we review developments in various forms of digital TV, including Digital Video Broadcast (DVB) and high-definition television (HDTV). U.S. TV networks have been participating in the technical development and political debate that concerns HDTV. It is the policy of the FCC that HDTV will be made available (which it is on a somewhat limited basis); however, the timetable was delayed in 2001 to allow more time for discussion and development of the market for home receivers. These are available at prices beginning around $1,000, although a really satisfactory projection system costs in excess of $5,000. As of September 1, 2003, approximately 1 million HDTV sets had been purchased in the United States.

In much of Europe, Latin America, and Asia, stations within a country can be controlled by the national network. Their primary support in the past has been from government funding, the motivation for which can be quite varied. As the premier government broadcaster, the BBC of the United Kingdom is acclaimed for the intelligence they put into the presented material. They have been free to use their judgment to decide what is best for the viewing public. Some of its programming makes its way to the United States to be enjoyed on public television and some cable networks. The BBC is under scrutiny to show that it continues to deserve subsidy from the U.K. populace. A similar situation exists in the United States with respect to PBS, which currently depends on U.S. government support. PBS is the national programming exchange and distributor for programming that is largely created by public TV stations. Ironically, PBS was the first to adopt satellite distribution at a time when the commercial network broadcasters relied on terrestrial microwave. Commercial TV thrives in the open economies of countries like the Philippines, Turkey, and now Poland. It is interesting to note that Turkey, now with its own satellites, is second only to the United States in the number of privately owned TV stations, and Polish households may actually watch more hours of TV than the people of any other country.

4.1.2   Cable TV

The cable TV industry was created by the simple need to improve TV reception in remote areas. Distant VHF and UHF broadcasts are received by high-gain antennas on mountaintops or tall towers, which can be shared by residents of a given community. In fact, the original name of the service was community-antenna TV (CATV), an acronym that is still in use. This type of business did not become an industry as it is today until the introduction of pay-TV services delivered privately to the cable TV system by geostationary satellite and on tape. The availability of a wider array of programming has made cable TV attractive to urban residents. Now, with a single cable access, we receive the local TV channels (including some that are just out of range or from nearby cities), national cable TV networks, pay-per-view (PPV) movies and events, and international programming as well. Video on demand (VoD) has been rolled out in some areas to allow subscribers to determine the starting time as well as the film.

The same cable plant is capable, with suitable upgrade, of two-way services, with the main application being Internet access using cable modems. This feature is

introduced by using the spectrum below 50 MHz (e.g., below Channel 2) on the cable and has allowed many cable systems to become very profitable while satisfying their customers' need for high-speed access. Facilitated by the Data Over Cable Service Interface Specification (DOCSIS) of Cable Labs, cable is possibly the most popular means of accessing the Internet at high speed. DOCSIS is also recognized as a global standard by ITU-T SG9. The downstream speed is approximately 1 Mbps and the upstream is typically more than 256 Kbps, making this service a benchmark of comparison for broadband communications to the subscriber.

4.1.2.1 Cable System Architecture

An example of the design and layout of a typical cable TV system is shown in Figure 4.3 [3]. The signals are distributed by a branching network composed of 75-Ω coaxial cable and wideband distribution amplifiers, which provide acceptable signal quality to homes. The topology of the network must be tailored for maximum efficiency of distribution and lowest investment and maintenance costs. The five major parts of this type of distribution network—head end, trunk cable, distribution cable, drop cable, and set-top-box—are reviewed next.

The *head end* is comprised of receiving antennas and equipment that provide TV signals to be distributed. Cable TV head ends include VHF and UHF antennas and C- and Ku-band satellite receiving dishes, along with the required number of TV receivers to demodulate the channels and transfer them to the cable network. For analog TV, each channel is modulated onto a separate frequency for transfer to the cable itself. The channel plan (e.g., specific placement of TV channels in the bandwidth of the cable) is designed to maximize capacity. This is aided by not having to avoid particular frequencies in the RF spectrum that were not assigned to TV stations. This is possible because of the electrical shielding on the cable that suppresses radio energy from the outside.

Trunk cables bring the collection of channels from the head end to the neighborhood to be served, usually on a point-to-point basis. *Amplifiers* are introduced at appropriate distances to maintain adequate signal strength and S/N. If the distance to be covered is extremely long, it may be preferable to use a terrestrial line-of-sight microwave link or fiber optic cable. Another use of trunk cable is to allow multiple head-end sites to connect to a common equipment room that houses the video receivers and distribution amplifiers.

Distribution cables (or feeder cables) branch out from the terminating point of the trunk cable and run past the homes in the neighborhood. There would be many distribution cables needed to cover a large community. This tends to maintain a better signal quality in terms of S/N and distortion because the video signals do not have to pass through as many individual amplifiers. Also, the failure of one feeder line or amplifier only affects a relatively small group of homes. Problems and power outages along the distribution cables are the principal causes of service degradation and interruption.

Drop cable to the home connects from the distribution cable to a tie point akin to the telephone box or electric power meter connection. From the home connection, internal coaxial cable is used to bring the signals to the TV set-top box. Cable TV systems differ in the number of channels that they are capable of carrying. This
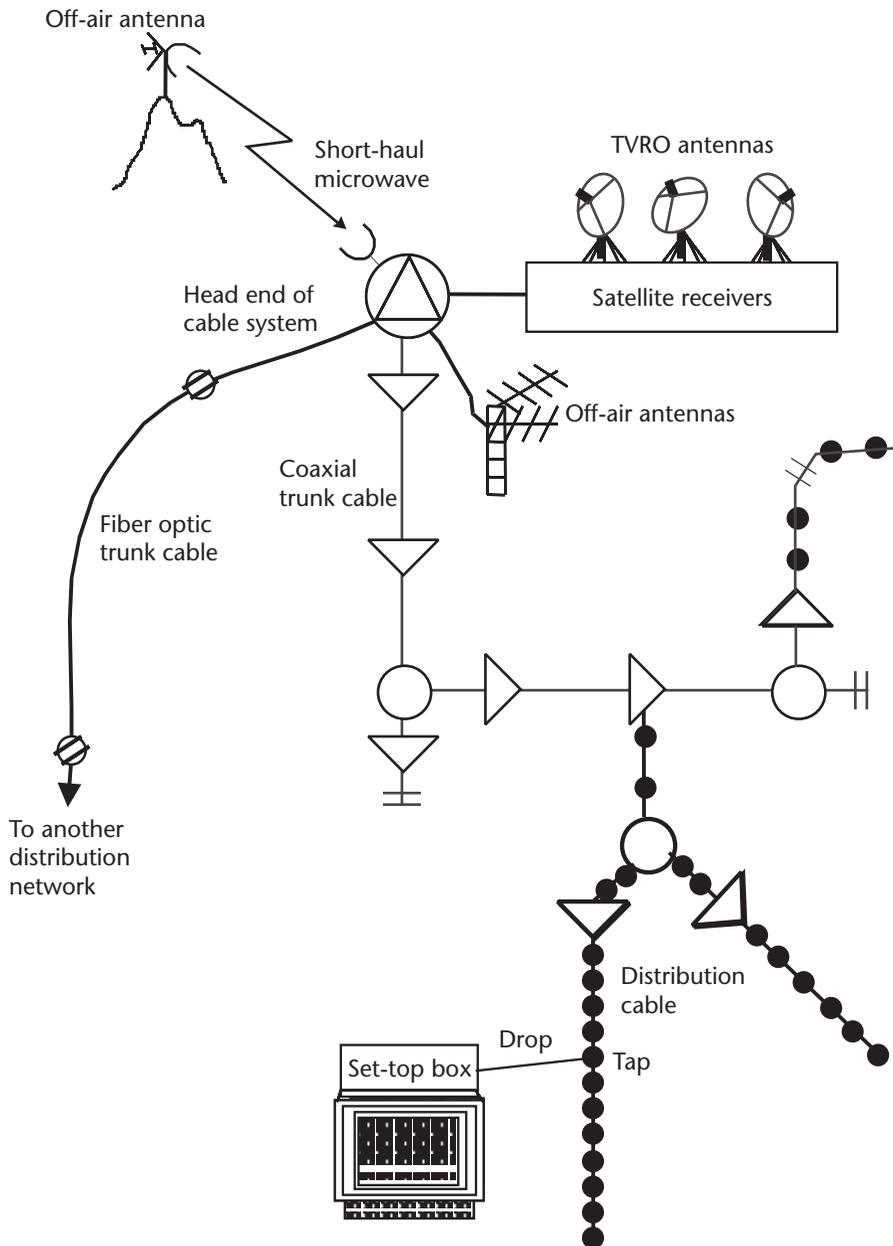
Figure 4.3   Typical layout of a cable TV distribution network within a local community.

depends, of course, on the bandwidth of the cabling plant and amplifiers as well as the tuning capability of the set-top box. The lowest capacity systems deliver about 50 channels as they only operate in a 350-MHz bandwidth, between 50 and 400 MHz. These systems are still found in rural areas and in small cluster communities. In developing countries, such a capacity would be viewed as substantial, so a design with this limitation would still be very attractive on a near-term basis. However, the clear trend is toward systems with 100 channels of capacity over a single cable, occupying a bandwidth of 500 MHz. Doubling the capacity to 200 channels is done

by adding a second cable so that each home is served with a pair of cables. What is attractive about the dual cable is that it can be upgraded for two-way communications.

In modern plant design, trunk and distribution coax are replaced by fiber optic cables and optical repeaters, as this technology is more reliable, lower in loss, and more cost-effective on a long-term basis. The first step is called hybrid fiber coax (HFC), by which the video from the head end is digitized and routed around the metropolitan area on the fiber backbone. The actual distribution to neighborhoods would continue to use analog coax. The HFC approach also allows the cable operator to share telephone company fiber plant, although the video codecs, which are very costly, have to be added. The next step in applying fiber optics is called "fiber to the curb," where the fiber extends through the distribution cable but not to or into the home. The transmission over the fiber distribution cable could either be analog or digital. This is a natural extension that leaves the expensive and difficult last step to new construction and future upgrades. A converter at each tie point transfers the channels from lightwave to the loop coax cable. This uses many of the benefits of fiber without having to replace existing loops and set-top boxes.

A *set-top box* (or consumer electronics unit) allows the subscriber to select particular channels for viewing and to control which channels are available under any particular service plan or option. The basic set-top box merely contains a downconverter that can translate one channel from the available spectrum to, say, channel 3 or 4, where it can be picked up by a standard TV set. This duplicates what goes on in a TV tuner and VCR but supports a more efficient spectrum arrangement than is used over the air. The next section discusses the evolution of this important element of cable TV.

4.1.2.2   Set-Top Box

The set-top box (STB) was originally introduced to provide a convenient interface between the cable drop and the home entertainment system (TV, VCR, surround sound system, DVD player, hard drive, and so forth). In addition, the STB improves quality through higher received signal level. Channel filtering may also be more effective. This is because the channels are adjacent to each other on the cable, whereas over-the-air broadcasts allow more guard band between channels within a common broadcast area. This scheme was adhered to in the United States so that the selectivity of the standard TV tuner could be relaxed [4]. The box also screens out over-the-air radiation that could interfere with cable signals. Cable signals are typically delayed in time due to the lower speed of propagation over cable and longer path as compared to the air medium. Functionality for descrambling and conditional access, discussed in Chapter 6, is incorporated for controlling service. Data can be delivered over the cable to the set-top box, either as a separate channel or through the vertical blanking signal of the video. The latter is limited by the speed and capacity that can be delivered.

The design and function of the digital STB box is shown in Figure 4.4. This model is in many ways similar to the DIRECTV Satellite System (DSS) used by DIRECTV in the United States and the European DVB standard. The video is digitized at the source and distributed over the satellite or terrestrial network. The
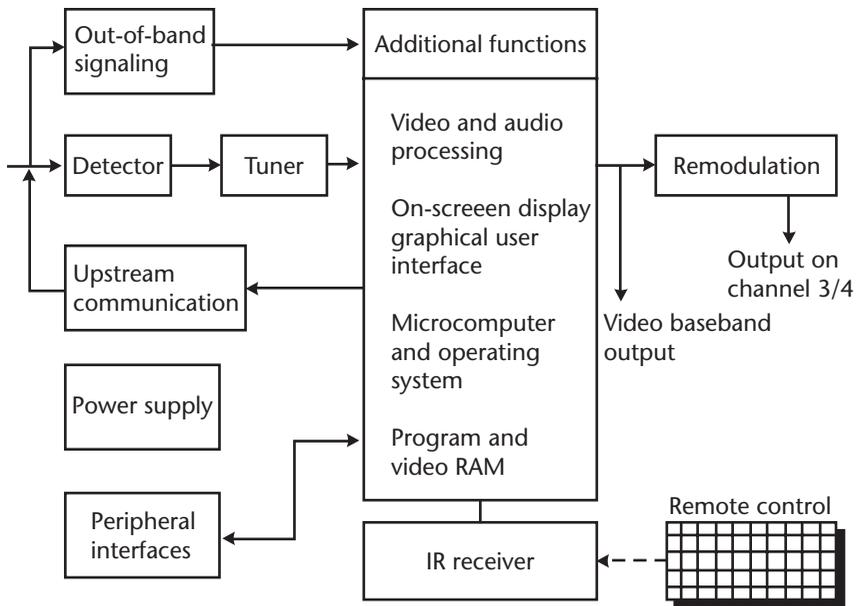
Figure 4.4   The next generation of the TV set-top box, including addressable features and digital decompression. An interface to an external device such as a PC is provided.

electronics that are needed to perform the decoding and demultiplexing functions are available through wide adoption of the MPEG 2 standard (discussed in Chapter 5). Produced in mass quantities using standard chips, the digital cable set-top box is used for one-way delivery. The additional functions in Figure 4.4 include the on-screen program guide and menus that allow the subscriber to customize the service. Two-way services such as data communication and telephone are usually introduced using a second cable that allows upstream communications. Digital sound is offered in the same manner, with the number of channels limited only by the availability of source material (one video channel is capable of supporting 50 stereo audio channels). A popular option in the United States is to include a hard drive that allows the subscriber to record one episode or all within a prescribed period. In this manner, the scriber gains full control of viewing, allowing the delay of a particular broadcast (to leave the room) and time-shifting for personal scheduling convenience. Introduction of HDTV into cable has begun in the United States and certainly impacts STB design. As will be discussed in Chapter 5, HDTV demands nearly 20 Mbps per channel but can be contained within the nominal bandwidth of a conventional analog TV signal.

4.1.2.3   Signal Quality and Security

Signal quality at the subscriber end is intended to be better than what reaches the typical home through over-the-air broadcasting. This, of course, depends on how well the home would be able to receive the TV signal with a roof antenna of some kind. The NTSC, PAL, and SECAM standards (defined in Section 4.4.1) have been around for a long time, and their quality is reasonably well understood. In 1959, the

U.S. Television Allocations Study Organization (TASO) devised a five-point scale to allow viewers to rate the quality of different levels of noise in the picture. The scale is described in Table 4.1, indicating a six-level scale that covers the range of (1) Excellent to (6) Unusable. In the context of cable TV, viewers expect to experience a level of fine to excellent—that is, TASO grades 2 and 1. Cable generally must deliver a S/N of 46 to 47 dB. The empirical data presented in Table 4.2 would suggest that this level of S/N would meet TASO 1 or 2, meaning that the picture is fine to excellent. This would be the value that arrives at the home receiver and therefore includes all contributions from the source material and origination processing, satellite link, Earth station, and cable plant. This approach was taken much further by the ITU in the creation of the ITU-R BT.500 series of recommendations. Further information on the digital aspects of BT.500 is reviewed in Chapter 5.

Distortion must be within acceptable limits, which is one of the more challenging aspects of cable design for analog TV. This is because the amplifiers along the cable each add intermodulation noise to the total distortion budget. A summary of noise and distortion requirements for analog cable plant design is provided in Table 4.3. These analog parameters make sense when the satellite and cable distribution systems provide an analog transmission channel. When we move to the digital format for everything but the TV receiver itself, it no longer makes any sense to view the problem in this manner. As discussed in Chapter 5, the process of digitizing and compressing the analog TV signal introduces impairments that cannot be removed. Based on human subjective evaluation, the viewing quality is equivalent to or better than what the standard analog cable system is capable of delivering. With the digital set-top box, all of the impairment is produced at the sending end and the resulting picture would be the same even if the encoder were directly connected without the satellite link and cable in the middle. The key parameter for the link itself is bit error rate as opposed to S/N. If the BER is less than about one error per 10 million information bits sent (10^{-7}), picture quality will be completely stable and acceptable. This will not be the case if more errors are produced than this threshold value.

Security in conventional cable TV systems is needed to prevent unauthorized viewing. This comes into play when the cable-operating company offers different levels of service: a basic level, which includes local TV stations plus 5 to 10 basic cable networks like CNN, TBS, ESPN, and the Weather Channel. The next level might add 10 more basic services, like The Discovery Channel, Arts and Entertainment (A&E), and MTV. There might even be a third level with basic services that

Table 4.1 The TASO Picture Quality Rating Scale Used in Subjective Evaluation of Video Signal-to-Noise Ratio

Number	Name	Description
1	Excellent	The picture is of extremely high quality—as good as you can desire.
2	Fine	The picture is of high quality, providing enjoyable viewing; interference is perceptible.
3	Passable	The picture is of acceptable quality; interference is not objectionable.
4	Marginal	The picture is poor in quality, and one wishes they could improve it; interference is somewhat objectionable.
5	Inferior	The picture is very poor but watchable; definitely objectionable interference is present.
6	Unusable	The picture is so bad, one could not watch it.

Table 4.2   The Relationship Between Measured
S/N of the Video Signal and the Viewing Quality

Video S/N (dB) (Weighted)	Viewing Quality
40	Snow objectionable in picture
42	Snow clearly visible
44	Snow just perceptible
46	No snow visible

Table 4.3   Video Quality Objectives for Analog Cable Plant Design (NTSC Video Standard)

Parameter	Abbreviation	Value
Signal-to-noise ratio	S/N	46.0 dB
Composite second-order distortion	CSO	−53.0 dB
Composite triple-beat interference	CTB	−53 dB
Signal into TV receiver		1 mV 0 dBmV

adds a regional sports channel and a comedy network like Comedy Central. Higher charges are incurred if the subscriber takes the premium (i.e., pay TV) channels, which do not carry advertising and therefore are paid for exclusively with sub-scriber revenues. The cable operator must be able to restrict the watching within the same home to allow only the subscribed-for channels to be available.

There are two security techniques in popular use, both of which are very low in cost. The cheapest approach is to insert a jamming waveform on the particular fre-quency channel and provide the paid subscriber with the ability to remove the signal within the set-top box. This type of protection is very poor since the means to evade it is very simple in electronic terms. The next in cost is the use of a negative trap, which is a band stop filter that removes or distorts a specific channel or a contiguous band of channels. The negative trap filter affects all TV sets in the same household since it is placed in the incoming line.

The direction that the industry is taking in security is to use signal scrambling and addressability of the STB. This is the same technique pioneered on satellite TV networks. The scrambling is accomplished either by (1) suppressing the synchroni-zation signal, which is needed in the TV set to lock on to the (analog) scanning waveform; or (2) baseband scrambling, which modifies the video and audio in a pseudorandom manner, requiring the set-top box to have prior knowledge of the random pattern of scrambling. Both of these approaches are combined with addressability, which allows the cable operator to activate a given channel on the set-top box. Upon customer order, the cable operator transmits a unique code over the cable that causes the set-top box to descramble the channel in question.

4.1.2.4   Cable Service and Programming

Cable subscribers pay a monthly access fee for use of the infrastructure. Rights to provide this type of service are usually granted by the local community government

as a monopoly franchise. In recent years, the monopoly privilege has come under attack as new entrants wish to offer a wider array of services, including interactive video and data. This aspect will be covered later in this chapter.

There are two classes of cable TV channels: advertiser-supported and pay (also called subscription or premium). These channels were selected based on their viewership and general appeal. An advantage of an advertiser-supported channel is that, in theory, the channel can be transmitted in the clear without scrambling. The only issue is to provide a measure of the audience size, which will satisfy potential advertisers. In reality, advertiser-supported channels must be scrambled to protect the copyrights of producers and owners of the programs. The producer charges the programmer such money for the right to distribute the movie to a specific market (say a country or even a city). Reception in a nonauthorized region must be curtailed. Also, some ad-supported channels, such as CNN and ESPN, are able to charge small but significant subscriber fees to defray some of the expense of producing programs. On the one hand, sports programming is less expensive than dramatic series to produce, but expensive on the other hand due to high royalty fees paid to professional sports leagues and universities.

Typical advertiser-supported channel formats include:

- *News (24-hour):* This can be either a general-interest service like CNN, MSNBC, Fox News Channel, and Sky News, or one that focuses on a particular aspect such as the Weather Channel or CNBC (financial news).
- *Sports:* This may be 24-hour like the news channels or limited in time to correspond to when certain types of sporting events are held. ESPN is by far the most popular sports channel in the world and happens to have the largest viewership in the United States as well. European sports networks include Sky Sports and Eurosport.
- *Super stations:* Originated by Turner Broadcasting as WTBS, a super station is principally a local TV station that is redistributed by satellite throughout the country or region. This is more of a novelty in the United States where TV stations are usually restricted to local service. In most countries, there are national stations that rebroadcast the same signal nationwide and, hence, by this definition, are already super stations.
- *Movie channel (old release):* Theatrical movies have long been a popular form of TV programming. The satellite-delivered cable channel dedicated to movies is popular as well. Through advertiser support, the subscriber does not pay an additional fee for the service but must endure the commercials. Examples of 24-hour movie channels that are advertiser-supported include TNT and USA, which are general interest, and American Movie Classics (AMC) and Bravo, which choose to display older films that have enduring appeal. Similar services are available in Europe through Sky and in Asia as part of the StarTV package.
- *Science and other special interest:* Cable TV has long had the promise of encouraging channel formats directed at selected audiences. This particular category is potentially of general interest but would not survive in over-the-air broadcasting except possibly in government-supported educational programs. The Discovery Channel (TDC) first appeared in the United States in 1984 and was an instant success. It selected program material that had originally

appeared on PBS. However, the difference here is TDC is advertiser-supported while PBS is not. TDC has since grown by adding new channel formats to focus on home improvement, travel, and nature. Another immensely popular special interest channel is Arts and Entertainment, a cable network offering cultural programming from the BBC, PBS, and internal production sources, augmented by reruns of popular TV network series like *Law & Order* and *NYPD Blue*.

- *Home and leisure:* This is a new genre, appearing and gaining popularity around 2000. Again a spinoff of the PBS network, home and leisure provides 24-hour access to programs relating to the home itself (Home and Garden Television, HGTV), cooking and food (The Food Network), travel (The Travel Channel), and home projects that almost anyone can do (Do It Yourself Network, DIY). Ziff Davis brought us ZDTV (now TechTV) in the late 1990s as a prime source for cutting-edge information about PCs and the Internet.

The advertiser-supported cable TV channel represents an interesting opportunity as a new venture. This is because there is literally an unlimited number of possible formats and subsequent subscriber markets. Any of those described previously can spin off new formats (like DIY from HGTV) and totally new ideas may sprout from the creative mind. The cable medium allows a new channel to reach a reasonably large viewer base that is hungry for new entertainment. There are a number of issues that a new entrant must address before attempting to get started in what has become a very competitive field. The programming idea itself is not the most difficult part. Rather, the first problem is startup capital to acquire the transponder space, uplink facility, and studio. This must be available 24 hours a day. The staff to run these facilities must be familiar with the business and able to manage this type of operation on a very professional basis. As a way to overcome this hurdle, AT&T Broadband and Ascent Media established media centers in Los Angeles, California, to allow a startup to get on the air with no direct investment.

Viewers will not accept anything less than the high production values that are standard in cable and over-the-air broadcasting. The other side of the coin is that conventional TV channels have more general appeal and can draw very large audiences. Their greater access and viewership means that they obtain higher advertising revenues from national advertisers like Coca-Cola, General Motors, and McDonalds. This then provides the money needed to produce the most attractive programming that, not coincidentally, makes the channel more attractive than cable to advertisers. Narrowly focused channels can exist on cable because some metropolitan areas have relatively high concentrations of specific viewers. The Chinese Channel, for example, is viable within major cities like Los Angeles, New York, and San Francisco, where the Mandarin-speaking population totals more than 1 million. Cable systems that serve these communities would offer the Chinese Channel, which delivers potential customers to a specialized type of advertiser.

Another major challenge that the new entrant faces is gaining access to the subscriber base. Cable subscribers cannot on their own cause a new channel to appear on their cable system. Rather, it is the cable system operator who controls technical access and the purse strings as well. In many developed countries, cable TV systems

are often massed under a national operator or holding company, called a multiple systems operator (MSO). This means that the national operator controls a substantial quantity of potential subscribers and hence can make or break the new programmer. For this reason, a new service will have to make it very attractive to the MSO by doing such things as allowing them to keep any subscription fees that are collected plus a portion of the ad revenue. It is even not without precedent for the fledgling programmer to have to give up some of its equity in order to reach the subscribers. Having gained access, the programmer can then approach the advertisers to start the flow of ad revenue.

Premium or pay channels are not advertiser supported but rather are paid for entirely by the subscriber through a monthly fee. The amount varies from as little as $5 to as much as $15 per month, depending on the quality of the programming and the ability of the subscriber base to pay. Delivery of the channel must be controlled through one of the security techniques described earlier in this chapter. The need for this was verified in the early 1980s with the development of the backyard dish industry in the United States. In that particular case, homeowners were encouraged to spend between $3,000 and $10,000 for a C-band receiving system because they could receive free cable channels, including the pay services like HBO and Disney. This was possible because none of the services were scrambled at the time. After HBO and others began scrambling (around 1986), backyard dish sales nearly halted and this new industry almost collapsed. The advent of Ku-band DTH services with their inherent quality and capacity has propelled the home dish business to new heights, and except in developing regions the "big ugly dish" (BUD) of the 1980s is declining.

Movie channels are clearly the most viable form of premium channel as they can attract a large audience and revenue. For HBO, the world market leader, premium service means gaining a significant percentage of cable subscribers. Not all subscribers, however, will choose to pay the significantly higher cost of this service, but for those who do, it delivers a continuous flow of movies that include at least one new hit a week as well as innovative series and special programs that cannot be found elsewhere. Leaders in the premium channel business also produce their own movies and specials. HBO, for example, is a credible movie studio of its own, having produced many movies for its cable services and for the box office as well. They are also in the business of producing special events such as musical concerts featuring famous stars. These are shown on the premium channel and find their way to other outlets such as theatrical movies and advertiser-supported cable channels. Examples of the series include *Band of Brothers*, an award-winning drama series about a real infantry company in the 101st Airborne Division during World War II, and *The Sopranos*, an extremely popular series about a modern-day mob family that combines humor with gore and obscenity (allowed because of the "private" nature of the broadcast).

Movie channels cost a lot of money to produce because of the high cost of acquiring the material and the need to collect revenues from every watcher on a monthly basis. The cable provides a degree of security itself with the cable operator taking responsibility for collecting the revenue. They keep a percentage of what the subscriber pays for the cable network service.

The cost of acquiring and/or developing programming for a pay TV channel far exceeds anything else, including the investment or operating cost of satellite

transponders. The purchase of recent-release movies dictates whether the channel can survive or even make money. In the early 1990s when British Satellite Broadcasting (BSB) was competing with Sky TV for market share and for movie material as well, the cost of programming skyrocketed. This severely weakened BSB, who subsequently merged with Sky to form British Sky Broadcasting, which is controlled by News Corp. (the force behind Sky).

The Walt Disney Company entered the subscription cable TV business in the early 1980s and achieved success with their Disney Channel. This concentrates on children's entertainment and family-oriented (Disney) movies. It is a premium service without advertising, something that many parents appreciate because their children are not bombarded with ads for toys and junk food. The Disney Channel is successful but has not reached the market size of HBO. Other Disney activities in TV include the production of TV series, game shows, and movies, as well as the sale of box office hits on videocassette. In 1996, Disney made its move into mainstream TV broadcasting by acquiring ABC with its network, TV stations in leading markets, and ESPN.

The final category of premium service is pay-per-view, a system for delivering single events to the subscriber for a specified one-time charge. PPV was introduced as a form of closed-circuit transmission of sporting events, particularly professional boxing. To view the event, customers purchase a ticket and then go to a designated viewing location such as a theater or bar to witness the event live on a screen or large monitor. Over time, the cable networks and cable TV systems figured out how to use the cable distribution system and the set-top box to permit PPV in the home. A popular PPV event may draw greater than 15% of the available subscribers [5].

Recent release movies were added in the mid-1980s over the first PPV network, Viewer's Choice. An issue at the time was the release "window" that defines the delay, measured in months, between when the film is first shown in the United States in regular theaters and when it is made available through cable's PPV facility. Tied to this is the release window for DVD and videocassettes. While the movie studios and distributors do not make this kind of information readily available, the sequence seems to be the theatrical release in the United States and other global markets, followed within 3 to 6 months by DVD and videocassettes, followed by PPV and then premium cable channels. In 2002, the DVD window has been reduced in some cases to little over 1 month due to its popularity. All three of these media provide the studios with a nice boost of revenue, where the ultimate viewer pays by the increment and the studio gets a discernible share. The total period is approximately 9 months, give or take a few months.

Not long after the U.S. theatrical release, the film is distributed to theaters around the world. Non-U.S. exhibition of films has become an important source of revenue to the studios, which may explain why the DVD, cassette, and PPV release windows can be delayed as much as another 6 months. The key point in all of this is that the studios and distributors optimize the release timing to obtain the maximum amount of revenue from a given movie in the shortest possible period of time. PPV networks, then, must compete with DVD and cassette rental and purchase, as well as the successive showing of the same movies over the premium cable channels, the latter obtaining their movies perhaps 2 months later. Not all movies find their way to PPV, for reasons that only the studios know.

There are a number of variants of PPV that are generally recognized. In its most primitive form, the subscriber must reserve the particular program well in advance and the service provider then provides a special access device such as a descrambler or inverse trap to be connected prior to viewing. An improved form results with an addressable set-top box, which allows the cable operator to activate the show remotely. The subscriber must still make an advance reservation over the telephone, anywhere between 1 month and 1 hour in advance of the event or show. The ordering and setup of program delivery is entirely manual.

The final variant is called impulse PPV (IPPV), which is made possible by the modern set-top box and some type of interactive connection between the subscriber and the cable system operator. The simplest scheme uses a telephone hookup whereby the box automatically dials the operator to request the PPV event. This can be done literally within minutes of the event. Interactive cable eliminates the need for the telephone call because the request is made using a data communication network. Alternatively, the set-top box can authorize and then descramble the show without going back to the cable operator. Subsequently, the box calls back to report an aggregate number of PPV viewings. This particular technique is used in DTH-delivered PPV, which is discussed in Chapter 6.

Vendors of cable equipment have found it difficult to implement true IPPV in many systems and so are offering variants that come under the category of "near" IPPV. The problem with this is, what exactly is near IPPV? One approach is to broadcast the same movie on several channels at the same time but to stagger the start times by between 20 minutes and 1 hour, a technique called *multiplexing*. For a 90-minute movie, for example, this would require either three or four channels. The viewer would only have to wait a maximum of 20 or 30 minutes to the start of the movie (the average wait time would be half this amount). Access control to this can be done by the set-top box without intervention from the cable operator. The movies could be delivered all the way from the cable TV network, there only being the requirement that several simultaneous channels be used over the satellite. As we move to digital compression and transmission, this multiplexed form of IPPV will no doubt become more affordable and popular, particularly for STBs with hard drives.

VoD, introduced earlier in this section, is a scheme that employs a suite of disc players and management equipment at the head end. Through a two-way cable system, a special STB at the subscriber can review what movies are available and initiate the request to start. The playback equipment needed to originate the movies can be located at the cable head end or studio. This could be attractive for a large system where the usage justifies the investment and operating cost. The more common but less interactive form is to have the origination point at the uplink to the program distribution satellite.

4.2   Educational TV and Distance Learning

Satellite delivery of educational TV programs and courses is a relatively small niche in the overall business of video distribution. While it got its start in the United States by the terrestrial medium of microwave transmission, its greater appeal is in the developing parts world where modern teaching and medical processes cannot

otherwise be offered to remote regions. One must also consider the impact of the Internet on learning, and in fact on-line training is likely to surpass the broadcasting approach. Satellite service can support both of these styles of distance education, and it is possible that a proper blend is going to produce the best results.

4.2.1   University Distance Education

Stanford University developed the first terrestrial closed-circuit education TV network using the instructional television fixed service (ITFS) frequency assignment at S-band [6]. This special allocation by the FCC allowed the Stanford Instructional Television Network to serve working professionals in the San Francisco area. Similar networks were created by UCLA, USC, and the California State University (CSU) system. Several universities in the CSU system followed suit, as California State University at Chico (Chico State) began to serve Northern California and California State Long Beach concentrated on Los Angeles and Orange Counties in Southern California.

Adding the interactive feature is feasible and, in fact, is done in a number of installations. The simplest and least expensive approach is to use a dial-up telephone connection and what amounts to a speakerphone. The incoming calls to the studio are bridged in a conferencing unit to allow all sites to hear the question and answer. Since the instructor is literally blind as to what is going on in any particular remote classroom, it is useful to include an indicator light to show from where a question or comment originates. More sophisticated conferencing systems include a data channel, which allows several useful features. One or more of the following facilities can be included, depending on the nature of the instruction:

- A readout that indicates the source of a question or comment;
- A mechanism to collect responses to multiple-choice questions from remote classrooms, useful for measuring the effectiveness of the teaching;
- Ability to open up (or close) the return sound channel on an individual classroom basis;
- A forward and return graphics capability to allow students to present their ideas (instead of a return video channel, which is usually inconvenient and prohibitively expensive as well); when necessary, this can be accomplished with standard fax machines;
- A computer networking function to allow exchange of text or files.

Another education network is operated by an organization called the National Technical University (NTU), which has neither a campus nor its own instructors. All of its classes are drawn from existing universities around the United States. NTU initiated operation with analog video but quickly switched to digital compression by adopting the spectrum saver system from CLI. Professors from more than 50 member universities offer courses leading to the M.S. degree in 14 fields, including business administration, chemical engineering, computer science, electrical engineering, engineering management, environmental systems management, management of technology, manufacturing systems engineering, materials science and engineering, mechanical engineering, microelectronics engineering, software engineering, and

systems engineering. Fifteen of the universities that support NTU have studios and uplinks, and more than 300 courses are offered each semester. In addition, there are literally hundreds of professional development courses. Delivery methods include digital satellite broadcast to closed receive points, video tape, CD-ROM, and Web/Internet. The thousands of students who attend NTU are provided with pre-pared materials ahead of class, and many of the classes are recorded on tape for closer study and first-time viewing when classes are unavoidably missed.

An entirely new style of university distance learning was introduced by the School of Engineering of the University of Wisconsin–Madison. Recognizing that practicing engineers in all disciplines desire a professional Master's degree but can-not easily make themselves physically available for class, UW–Madison developed the Masters of Engineering in Professional Practice (MEPP). Unlike other Masters delivered by TV or correspondence, the MEPP program combines an annual week on campus with a well-disciplined remote-learning course schedule throughout the year. Students may be literally anywhere in North America or the world, having to attend a weekly teleconference and to accomplished their lessons through the medium of the Internet (even some Madison residents have been participants). The teleconference combines audio with the use of a PowerPoint presentation pushed through the Internet to each attendee. All the class-members need is to dial into a standard audio bridge to hear the professor and classmates, and simultaneously con-nect through the Internet to the UW presentation Web site. By 2003, the program had graduated three classes, and the evaluations have been excellent.

The experience with university distance education has demonstrated that satel-lites and telecommunications can support viable programs that meet the needs of modern-day students. In time, the best features of TV, telephone, and the Internet will merge into a powerful medium that will expand rather than contract what the best universities have to offer.

4.2.2 Corporate Education and Interactive Learning Networks

Distance learning has an important place in business, both to facilitate the sale and service of products and as a revenue generator in its own right. One of the best examples of this was the Interactive Satellite Education Network (ISEN), operated for IBM by Hughes Communications between 1983 and 1993. It contained all of the features in the previous list, employing leading-edge satellite technology (of the time) for both directions of transmission. As shown in Figure 4.5, it included four instruc-tor studios and 20 classroom sites around the continental United States. A typical video receive site (which can transmit voice and low-speed data as well) can present all four classes at the same time. A typical classroom can hold 16 attendees and has two monitors to display a live instructor along with a transparency (or, alterna-tively, any combination of these and a 35-mm slide, computer graphic, or video tape). Note that this was before the day of PowerPoint and the Internet. The basic arrangement of these two facilities is shown in Figure 4.6. Attendee access to the return channel is through a *student response unit* (SRU) on each desk, which con-tains the microphone, activation switch, indicator light (showing if the instructor has put this position "on the air"), and a set of five radio buttons to allow each stu-dent to indicate a selection to a multiple-choice question from the instructor. The

Figure 4.5   Geographic locations of studio and student receive sites in the Interactive Satellite Education Network.

SRU approach has been adopted by many other networks now using the PC and mouse and can be applied even if the return channel is over a terrestrial network rather than the satellite.

The success of each class and the network as a whole depend on how well the service is organized and the resulting impression this makes on students. The IBM approach was to have a qualified ISEN specialist at each location that would assist the students with administration of the class, local problems, and equipment operation. Depending on the frequency of use of a particular location, the administrator's role could be a full-time job, a part-time job, or an additional assignment. Management of the instructor studio location is even more critical, as will be discussed later in this section.

Regarding the ISEN technical design, IBM set a very high objective for class availability. It was the view of management that an outage at one site in 20 would cause a delay or cancellation of the entire class. To minimize this possibility, C-band was selected for its lower incidence of rain fade, particularly in the eastern part of the United States. One 36-MHz transponder on the Galaxy 2 satellite at 74° WL was sufficient to carry the four video carriers from the studios and the 20 audio/data carriers from the remote sites. This arrangement is indicated in the spectrum plot in Figure 4.7. Each video required a full T1 and used discrete cosine transform (DCT) digital compression (discussed in the next chapter), supplied by NEC. Video was therefore near-broadcast quality, desired for the premium nature of the service to IBM customers. ISEN found additional use as a medium for announcing IBM's new products to the media and customer community. The audio/data carriers time-share the frequency slots (shown by the thin carrier lines) while the video carriers, each of which transfers two TV channels at 1.5 Mbps each, are constant. The display also

Figure 4.6   Configuration of studio uplink and one remote ISEN site.

Figure 4.7   Spectrum analyzer frequency display of the ISEN studio outbound carriers; the narrowband carriers are for the inbound audio and data from the remote sites.

shows the "humps" of intermodulation noise in the transponder, which result from operating the transponder as close to saturation as possible.

ISEN met all of its technical and operational requirements during its 10-year lifetime. Several of the Earth stations were located away from the classroom site in order to avoid terrestrial interference. Most of this backhaul transmission was obtained from the local telephone company in the form of multiple T1 circuits—private microwave was used in one instance since T1 circuits were not available at the time. The network achieved very high availability, usually in excess of 99.95% for class delivery, because of the excellent propagation characteristics of C-band. In fact, the majority of outages were due to equipment failures and interruptions of the terrestrial links between the classroom sites and Earth stations.

The ISEN approach has found its way into literally every interactive distance learning system. Such networks that employ satellite, dedicated terrestrial lines, or the Internet have adopted the SRU approach (typically done on a PC rather than a separate device), and a display of presentation slides and possibly the instructor as well. The latter is often limited due to availability of consistent bandwidth through the channel of the Internet. In time, we would expect to see better transmission services appear at reasonable prices using facilities such as VSAT networks, described in Chapter 9, and videoconferencing with the H.323 standard, discussed later in this chapter.

4.2.3   Guidelines for Effective Distance Learning

Creating the education network is relatively straightforward. However, more challenges lie in employing the technology effectively. The following guidelines can improve the effectiveness of distance learning:

- Introduce the instructors to the network before they are expected to go "on camera." Provide a training session so that they understand how their presentation style must be modified to the requirements of the new medium. Begin the indoctrination early by bringing the instructors together with the administrators who will operate the network. Discuss any special arrangements that are needed on both sides, such as scheduled break times, delivery of printed support materials, attendance recording, and makeup of the student body.
- Reassure instructors that they can adapt their teaching styles to meet the new medium of television. The lack of visual feedback from students will give some instructors difficulty, but this can be overcome after a few exposures to educational TV. Audio return and interactive data can give a reasonable facsimile.
- Demonstrate all of the facilities ahead of time so that instructors become comfortable. Give them time to practice with people at the other end who can provide feedback. An effective way to practice is to include small groups of peer instructors who teach in the same field. This reduces some of the uncertainty in the dialog, which must traverse a considerable distance.
- Provide each new instructor with a mentor or course development expert who can assist with the mechanics of creating the course. This kind of help removes

much of the uncertainty in the instructor's eyes and provides insurance that the course will be developed consistent with the overall principles and style of the program.

- Plan and schedule the classes so that there are no surprises on the day of the class. Make instructors and administrators aware of the schedule and procedures that must be followed. Typically, the equipment and satellite capacity will need to be scheduled ahead of time, and there will be little if any flexibility. Supporting materials must be available at the time of the class and hence will have to be delivered ahead of time (by e-mail, mail, or overnight delivery service).
- Communicate to all participants the expected benefits of using this medium rather than focusing on its limitations. Promote the system and the session so that there is maximum chance for success.

Here are some additional considerations and concepts:

- Current best combination: video outbound, voice and data inbound;
- Observation: satellite-delivered education is a marriage between education and live TV;
- Requires TV production skills as part of management and support;
- Careful design of "telecourse" content as well as preparing the instructor to deliver "on air";
- Graphic aids are important: must have better quality;
- Instructor guidelines:
  - Ability to effectively communicate, relate to, and interact with their students, and have high level of subject-matter expertise;
  - Credibility in the eyes of the target student group;
  - An open mind and willingness to learn how to apply this medium;
  - Have had at least one successful on-camera experience (this can be in the form of a course pilot in front of a friendly group).

4.3   Business TV

Entertainment and educational TV provide the foundation for the business application of the video medium. Businesses can employ private broadcasting, which relies on the point-to-multipoint nature of the satellite delivery medium, and video teleconferencing (VTC), which uses point-to-point two-way links to add the visual element to the standard interactivity of voice telephony. These techniques are used widely in the United States, Europe, and leading Asian nations, although their growth has been restrained by the cost and complexity of operating the equipment and arranging private broadcasts, digital long-haul circuits, and VTC events. The U.S. military services and some government agencies are heavy users of VTC due to the immediacy and security needs of their communication.

There was an expectation for more rapid adoption of this technology in 2000 because of lower equipment prices and the next generation of digital satellite and

terrestrial networks that overcame many operational limitations. However, limitations on bandwidth and costs that are still high (in comparison to e-mail and telephone calls) have hampered more widespread use. Adoption will increase with available bandwidth, as equipment and software suppliers reduce prices to garner larger markets. Compression and network operating standards also play a role, as will be discussed later.

4.3.1   Private Broadcasting

Private broadcasting is no different technically from video distribution and one-to-many educational TV. The originator of the program uses a TV studio and uplink to create the broadcast, and the signal is received at multiple sites that have simple TV receive-only antennas and electronics. The broadcast is viewed on TV monitors in conference rooms and, ultimately, the desktop. Private broadcasts are often scheduled and may be employed almost daily in the routine of business. Examples include:

- Announcements of new product introductions and marketing campaigns, in heavily marketing-oriented organizations like Frito-Lay and Microsoft;
- Distribution of financial or critical business news that can impact the company or its customers, which is popular in the financial services industry for stock brokers like Merrill Lynch and large investment banks like CS First Boston;
- Instructions on product application and display for merchandising for retail store chains like Wal-Mart and Sears;
- Public relations–oriented communication between the government and the press;
- Instructional information and product guidance for representatives and dealers, by major manufacturers like General Motors and IBM.

Any of these applications could justify a dedicated private broadcasting network that operates daily or even several times per day. If the need is less frequent—perhaps once per week or month—then the network can be put together on an ad hoc basis by renting the studio. This reduces the capital commitment but increases the operating cost. The only problem with this approach is that, due to the higher operating cost, it becomes a candidate for cutting when times get bad. On the other hand, depreciation and maintenance changes, as well as the cost of acquiring long-term satellite capacity, can be a heavy burden in times of financial need. Occasional capacity and other innovative network offerings that include dynamic bandwidth allocation should help overcome this issue.

Many applications for private broadcasting can be satisfied on an ad hoc basis—that is, without the acquisition of a dedicated studio facility, uplink, and transponder capacity. Downlinks, on the other hand, would have to be installed on a more permanent basis and consideration given to which satellite would be the focus of the ad hoc network. The following are some examples of ad hoc private broadcasts:

- Press conferences of top executives who must inform the public of a major change in strategy or financial performance;
- Announcements by chief executives to the entire employee population across a wide geographical area, which may occur when there is a change of leadership or a major acquisition;
- Interviews with political candidates during a national campaign, for distribution to local TV stations and eventual rebroadcast (the reason why this is private broadcasting and not program origination is that the candidate usually pays for the event from campaign funds);
- Marketing presentations on major new products such as the Microsoft Windows XP PC operating system or the newest Lexus motor car;
- A private TV channel for viewing in retail stores (Wal-Mart) or airports (CNN).

At a private broadcast, there is usually not a requirement for interactivity because of the potentially large number of remote locations and attendees. Also, the person doing the talking is almost always following a written script that does not allow for interruptions. There may be a question-and-answer period at some point in the broadcast, which could involve either people in the studio or call-ins over a return channel (almost always through dial-up telephone). The problem here is the unpredictability of call-ins, which can put the presenter at a major disadvantage in front of a potentially large audience that is the target of what is otherwise a well-prepared presentation (of course, it might be better to use precleared questions that the presenter is already prepared to answer).

Private broadcasting received a boost by the rising popularity of VSAT networks in the United States and Europe. As discussed in Chapter 9, a VSAT can be equipped with either an analog or digital integrated receiver-decoder (IRD) to receive private TV broadcasts. The cost of this upgrade is small compared to the cost of the VSAT network and represents an excellent way to increase the return on investment. The IRD is connected to the downlink using a simple power splitter. The outputs of the IRD are connected to one or more video monitors located in conference rooms. The quality of reception is not affected in any way by data or voice services that are provided by the VSAT indoor unit. The only consideration is that the antenna be of sufficient size to provide an adequate link C/N, derived by a link budget calculation as discussed in Chapter 2.

There are two options for the uplink for the VSAT private broadcast: either a dedicated facility that is owned and operated by the corporation or a rented facility used occasionally. If a separate video carrier is used, the uplink can be completely separate from the VSAT hub coming from a totally different part of the country. Some form of scrambling or encryption could be used to secure any proprietary content. The only condition is that the uplink be capable of transmitting to the same satellite where the VSATs are pointed. Transponder capacity for the ad hoc event would be rented from the appropriate satellite operator.

The bottom line in private broadcasting is that it is a traditional TV medium, where the presenter is a star on the screen. Consequently, considerable effort must be placed on the visual impression of the scene and the presenter(s). This should be organized and directed by someone who has experience in TV production.

4.3.2 Video Teleconferencing

Video teleconferencing links and networks were touted in the 1980s as attractive ways to reduce business travel costs and improve organization performance by increasing communication among distant groups. Major U.S. corporations like ARCO and Citibank invested millions of dollars on special VTC-equipped conference rooms, video compression codecs, and Earth stations needed to deliver adequate bandwidth. These pioneers demonstrated that it was feasible and that those participating could fulfill many useful purposes. As time progressed, the cost of the rooms and codecs came down, along with increased availability of much cheaper terrestrial communications using the fiber optic networks of the long-distance carriers. The innovator and leader in this field is Sprint Communications, which was first to go totally fiber in its national long-distance network and continues to lead the market in providing connectivity for VTC users. Satellite links were subsequently adopted to the unique needs of VTC when HNS and Spar Telecommunications (now EMS) began to offer mesh networking systems using larger-sized VSATs (1.8m to 2.4m). The advantage of this approach is that a VTC can be scheduled and activated by central control, even though the sites involved are located elsewhere in the network. The partial transponder bandwidth needed for such a private network would be leased from a satellite operator, perhaps on a long-term basis. During the 1990s and through 2002, VTC codec equipment became almost a commodity item, with functionality available in PC software. However, high-quality VTC (an oxymoron to some people) still demands the power of the dedicated codec that sells for prices in excess of $1,000.

Many of the most popular applications for VTC are summarized as follows.

- *Routine meetings between members of a team that is engaged in a very large project:* Groups in different locations can interact as frequently as daily, which can be vital if the project is moving quickly. Projects of this type are very high valued and often are for a government agency such as the U.S. Defense Department or the national telecom operator of a country like China.
- *Coordination meetings of a joint venture involving groups in different countries:* In this way, the combined organization can cooperate and collaborate better because they see each other more frequently than they would if face-to-face meetings were relied upon. This tends to build trust and improve communication, which are vital for the success of a joint venture business activity particularly in its formative phase. Likewise, VTCs can deal more effectively during a period of difficulty by allowing issues to be aired and discussed.
- *Routine financial reviews of a multinational corporation that involve many remote locations:* Headquarters financial managers can speak to their counterparts at remote locations, either collectively or one at a time. Not only the operating numbers but their meanings can be discussed. Any new policies or practices would be reviewed and comments collected for consideration.

While the majority of VTC installations rely on terrestrial networks, those who invest heavily in the rooms and equipment can provide a satellite and terrestrial access. This increases the versatility of the systems, making possible internal as well as external conferencing. In the case of the latter, the most effective approach is to

connect to a terrestrial network on both sides. Sprint, for example, serves the United States, Canada, and many points overseas through connectivity with a large number of counterpart national network operators. This author, for example, has participated in such conferences with points in Japan and Singapore.

The standard arrangement of VTC is for point-to-point connectivity as presented in Figure 4.8 for a typical system. This produces a two-way service where both sides of the conference can see and hear each other. Each end of the connection is equipped with cameras, microphones, monitors or TV projection systems, a digital video compression codec, and a controller. The most expensive item in the system

Figure 4.8 Typical arrangement of a two-location VTC network with the ability to use either satellite or terrestrial links.

and the one that is most critical to the operation and performance is the codec. To reduce cost, the codec has been minaturized and integrated with the camera (Figure 4.9). The leading suppliers of VTC codecs in 2003 were Vtel, PolyComm, and Sony. The standard VTC codec digitizes and compresses the video signal and performs the reverse function as well. In addition, a typical codec provides separate inputs for audio, data, and control. The user can interact with the device using either a separate control box or a special type of handheld remote controller and on-screen display. This is advantageous because the older control boxes are not intuitive and usually do not have on-line instructions available.

The telecommunication aspects of VTC can be provided either by satellite or by a terrestrial digital network. As shown in Figure 4.8, rooms A and B are to be connected by VTC so that a single meeting will involve both locations. The conference table is arranged so that attendees can see the video monitors and can be seen on the other end through the cameras that are mounted on top of the monitors. One camera could provide a wide-area view of all in the room, while a second would point to the speaker. Microphones on the conference table carry speech and activate the

(a)

(b)

Figure 4.9   Examples of VTC devices: (a) Tandberg 1000, and (b) Polycom Viewstation SP.

appropriate camera. The two ends of the conference would appear either on separate monitors or on a split screen of a single monitor. Another feature is the use of a still projector or computer display to add a graphic capability to the meeting. This could either be substituted for the live picture or sent simultaneously over the data channel that is multiplexed with the digital video.

The cameras and monitor of the VTC in Figure 4.8 are connected to a digital compression codec that performs the processing and multiplexing of all of the inputs and outputs. Overall operation of the room equipment and the telecommunication links is managed by a controller, which could be part of the codec or a separate unit. One function of the controller is to allow the VTC to use either a satellite link or a terrestrial network, depending on what is available and what is most cost effective for the particular meeting.

Digital compression of video signals is discussed in detail in Chapter 5. Briefly, the video signal is first converted from analog to digital format, resulting in a high bit rate data stream at approximately 100 Mbps. This is substantially higher than the rate to be transmitted over the link. The data is first processed on a single-frame basis; that is, spatial compression is performed on the first frame to reduce the quantity of bits used to represent the image. Frame-to-frame processing (temporal compression) then causes only the changes between adjacent frames to be transmitted. The most popular spatial compression technique is based on the DCT, a mathematical conversion algorithm that takes the scanned image and produces a set of coefficients from a corresponding mathematical series. The algorithm is now standardized as part of the H.320 series of ITU-T specifications, which are discussed later in this section. These coefficients are then compressed to further reduce the bandwidth. Instead of sending the picture image, the coefficients are transmitted to the other end where the image can be recreated. The combined effect of spatial compression and frame-to-frame compression produces a moving picture image that can closely approximate the natural motion performance of the original analog TV signal. The amount of naturalness and the ability to track fast movement is directly dependent on the degree of compression. In other words, the more compression, the less natural the resulting TV images but the lower the required transmission speed.

Because of this trade-off, developers of VTC networks must examine the codecs carefully before committing to a particular brand of equipment and transmission data rate. Most codecs sold today comply with two ITU-T standards: H.320 for use with constant bit rate links such as T1/E1 and ISDN basic rate interface (BRI), and H.323 for use with the Internet Protocol (e.g., the Internet or an intranet). It is highly desirable to expose prospective users to a typical VTC link of the same design before making this commitment. Otherwise, it is possible that a great deal of time and money could be spent on a network that users find unacceptable for their intended purpose. Vendors often provide demonstrations at expositions and trade shows, but prospective buyers must be sure that the demonstration is for the same arrangement that they intend to purchase, such as monitor size and quality, cameras, codec (including features and standards support), and transmission data rate.

A fully functional codec has multiple line speeds available and, therefore, can provide different levels of video quality in relation to how much the user is willing to pay for transmission. These can step through 128 Kbps, 256 Kbps, 384 Kbps, 728 Kbps, and 1.544 Mbps. The general reaction to the motion quality, color, and

resolution of these various speeds is that 128 Kbps is barely acceptable and that 384 and 728 Kbps are preferred for typical meetings. If the VTC can transmit full-motion material like movies and TV spots, then 1.544 or 2.048 Mbps is required. Teleconferencing codecs are designed for meetings and should not be used to transmit fast-action material like automobile testing and live sporting events.

Since the most popular speeds, and the speeds most available on a national and international basis, are 128 and 384 Kbps, less expensive boxes that fix these rates are appearing. The cost differential between the fully variable and the fixed rate codec is nearly two to one. The codec can also provide transmission security using a symmetrical encryption algorithm. This would be important for very private transmissions that would be transmitted over satellite links or public networks. Use with the Internet Protocol subjects the service to a variable bit rate and the possibility for lost or dropped packets. This produces a quality that is potentially lower than from a dedicated constant bit rate connection, but has the advantage of allowing the service to be combined with other data in the total IT environment.

Satellite-based VTC remains a useful application because it bypasses the remaining limitations of terrestrial networks. The HNS approach, called Intellivision, is based on FDMA with each station activated on an assigned channel at the time of transmission. A common control station is used to schedule the conferences and to control the remote equipment.

Another network uses ViaSat manufactured equipment based on the TDMA access protocol so that only one frequency is employed. However, the bandwidth of this channel is proportionately greater because it must support multiple stations in a burst transmission mode. Both systems are relatively user-friendly, allowing a non-technical administrator to arrange and manage teleconferences across the diverse network.

Satellite transmission of two-way VTC raises the interesting possibility of point-to-multipoint communication (e.g., private broadcasting) and true multipoint conferencing (e.g., many-to-many connections). The former is obtained by having only one site transmit and the other sites operate in the receive mode. The audio and control of the transmission would be in one direction as well. There could be audio return either over the satellite using some form of multiple access or by separate telephone dial-up connections. Through the concept of a video bridge, the multipoint VTC allows sites transmit video and audio simultaneously. This means that at a given location, all participants from all locations can be seen and heard. The video bridge display looks like the TV game show *Hollywood Squares*, where the screen is divided up into a matrix of boxes with separate pictures, one for each location. You would see each location as a tiny picture among many on the screen. The trouble with this, or course, is that it would be relatively hard to tell who is doing the talking at a given time unless the square containing the active talker is highlighted or expanded in size.

The previous discussion focused on the use of VTC for meetings between groups at different locations. This has been and continues to be the most valuable application of the technology. Furthermore, the H.323 standard has been extended to the desktop through add-on codec equipment from Vtel and Polycom, as well as with downloadable software by Microsoft (e.g., NetMeeting). The former again uses the codec integrated with the camera, while the latter relies only on an

inexpensive Web-cam type of device (such as the Quickcam by Logitech). A difficulty with NetMeeting, however, is that the user must employ a static IP address such as associated with a private intranet. In the public Internet, IP addresses are typically assigned dynamically by the ISP. The way around this problem is to have both users in the VTC connect to a common server, which takes care of the transfer of H.323 packets. This works but the performance is highly dependent on the loading of the common server as well as the local access service. In time, low-cost desktop VTC should increase awareness of the medium and substantially reduce barriers now prevalent among the user community.

4.4   Analog TV Standards

Television standards define the format and quality of video signals that are intended for viewing by the general public. They are applied at the origination point where the picture is acquired and subsequently contributed to the program, to the studio where programs are prepared, and the distribution link to the broadcast station or cable TV system that transfers the signal to the ultimate viewer. There is little doubt that eventually analog standards will give way to digitally based processing, transmission, and display. The current status of the digital standards is covered in Chapter 5. However, the majority of the existing infrastructure of TV sets, local stations, and cable TV systems is still analog. Anyone contemplating a new TV application or network will have to consider this factor when determining how they will reach enough potential viewers to make their venture a success.

Analog TV standards exist for every phase of the process of creating and distributing video programs to the public. We use the following definitions with regard to each of these aspects of program delivery.

Origination defines standards for the video format that the camera uses to present the video image. These fall into the three recognized color systems: NTSC, PAL, and SECAM. The standards are defined further, and there are differences that make the specific details differ from country to country, in some cases.

Transmission defines standards that quantify the allowable distortion and degradation due to carrying the signal from the camera or studio to the point of distribution to the viewer. This format is usually not intended for direct reception by the public but is designed for minimum degradation in signal quality. Point-to-point transmission is the normal mode, using fiber optic or coaxial cable, microwave radio, or satellite links. Point-to-multipoint transmission via satellite or terrestrial microwave radio towers can be employed to reach the public directly using the same standard, provided that the end user has an appropriate converter box.

Contribution refers to the original content that is created at a site such as a sports stadium, meeting or convention hall, or remote studio. The video and audio is in its raw and unedited form; therefore, it is unsuitable for viewing directly by the public. For it to be considered a program, the contribution feed is transmitted to the central broadcast center (Figure 4.2) where it is edited, recorded, and assembled into the final program. Quality at this point is the best attainable under the circumstances. If contribution is from a professional facility of some type, then the format and quality are comparable to what is produced in a proper studio facility. On the

other hand, much content originates from portable cameras and highly compressed digital transmission links that results in severely degraded quality. In this case, it is the urgency and interest that determines value in the context of the final program.

Distribution defines standards for the allowable degradation as the signal is carried to the ultimate viewer. Originally, this considers the radiation of the TV channel from local broadcasting stations, directly over the air at VHF and UHF frequencies. More recently, digital formats like Motion Picture Experts Group (MPEG) are the popular system for delivery, where each viewer has a set-top box to convert from the unique distribution format to one of the standard analog formats.

These aspects of the analog TV standards are covered in the following sections.

4.4.1 Video Format Standards

The 1950s and 1960s saw the adoption of analog color TV standards: National Television System Committee (NTSC) in the United States and Japan and Phase Alternation Line (PAL) and Séquentiel Couleur Avec Mémoire (SECAM) in Europe. These standards are used worldwide, and in some cases the same TV receiver is capable of displaying more than one. For many years, the electronics in receivers employed vacuum tubes, which are relatively expensive and less stable than solid-state equivalents. Therefore, first generation TV sets were designed for a minimum number of parts. Transistor circuits began to replace tubes in the 1960s, greatly improving both the stability and reliability of home receivers. The first role of digital circuitry in the 1970s was in the form of digital channel display and remote control. Later, integrated circuits were introduced to replace nearly all of the active electronic elements, producing very low cost receivers with far more complexity and sophistication than the original developers might have thought possible. Picture quality is generally felt to be as good as can be obtained with these analog systems, which is still quite acceptable for comfortable viewing of entertainment TV and many business video applications.

The analog standards are divided according to two properties: the basic black-and-white signal, also called the *luminance*, which existed before color was added; and the technique for adding color (*chrominance*), namely, NTSC, PAL, and SECAM. North America and much of South America, as well as Japan and Korea, support NTSC with 525 lines and 60 fields/30 frames per second. Europe and the rest of the world adopted 625 lines and 50 fields/25 frames per second and are split between PAL and SECAM. The luminance creates the black-and-white image during the scanning process over the screen of the picture tube. Color is added by phase modulating a subcarrier frequency that occupies a position within the luminance baseband frequency range. Table 4.4 summarizes the key parameters for worldwide TV systems, where the capital letter indicates the international standard designation (discussed later in this chapter). The table is abbreviated, since it requires more than 50 individual technical characteristics to specify each system properly.

4.4.2 Analog Transmission Standards

The purpose of analog transmission standards is to provide television engineers and specialists with standardized performance objectives and measurement methods to

Table 4.4   Summary of Analog TV Standards That Apply Throughout the World, According to the ITU-R

Basic TV Standard	M	M	N	B,D,G,H,N	I	D, K, K1, L
Color system	NTSC	PAL	PAL	PAL	PAL	SECAM
Video bandwidth (MHz)	4.2	4.2	4.2	5	5.5	6
Broadcast TV channel bandwidth (MHz)	6	6	6	8 (B:7)	8	8
Field frequency	60 (59:94)	60	50	50	50	50
Line frequency	15,750	15,750	15,625	15,625	15,625	15,625
Sound subcarrier frequency (MHz)	4.5	4.5	4.5	5.5	6	6.5
Video levels (%)						
Blanking	0	0	0	0	0	0
Peak-white	100	100	100	100	100	100
Sych tip	−40	−40	−40	−43	−43	−43
Difference between black and blanking	7.25	7.25	0	0	0	0

determine signal quality. In this instance, we are concerned with the link between the studio and either the local TV station or cable TV head end. Transmission systems that are used for this purpose include microwave radio, satellite links, fiber optic cable (used in an analog manner), and coaxial cable. The signal is not normally available to the public along one of these systems and so analog transmission standards are not designed for the minimum cost of reception. Rather, they emphasize the quality of the resulting delivered signal with a minimum of added distortion and interference noise.

An example of a typical transmission system for point-to-point video transfer is shown in Figure 4.10. The studio delivers separate video, audio, and data outputs to a video exciter that is associated with, in this case, a TV uplink Earth station. Point A

Figure 4.10   A typical transmission system for point-to-point video transfer.

represents where the video portion of the information is essentially perfect in a technical sense. (This may not be the case in practice as source material may be contaminated for other reasons.) The transmission system extends from the exciter, which produces a modulated carrier at the IF frequency (typically 70 MHz), containing the three components. Translation to the RF transponder channel is performed by a separate upconverter or as an integral part of the exciter.

The most popular analog technique is to employ FDM to combine the video with the associated audio channels, as shown in Figure 4.11. The video is transferred directly across to the low end of the baseband and stops at frequency f_m. Each audio channel is frequency modulated onto a subcarrier on an upper baseband frequency. This particular example provides two audio channels on separate subcarriers at f_{c1} and f_{c2}, for stereo audio in the primary language. Audio channels for multiple languages can be included by adding subcarriers into the baseband. Up to a total of 10 such subcarriers have been used in practice.

The third subcarrier (at f_{c3}) is for a broadcast data channel to be received at the remote stations or by other downlinks. Some of the possible applications for this data broadcast include:

- Network control and coordination, for automated operation of remote antenna, transmitters, receivers, and studio equipment;
- Program information and verbal instructions to allow the distant stations (in the case of TV broadcasting) to be aware of upcoming events and any special requirements;
- Data services like teletext that can be delivered to the public along with the video or offered as a totally independent service for additional revenue;
- A paging channel to provide nationwide paging services through the facilities of the local broadcast station or an auxiliary transmitting tower.

The composite baseband containing the video and all of the subcarriers is transferred to the IF carrier using FM. Typical baseband and modulation formats are shown in Table 4.5, with separate listings provided for NTSC, PAL, and SECAM. These values are not formal standards but are used in practice around the world on

F_m Highest video baseband frequency
F_{c1} First subcarrier (program sound)
F_{c2} Second subcarrier (auxiliary sound)
F_{c3} Third subcarrier (orderwire)
F_{c4} Fourth subcarrier (data)

Figure 4.11   Arrangement of the video baseband used in analog FM transmission over satellite links.

Table 4.5   Recommended Baseband Subcarrier Frequencies Used in Satellite Video Transmission to Convey Program Audio, Stereo, and Data (All Frequencies in Megahertz)

Frequency (MHz)	NTSC	PAL	SECAM
Highest baseband frequency, f_m	4.2	5.0	6.0
TV channel width	6.0	6.0	8.0
Primary audio, f_{c1}	6.8	—	—
U.S. cable TV			
U.S. broadcast	6.2/6.8	—	—
SES Astra	—	6.5	—
EUTELSAT	6.6	6.6	6.6
INTELSAT	6.6	6.6	6.6
Stereo audio, f_{c2}	(left/right)		
U.S. cable TV	(encrypted)	—	—
U.S. broadcast	5.94/6.12	—	—
SES Astra	5.94/6.12	—	—
EUTELSAT	—	7.02/7.20	—
INTELSAT	6.65	6.65	6.65

the systems indicated in the first column. They are incorporated in commercial video exciters and receivers from a variety of suppliers in the United States, Europe, and Japan.

Television engineers who work with transmission systems and standards have the intention of not degrading the signal in a significant way, expecting that the eventual distribution process will be less controllable and therefore tend to provide the bulk of the degradation to quality.

The performance requirements for the video signal, as received at point B in Figure 4.10, are contained in widely recognized and supported standards of the U.S. ANSI and the ITU. The Radiocommunication Bureau of the ITU (ITU-R), formerly known as the CCIR, issues recommendations that apply to the international transmission of TV signals over satellite links. While accepted among nations, the ITU-R recommendations are not detailed enough to assure commercial service quality and therefore must be augmented by the specifications discussed in this section. Other organizations like the EBU have specific standards that apply within a given country or region.

4.4.2.1   NTSC Transmission Requirements

The requirements for the transmission of the NTSC signal are specific and detailed. The Electronic Industries Association (EIA) and the Telecommunications Industry Association (TIA) have produced a well-known standard: EIA/TIA-250-C, "Electrical Performance for Television Systems" [7]. This is an updated version of RS-250-B, which had been the measurement standard for all North American TV transmissions up until 1990 when EIA/TIA-250-C was formerly issued. The latest standard, effectively applied anywhere in the world where the NTSC system is used,

is used to evaluate the performance of short-, medium-, and long-haul microwave links, satellite links, and various end-to-end combinations thereof.

The following basic definitions are essential to understanding the role and application of 250-C.

A short-haul transmission system is usually a simple point-to-point transmission link (also called a hop) that is of the order of 30 km in length. These links are used to connect the studio to the broadcasting tower or to a local transmitting Earth station.

A medium-haul transmission system is a microwave or cable relay system consisting of more than one hop over a distance of between 200 and 4,500 km. Such systems were popular before the age of satellite transmission and have reappeared through the introduction of long-haul fiber optic systems.

A satellite transmission system is a single-hop satellite link between a transmitting Earth station and a receiving Earth station through a bent-pipe satellite repeater. A typical example is shown in Figure 4.10.

An end-to-end network is an interconnection of multiple transmission systems, consisting of, for example, a satellite transmission system with short-haul microwave transmission systems on both ends. This is the typical case, using various combinations of transmission systems that depend on the requirement.

IRE units measure the TV signal, where one IRE unit is 0.01 times the range of the luminance signal. IRE is the abbreviation for the Institute of Radio Engineers, an organization that merged with the American Institute of Electrical Engineers to form the Institute of Electrical and Electronic Engineers (IEEE). The typical time waveform in Figure 4.12 displays an NTSC signal showing with the blanking level at 0 IRE units, the maximum (white) video level at 100 IRE, and the synchronization waveform negative pulse extending to –40 IRE. In total, the video signal ranges 1V, peak-to-peak. IRE units are only used for 525-line systems (NTSC), while the 625-line systems (PAL and SECAM) refer either to the percentage of maximum video level or to the actual voltage.

Figure 4.12   The waveform of a typical line of the NTSC signal.

Table 4.6   NTSC Television Transmission Performance Requirements (EIA/TIA 250-C, 1990)

Test Parameter	Waveform	Short Haul	Satellite	End-to-End
Baseband frequency response	Multiburst	±2.5 IRE	±7 IRE	±12 IRE
Chrominance to luminance gain inequality	Modulated stairstep	±2 IRE	±4 IRE	±7 IRE
Chominance to luminance delay inequality	Modulated stairstep	±20 ns	±26 ns	±60 ns
Differential gain	Modulated stairstep	2 IRE (2%)	4 IRE (4%)	10 IRE (10%)
Differential phase	Modulated stairstep	0.7°	1.5°	3°
Luminance nonlinearity	Modulated stairstep	2%	6%	10%
Chrominance to luminance intermodulation	Three-level chroma signal	1 IRE	2 IRE	4 IRE
Chrominance nonlinear gain	Three-level chroma signal	1 IRE	2 IRE	5 IRE
Chrominance nonlinear phase	Three-level chroma signal	1°	2°	5°
Dynamic gain of video signal	Stairstep with variable APL	2 IRE	4 IRE	6 IRE
Dynamic gain of sync signal	Stairstep with variable APL	1.2 IRE	2 IRE	2.8 IRE
Signal-to-noise ratio	10 kHz to 4.2 MHz, weighted (any)	67 dB	56 dB	54 dB

The average picture level (APL) is an average taken of the signal level during the active scan time, not including the blanking and synchronization. In other words, it is the integrated average of the picture waveform itself for one horizontal line (e.g., 33.4 ms for NTSC), over the range of 0 to 100 IRE units.

A sample of the specified values for each type of transmission system, summarized from [7] is provided in Table 4.6. This demonstrates the extent and depth of 250-C, which can be difficult to meet unless high-quality equipment (particularly exciters and receivers) is used. Considerable debate erupted during the early 1980s over whether digital video links can or should meet RS 250-B. This was first fueled by the introduction of high-quality, low error rate fiber transmission systems within major U.S. cities like Los Angeles (in fact, it was within Los Angeles that the ABC network first experimented with fiber for the 1984 Summer Olympic Games) and later by the availability of compression systems based on the MPEG-2 standard.

In the interest of brevity and clarity, we have limited Table 4.6 to the cases of short-haul terrestrial, satellite, and end-to-end transmission systems. The general trend is that the longer the transmission system and the more components that the signal must traverse, the larger the specification range of allowed performance. The three specifications that are emphasized—differential gain, differential phase, and S/N—are the most critical to the visual quality and color of the received analog signal and for that reason they have been adopted by the ITU-R for their recommendations on television transmission over satellites.

Actual measurement of the specifications is accomplished using standard test waveforms, which are listed in the table and shown in Figures 4.13 through 4.15. The signals work in the following manner.

Figure 4.13 The multiburst signal used to measure baseband frequency response as part of EIA/TIA 250 LC testing.

A *multiburst* signal (Figure 4.13) measures the frequency response at six discrete frequencies over the video baseband range.

A *stairstep* signal (Figure 4.14) measures the gain at the color subcarrier frequency at six different brightness levels, from zero (black) to 90% of the maximum white level.

Three-level chrominance (Figure 4.15) detects any change in the phase of the color subcarrier (which produces the color or hue) as a function of the amount of color saturation.

Figure 4.14 The modulated stairstep signal used to measure amplitude and phase nonlinearity as part of EIA/TIA 250-C testing.

Figure 4.15 The three-level chrominance signal used to measure chrominance performance as part of EIA/TIA 250-C testing.

These and other test waveforms are inserted into horizontal lines during the nonvisual portions of the vertical blanking interval. Alternatively, the normal TV signal can be interrupted to allow near-continuous transmission of a particular test signal.

Television standards also consider the quality of the audio portion of the program. These can be stated more succinctly as the required S/N and the allowable amount of audio distortion in the received signal. Also, satellite networks that deliver multiple video channels from the same orbit position must also adopt a standard audio level to prevent contrast between video channels as the viewer tunes the home receiver across the transponders. This is mostly a concern in analog transmission systems where levels can drift over time. Standard 250-C specifies that the audio S/N must be greater than or equal to 66, 58, and 56 dB, for the short-haul, satellite, and end-to-end cases, respectively, presented in Table 4.6. There is also a requirement that the time differential of the audio channel with respect to the video channel fall within the range of +25 to –40 ms.

The profile of 250-C testing is lengthy and complicated, using a test signal generator and a number of receiving measuring devices. Among the receiving test devices are the video analyzer and the vector scope. This equipment has been incorporated into an automated test system to both speed up and make consistent the entire procedure. This allows the full suite of 250-C tests to be performed at the touch of a button (or return key, as the case may be).

4.4.2.2 PAL and SECAM Transmission Requirements

Analog transmission systems that are designed for NTSC already meet many of the requirements for PAL and SECAM. This is not surprising because both PAL and SECAM are derived from NTSC with respect to the use of interlaced scanning and

the manner in which the luminance (black and white) information is carried. As shown in Table 4.5, PAL and SECAM differ from NTSC in that they use a basic frame rate of 25 per second as opposed to 30 for NTSC. In addition, with 625 lines per picture instead of 525, PAL and SECAM actually have better resolution. Because of these basic differences, PAL and SECAM require more baseband bandwidth, in the range of 5 to 6 MHz, as compared to 4.2 MHz for NTSC. This has a direct impact on the transmission system, which must be designed to carry the greater bandwidth and to accept the 25-Hz frame rate and corresponding field frequency of 50 Hz. Since we do not increase RF power or bandwidth to compensate, the resulting S/N that can be achieved with PAL and SECAM is approximately 4 dB less than for NTSC.

Transmission standards for PAL and SECAM are generally covered under the ITU-R and its series BT recommendations. The most popular color TV system in the world, in terms of the number of countries and total population, is PAL. These countries have the 625-line, 50-Hz setup for the basic TV signal. The only country in the world with the combination of PAL and a 525-line/60-Hz standard is Brazil (all neighboring countries in South America have the same NTSC setup as the United States).

PAL is closest to NTSC in the manner in which the color information is transmitted, using the phase angle of the color subcarrier. Instead of requiring that the absolute value of phase be carried from camera to TV set, PAL improved upon the technique by doing it with a phase shift between alternating lines (hence the name, phase alternation line). This renders PAL less sensitive to several of the impairments found in the typical analog transmission system. For example, the 250-C specification for differential phase on a satellite link is 1.5°, which while significant for NTSC has effectively no meaning for PAL (or SECAM, for that matter). Both systems are equally susceptible to differential gain and noise.

The performance requirements for PAL can be determined with standard test equipment, similar in operation to that for NTSC. Table 4.7 summarizes the most important requirements for a typical combination of a short-haul microwave and a point-to-point satellite link. Other requirements apply to PAL and SECAM transmissions between studio and distribution point.

This chapter has taken us through common systems for TV distribution and many of the standards that apply. In Chapter 5, we review the developments in digital video, particularly compression systems that increase the channel capacity of the satellite by a substantial factor without reducing the visual quality as seen by the public. Chapter 6 covers how the foundation of satellite transmission, cable programming, and digital compression are producing broadcasting satellite systems.

Table 4.7   Signal-to-Noise Requirements for PAL Transmission Systems (CCIR Recommendations 421-2 and J22)

Type of Link	Video	Audio
Short-haul microwave	62 dB	67 dB
Satellite link	52 dB	57 dB
End-to-end link	50 dB	54 dB

References

[1]   Vogel, H. L., *Entertainment Industry Economics: A Guide for Financial Analysis*, 5th ed., Cambridge, England: Cambridge University Press, 2001.

[2]   *Entertainment Weekly*'s EW.com, September 3, 2002, TV Chart.

[3]   Ciciora, W. S., "Cable Television Systems," *Engineering Handbook*, 9th ed., Washington, D.C.: National Association of Broadcasters, 1999, p. 1339.

[4]   Ciciora, W. S., "Inside the Set-Top Box," *IEEE Spectrum*, April 1995, p. 70.

[5]   Sellgren, Cable & Satellite Express, February 6, 1997, about Viasat getting a 16% buy rate for the Nielsen-Holmes match.

[6]   Pelton, J. N., *Space 30—A Thirty-Year Overview of Space Applications and Exploration*, Alexandria, VA: Society of Satellite Professionals International, April 1999, p. 103.

[7]   EIA/TIA, "Electrical Performance for Television Relay Facilities," EIA/TIA-250-C, 1977.

Digital Video Compression Systems and Standards

The television signal conveys a lot of information originating from the analog camera and microphone. As discussed in Chapter 4, the TV systems of the world employ about 5 MHz of baseband bandwidth. Satellite transmission using FM requires that this bandwidth be multiplied further to occupy between 27 and 36 MHz. This amount of bandwidth results in a high-quality signal that can be recovered with relatively inexpensive receivers. However, the real cost of the analog baseband and analog FM comes in the inefficient use of space segment. Digital video compression technology provides the means to greatly reduce this occupied bandwidth. The trick is to do it without degrading the enjoyment of the recovered signal.

Digital compression plays a very important role in modern video transmission. Its principal benefits are:

- Reduced transmission bandwidth, which saves space segment costs and reduces the amount of power needed to transmit an acceptable signal;
- More channels available per satellite, which greatly increases the variety of programming available at a given orbit position, in turn promoting expanded services like impulse PPV and home education and making it feasible to expand a programming service through tailoring (e.g., packaging several different feeds of the same material with different advertising or cultural views) and multiplexing (i.e., sending the same channel at several different times);
- The potential of using a common format for satellite DTH, cable TV, and terrestrial broadcasting;
- Provision of a base for HDTV in the digital mode because the number of bits per second of a compressed HDTV signal is less than what was previously required for a broadcast-quality conventional TV signal.

Digital compression was first developed 30 years ago to save bandwidth by a factor of two on international satellite links. Some of the motivation also came from the sheer excitement of dealing with a challenge given the technology available at the time. A system developed by COMSAT Laboratories in 1970 could multiplex two NTSC signals within the bandwidth of one transponder using a hybrid approach of analog filtering and digital processing. The system was not adopted because of the high cost of codecs, even though these researchers proved that a compressed signal could provide acceptable viewing.

Work in digital image processing continued for a very long time, yielding many innovations in the theory of digital video representation and electronic digital signal

processing. The implementation of video compression has gone through a number of iterations, resulting finally in very affordable and usable consumer equipment. An example of this type of equipment is shown in Figure 5.1, produced the EchoStar DISH Network.

Compression systems that were marketed in the 1980s met a variety of needs, such as video teleconferencing, PC videophones, distance education, and early introductions of narrowband ISDN. Some examples of these early applications of digital video compression are listed in Table 5.1. The quality of the video portion is generally unacceptable for entertainment programming but probably adequate for a specific business purpose. For example, the H.320 and H.323 systems are extensively used for point-to-point meetings to serve the needs of business and government users (see Chapter 4). The locations can be separated by a few hundred kilometers (as in the case of communication between subsidiaries located in different cities of the same state, province, or region) to thousands of kilometers (when international service is needed). People who use videoconferencing find it convenient because no significant travel is required and more people may participate. Generally, these people

(a)

(b)

Figure 5.1   TV receiving equipment for the DISH Network: (a) DTH dish installation, and (b) EchoStar personal video recorder.

Table 5.1 Examples of Limited-Motion Digital Video Compression Applications in Use on Satellites and Terrestrial Networks

Line Characteristic	Typical Line Speed	Applications
Digital carrier	3 Mbps	Distance education
T1, E1	1.544 Mbps	Meetings; distance education
P*64	384 Kbps	Teleconferencing; integrated networks
ISDN	128 Kbps	PC conferencing; videophones
Voice band dial-up	19.2 Kbps	Technical communications, PC conferencing

already know each other and so can recognize who is doing the speaking and even pick up nonverbal clues from body language. Services that involve basic rate ISDN and analog dial-up are not attractive in the meeting situation but would prove useful for desktop applications, as suggested in the table.

A wide range of performance of compression systems results from the relationship between the data rate (which is proportional to the occupied bandwidth) and the quality of the picture. Data rates below 1 Mbps are possible when quality can be sacrificed. On the other hand, if the intended application is in the field of education or entertainment, then significantly more than 1 Mbps is dictated. The first introduction of compression equipment with adequate quality for education applications was the Spectrum Saver system, mentioned in Section 4.2. With a selectable data rate of either 3 or 6 Mbps, the user can determine the level of absolute quality against the cost of satellite transmission. Terrestrial transmission of the Spectrum Saver was not considered in the development of the facility.

Table 5.2 gives an indication of the relationship between bit rate and application in commercial broadcasting. A perfect video reproduction of analog TV standards (e.g., NTSC, PAL, and SECAM) is achieved with rates of 90 Mbps or greater. Typical viewers cannot usually tell that anything is impaired when the signal is compressed to a rate of 45 Mbps. Below this value, it becomes subjective. For movies, a rate as low as 1.5 Mbps, the standard T1 in North America and Japan, is sufficient. However, for any live action as used in sports, at least 3 Mbps will be needed.

Compression systems that operate at 45 Mbps or greater are designed to transfer the signal without permanent reduction of resolution and motion quality. They are said to be lossless in that the output of the decoder is identical to the input to the encoder. In contrast, operation below about 10 Mbps is lossy in that it introduces a change in the video information that cannot be recovered at the receiving end. Lossy compression can produce a picture of excellent quality from the viewer's point of view at data rates above about 1.5 Mbps. There is an intermediate position called quasi-lossless wherein a lossy compression service is augmented with the parallel

Table 5.2 Typical Data Rate Requirement for Production and Distribution of Network TV Signals

Purpose	Data Rate (Mbps)
Acquisition (camera)	150
Production (studio)	150
Transmission (distribution)	30–45
Reception (direct-to-home)	3–6

transmission of an error signal that contains correction data to recreate a lossless image at a compatible receiver. The application of lossy transmission with reduced picture quality may be attractive since it can reduce transmission costs (or allow the user to employ an existing communications system such as a VSAT network).

Our focus in the chapter is on modern compression technology and standards that are being applied to the consumer marketplace. While drawing on previous experience, the new approaches provide high-quality images and employ low-cost set-top equipment. This breakthrough in applying technology revolutionized the satellite TV industry and caused a major shakeup in cable television (the latter has had to adopt the digital channel approach of DTH to maintain its dominant position).

5.1   Compression Technology

The analog waveform of the NTSC or PAL standard is very effective in its ability to provide entertainment and business communications. Enjoyment has been further enhanced with the addition of stereophonic and surround sound; and additional services like closed caption and second language are available as well. These systems are also relatively simple in terms of generation and display. The transmission of the video signal is relatively straightforward, provided that the link has adequate bandwidth and good linearity (reviewed in Section 4.4.2). One advantage of modern digital modulations is that excellent linearity is not demanded as the information is conveyed in a bit stream rather in its original analog form.

From a technical standpoint, video sequences scanned at the rate of either 30 or 25 frames per second with 525 or 625 lines, respectively, contain a significant amount of redundancy both within and between frames. This provides the opportunity to compress the signal if the redundancies can be removed on the sending end and then restored on the receiving end. To do this, the encoder at the source end examines the statistical and subjective properties of the frames and then encodes a minimum set of information that is ultimately placed on the link. The effectiveness of this compression depends on the amount of redundancy contained in the original image as well as on the compression technique (called the compression algorithm).

For TV distribution and broadcast applications over satellites, we wish to use data rates below 10 Mbps per video stream in order to save transponder bandwidth and RF power from the satellite and Earth station. This means that we must employ the lossy mode of compression, which will alter the quality in objective (numerical) and subjective (human perception) terms. A subjective measure of quality depends on exposing a large quality of the human viewers (subjects) to the TV display and allowing them to rate its acceptability. The TASO scale shown in Table 4.1 is an excellent example of such a subjective scale for measuring quality. In the case of digital TV, the ITU-R has adopted a series of recommendations and reports that further delineate the process (discussed later in this chapter). It turns out that exposure to the better perceived quality of good digital video has a tendency to raise expectations on the part of viewers. This is similar to how telephone subscribers now expect all voice calls to sound like a terrestrial fiber optic connection.

The ultimate performance of the compression system depends on the sophistication of the compression hardware and software and the complexity of the image or

video scene. For example, simple textures in images and low video activity are easy to encode and no visible defects (called artifacts) may result even with simple encoding schemes. The real test is for scenes with a great deal of detail, including varying textures, and fast-moving live action. Conventional movies that were filmed at 25 frames per second do not represent a challenge; however, TV coverage of live sporting events will severely test any compression system.

5.1.1   Digital Processing

Any analog signal can be digitized through the two-step process of sampling at discrete time intervals, followed by converting each sample (usually a voltage value) into a digital code. The latter process is also called *quantization* because it involves forcing the measurements to fit onto a scale with discrete steps. The example in Figure 5.2 shows a simple analog waveform on the top and its quantized version on the bottom. There are only eight quantization levels, which correspond to a digital representation of 3 bits per sample (because the numbers 0 through 7 are represented by the binary numbers 000 through 111). The number of bits determines the total number of levels that are available to show detail in the image. A superior code for doing this is the Gray code, wherein there is only a 1-bit change corresponding to a change of one level [1].

The resulting quality of reproduction can be specified in terms of the signal-to-quantization noise ratio (S/N_q). The more bits per sample, the better the reproduction, as evidenced by the equation

Figure 5.2   A simple analog waveform and its quantized version.

$$S/N_q = 3M^2 \tag{5.1}$$

where M is the number of bits per sample. Equivalently, in terms of decibels,

$$S/N_q = 4.8 + 20\log(M) \tag{5.2}$$

Typical values of M are in the range of 6 to 12, with 8 being the most common. At this level, the S/N_q is equal is 22.8 dB. This relation indicates that doubling the number of bits per sample reduces the quantization noise by 6 dB.

A further refinement is to compress the high end of the scale to emphasize the lower levels and de-emphasize higher levels, a procedure that is inherently easier to accept by the human watcher or listener. This is the familiar process of companding used in PCM applied in digital telephone networks. In companding, the sending end of the link *com*presses the extremes of the voltage scale (positive and negative) by applying a nonlinearity before transmission. This nonlinearity is removed on the receiving end where the inverse process (*expanding*) is performed. Companding provides a subjective improvement to the basic quality provided by the quantization process. If this companding advantage is 30 dB (a typical value), then the S/N_q of 22.8 dB in the previous example increases to 52.8 dB in terms of its subjective effect on humans. At this level, it would be comparable to a high-quality analog telephone line that has a measured value of S/N above 50 dB. In fact, listeners find the digital voice channel to be superior to the analog equivalent due to the removal of background noise and interference. A subjective evaluation of telephone communications is discussed further in Chapter 11 for mobile satellite service. In the following, we consider how this performance will affect video images.

According to the Nyquist sampling theorem of communications engineering, lossless sampling requires that samples be taken at a rate that is twice the highest baseband frequency. Therefore, for a typical video signal of 5-MHz bandwidth, the sampling rate would have to be at least 10 million times per second (e.g., 10 MHz). In the technique of subsampling, the analog information is sampled at a rate lower than that prescribed by the Nyquist criterion. This causes the spectrum to fold back over itself, which could produce in-band interference in the recovered signal. However, if the process is controlled correctly, the folding back can be completely corrected with no interference or information loss. Another approach is oversampling, which samples at a rate higher than Nyquist's original criterion. Oversampling is used in digital processing systems to increase dynamic range and reduce in-band interference. It would not be recommended for encoding of signals such as video due to its inefficient use of bandwidth.

Essentially all of the practical encoding and compression systems use subsampling and quantization prior to compression. Subsampling is applied by first reducing the horizontal and/or vertical dimension of the input video, which in turn reduces the number of picture elements (pels) that must be coded. At the receiving end the images are smoothed using the mathematical process called interpolation. This produces a more natural look to the image. Although done at the receiving end, interpolation is the first level of compression, ahead of the steps to be taken in the coding of each image and the compression from frame to frame.

Compression need not be applied uniformly across the standard color television signal. The black-and-white (luminance) component of the picture has more

information content and hence requires more bandwidth than the color component (chrominance). As a result, subsampling is not applied equally to the luminance and chrominance parts of the picture. Since the human eye is much less sensitive to chrominance detail, chrominance may be sampled at half the interval of luminance. Since there are two dimensions, this means that the two chrominance components together require only half the samples of the luminance. Furthermore, the number of bits needed per sample for the color information is slightly less than for luminance. These sampling and quantization aspects are reflected in all of the video compression standards, particularly MPEG.

5.1.2 Spatial Compression (Transform Coding)

Transform coding is the most popular technique for reducing the number of bits required to represent a digital image. The basic idea is to replace the actual image data with a mathematically derived set of parameters that can uniquely specify the original information. The parameters, which are the coefficients of a mathematical transform of the data, require fewer transmitted bits to be stored or sent than the original image data itself because they can be compressed further. Examples of mathematical transforms include Fourier, Laplace, Z, and wavelet. The Fourier transform is one of the most versatile in communications engineering because it provides the means to convert between a time waveform representation of a signal (the time domain) and an equivalent frequency spectrum (the frequency domain). The math is somewhat complex because it involves integration over an infinite range. Here are the two key integral formulas that convert time-ordered data to frequency-ordered data, and vice versa:

$$F(s) = \int_{-\infty}^{\infty} f(x) \exp(-i2\pi xs) dx \tag{5.3}$$

where s is the frequency in hertz and i is the square root of -1 (which identifies the imaginary part in complex algebra).

Applying the same transform to $F(s)$, the time domain representation of the same signal, we get:

$$f(w) = \int_{-\infty}^{\infty} F(s) \exp(i2\pi ws) ds \tag{5.4}$$

These formulas are useful for analog signals that are continuous in time. For signals that are first digitized, there are more effective versions of the Fourier transform that allow efficient computation. The fast Fourier transform was an important innovation and is still regarded as the basis of comparison for all digital transforms used in communications engineering. Over the years, the discrete cosine transform has proven to be the most popular mathematical procedure and is now part of the JPEG and MPEG series of standards, which will be discussed later in this chapter. The DCT maintains all calculations in the real domain; that is, it does not require complex algebra in its application. The mathematical formulation of the DCT in two dimensions and in the forward direction is

$$F(u,v) = \frac{4c(u)c(v)}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j) \cos\frac{(2i+1)u\pi}{2N} \cos\frac{(2j+1)v\pi}{2N} \qquad (5.5)$$

and in the inverse direction is

$$f(i,j) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} c(u)c(v)F(u,v) \cos\frac{(2i+1)u\pi}{2N} \cos\frac{(2j+1)v\pi}{2N} \qquad (5.6)$$

where

$$c(w) = \frac{1}{\sqrt{2}} \quad \text{for } w = 0$$
$$= 1 \text{ for } w = 1, 2, ..., N-1 \qquad (5.7)$$

The DCT is similar to the FFT in that both allow a computer to generate the frequency spectrum for a time waveform and vice versa. Because of this, the DCT could be described as a way to convert from linear graphic display (which is in two dimensions) to an equivalent series of spatial frequency components. A spatial frequency is measured in cycles per unit of linear measure (cycles per inch, for example) instead of cycles per unit time (cycles per second, or hertz). A square wave in terms of a spatial frequency would look like an alternating sequence of black and white squares, like on a chessboard. In fact, a chessboard image can be transferred very efficiently using the DCT with a minimum number of bits. Any picture can then be viewed as the overlay of these frequencies, much the way an analog time signal can be viewed as the combination of frequency components. The difference with image data is that the time axis of the signal is replaced with a two-dimensional distance axis, as one scans across and down the particular picture element.

The basic concept of how the DCT is applied to two-dimensional compression of an image is shown in Figure 5.3. The image is divided into square or rectangular segments, and then the transform is applied to each individually. In this particular example, the image is split into blocks that are $N \times N$ pels on each side. If you examine one of these squares, you can see that a given horizontal string of pels can be represented by a combination of frequency components in the same way that a time waveform can be expressed by a combination of sine waves (a Fourier series). The first step taken by the DCT coder is to represent the block in the form of an $N \times N$ matrix of the pels and then apply the DCT algorithm to convert this into a matrix of coefficients that represent the equivalent spatial frequencies. The quantity of bits is further reduced by limiting the number of quantization steps and by removing some of the obvious redundancy. For example, coefficients that are zero are not transmitted.

Using frequencies to represent the individual blocks that comprise the image is very effective from the standpoint of compression and transmission. Information loss increases as the relative size of the block increases and the number of frequencies used to represent the content of the block decreases. What you see is that some elements of the image show up as small squares (like a cubist painting), and some shaded areas have odd patterns (like herringbones) running through them. In the worst case, the squares themselves are clearly apparent in the picture. Of course,

Figure 5.3   The zigzag scanning pattern used to collect DCT coefficients with an 8 × 8 block, selecting nonzero values at the locations of the dots.

when the system is properly optimized and the error rate is acceptable, the DCT performs very well and the image is natural. On close examination of the original as compared to the previously compressed image, the modifications are clearly discernible. This does not detract from the enjoyment or utility of the video service, unless of course one demands a perfect reproduction (which is also impossible in the case of any of the standard analog TV systems).

5.1.3   Temporal Compression (Frame-to-Frame Compression)

The types of video sequences involved in NTSC, PAL, and SECAM are statistical in nature and, of course, contain a high degree of redundancy. Stable sequences, such as what happens when the camera is held in a fixed position during a scene, are highly correlated because only a few aspects change from frame to frame. Consider a video segment of the nightly news with a reporter at her chair behind a desk. During the entire time that she is speaking, the foreground and background never change; in fact, the only noticeable motion is of her head, mouth, and perhaps her upper body and arms. The result of this is that only the first frame needs to be encoded in its complete form; the remaining frames must only be updated with information about the changes. This is possible because interframe (frame-to-frame) correlation is high and, in fact, two or more consecutive intermediary frames can be predicted through interpolation (tracing the line from start to end of sequence and choosing the appropriate intermediate point). The formal way to state this is that an approximate predication of a pel can be made from the previously

Figure 5.4   A technique for frame-to-frame projection using motion compensation. The motion vector, mv, indicated in frame *N* – 1 points toward the new location to be used in subsequent frame *N*.

coded information that has already been transmitted. For greater resolution, the error between the predicted value and the previous one can be sent separately, a technique called differential pulse code modulation (DPCM). Both DCT and DPCM can be combined to provide a highly compressed but very agreeable picture at the receiving end.

5.1.4   Motion Compensation

The technique used to reduce the redundant information between frames in a sequence is called motion compensation. It is based on estimating the motion between video frames by observing that individual elements can be traced by their displacement from point to point during the duration of the sequence. Shown in Figure 5.4, this motion can be described by a limited number of motion parameters that are defined by vectors. For example, the best estimate of the motion of a given pel is provided by the motion-compensated prediction pel from a previously coded frame. To minimize error, both the motion vector and the prediction error are transmitted to the receiving end. Aggregating nearby pels and sending a single vector and error for them as a group is possible because there is a high degree of correlation between adjacent pels as well.

Motion compensation is computation-intensive process, and therefore, it only became practical for commercial video compression systems in the 1990s. The MPEG standard was developed in such a way that the encoder provides most of the computation needed for compression and motion compensation; the receiver is relatively dumb, responding to commands from the encoder on how to perform the details of the inverse function. In time, the quality of compressed digital video will improve due to advancements principally on the encoding side. One of these is

statistical multiplexing of several video channels, which removes excess bits from relatively slow-action video and making those available to high-action channels demanding more throughput.

5.1.5   Hybrid Coding Techniques

Two or more coding techniques can be combined to gain a greater advantage from compression without sacrificing much in the way of quality. A technique commonly applied is to combine frame-to-frame (temporal) DPCM with the spatial DCT method. Temporal correlation is first reduced through prediction, and then the DCT is applied to the prediction. The DCT coefficients in the $N \times N$ matrix are quantized and compressed further. This method is central to the MPEG standard to be discussed in the next sections.

Hybrid coding can be applied in levels as a way to enhance the quality of a service that is tailored to a particular requirement. The basic service using DCT may be transmitted with maximum compression for minimum cost of transmission. Enhancement through a second channel that adds the hybrid coding feature would require more capacity and processing on the sending and receiving ends. This opens up the possibility of offering a higher quality service with better resolution and motion performance where bandwidth is available. It is a way of making a service backward-compatible where the first service is based on the current standard for MPEG 2 and the next generation is implemented through hybrid coding. However, there are serious theoretical issues with realizing the benefits of hybrid coding for providing different levels of viewer quality off the same signal. (Here is a more technical view of this difficulty: the idea of such a multioutput hierarchical coding is certainly appealing, but in general there are constraints and suboptimalities that negate the prospective efficiency gains. The sampling and quantization of a higher resolution picture by itself may result in a certain set of pixel values. However, to get that set from a previously coded lower resolution master picture requires many more details, possibly even more data to transmit than the cleanly encoded higher resolution picture. In contrast, *hierarchical modulation* may have a real benefit because different classes of users can have different capabilities of receiving equipment. Here, a standard user would receive QPSK, whereas the premium user would pick from four symbols at each QPSK corner, getting a constrained 16 QAM constellation.)

The other area of potential improvement in compression performance is through transforms other than DCT. There are currently two candidates: fractal coding and wavelets. The details of these methods are beyond the scope of this book, but their introduction is certainly of interest. Fractal compression has been investigated for some decades and holds promise for another order of magnitude reduction in the number of bits per second as compared to the DCT. The basic principle is that any graphic object can be divided down (fractured) into elements of a particular shape (a line, square, star, or irregular shape). It is recognized that fractal compression is lossy because of the type of decomposition of the image that is applied, but it will likely find some significant applications in telecommunications and information storage and retrieval. The wavelet, on the other hand, is a transform related to Fourier. Unlike the Fourier transform, which converts the time

waveform into a spectrum of frequencies (which are individual sinusoids with infinite time duration), this approach is to use time-limited little waves (wavelets) as components. The wavelet has already found application on the Internet as a more efficient compression algorithm than JPEG for images. The advantage of wavelet compression is that, in contrast to DCT, this algorithm does not divide each image into blocks, but rather analyzes the whole image. The characteristic of wavelet compression is to get the best compression ratio, while maintaining quality. The JPEG 2000 image file standard includes optional use of wavelet compression.

5.2   ITU Recording and Transmission Standards

The Radiocommunication Sector of the ITU (ITU-R) has long played a role in television standards, particularly on the analog side. In moving to digital, their first consideration was in two areas: the recording studio and then digital videoconferencing (really a telecommunications service, but one that ties back into broadcasting, as we shall see later). To enter into this particular field, we start at the studios where digital was first employed and discuss a video standard that predates MPEG, namely, ITU-R BT.601 (formerly CCIR Recommendation 601). The significance of this standard is that it was adopted by the international TV production community and therefore provides an agreed baseline for digital encoding and interconnection. After this, an overview of the first widely adopted digital TV transmission standards, the ITU H. series, is presented to provide the requisite background for proper understanding of modern digital video.

5.2.1   ITU 601 Uncompressed Digital Television

ITU 601, short for ITU-R BT.601-5, which is now up to revision 5 (i.e., BT.601-5), is an international standard for component digital television from which was derived the Society of Motion Picture and Television Engineers (SMPTE) 125M and EBU 3246E standards. ITU 601 defines the sampling systems, matrix values, and filter characteristics for both Y, B-Y, R-Y and RGB component digital television. A basic agreement on this recommendation was reached in 1981 between SMPTE and the EBU through the efforts of then CCIR Study Group XI. MPEG 1 operates on noninterlaced video inputs but has been adapted to normal TV for 525- and 626-line systems. The intersection of MPEG and ITU 601 comes in the form of a transcoder between the two systems. The focus of ITU 601 is on studio and broadcast center applications where the bandwidth of communication is less of a problem. The exception to this, of course, is with respect to storage of video programs for editing and archival purposes.

ITU 601 is a standard digital video format that can accommodate the three analog TV systems in use throughout the world [2]. The application as a primary studio format for program acquisition is based on the principle of component coding and its extensibility to various analog formats. It is a family of standards rather than a single unified format. Sampling is accomplished with a common sampling rate of 13.5 MHz, which is 858 times the 525 horizontal frequency and 864 times the 625 horizontal frequency. The specific parameters are summarized in Table 5.3. Color-

Table 5.3   Main Parameters of ITU-R Recommendation 601 (4:2:2)

Analog Standard Input	*525 Line/60 Hz*	*625 Line/50 Hz*
Samples per line		
Luminance component	858	864
Color component (each)	429	432
Sample frequency		
Luminance component	13.5 MHz	13.5 MHz
Color component (each)	6.75 MHz	6.75 MHz
Samples per active digital line		
Luminance component	720	720
Color component (each)	360	360
Correspondence between number of quantizing bits and signal level	on a scale of 0–255	on a scale of 0–255
Luminance component	16 (black) to 255 (white)	16 (black) to 255 (white)
Luminance component	128 (no color) ±112 (16 to 240 full saturation)	128 (no color) ±112 (16 to 240 full saturation)

difference signals are encoded at half this rate because the amount of information contained in the luminance signal is at least twice that of each of the two color-difference signals. Frame repeat rates differ for the 525- and 626-line systems, being 29.97 and 25 Hz, respectively. These properties are reflected in ITU 601 by the relative sampling rates and the corresponding effective bit rates; the ratios that apply to the luminance channel and the U and V color-difference channels are 4 to 2 to 2, respectively (designated 4:2:2).

The standard manner by which to transfer ITU 601-encoded video is via the serial interface, which is lossless and has a bit rate of 216 Mbps. The engineering community has determined that a gross rate of 243 Mbps is to be employed to allow other forms of information to be added, particularly stereo audio. The advantage of using this format is that it is an international standard to which manufacturers and operators can comply.

ITU 601 is important because of its pioneering status as a worldwide digital video standard. With a serial information transfer rate of almost 250 Mbps, it is unlikely that ITU 601 signals will find their way into the home. Rather, it is an interface standard or baseline upon which practical digital compression systems will be evaluated.

5.2.2   The ITU H. Series Standards

Much of the development of digital video compression, particularly through the application of DCT, has been for videoconferencing and video telephony applications [3]. The greater penetration of conferencing technology in industry and government provided the motivation for the ITU to define standards to interface the codecs and transmission systems from different countries. This has lead to the H series of specifications, and in particular, the following two popular examples: H.320 Videoconferencing Using Narrowband Integrated Services Digital Networks (N-ISDN), and H.323 Videoconferencing and Telephony Using both N-ISDN and

the Internet, allowing both constant bit rate transmission and packetized transmission, based on the means available.

H.320 has promoted the use of videoconferencing throughout the developed world. The standard fixes the video and audio compression approach so that codecs supplied by different manufacturers can produce the appropriate picture and sound over a medium data rate connection (less than 2 Mbps, but typically between 128 and 512 Kbps). In addition, H.320 includes the intelligence to establish a bidirectional connection at a constant bit rate using standard N-ISDN service from the digital telephone network. The DCT structure and means of frame-to-frame (temporal compression) are very similar to that of MPEG, discussed in Section 5.3. The audio would be compressed from 64-Kbps PCM to a hybrid system such as the code-excited linear predicted (CELP) algorithm (reviewed in Chapter 11). Another function of H.320 is the provision of data for use in managing the connection and transferring other information such as graphics.

H.323 is more revolutionary in that it allows videoconferencing to be delivered through the Internet Protocol. It includes H.320 services as a subset, but goes further to provide a superior video compression technique, still based on DCT. Packetization of the video and audio is what allows the connection to be provided over the Internet. As discussed in Chapter 8, VSAT networks are a popular means of providing H.323 teleconferencing. The ability to establish end-to-end connections for video and audio is a powerful function of H.323, which allows it to provide VoIP telephone service as well.

5.3   Motion Picture Expert Group

The Motion Picture Expert Group, also affiliated with ITU-T and ISO, provides us with a solid and stable standard for full motion pictures, making use of frame-to-frame compression with associated sound and ancillary data. MPEG permits transmission of full-color, full-motion TV images at a rate as low as 1.5 Mbps. The breath of the standard is indicated in Figure 5.5. A key point is that the compression is done in real time, although there is a delay due to the need to analyze several frames of information before a sequence can be transmitted. This is not unusual for any compression scheme.

Figure 5.5   MPEG standards and application areas.

The MPEG series of standards for motion pictures and video provides many desirable features.

- The MPEG series of standards supports a wide variety of picture formats with a very flexible encoding and transmission structure (see Figure 5.5).
- It allows the application to use a range of data rates to handle multiple video channels on the same transmission stream and to allow this multiplexing to be adaptive to the source content.
- The algorithms can be implemented in hardware to minimize coding and decoding delay (typically less than 150 ms for processing).
- Developers can include encryption and decryption to comply with content restrictions and the needs for business integrity.
- Provisions can be made for an effective system of error protection to allow operation on a variety of transmission media such as satellite and local microwave links (typically handled at the application layer or physical layer).
- The compression is adaptable to various storage and transport methods, an excellent example of which is the DVB standard, discussed later in this chapter.
- The frame-to-frame compression approach with the use of intra (I) pictures (discussed later in this chapter) permits fast-forward and reverse play for editing and CD-ROM applications, impacting the degree of compression since frames cannot be interpolated if used for these features.
- Transcoding, the digital process of converting from one format into MPEG and vice versa, permits conversion between other compression formats like H.323 into MPEG.
- MPEG-processed videos can be edited by systems that support the standard.
- Random access can be allowed using the standalone frames that are DCT encoded (called I pictures—discussed later in this chapter).
- The standard will most probably have a long lifetime since it can adapt to improvements in compression algorithms, VLSI technology, motion compensation, and the like.

These properties relate to the evolving MPEG family of standards. As of the time of this writing, two of these standards were complete and already available on commercial markets. These include the MPEG 1 standard, which provides for encoding video sequences intended for CD-ROM and other multimedia applications, and MPEG 2, which is the standard for commercial digital television. The obvious focus of this book is MPEG 2, but MPEG 1 is important because it is the predecessor and provides an important technical foundation. We review each of these standards in the following sections.

5.3.1   MPEG 1

The first full-motion compression standard to be produced was MPEG 1, which is aimed at nonbroadcast applications like computer CD-ROM and Internet download. It draws from JPEG in the area of image compression using DCT and provides

a broad range of options to fit the particular application. For example, there are various profiles to support differing picture sizes and frame rates and it can encode and decode any picture size up to the normal TV with a minimum number of 720 pixels per line and 576 lines per picture. The minimum frame rate is 30 (noninterlaced) and the corresponding bit rate is 1.86 Mbps. Principal technical parameters for MPEG are listed in Table 5.4.

Because MPEG 1 is aimed at multimedia applications, there was a need to allow convenient fast-forward and fast-backward capability. This means that complete frames are needed at relatively frequent intervals to permit scanning by the user. Otherwise, the various forms of frame-to-frame compression and motion compensation would have greatly reduced the possibility to cue material at intermediate stages in a video sequence. The special multimedia features of MPEG 1 include:

- Fast-forward and fast-reverse (FF/FR);
- Reverse playback;
- Ability to edit a compressed bit stream.

These features are provided by encoding two types of pictures: intraframe (I) pictures and interpolated (B) pictures. As mentioned at the beginning of this section, I pictures are encoded individually using DCT without considering other pictures in a sequence. This is analogous to taking an image and compressing it by itself with JPEG. These pictures can therefore be decompressed individually in the decoder and displayed one-by-one, providing random access points throughout a CD-ROM or video stream. They are also the preferred format when editing of a movie is involved. Predicted (P) pictures, on the other hand, are predicted from the I pictures and incorporate motion compensation as well. They are therefore not usable for reference because they only exist in the decoder and require I pictures to be recovered properly. Another aspect is that it takes a sequence of I and P pictures to reproduce a video sequence in its entirety, which causes a delay at the receiver. If the sequence had been encoded exclusively with I pictures, then the delay is minimal, amounting

Table 5.4   Summary of MPEG 1 Technical Characteristics

Characteristic	Value or Description
Type of coder	Hybrid
Spatial transform	DCT, 8×8 block
Quantization	Separate luminance and chrominance matrices Can be user supplied User-supplied scaler used to adjust absolute level Can be varied macroblock by macroblock
Variable length code	Default: two-dimensional Hoffman Can be user supplied
Temporal compression	Motion compensation by motion vectors ±15 pixel search in both axes Difference image coding
Active pixels	Up to 4,096 by 4,096
Rate	8 frame rates, up to 72 frames per second
Raster	Progressive scan

only to the time needed to convert back from the DCT representation to the equivalent uncompressed sequence.

An example of an MPEG frame sequence is shown in Figure 5.6, displaying the relationship between three classifications of pictures (I, B, and P). As stated previously, the I pictures are standalone DCT images that can be decompressed and used as a reference. The B pictures are interpolated between I pictures and are therefore dependent on them. The P pictures (discussed under MPEG 2 in Section 5.3.2) are computed from the nearest previously coded frame, whether I or B, and typically incorporate motion compensation. Interpolated B pictures require both past and future P or I pictures and cannot be used as reference points in a video sequence.

The temporal compression of MPEG 1, based on interpolation, means that one can no longer transmit a true time-sequential stream. This is shown in Figure 5.6 for one complete intraperiod of a digital video sequence. Every transmitter and receiver requires adequate video frame memory to hold the forward I pictures and computational power to calculate P and B pictures from them. With high-speed VSLI and ASICs, the cost and complexity of accomplishing this have been reduced to a very accessible level. MPEG 1 decoders are available on chips within a CD-ROM drive itself or the display device. MPEG 1 encoders and decoders are available at low cost as computer plug-in boards for low-cost multimedia applications in education and business.

Many specific trade-offs are possible because the three types of pictures can be introduced and arranged to suit the needs of the application. The most flexible approach is to employ the I picture format exclusively because it will provide random access to any frame in the sequence. However, it is also the most expensive in terms of storage or bandwidth. A sequence with a moderate quantity of P pictures provides a degree of random access and fast-forward/fast-reverse functionality. If storage and access are not contemplated, then B pictures may be used along with the other types as this provides the greatest opportunity for bandwidth reduction.

The MPEG 1 standard was clearly a pioneering effort in the journey to a highly efficient digital TV system. It grew out of the work on computer images, which is clearly a mass market of the type to capture the attention of the leading electronics and media companies. The next step was to recognize the special needs of the broadcasting industry. However, the participants in conventional TV have long resisted change. In comparison, those in satellite and computer communications have

Figure 5.6   An example of an actual sequence of pictures used in an MPEG sequence.

experimented with digital technology for decades in the hopes of finding that killer application that would allow the development of new markets.

5.3.2   MPEG 2

The second phase of consumer digital video standards activity took the innovations of MPEG 1 and added features and options to yield an even more versatile system for broadcast TV. From standard-setting activity begun in 1992, MPEG 2 quickly became the vehicle for bringing digital video to the mass market. The purpose of this standard is to provide lossy video quality equal to or better than NTSC, PAL, and SECAM, along with the facility to support the lossless performance of ITU 601. The developers of MPEG 2 had in mind the most popular applications in cable TV, satellite DTH, digital VCRs, and terrestrial TV networks as well. In 1994, the draft specification was produced; yet, an operational system based on this standard was already being introduced in the United States by DIRECTV, Inc., a subsidiary of Hughes Electronics Corp. MPEG 2 was subsequently adopted for all digital TV broadcasting and consumer delivery throughout the world.

MPEG 2's important contribution is not in compression (that was established by JPEG and MPEG 1), but rather as an integrated transport mechanism for multiplexing the video, audio, and other data through packet generation and time division multiplexing. It is an extended version (or superset) of MPEG 1 and is designed to be backward compatible with it. The definition of the bit structure, called the *syntax*, includes a constant-rate bit stream, a set of coding algorithms, and a multiplexing format to combine video, audio, and data (including Internet Protocol data). New coding features were added to improve functionality and enhanced quality in the conventional video environment of interlaced scanning and constrained bandwidth. The system is scalable for lossless and lossy transmission, along with the ability to support HDTV standards. Robust coding and error correction are available to facilitate a variety of delivery systems including satellite DTH, local microwave distribution (e.g., MMDS), and over-the-air VHF and UHF broadcasting.

Because delivery systems and applications differ widely, MPEG 2 provides a variety of formats and services within the syntax and structure. This is the concept of the *profile*, which defines a set of algorithms, and the *level*, which specifies the range of service parameters that are supported by the implementation (e.g., image size, frame rate, and bit rate). The profiles and levels are defined in Tables 5.5 and 5.6, respectively. Table 5.5 begins at the lowest profile called *SIMPLE*, which corresponds to the minimum set of tools. Going down the table adds functionality.

The *Main* profile is the current baseline for MPEG 2 applications and has been implemented in a number of DTH systems. As suggested in Table 5.5, this profile does not include scalability tools and therefore is a point of downward-compatibility from the higher levels that provide scalability. The scalability tools for SNR and Spatial profiles add power to the standard for applications. A general implementation of a scalable MPEG 2 system is diagrammed in Figure 5.7. The other dimension of MPEG 2 takes us through the *Levels*, which provide a range of potential qualities from standard definition (SDTV) to HDTV. There is an obvious impact on the bit rate and bandwidth. These levels are reviewed in Table 5.6.

Table 5.5   Profiles and the Associated Algorithms for the MPEG 2 Standard

Profile	Algorithms
Simple	Provides the fewest tools but supports the 4:2:0 YUV representation of the video signal.
Main	Starts with Simple and adds bidirectional prediction to give better quality for the same bit rate. It is backward compatible with Simple as well.
Spatial Scalable	This profile includes tools to add signal quality enhancements. By Spatial it is meant that the added signal complexity allows the receiver to improve resolution for the same bit rate. There would be an impact on the receiver in terms of complexity and hence cost. It can also be a means to add HDTV service on top of conventional resolution (i.e., only the appropriately designed receivers can interpret and display the added HDTV information).
SNR Scalable	The added signal information and receiver complexity improve viewable S/N. It provides graceful degradation of the video quality when the error rate increase.
High	The is includes the previous profiles plus the ability to code line-simultaneous color-difference signals. It is intended for applications where quality is of the utmost importance and where there is no constraint on bit rate (such as within a studio or over a dedicated fiber optic link).

Table 5.6   Levels and the Associated Parameters for the MPEG 2 Standard

Level	Sample/Line	Lines/Frame	Frames/sec	Mbps	BW (MHz)*
High	1,920	1,152	60	80	54
High 1440	1,440	1,152	60	60	40.5
Main	720	576	30	15	10.125
Low	352	288	30	4	2.7

*The bandwidth (BW) indicated in the last column is based on direct QPSK modulation of the bit stream and does not include forward error correction coding.

These levels are associated with the format of the originating source video signal and provide a variety of potential qualities for the application. It ranges from limited definition and the associated low data rate all the way up to the full capability of HDTV. Another feature of the standard is that it permits the normal TV aspect ratio (width to length) of 4:3 as well as the "letter box" movie screen or HDTV aspect ratio of 16:9. This particular part of MPEG 2 covers the base input and does not consider the degree of compression afforded by the profiles covered in Table 5.5. The basis of each of the levels is as follows.

- The Low level is an input format that is only one-quarter of the picture defined in ITU 601.
- The Main level has the full 601 input frame format.
- The High-1440 level is the HDTV format with 1,440 samples per line.
- The High level is an even better HDTV format with 1,920 samples per line.

The tools and formats of MPEG 2 allow as many as 20 different combinations, which the standard calls *convergence points*. As with any such standard, not every combination is either useful or viable. At the time of this writing, the Main profile and Main level represent the convergence point of all practical implementations. This is the case in North America with the various systems already in use as well as with the European DVB standard, which will be discussed in another section.

Figure 5.7   Conceptual configuration of scalable video encoding, to allow both a high-resolution and a low resolution mode of transmission and decoding. For low-resolution, only the base layer bit stream is required at the receiving end. Enhancement layer bits are transmitted to permit compatible receivers to display full quality video at high resolution.

To summarize, MPEG 2 is an attractive digital video standard that was developed for wide consumer application. Its core algorithm at the Main profile features nonscalable coding for both progressive and interlaced video information sources (e.g., broadcasting and specialized applications like video teleconferencing and multimedia). The Main level further specifies an input, which meets the needs of commercial television at 25 or 30 per second with an input rate of 15 Mbps. This, or course, is compressed down to as little as 1.5 Mbps, based on the trade-off between transmission cost and application quality.

5.3.3   MPEG Audio

The other important element of MPEG 2 is the provision of stereo audio. Most quality audio compression systems are based upon one of two basic technologies:

- Predictive or adaptive differential PCM (ADPCM) time domain coding;
- Transform or adaptive PCM (APCM) frequency domain coding.

PCM and ADPCM have been with us for many decades and provide little effective compression for high-quality audio. Transform coding turns out to be as effective for sound as it is for picture; however, the human subjective aspects are considerably different. In this case, we are concerned with the way the sound is received and then interpreted by the brain. We must consider both the intelligibility of voice communication along with the enjoyment of other sounds, particularly music. This is the general subject of psycho-acoustics, which is the process by which

the human brain is able to choose between the multitude of sounds that excite the ears. Figure 5.8 presents how humans react to sound over the frequency range of hearing: approximately 20 Hz to 20 kHz. Within this range, we are most sensitive to the frequencies in the center and relatively insensitive to sounds at the low and high ends. This means that anything at the extremes has less of an impact and therefore need not be as precisely conveyed as those in the center. Likewise, a loud sound near the center will mask those that are to the extremes in frequency. The other aspect relates to the loudness of sound, where our hearing is somewhat deafened for a brief period by a loud and abrupt sound. This range of algorithms includes the industry standards ISO/MPEG 1 and 2; Layers 1, 2, and 3 (the highly-recognized MP3 format); MPEG AAC; MUSICAM; Dolby AC2 and AC3, apt-Q, and others [4].

The audio compression system employed in MPEG 2 is based on the European MUSICAM standard as modified by other algorithms [2]. It is a lossy compression scheme that draws from techniques already within MPEG. Like differential PCM, it

(a)

(b)

Figure 5.8   (a) Human hearing response to threshold in quiet and spectral masking; and (b) temporal masking effect by human hearing.

transmits only changes and throws away data that the human ear cannot hear. This information is processed and time division multiplexed with the encoded video to produce a combined bit stream that complies with the standard syntax. This is important because it allows receivers designed and made by different manufacturers to be able to properly interpret the information. However, what the receiver actually does with the information depends on the features of the particular unit.

5.3.4   Assessing MPEG 2 Video Quality

The quality of the video produced by the MPEG 2 standard can be as good as the provider wishes it to be. For reasonable data transfer rates, between 3 and 6 Mbps, the SD version MPEG 2 generally wins acclaim, particularly when compared to what existing analog cable and over-the-air channels were capable of providing. The old TASO grade scale was really invented for these analog systems, where S/N was the primary measure. However, in digital television, the quality is actually set at the source during the encoding and compression process—the transmission link has little to do with it.

The DCT approach to compression produces a well-understood "blockiness" to the picture, akin to what we are accustomed to with JPEG images. Use of motion compensation in conjunction with B and P pictures adds another level of distortion to what the viewer experiences. With greater compression comes increased incidence of jerkiness to the picture sequence. On occasion, the decoder may be unable to stay up with the stream and actually break lock. These issues are being addressed through improved processing on the sending end.

To evaluate digital television, the ITU-R has come up with more standards for subjective evaluation in the form of Recommendation ITU-R BT.500-10, "Methodology for the Subjective Assessment of the Quality of Television Pictures." This detailed document describes both the subjective criteria and the procedures for gathering the data using human observers. As with similar approaches for subjective evaluation of fixed and mobile telephone, it is complex and sometimes inconclusive. The ITU-R continues to work to resolve uncertainty in the methods. The traditional approach is to have human subjects experience two different test exposures—perhaps one with the system being evaluated and the other with some kind of control. This is called dual stimulus (DS). Currently, they suggest the use of a single stimulus continuous quality evaluation (SSCQE) approach:

> The introduction of digital television compression will produce impairments to the picture quality what are scene-dependent and time-varying. Even within short extracts of digitally-coded video, the quality can fluctuate quite widely depending on scene content, and impairments may be very short-lived. Conventional ITU-R methodologies alone are not sufficient to assess this type of material. Furthermore, the double stimulus method of laboratory testing does not replicate the SSCQE home viewing conditions. It was considered useful, therefore, for the subjective quality of digitally-coded video to be measured continuously, with subjects viewing the material once, without a source of reference.

Compared to dual stimulus, use of a single stimulus (SS) approach, increases the number of test subjects and tests. This increases cost and tends to extend the test program. Another issue identified by the ITU-R is that once a subject has

experienced digitally encoded video, their personal standard of comparison goes up. In other words, once they get used to digital TV, they do not so easily accept analog.

There is still value to testing the impact of different transmission rates and compression systems. For example, a DTH operator would want to know at what rate it should encode film movies as opposed to American football or European soccer. To this end, Tektronix has developed the PQM 300 video evaluation equipment to aid in the following:

- Detection of picture defects such as:
    - MPEG blockiness;
    - Repeated frames;
    - Uncorrelated Gaussian noise.
- Aids in detecting frozen frames and loss of service;
- Facilitate consistent levels of performance at any point within the network;
- Helps to manage available bandwidth.

This device came out of research supported by the ITU-R and could be valuable for fine tuning the data rates and multiplexing of a large quantity of video channels (a process called *grooming*).

Quantitative measurements of MPEG 2 quality amount to assessing the digital transmission characteristics of the bit stream and the detailed performance at higher layers of the protocol stack. This is not a new area in telecommunications as techniques for this sort of evaluation have been available since the early days of digital telephone and data communications. However, what is new is that the techniques are applied in commercial TV, an area that has been heavily analog. The properties that matter are:

- Bit error rate, which defines the floor where the decoder can function properly;
- Timing jitter, which like bit error rate can cause the decoder to malfunction. This can be occasional (causing blips in the picture) or absolute disruption;
- Improper actions of the MPEG protocol that can only be resolved through detailed review of packets and the information contained therein.

BER is traditionally measured by originating a repetitive pseudo-random-noise (PRN) bit sequence of a fixed duration through the communication channel. At the receiving end, the already-known PRN sequence is compared with what was sent (e.g., since it was known ahead of time) and the errors counted. The BER is continuously computed by dividing the number of errors by the associated increment of time. At a bit rate of 10 Mbps and an error rate of 10^{-7}, we would receive one error per second, on average. A lower error rate would reduce the average by less than one error per second; therefore, the period of observation would have to be increased. For this reason, one must allow sufficient time to collect statistics in order to obtain a meaningful result.

Errors are produced by a few fundamental mechanisms in the satellite link:

- Noise in receivers (e.g., thermal noise);

- Interference (adjacent channel, cross-polarization, intermodulation distortion, adjacent satellite, and terrestrial);
- Distortion of the signal by the uplink equipment and the satellite transponder or OBP (discussed in Chapter 3).

BER can only be measured through the end-to-end channel, making it useful for on-line service monitoring. Noise and interference can be observed using a conventional spectrum analyzer. The distortion properties of the channel can be isolated down to the uplink, satellite, and downlink equipment using the technique of the *eye diagram*, shown in Figure 5.9. This is nothing more than a waveform of a symbol captured by an oscilloscope. One can observe the actual pulse shaping of individual signals (e.g., encoded bits on the carrier waveform) along with the distortion and noise that disruption proper detection in the receiver. The clear area within the curve is called the *eye opening*: the larger the opening, the better the quality of transmission. This picture provides a qualitative and even quantitative measure of service quality of the analog IF and RF transmission link.

The other important measurement parameter is timing jitter. Because MPEG uses a synchronous transmission plan (e.g., a constant packet size and bit rate), any short-term variation can be cause for concern. The key characteristic in this case is the program clock recovery (PCR) function of the decoder, which must stay within a strict bound. Once PCR timing varies outside of this allowable range for the particular service, decoding is disrupted and the picture and audio can break up. This unacceptable variation in timing is called PCR jitter. A single-channel MPEG 2 stream that is properly encoded will have little if any PCR jitter. However, jitter is introduced through subsequent multiplexing and data insertion, if applied. Steps such as reclocking must be taken during such processes to control PCR jitter.

As one would expect, the test and measurement industry, led by companies like Tektronix, Agilent, and Thomson Grass Valley Group, has produced very capable instruments. For example, the Tektronix MTS 300 portable tester provides all of the following features:

- Real-time monitoring and compliance testing of MPEG, DVB, ATSC, and ISDB transport streams;
- Status and error logging to capture intermittent problems or create test records;

1 ns/div

Figure 5.9   Typical eye diagram.

- Dolby digital AC-3 compliance testing and AAC stream monitoring for testing advanced audio;
- PCR overall jitter, drift, and offset measurements;
- Detailed off-line analysis of transport streams, program streams, and elementary streams;
- Capture, playback, and on-line storage of transport, program, and elementary streams.

5.3.5   MPEG 4

MPEG established strong credibility in the digital video and content field; so, the question is, what do they do for an encore? This is being answered by a new standard called MPEG 4, first introduced in 1991, which has the formal name of "Coding of Audiovisual Objects." An example of such an object would be the oft-seen announcer's face, which it turns out can be very efficiently coded with a few wireframe-based facial animation parameter (FAP) words. Other examples would be common objects such as vehicles and buildings.

The expectation of this object-oriented standard was to overcome the limitations in bit rate and interactivity that MPEG 1 and MPEG 2 had offered. MPEG 4 is much more than just data compression. It was primarily aimed at low bit rate communications; however, its extent was expanded further to have a number of different technologies and cover a broad range of applications. MPEG 4 gives the user the ability of manipulating the audiovisual objects in a scene. These objects could be parts of video- or computer-generated models. The user can actively interact with the objects and modify scenes by adding, removing, or repositioning them in a scene. The standard uses a language called binary format for scenes (BIFS) for scene composition. This language is based on the concepts developed in virtual reality modeling language (VRML), and includes features such as facial animation, native two-dimensional primitives, and streaming protocols. Using BFIS for real-time streaming allows scenes to be built up on the fly with no need to be downloaded in full before their display. Like MPEG 2, MPEG 4 conformance is defined in terms of profiles and levels depending on the visual object types supported and the bit stream parameters.

Because MPEG 4 coding is descriptive it also provides the standardized elements enabling the integration of production, distribution, and content access paradigms. As descriptive coding is typically an order of magnitude more efficient than DCT coding, MPEG 4 is positioned to achieve for the Internet what MPEG 2 did for broadcasting. It supplies tools with which to create uniform audio and video encoders and decoders that compete effectively with proprietary schemes, including Quick Time (from Apple Computer), AVI (from Microsoft), and RealOne (from Real Networks).

MPEG 4's general video compression schemes have already been discussed in the context of the DCT, fractals, and wavelets. It enhances these with different types of "objects" in much the same way as a Web page is composed using HTML and other Web authoring tools. Figure 5.10 provides a simple graphic of how various types of content (video, audio, graphic, text) are pulled up to the PC client through a variety of different networks—the Internet, an MPEG 2 broadcast, a broadband

Figure 5.10 MPEG 4 transport protocols and interface layers. (*From:* [5]. © 1999 IEEE. Reprinted with permission.)

ATM network, the telephone network, and so forth [5]. As one might expect, this is a tall order as there are many issues concerning the interfaces among these facilities and the manner in which the content is put together into something useful and enjoyable by the public. As a result of the complexity of the problem, MPEG 4 has just finished development at the time of this writing. The promise is of a style of Internet-motivated content distribution that offers greater meaning than current static Web pages, that employs any network bandwidth efficiently and that is as open as the Internet Protocol suite itself.

The first introduction of MPEG 4 was slow in coming; some of the hesitancy was due to a lack of a profound application. In the late 1990s, Lockheed Martin, EchoStar, and others supported demonstrations of MPEG 4 applications via satellite. These were well publicized but not made available to public audiences. More recently, a number of small technology companies that are members of the MPEG 4 Industry Forum (http://www.mp4.com) have introduced a variety of solutions based on the standard. These are software tools and chip sets that can be made part of complete systems to originate and delivery MPEG 4 content. Gaining initial favor was broadband wireless, where several companies have developed chip sets to provide simultaneous voice and video for such applications as the wireless video phone. Among the large companies, Toshiba engineers in the semiconductor division have worked with the standards body to come up with software and silicon that provide both the encoding and decoding functions. Toshiba offers an MPEG 4 video encoder and decoder (CODEC) with 12 MB of embedded DRAM to deliver a low-power, end-to-end solution with encoding, transmission, and decoding functionality. To support these products, the company has developed firmware, drivers, middleware, and development tools that include reference boards. Toshiba's single-chip MPEG 4 products are intended for wireless systems including video phones, streaming wireless media players, security/surveillance equipment, wireless LAN applications, and other emerging Internet, communications, or multimedia storage areas.

Some interesting chip hardware is offered by Amphion (formerly Integrated Silicon Systems), a small system on a chip (SoC) developer based in Northern Ireland. Using the standard approach to MPEG and relying on the DCT as video compression algorithm along with motion estimation, Amphion makes available both encoding and decoding MPEG 4 chip designs. Intended applications are:

- 3G mobile phones:
  - Short video clips;
  - Multimedia Messaging Service (MMS).
- Wireless personal digital assistants (PDAs) with video e-mail;
- Personal digital video recorder;
- Digital cameras;
- Hard-disk camcorders;
- Videoconferencing;
- Surveillance;
- Video phones;
- Mobile multimedia.

MPEG 4 is very different from MPEG 2 because it is object oriented to suit systems that are inherently low power. The latter refers to uses in handheld devices like PDAs and advanced cell phones.

More recently, MPEG 4 has been identified as a platform for new services via cable TV. This requires development of a new generation of STB to employ the appropriate software for VoD and interactive services. A joint venture of several

Japanese consumer electronics firms along with a U.S. startup called iVAST and others intends to make the technology operational by 2004. According to a March 2002 press release:

> Seven leading consumer electronics and technology companies today announced the formation of a new corporation in Japan, e-BOX, that will enable cable television system operators to deliver enhanced video on demand and interactive TV services to subscribers. Based in Tokyo, the joint venture includes five publicly traded companies: Pioneer Corporation, Sharp Corporation, National Semiconductor Corporation, Sigma Designs, and CMC Magnetics; and two privately held companies, iVAST, Inc. and Modern VideoFilm Inc.

Comcast Cable Communications, Inc., the largest cable company in the U.S., is advising the joint venture partners on the technical requirements of the system and intends to conduct field trials of the new services. Targeting North American and Asia-Pacific cable operators, the joint venture partners will deliver a complete infrastructure that includes head-end equipment, system software, content-protection systems and digital set-top boxes to provide scheduled and on-demand MPEG-4-encoded content and interactive services.

iVAST is based in the Silicon Valley of California and offers software tools for MPEG 2 production and distribution. iVAST systems can deliver MPEG 4 content over diverse networks—Internet or intranet, wireless or wireline, broadcast or broadband. iVAST solutions are compatible with industry-standard network and messaging protocols.

The rather long preapplication period of MPEG 4 has allowed a new approach to rise up and gain attention. Known as ITU-T standard H.23L, it relies on improved motion compression to further reduce the required bit rate. This has resulted in the MPEG to create yet another version of itself that employs H.23L: MPEG 4 part 10. There appears to be a split developing in the cable STB sphere. Some have decided to proceed with the development of straight MPEG 4 solutions, based on technology supplied by iVAST. Others, however, are taking a wait and see approach while the details of part 10 are worked out. At the time of this writing, a satellite-based application was not on the immediate horizon. If and when MPEG 4 gains a base in either the 3G wireless or cable TV markets, one can expect to see it attach to services such as DTH, broadband VSAT, and MSS.

5.4   Digital Video Broadcasting Standard

The MPEG series of standards was the result of international cooperation among world-class organizations from several continents. From this, engineers and manufacturers must create the specific implementations and products that allow the public to enjoy the versatility of digital video. The first MPEG-based products were created for the U.S. market, but, as has been the usual case, these systems are incompatible with each other. At the same time that pioneers were addressing the most attractive single consumer market in the world, their counterparts in Europe set out to build a better mousetrap and to do it in a way that a unified approach might be produced. This is the effort that has resulted in DVB, a family of standards for DBS,

cable TV, and over-the-air broadcasting that dominate the global landscape in digital video as GSM does in mobile telephone.

5.4.1   DVB Requirements and Organization

We cover DVB in some detail because of its technical relevance, openness, and success in the satellite communication field. The DVB system is intended as a complete package for digital television and data broadcasting [6]. It is built on the foundation of the MPEG 2 standard, providing full support for encoded and compressed video and audio, along with data channels for a variety of associated information services. The MPEG standard provides for data stream syntax, discussed in the previous sections, to multiplex the required functions together. On top of this, the DVB standard considers the modulation and RF transmission format needed to support a variety of satellite and terrestrial networking systems.

The overall philosophy behind DVB is to implement a general technical solution to the demands of applications like cable TV, as discussed in Chapter 4, and DTH, to be discussed in Chapter 6.

It includes the following features:

- Information containers to carry flexible combinations of MPEG 2 video, audio and data;
- A multiplexing system to implement a common MPEG 2 transport stream (TS);
- A common service information (SI) system giving details of the programs being broadcast (this is the information for the on-screen program guide);
- A common outer block coding scheme using the Reed-Solomon (RS) forward error correction system that improves the reception by providing a low error rate;
- Inclusion of energy dispersal to maintain spectral spread and interleaving to improve performance in the presence of burst errors;
- A flexible inner convolutional coding scheme using primarily the Viterbi algorithm with the ability to adjust the code rate between R = 1/2 and R = 7/9 (rates higher than 1/2 are achieved through puncturing);
- Modulation and additional channel coding systems, as required, to meet the requirements of different transmission media (including FSS and BSS satellite delivery systems, terrestrial microwave distribution, conventional broadcasting, and cable TV);
- A common scrambling system;
- A common conditional access (CA) interface (to control the operation of the receiver and assure satisfactory operation of the delivery system as a business).

The origin of DVB is a pan-European program of industrial and government cooperation that began in 1990. Over the course of a year, the group expanded to include consumer electronics manufacturers and common carriers. A Memorandum of Understanding (MoU) was signed in 1993 that established DVB as a set of standards, with digital satellite and cable TV drawing the most immediate attention.

As of May 1995, almost 200 companies and agencies had signed the MoU, many from the United States (including AT&T, CLI, DEC, General Instruments, Hewlett Packard, Hughes Electronics, Motorola, and Texas Instruments), Japan (including NEC, Mitsubishi Electric, Pioneer, and Sony), and of course Europe (including ALCATEL, the BBC, EUTELSAT, France Telecom, News Corp., Nokia, RTL, Thomson, and ZDF). While its use in Europe was a given, a big step for DVB-S was its U.S. adoption for the DISH Network of EchoStar.

The basic transmission design of DVB-S has proven to be robust and economical in use. As a result, DVB-S modulators and ancillary equipment have found applications outside of the strict definition of broadcasting. One of these is as the outbound transport of wideband data at speeds between 1.5 and 155 Mbps. In addition to the inherent technical features of this standard (such as the use of concatenated coding and QPSK), the property of MPEG 2 to transfer Internet Protocol data efficiently and transparently has gained a large following. This will be addressed further in Chapters 8 and 9.

5.4.2 Relationship Between DVB and MPEG 2

From the outset, DVB followed the spirit and the letter of MPEG 2. This meant that a close tie was needed, which was possible due to the association among the organizations that signed the MoU and those that were part of MPEG. The group specified a family of DVB standards, including the following.

- DVB-S: the satellite DTH system for use in the 11/12-GHz BSS band, configurable to suit a wide range of transponder bandwidths and EIRPs (the standard is also applied in C, Ku, and Ka FSS bands);
- DVB-C: the cable delivery system, compatible with DVB-S and normally to be used with 8-MHz channels (e.g., consistent with the 625-line systems common in Europe, Africa, and Asia);
- DVB-CS: the satellite master antenna TV (SMATV—pronounced "smat-vee") system, adapted from the above standards to serve private cable and communities;
- DVB-T: the digital terrestrial TV system designed for 7- to 8-MHz channels;
- DVB-SI: the service information system for use by the DVB decoder to configure itself and to help the user navigate the DVB bit streams;
- DVB-TXT: The DVB fixed-format teletext transport specification;
- DVB-CI: The DVB common interface for use in CA and other applications;
- DVB-RCS: The return channel by satellite scheme being advanced as a mechanism for two-way interactive services within a general broadcast context (further discussion of DVB-RCS can be found in Chapter 8).

5.4.3 The Satellite Standard

Since the topic of this book is satellite communication applications, the basic standard of most interest is DVB-S. (The requirements for DVB-SI are covered later in this section.) It provides a range of solutions that are suitable for transponder

bandwidths between 26 and 72 MHz, available in BSS and FSS satellite systems. The basis of transmission is a single carrier that has multiple digital video and audio channels multiplexed onto it. There is nothing to preclude DVB-S from being used for single video carrier per transponder services, however.

DVB-S is a layered transmission architecture [7]. At the highest layer we find the MPEG 2 payload, which contains the useful bit stream. As we move down the layers, additional supporting and redundancy bits are added to make the signal less sensitive to errors and to arrange the payload in a form suitable for broadcasting to individually owned IRDs. The system uses QPSK modulation and concatenated error protection based on a convolutional code and a shortened RS code. Compatibility with the MPEG 2-coded TV services, with a transmission structure synchronous with the packet multiplex, is provided. All service components are time division multiplexed on a single digital carrier at a constant bit rate. Bit rates and bandwidths can be adjusted to match the needs of the satellite link and transponder bandwidth and can be changed during operation.

The video, audio, and other data are inserted into payload packets of fixed length according to the MPEG Transport System packet specification. This top-level packet is then processed as follows (adding additional bits).

- The payload is converted into the DVB-S structure by inverting synchronization bytes in every eighth packet header (the header is at the front end of the payload). There are exactly 188 bytes in each payload packet, which includes program-specific information so that the standard MPEG 2 decoder can capture and decode the payload. These data contain picture and sound along with synchronization data for the decoder to be able to recreate the source material.
- The contents are then randomized according to a predetermined PRN code. This assures that the resulting RF spectrum is always spread smoothly across the occupied bandwidth.
- The first stage of FEC is introduced with the outer code, which is in the form of the RS fixed-length code. This adds 12% of overhead bits. RS is the most common outer code in use in this type of application.
- A process called convolutional interleaving is next applied, wherein the bits are rearranged in a manner to reduce the impact of a block of errors on the satellite link (due to short interruptions from interference or fading). Convolutional interleaving should not be confused with convolutional coding, which is added in the next step. Experience has shown that the combination of RS coding and convolutional interleaving produces a very robust signal in the typical Ku-link environment. The end-to-end process of the DVB-S system is illustrated in Figure 5.11.
- The inner FEC is then introduced in the form of punctured convolutional code. The amount of extra bits for the inner code is a design or operating variable so that the amount of error correction can be traded off against the increased bandwidth. This permits the coding rate, R, to be varied between 1/2 and 7/8 to meet the particular needs of the service provider. Specified values of R and the typical values of required E_b/N_0 are indicated in Table 5.7.

Table 5.7 Coding Rate, *R*, and Typical Values of Minimum E_b/N_0 for Use in the DVB-S Standard IRD (the Link Is Assumed to Operate at a 2×10^{-4} Error Rate)

Inner code rate, *R*	1/2	2/3	3/4	5/6	7/8
E_b/N_0 (dB)		3.3	3.8	4.3	4.8	5.2

Note: Implementation margin of 1 to 3 dB is required to account for nonlinearity.

- The final step is at the physical layer where the bits are modulated on a carrier using QPSK using root-raised-cosine filtering to set the bandwidth at 1.35 times the symbol rate. The variables available to the service provider under DVB cover the multiplexing of individual video and audio channels along with key aspects of the link (i.e., coding and interleaving). Burst errors are compensated for through the randomization process, and the amount of FEC is adjusted to suite the frequency, satellite EIRP and receiving dish size, transmission rate, and rainfall statistics for the service area. The system, therefore, can be tailored to the specific link environment, which was discussed in Chapter 2. The standard DTH operator would consider a range of availability requirements, including 99.7%, 99.9%, and 99.99%.

Block diagrams of basic DVB transmitting and receiving (IRD) systems are provided in Figure 5.11. An example of a typical transmission design, indicating all of the features needed to encode, process, filter, and modulate the composite MPEG 2 signal, is outlined in Table 5.8.

Further details concerning the options for bit rates and transponder bandwidths are provided in Table 5.9. Figure 5.12 gives a typical example of the coding performance on the satellite link from end to end.

There is an obvious relationship in a business sense between DVB-S and the DVB-C standard that applies to cable TV delivery. This is because both services rely on much the same programming and address very similar markets. In DVB-C, the physical layer is different because of the nature of the cable environment as compared to satellite transmission. Here, there is essentially no fading but the bandwidth available per channel is potentially less. Cable also can pass amplitude variations and therefore can support hybrid modulation systems.

The designers of DVB-C therefore chose a hybrid modulation method that combines both phase and amplitude modulation. The modulation is based on quadrature amplitude modulation (QAM) and no inner FEC is applied (because the error rate on cable would be lower and more stable as well). The system can support varying levels of QAM, including 16, 32, and 64 QAM. A typical cable system with 8-MHz video channels can accommodate a 38.5-Mbps information rate with 64 QAM.

5.4.4 Supporting DVB Services—Sound, Service Information, and Conditional Access

There are three supporting areas of DVB that are particularly important to the operation and success of a DTH system. These are the arrangements for high-quality stereo audio, provision of service information (DVB-SI), and the common interface

Figure 5.11   The DVB-S baseline system: (a) transmitter and (b) receiver.

Table 5.8 The DVB-S Baseline System

a)	Synchronization inversion and energy dispersal -Synchronous scrambling -Every eighth synchronization byte is inverted to synchronize the deinterleaver and RS-decoder -No additional synchronization byte -Phase ambiguity recovering
b)	Outer coding -Shortened Reed-Solomon code: RS (204, 188)
c)	Interleaving -Forney interleaver -Interleaving depth: I = 12
d)	Inner coding -Mother convolutional code: 1/2 -Punctured convolutional code: 2/3; 3/4; 5/6; 7/8 -Constraint length K: 7
e)	Baseband filtering -Square-root-shaped cosine filter, 35% roll-off
f)	Modulation -QPSK Gray coded

Table 5.9 Examples of Bit Rates Versus Transponder Bandwidth

BW (at −3 dB) (MHz)	BW (at −1 dB) (MHz)	R_s (for BW/R_s = 1.28) (Mbaud)	R_u (for QPSK+ 1/2 convol) (Mbps)	R_u (for QPSK+ 2/3 convol) (Mbps)	R_u (for QPSK+ 3/4 convol) (Mbps)	R_u (for QPSK+ 5/6 convol) (Mbps)	R_u (for QPSK+ 7/8 convol) (Mbps)
54	48.6	42.2	38.9	51.8	58.3	64.8	68.0
46	41.4	35.9	33.1	44.2	49.7	55.2	58.0
40	36.0	31.2	28.8	38.4	43.2	48.0	50.4
36	32.4	28.1	25.9	34.6	38.9	43.2	45.4
33	29.7	25.8	23.8	31.7	35.6	39.6	41.6
30	27.0	23.4	21.6	28.8	32.4	36.0	37.8
27	24.3	21.4	19.4	25.9	29.2	32.4	34.0
26	23.4	20.3	18.7	25.0	28.1	31.2	32.8

Note 1: R_u stands for the useful bit rate after MPEG 2 MUX. R_s (symbol rate) corresponds to the −3-dB bandwidth of the modulated signal.

Note 2: The figures of column 1 correspond to the E_b/N_0 degradation of 1.0 dB [with respect to additive white Gaussian noise (AWGN) channel] for the case of 0.35 roll-off and 2/3 code rate, including the effects of IMUX, OMUX, and TWTA.

for conditional access (DVB-CI). The following sections review each of these areas and how they relate to the overall system.

5.4.4.1 MPEG 2 Sound Coding

Sound coding in DVB provides program audio at bit rates reduced over raw digitized audio. Claims that it is of CD quality are perhaps exaggerated, but the notion is supported by its acceptance by users as an improvement over what analog TV offers. It is based on the MPEG Layer II MUSICAM standard, which is being applied to a variety of digital audio products from manufacturers in the United States, Asia, and Europe. The digital compression of audio takes advantage of the fact that a sound

Figure 5.12   Demodulation and error correction performance of a typical DVB-S system.

element will have a masking effect on other nearby sounds that are at a lower level of volume. White noise has the same effect. This is used to increase the compression by not sending this unheard information. Thus, MUSICAM provides sound quality that is very close to that of the familiar audio CD and can be used for digital channels that provide stereo, mono, multilingual sound, and surround sound.

5.4.4.2   DVB-SI Service Information

The DVB-SI portion of the transmission adds information that groups the individual video/audio services into categories and allows the IRD to tune to a particular service. This facilitates the creation of the now-familiar on-screen menu of programming, also called the electronic program guide (EPG). Relevant schedule information and descriptions of the programs are broadcast over the same link with the video and thereby provide the EPG directly to viewers. The typical DVB environment will support hundreds of video channels and other options, so the DVB-SI standard and resulting EPG are vital to delivering a service that subscribers will find both entertaining and usable.

5.4.4.3   Conditional Access

Any DTH system may achieve its goals in a business sense only if set-top boxes are controlled so that users get those programming and information services that they are authorized to receive. CA starts with the initial granting of access when the user first subscribes but must be extended to cover the particular set of services. These can change from time to time as the user adds and delete services, including PPV movies and events and e-mail services. Another CA element to this is support of a

variety of controls and restrictions that depend on intellectual property rights (copyrights in particular), local government regulatory controls, and limits that are self-imposed by the subscriber (such as blocking adult material in a family environment). A particular threat to the financial integrity of an otherwise viable DTH service is piracy. These aspects are covered next as they have been addressed in the DVB CA package.

The committee that defined DVB came to a consensus on the following seven points that the CA package was to address.

1. Two basic options are available, namely, a single CA system (the "Simulcrypt" route) and a common interface that allows for the use of multiple CA systems (the "Multicrypt" route). A key difference between the two schemes is that in Multicrypt the entire video stream goes through a removable smart card, whereas in Simulcrypt the video stream is decoded within the installed circuitry of the set-top box. An IRD produced with Simulcrypt would only work on a network that is set up for this CA arrangement, whereas the Multicrypt route permits an IRD to be able to work with several different networks through the facility of the smart card;

2. The definition of a common scrambling algorithm and its inclusion, in Europe, in all receivers, which enables the concept of a single receiver per home even if different services are to be subscribed to;

3. The drafting of a code of conduct for access to digital decoders, applying to all CA providers;

4. The development of a common interface specification;

5. The drafting of antipiracy recommendations to help track down and prosecute pirates;

6. The licensing of CA systems to manufacturers should be on fair and reasonable terms and should not prevent the inclusion of the common interface;

7. The CA systems should allow for simple transfer of control so cable operators can replace the CA data with their own.

The ability of the CA system to control and regulate usage is highly dependent on the security offered by the DVB common scrambling system. At the IRD, this consists of two parts: decryption and descrambling. The purpose of decryption is to translate the scrambling key that is transmitted over the satellite link along with the programming and other data. This is the most efficient and effective means of delivering the scrambling key. However, using the same broadcast link for this purpose greatly simplifies the pirate's job, since they have continuous access to this critical information. The encryption/decryption process was designed to make its compromise as difficult as possible. The ultimate success depends on the strength of the system coupled with the ability of the operator to modify the technique in response to a compromise of security. Because of the sensitive nature of this information, the technical details of decryption are tightly controlled and are only available through a rigorously enforced system of confidentiality agreements and custodians.

Simulcrypt allows the delivery of one program to a number of different decoder groupings that contain different CA systems. It also provides for the transmission

between different CA systems in any grouping, in particular for the recovery after compromise of the particular implementation by pirates. Multicrypt is the approach that puts the intelligence of the encryption system on a separate module such as a smart card. The CA data is broadcast in the MPEG 2 syntax structure through the common interface (DVB-CI). This interface physically lies between the DVB decoder that recovers the payload and a CA module within the IRD that provides the decryption and descrambling.

When we examine the approach taken by the developers of the DVB family of standards, we come to realize that an excellent structure has been created. In many ways, it mirrors the popularity of the GSM digital cellular standard, which resulted from a lot of hard work by European organizations. DVB facilitated the early introduction of digital video throughout the satellite communications industry. Work from this point involves incremental improvements in transmission performance through turbo coding as well as the long awaited rollout of the DVB-RCS standard for interactive multimedia via satellite.

5.5   Data Broadcasting and Internet Protocol Encapsulation

The dominance of the Internet Protocol in the world of data communications has also made it the preferred mechanism for data broadcasting over satellites. Historically, data broadcasting was a specialized application, used in expensive subscription services and by financial institutions of various types. Using proprietary standards or just plain ASCII character data, data broadcasting has been used to deliver the stock market ticker, cattle and precious metals prices, elevator music, and nationwide paging messages. The developers of MPEG made the excellent observations that its efficient digital transport scheme could be adapted to carry IP (or vice versa, as the case more accurately can be described). IP through MPEG 2 is fast reaching critical mass as a key satellite application, used for broadband access services and, of course, the old standby of data broadcasting. We review the principles of IP Encapsulation (IPE), the process and standard that is available through MPEG 2, and almost automatically, with DVB as well.

5.5.1   IP Encapsulation in the MPEG Transport Stream

The process of IPE is simple on the surface: take IP packets and insert them into the MPEG time frame. This is made complicated by the following aspects of the problem:

- MPEG employs a constant bit rate frame, with data divided into fixed packets of 188 bytes each;
- IP employs variable length packets whose size depends on the nature of the information transfer as well as the current properties of the communications channel (error rate, time delay, lost packet rate, and so forth).

There are some additional factors that make this a collaborative union. While MPEG employs a constant bit rate and fixed packet length, the video and audio

information it transfers actually is highly variable in nature. When on-screen action is slow, the effective data rate may be reduced; the rate increases when there is a need to track fast motion. A single video channel such as a movie will have a variable demand for data transfer. It is the job of the MPEG encoder to regulate the flow to produce a constant bit rate—this amounts to adding dummy bits in the form of "null" packets when demand is low and to drop bits when the demand exceeds the instantaneous supply. This applies even more when the process of statistical multiplexing is used to combine several video channels into one stream.

The other side of the coin is the nature of IP and the applications it supports. Applications are those data communications functions familiar to users of PCs on LANs, WANs, and the Internet as a whole. Whether we are talking about e-mail, file transfer, VoIP, or streaming (IP) video, the actual data rate offered by a server or PC is highly dynamic as to time. IP is also accustomed to links that are imperfect, including such disturbances as dropouts, variable time delay, and timing jitter. Flaws such as these are extremely unwelcome for MPEG encoded video and audio; hence, when an MPEG link is working properly, IP data has a rather smooth ride. A DVB-S transmission, with its point-to-multipoint connectivity, exhibits the following properties with respect to data broadcasting:

- Point-to-multipoint connectivity (low cost per added site);
- Long propagation delay (nominally 260 ms);
- Stable link (assuming line of sight);
- Fewer elements and nodes than the terrestrial Internet;
- Greater bandwidth than PSTN local loop;
- Lack of a return channel (unless implemented terrestrially or with a return channel system);
- Eased regulatory environment as most countries allow installation of receive-only dishes without a license (local zoning restrictions may apply, however).

A typical arrangement for data broadcasting with IPE is illustrated in Figure 5.13. The originating server at the right contains multimedia content to be delivered by satellite. It is connected either directly by access line or through the Internet to the

Figure 5.13   Properties of a data broadcasting link based on IP encapsulation.

broadcast uplink. The latter is nothing more than a basic MPEG 2 transmission Earth station with the addition of the device to provide IPE (discussed in the next section). The IP packets are inserted into the MPEG stream and uplinked by a typical baseband to RF chain, likely using a DVB-S modulator, upconverter, HPA, and antenna. The satellite link produces the typical broadcast throughout the downlink footprint, where the home or office receiver is located. This consists of a standard DTH antenna, LNB, and cable to bring the DVB-S channel indoors. We can assume that the user on the left is employing a PC with the appropriate PCI card to demodulate the carrier and restore the MPEG 2 data stream. These are the same functions as found within the STB; however, there is the added feature of packet capture and an IP protocol stack that interfaces with the Web browser, e-mail client, or other application software.

Alternatively, the data broadcasting receiver can be contained in a separate box that connects to the PC or LAN via an Ethernet cable. An example of such a unit is the Edge Media Router (EMR), by SkyStream Networks of Mountain View, California. The SkyStream 2000 Edge Media Router is a content router that receives streaming data in MPEG 2 transport format and delivers it over the IP network. It receives data with up to 68 Mbps while supporting up to eight simultaneous packet identifications (PIDs). Streaming Internet and other IP content can be delivered to multiple users via standard 10/100 Ethernet LAN with up to 68-Mbps data throughput.

5.5.2   Packet Identification

The basic data transport protocol of MPEG is the 188-byte (B) packet. As a result of this structure, digital information such as IP protocol data units (PDUs) must be segmented into 188-B fixed length packets prior to transfer on the MPEG constant bit rate stream. The basic MPEG packet, shown in Figure 5.14, consists of a header, the adaptation field, and the payload. The function of each of these is a follows:

- Header, containing the all-important PID (pronounced "pid," as in the name Sid). With a field length of 13 bits, there are a maximum of 8,192 possible PID addresses.

Packet ID (PID)-Unique address per application or network, maximum 8,192

Figure 5.14   MPEG 2 datagram structure.

- Adaptation field, which is used to specify the type of information being transferred within the payload field of the packet. This is necessary to allow the opposite end to recover the information in its original format (e.g., to reassemble the IP packet, which typically has a variable length).
- Payload field, containing one packet worth of the originating user data.

There has been much interest in the PID field of the MPEG packet. This is because it may be used as a device and application identifier to allow the user equipment to select particular packets for reception. With only 8,192 possibly unique addresses, the PID is limited in its potential as a unique designator. Therefore, the most common approach in data transport over MPEG is to use an Ethernet MAC address instead. This is entirely consistent with how IPE may be employed in the context of an Intranet structure, where routing devices take care of the translation from IP address to MAC address, and vice versa. The PID may then be used to uniquely identify the overall applications, such as video, audio, program channel identification, and IP application.

5.5.3   Performance of IP Encapsulation

The transfer of MPEG packets containing IP data is implemented in a straightforward manner through the principle of IPE with a packet multiplexer. In the simplified multiplexing diagram in Figure 5.15, individual packets containing various types of data/content are multiplexed in time, producing a serial stream of data at a constant bit rate. To cause the bit rate to be a constant, the packet multiplexer inserts null packets as necessary. NULL packets represent lost data capacity but cannot be avoided in a dynamic multiplexing scheme such as this. If, on the other hand, the instantaneous demand for packet transfer cannot be satisfied due to all available time slots being occupied, the packet multiplexer will actually drop incoming packets. This can be done according to a predetermined priority that is specified with each packet. The selection and implementation of a priority scheme has to consider the value or importance of the associated application and its sensitivity to occasional loss of packets. Many IT applications incorporate an automatic retransmission

Figure 5.15   Multiplexing of IP data in MPEG 2.

request protocol (ARQ) that detects a missing packet and requests its retransmission from the origination end. For this to work, there must be a return channel of some type.

Figure 5.16 presents an example of a working IPE system that allows IP packets to be inserted into an existing MPEG video stream. In this case, the latter is obtained through a downlink channel from a separate satellite broadcast. The teleport equipment consists of a commercial IP encapsulator and multiplexer. IP encapsulators are produced by companies like SkyStream Networks, Logic Innovations, Harmonic Data Systems, and International Data Casting. Characteristics of an off-the-shelf unit from Logic Innovations are as follows:

- Fast transfer of IP packets to MPEG 2 multiprotocol encapsulation (MPE);
- Up to 100 Mbps with less than 5-ms latency;
- Supports 8,192 PIDs;
- SNMP support and GUI control;
- QoS management;
- Compatible with other vendors' DVB-S IRDs.

These properties make the device effective and nearly transparent to video transmission. The entire installation produces two main benefits:

- An MPEG video signal can be "turned-around" from one satellite to another, to extend coverage to another continent or hemisphere.
- IP data may be introduced to provide data broadcasting service into the new coverage area.

The capability of IPE has been boosted through a feature called opportunistic data insertion. Recall the null packets inserted by the originating MPEG multiplexer

Figure 5.16   IP encapsulation at a TV teleport.

to produce a constant bit rate. With opportunistic data insertion, null packets are replaced with IP packets on the fly. IP data is tolerant of variable data rate transfer, which is inherent in the opportunistic principle (the average data rate produced can be measured in megabits per second, but may be squeezed to a trickle at times). Figure 5.17 provides an illustration of the principle, based on the Satellite Media Router (SMR) of SkyStream Networks.

5.6   Digital Video Interface Standards

In digital TV, audio and data signals contained within an MPEG stream must be interfaced properly for their transfer between equipment and networks. Discussed below are the Serial Digital Interface (SDI) and DVB-Asynchronous Serial Interface (ASI), two standards which both support connection via coaxial cable. Generally speaking, SDI is the older standard that exists in legacy systems and has general applicability in digital TV for production, contribution, and distribution. In contrast, ASI is part of the DVB family and is therefore designed expressly for broadcasting and distribution services.

5.6.1   Serial Digital Interface

The SDI was first on the scene for interconnection of MPEG 2 devices and links. The American National Standards Institute (ANSI)/SMPTE 259M-1997 standard specifies a SDI for digital video equipment operating at either the 525-line, 60-Hz video standard or the 625-line, 50-Hz video standard. Another standard, SMPTE 292M, defines a serial digital interface standard for high-definition digital video, commonly called HD-SDI. The bandwidth requirements for high-definition video are significantly higher than for standard definition video. Also, the HD-SDI standard differs

Figure 5.17   Dynamic Internet Protocol encapsulation using the SkyStream Satellite Media Router. (Courtesy of SkyStream.)

from the SDI standard in the way that the video components are interleaved. Any of the digital video formats supported by the SDI standard use either 8 or 10 bits per data word. The SDI standard is natively a 10-bit format, but allows 8-bit video to be transported across the interface.

5.6.2   DVB Asynchronous Serial Interface

The DVB family of standards identifies an efficient baseband interface called the Asynchronous Serial Interface. Using a 75-Ω standard cable BNC connector, ASI has done a lot to simplify matters. Some of the basic characteristics of ASI are:

- Delivery of a 270-Mbps serial bit stream for the MPEG 2 protocol set;
- High-speed transport-stream input, compliant to DVB/ASI as defined in DVB document A010 rev1 and EN50083-9;
- Support for full DVB/ASI bit-rate range from 0 to 214 Mbps.

The popularity of the DVB family is making ASI the interface standard of choice. Consequently, it is a standard feature in later models of multiplexers, IP encapsulators, modulators, demodulators, and other ancillary devices that interface at baseband. This would include the equipment discussed in the next section that is used to interconnect the studio and teleport with terrestrial backhaul services. Another important reason for selecting ASI is its presence on test equipment provided by major suppliers such as Tektronix, Agilent, and Thomson Grass Valley Group.

5.7   Terrestrial Backhaul Interfaces

The interfaces defined in Section 5.6 address the interconnection of equipment within the environment of the broadcast center or teleport. Backhaul transport of the constant bit rate MPEG stream requires one of the standards employed in the broader telecommunications industry. The conventional digital hierarchy in terms of the T1 (1.544 Mbps) and E1 (2.048 Mbps) is certainly popular for this purpose and is widely available. This follows what has become known as the Pleisiochronous Digital Hierarchy (PDH), a somewhat archaic but very effective structure that allows for differences in timing at end points of the network. Within North America, the DS-3 (44.736 Mbps) is a popular digital channel for point-to-point transfer of full-motion television. The corresponding European PDH standard at this level, the E3, can only convey 34.368 Mbps, possibly requiring a user to double-up. Since the PDH is generally known and covered in other texts, we will not go into detail here. What follows is a brief review of current high-speed terrestrial backhaul interfaces and services that can transfer one or more MPEG streams on an efficient and cost-effective basis. They are commonly used for point-to-point backhaul applications, such as between a studio and teleport, or for studio-to-studio contribution transfers. In addition, we are beginning to see how terrestrial backhaul might replace the point-to-multipoint broadcast link to deliver the same content to multiple TV stations or even cable TV systems.

5.7.1   Fiber Optic System Interfaces—Synchronous Optical Network and Synchronous Digital Hierarchy

Fiber optic transmission systems have certainly grown in coverage and capacity. Since the first edition of this book, fiber has made substantial inroads in the television broadcasting field and now provides much of the contribution function for programming. This, or course, needs to be qualified as fiber is more readily available in and between major cities in developed economies like North America, Western Europe, Japan, Korea, Singapore, Hong Kong, Australia, and the Middle East. The rapid buildout of fiber across the Atlantic and Pacific Oceans that we experienced during the late 1990s provided much fiber bandwidth for video as well.

Fiber can be obtained in one of three forms:

- *Dark fiber:* unused glass fiber pairs that are available for purchase or lease (equipment must be introduced on both ends to "light" the fiber). Telecommunication operators resell dark fiber to each other to create larger and more diverse networks; individual users may, on occasion, obtain an attractive deal from the operator who would otherwise have no revenue from these glass strands. To use dark fiber, one must be able to take advantage of the particular physical path between the end points, and then invest in the equipment and its maintenance.
- *Analog transmission:* lighted fiber that transfers analog information that is continuously modulated on a particular wavelength. This type of service is somewhat similar to dark fiber since the particular pair cannot be used for anything else by the operator.
- *Digital transmission:* the modulation is digital and hence the service is used to transfer digital information. All of the fiber used within the telephone networks employs digital transmission, as this is the most efficient manner of aggregating the necessary channels. Since television signals are now digitized at the source, it is a common practice to hand the programming channel over to the fiber operator using the Synchronous Optical Network (SONET)/Synchronous Digital Hierarchy (SDH) family of standards (discussed next).

Originally developed by AT&T and rolled out in the United States, SONET provides a means of multiplexing a large quantity of channels onto an optical carrier. The adoption of SONET led to the establishment of an ITU standard dubbed Synchronous Digital Hierarchy to differentiate it from earlier digital standards that were not purely synchronous. In earlier years, the digital networks of different countries (and even cities within the same country) could not be synchronized in terms of their respective bit clock timing. As a result, the PDH multiplexing allows sloppiness at the interface, via a process called bit stuffing or buffering. With the advent of highly accurate but low-cost atomic clock standards and the Global Positioning Satellite (GPS) system of the United States, networks could be depended upon to have very close timing. Both SONET and SDH, therefore, rely on tight synchronous timing to achieve high throughput without buffering.

SONET/SDH use a fixed 90-byte packet length and strict framing structure that lend themselves to point-to-point transfer of data channels of literally any speed. The hierarchy is arranged according to the input bit rate. The starting point for

SONET is the Optical Carrier–1 (OC-1), which has a raw bit rate of 51.84 Mbps; this is slightly higher than the older DS-3 at 45 Mbps, which it was meant to accommodate. However, OC-1 is not actively used in the network, which instead starts with OC-3 (three times 51.84, or 155.52 Mbps). This is also the starting point for SDH, termed the Synchronous Transport Mechanism–1 (STM-1). From here, the equivalence is indicated in Table 5.10.

A given fiber optic pair can support multiples of these rates through the technique of dense wave division multiplex (DWDM), making it possible to transfer substantial quantities of digital video channels. This transmission is on a point-to-point basis through the structure of SONET/SDH as applied within public telecommunications networks around the world. To actually interface on the user side, one must add another internetworking layer such as ATM or Gigabit Ethernet, which are discussed next. There is also consideration being given to transferring packet protocols like the Internet Protocol and MPEG directly to SONET/SDH.

5.7.2 Asynchronous Transfer Mode

ATM has been available in a number of applications for approximately 10 years at the time of this writing. Its well-known 53-byte fixed length packet, called a cell, is recognized for its ability to combine voice, video, and data within the same network structure. ATM packet data "payloads" are 47 bytes, with the other 5 bytes devoted to routing and overhead. While ATM's original intention was to replace the fixed time-division switching structure of earlier digital telephone networks, its main role is now within the backbone of the Internet and private intranets. The latter allow organizations, including broadcasters, to acquire digital capacity on a relatively low-cost basis and pay for only what is actually used. ATM rides on top of other networks that employ SONET/SDH, Ethernet, and satellite communications links. Providers of ATM services include all of the major international carriers, such as AT&T, Worldcom, Sprint, British Telecom, France Telecom, Deutsche Telekom, KDD, and Telstra. Acquiring this capacity amounts to a commercial transaction with the appropriate carrier or carriers. There are also resellers in this market who arrange for both the long-haul and short-haul connections, since any service must be connected to the ultimate application at the source and destination.

The forwarding mechanism of ATM is the virtual circuit (VC), which is basically the same scheme used in Frame Relay and TCP. A VC is like a point-to-point connection between the two ends of the circuit and must be established prior to the transfer of data. There are two forms of VC: the permanent virtual circuit (PVC) and the switched virtual circuit (SVC). ATM supports both types of connections:

Table 5.10 Transport Bit Rates in SONET/SDH

Raw Bit Rate (Gbps)	SONET Designation	SDH Designation
0.15552	OC-3	STM-1
0.62208	OC-12	STM-4
2.488	OC-48	STM-16
9.953	OC-192	STM-64
13.271	OC-256	—

one must make prior arrangements for PVCs (e.g., the specified end points), while an SVC can be established on demand, like a dial-up phone call. The function of transferring information on the user-to-network interface (UNI) into ATM cells is provided by the ATM adaptation layer (AAL). The rules and procedures that are contained within the AAL are rather complex as they must consider a multitude of characteristics of the data, such as allowable time delay for transfer, treatment of header information, and steps to take if the VC is disrupted in some manner. For MPEG over ATM, two AAL versions are available: AAL-1 and AAL-5 (details can be found on the Web site of the ATM Forum, http://www.atmforum.com).

In the context of television service in general and MPEG in particular, the user must typically take care of adaptation. This is the technical process of subdividing the 188-byte MPEG packets into the 53-byte ATM cells. The details of MPEG-to-ATM adaptation are complex and beyond the scope of this book. Fortunately, the adaptation units are commercially available from companies like Tandberg Television of the United Kingdom and Thomson of France. Table 5.11 provides a summary of the characteristics of one of these units. Users can control and monitor adaptation units using the Simple Network Management Protocol (SNMP), the management protocol provided for the Internet.

As with other network technologies, an important issue in ATM is that of QoS. One valuable property of ATM is that QoS facilities are integral to its structure, allowing the user and service provider to specify such characteristics as priority, delay, and delivery guarantee. Telecommunications service providers may be held to standards of quality and reliability through a contractual SLA; however, SLAs can vary widely in terms of the promises for maintaining the service and responding to outages.

5.7.3   Gigabit Ethernet (IEEE 802.3z)

Traditional Ethernet LANs are with us in literally every corner of the data communications market. However, until the raw data rate could be scaled up above 100 Mbps, it could not address the demands of television broadcasters and the related service providers. Gigabit Ethernet (GE) provides the boost that makes this medium a serious candidate in the LAN and metropolitan area network (MAN) environment. The MAN is more or less defined to be a radius of approximately 100 km and is limited by the timing characteristics of Switched Gigabit Ethernet. The range may be extended to a WAN using the principle of Layer 3 routing, which allows Ethernet frames to be segmented into IP packets. This allows the frames to be transferred across the long distance networks within a country and internationally as well.

The GE solution sits on top of a structure that has taken almost 20 years to reach maturity. This is a good thing, which, coupled with the popularity of Ethernet in general (most PCs come equipped with it or can be upgraded very cheaply), makes it the lowest cost data communications access technology. The network environment for GE builds upon what is already in place: unshielded twisted-pair copper wire called Category 5 and GE network interface cards for copper cable interconnection priced under $100. However, to provide the greatest LAN extension of GE, the wiring should employ dual twisted-pair with differential signals (Cat-5E). To make this work adequately (in the presence of other copper wires and over distances up to 100 km), much development work was necessary. To fit 1 Gbps on a Cat-5 cable

Table 5.11 Characteristics of the Tandberg Television MPEG-to-ATM Adaptor

Standard Features	
	Dynamic selection of adaptation layer (AAL-1+FEC and AAL-5)
	PVC/SVC operation, point-to-multipoint SVC
	Robust clock recovery
	IP routing over ATM (classical IP)
	IP routing over ATM using PVCs
	Supports standard and enterprise specific SNMP
	UNI 3.0, 3.1, and 4.0 compliant
	Open standards based on ITU-T.J82 and ATM Forum specifications
	SNMP traps
Optional Features	
	STM-1 (multi, single, and electrical mode), DS3, and E3
	Up to three network interface cards
Inputs and outputs	DVB ASI copper input
	DVB ASI copper output
	Full-duplex ATM input/output dependent upon physical interface
Physical interfaces	STM-1/SDH, SONET multimode and single mode, STM-1 electrical
SC connector type	DS3/PDH, G703 BNC connector
	E3/PDH, G703 BNC connector
Alarm Relay Contacts, Nine-Way D type (male)	
Features	Bidirectional operation
	ATM/ASI bridge, ASI/ATM bridge
	Dynamic selection of adaptation layer, AAL-1 +
	FEC or AAL-5
	Supports FEC/interleaving in AAL-1 mode
	Supports SNMP
	RS-232 user diagnostics
Control	Local control through 10baseT Ethernet port
	Remote control through RS232 port
Physical and Power	
	Input voltage: 110–120Vac/220–240Vac (single phase)
	Approx. dimensions: (W D H) 442 499.5 44.5 mm (17.5" 19.7" 1RU)
	Approx. weight: 9 kg

requires using all four of the wire pairs, each pair capable of nominally 100 Mbps. Additional capacity beyond 400 Mbps is obtained by what is called multiple-level signaling, where bits are transferred in groups of up to four per signal. Forward error correction is included to overcome the additional attenuation, noise, and cros-stalk that are present as one pushes up in bandwidth and uses all of the wires for signal transfer. As a result, the range of operation of GE over Cat-5 is only about 100m. This is sufficient for connections within a server farm or between a switch and multiple desktops. The range can be increased to hundreds or even thousands of kilometers using a device called a GE extender, discussed next.

The Cat-5 approach is designed to protect the large installed base of copper cabling in and between buildings. This is not a concern in the broadcasting environment, where digital facilities rely on wideband coaxial cable and fiber optics. The real benefit of GE comes about in the MAN as a relatively low-cost means of delivering high-quality video between locations. Making this work correctly will require the following facilities:

- Adaptation of MPEG 2 fixed length packet transport stream to Gigabit Ethernet variable length frames;
- Extension of GE to fiber optic service using the appropriate optical carrier modulation and demodulation (basically the conversion from IEEE 802.3x to SONET);
- Availability of a MAN GE service.

The role of GE in the MAN can be to facilitate a concept called the Virtual Teleport, suggested by Luann Linnebur, executive director, sales and business development of Path 1. This would permit moving a video stream from one crowded teleport facility to another distant teleport with available bandwidth. The following is an example of how this is made available in a metropolitan area. For this, we review adaptation equipment from Path 1, based in San Diego, CA, along with GE fiber optic extension from JDS Uniphase, and the GigaMAN service offering of SBC Communications. The Path 1 gateway follows an Internet standard protocol structure to adapt MPEG 2 to IP. In particular, MPEG 2 packets are transferred using the Real Time Protocol (RTP) on top of the User Datagram Protocol (UDP). The following provides top-level capabilities of the Path 1 Cx1000 gateway product.

> The Cx1000 is interoperable with most IP switching equipment and can compensate for network delay and jitter introduced by such equipment. A built-in program clock recovery (PCR) correction mechanism supports multiplexing of several single program transport streams (SPTSs) into a multiple program transport stream (MPTS). In addition, the Cx1000 automatically regenerates MPEG-2 tables in accordance with the MPEG-2 program composition at the output streams. The mapping of the IP encapsulated MPEG-2 SPTS streams to the output MPTS streams can be accomplished automatically via translation of UDP port numbers or manually via a user program table. The Cx1000 complies with MPEG-2 and DVB, as well as 10/100 Base TX Fast Ethernet (twisted-pair) and 10 SX Gigabit Ethernet (optical fiber). Both the SDI and ASI digital video interfaces may be used and MPEG 4:2:2 and 4:2:0 are supported. Gateway traffic can be monitored locally from a front panel and remotely via SNMP.

The Cx1000 provides the key interfacing function between ASI and IP. The next step is to perform the handover from GE to fiber, which typically requires a fiber optic extender. One of these devices is connected on each end to perform the adaptation from GE to SONET, as well as the function of optical modulation and demodulation. The following description is for a typical product used for this purpose, provided by JDS Uniphase:

The Model 1280 GbX Gigabit Ethernet Fiberoptic Extender increases the maximum interconnect distance of a Gigabit Ethernet switch from the standard IEEE 802.3z distance (220 or 5000 m) to distances of up to 100 km (67 miles). The GbX allows LAN technology to be extended to the MAN. Two standard product ranges of long distance links are offered: up to 25 km and up to 100 km. The GbX is fully compatible with Gigabit Ethernet IEEE 802.3z/D5 standards for fiberoptic interfaces. The GbX currently supports short wavelength (1000Base-SX) optical interfaces. Network management and loopback control functions are provided via an RS232C port. The GbX link allows Gigabit Ethernet users to take full advantage of the available 2 Gb/s full duplex bandwidth, as specified in IEEE 802.3z/D5. The link provides low latency by allowing simultaneous transmit and receive functions.

The actual MAN GE service would be provided by a local telecommunications carrier who owns and operates the requisite fiber infrastructure. In southern California, SBC Communications offers the GigaMAN service to major users who wish to create a GE MAN within a range of approximately 100 km. The following describes the service as it was introduced in 2002:

GigaMAN service is a high-speed, fiber-based, transport service designed to offer transparent interconnection of customer local area networks (LANs). By combining traffic types over a single, high-speed ring network customers will realize even greater savings over multiple lower speed services. GigaMAN service consists of two dedicated single mode fibers between the customer sites and specialized fiber repeaters that are placed at the customer's premises as Network Terminating Equipment, providing one Gigabit Ethernet handoff interface to the customer. The customer's Ethernet LANs incorporate Gigabit Ethernet switches which may be purchased through SBC DataComm or through another provider. Examples of bandwidth intensive applications suited to this product include WAN service for LAN interconnectivity, large file transfers, distance learning, medical imaging, multimedia, CAD/CAM, and video that would benefit from connectivity at the native 1.25 Gbps rate.

Our purpose for reviewing the status of GE was to highlight its use in the TV production and distribution environment. It is unlikely that GE will find its way directly into satellite service; rather, it is an economical broadband approach for taking an MPEG stream at one location and moving it to another. This is constrained by distance; however, if the limits are workable, then the GE approach represents a very attractive solution.

References

[1]   Couch, L. W., II, "Pulse Code Modulation," in *The Communications Handbook*, 2nd ed., J. D. Gibson, (ed.), Boca Raton, FL: CRC Press, 2002.

[2]   Kretz, F., and D. Nasse, "Digital Television: Transmission and Coding," *Proc. IEEE*, Vol. 73, No. 4, April 1985, p. 575.

[3]   Schaphorst, R., *Videoconferencing and Videotelephony—Technology and Standards*, Norwood, MA: Artech House, 1997.

[4]   Wylie, F., "Digital Audio Compression Technologies," in *National Association of Broad-casters Engineering Handbook*, J. Whitaker, (ed.), Washington, D.C.: National Association of Broadcasters, 1999.

[5]   Koenen, R., "MPEG 4, Multimedia for Our Time," *IEEE Spectrum*, February 1999, p. 28.

[6]   *Digital Video Broadcasting – Television for the Third Millennium*, Geneva, Switzerland: DVB Project Office, European Broadcasting Union, May 1995.

[7]   "Digital Broadcasting Systems for Television, Sound and Data Services; Framing Structure, Channel Coding and Modulation for 11/12 GHz Satellite Services," *European Telecommu-nication Standard ETS 300 421*, Valbonne, France; European Telecommunications Standards Institute.

Direct-to-Home Satellite Television Broadcasting

DTH systems are designed to transmit entertainment TV programming to home-receiving Earth terminals (or, simply, home receivers). This is a natural extension of TV distribution by satellite, utilizing the area-coverage and single service provider features of the technology. DTH systems, also called Direct Broadcast Satellite, employ either the BSS allocations, which are intended for this use, or the FSS allocations as one of a number of possible applications. As discussed later in this chapter, this choice has some important impacts, yet the end result is the same to the user.

This chapter focuses on the nature of these services and the various factors that must be addressed to assure a successful introduction. Among the latter are:

- The programming mix, for example, the quantity, variety, language options, and degree of interactivity, which must compete with other DTH systems and delivery mechanisms (e.g., cable TV, AM and FM radio, audio CDs, Internet delivery of MP-3 files, multichannel microwave distribution service, cassette and DVD rental);
- Receiving equipment—that is, its affordability, convenience of installation and use, integration with other video and audio devices, and aesthetics;
- Acceptability of the service price and an effective means to collect payment;
- Incompatibilities with the other DTH, radio, and cable TV systems, which are dependent on the nature of the business plan;
- Conditional access and scrambling in order to deal with copyrights, privacy, collection, regulations, and content rules (which may exist in the country markets of interest);
- Uplinking system, including the redundancy, strength, and program development and contribution facilities.

The major elements of a DTH system are shown in Figure 6.1. Experience with DTH systems has shown that the service must be attractive as compared to other forms of video distribution; and access to the programming by the consumer must be properly controlled. The competition between delivery media is highly variable between countries. The U.S. market is the most competitive in the world, with high cable TV penetration and two DTH systems in operation and a third to start as of the time of this writing; newer broadband options continue to be investigated as well. Other environments like the United Kingdom or Japan are less dependent on cable, so DTH has greatly improved prospects. In developing countries, the

Figure 6.1   Major elements of a DTH system.

limitations have less to do with competition and relate mainly to the ability of con-
sumers to afford the price of the equipment and the service.

6.1   Relative Cost of Satellite DTH Versus Cable

Cable TV is a viable technology for providing a wide array of programming services
to urban and suburban homes. In countries like the United States, Canada, Belgium,
and Germany, the percentage of homes that actually receive cable services is more
than 75%, in relation to the total that have access to cable service (i.e., the homes
passed by cable). In only a few places in Asia does cable penetration approach this
number, such as in major cities in Japan and Korea and in Singapore. For the highly
populous countries in Latin America, Asia, Eastern Europe, and Africa, cable TV is
simply much less economical in relation to digital DTH (DDTH).

In the following analysis, it is assumed that both the cable TV and DDTH serv-
ices provide in excess of 50 channels. The cost of providing cable service to a typical
single-family home is approximately $2,000, which includes the cable, set-top box,
its installation, as well as the head-end equipment needed to receive satellite and ter-
restrial TV signals. Also required are administrative offices, maintenance facilities,
and a studio to develop local advertising and program content. For a DDTH satellite
system, the cost per subscriber includes a share of the cost of the satellites and
uplinking facilities added to the cost of maintaining the subscriber base. For a system
of three satellites and two uplinks, the investment is assumed to be approximately $1
billion. The lower curve in Figure 6.2 shows the investment cost per subscriber,
which is obtained by dividing the total investment by the number of subscribers.
Added to this is the installed cost of the home-receiving system, which is assumed to
be $400 (this can be viewed either as the purchase price with installation or the DTH
operator's average subscriber acquisition cost). In the DDTH model, the satellites,
uplinking facility, and administrative systems are shared by between 1 and 10

Figure 6.2   Investment cost per DTH subscriber as a function of the number of subscribers. System capital includes two in-orbit BSS satellites and one broadcast center.

million subscribers. The upper curve in Figure 6.2 shows that the total investment per subscriber decreases to $500 when a subscriber base of 10 million homes is reached. For both the cable TV and DDTH TV broadcast systems, the cost of programming is assumed to be the same and is born by the subscriber through monthly access fees.

6.2   DTH System Architecture

The discussion of DTH architecture has at its foundation the matter of the type of video processing and compression, along with the supply and cost of the set-top box. In this section, we briefly summarize the overall architecture of a DTH program delivery system such as that used for commercial purposes. It encompasses the uplink systems for digitizing, compressing, and transmitting multiple television programs using the DVB-S standard, for example. Other elements are required for the contribution of the programs, storage and switching of video signals, and the management of DTH as a customer service. Provided here are some of the technical approaches to the uplink side of the network and a review the capabilities of some of the key suppliers of the compression and processing equipment. The market continues to evolve, so the names, owners, and ultimate suppliers of critical elements will change over the coming years.

6.2.1   Basic Elements and Signal Flow

The major elements of DTH system are listed as follows and are indicated in Figure 6.1:

1. DTH satellites in GEO (one or more):
   • Spacecraft construction;

- Launch services;
- Launch and on-orbit insurance.
2. TT&C:
   - Controls the space segment and monitors spacecraft health;
   - Verifies that transmissions to satellite do not cause interference;
   - Provided by satellite operator (usually a separate company);
   - Limited communication required between DTH network operator and satellite operator.
3. Broadcast center:
   - Originates, acquires, and transmits program material;
   - Generally centralized, with no or limited backup;
   - Part of conditional access system.
4. Customer service:
   - Billing and customer turn-on-off;
   - Customer assistance.

These are the major elements, and there are many vital components and functions hidden within each. For example, customer service is involved with connecting and controlling individual subscribers. However, how they obtain their equipment in the first place and have it installed has turned into an industry all its own. Ownership and operation of the satellites can be internal or taken as a service from a professional satellite operator. The DTH companies reviewed previously in this chapter have followed both approaches. The design and manufacture of the IRDs, discussed in another work [1], represents the largest satellite consumer segment, having produced tens of millions of units now installed around the globe. The programming for the system is probably the most important as this is what the subscriber is after. At any given time, the most attractive programming in terms of consumer appeal brings with it the highest price (see discussion in Chapter 4). A successful DTH business must address and integrate all of these aspects.

A more detailed configuration of the operating components is presented in Figure 6.3. At the top of the diagram we find the service management functions of the network. These manage the interaction with the customer over the telephone and Internet, and provide the means to download PPV movie selections on a monthly basis. It also ties into the CA segment, which authorizes individual IRDs over the space segment. The technical functions at the bottom of the diagram show the physical production and transmission facilities, from content input through baseband processing and on up to the satellite.

6.2.2 Compression System Arrangement

The basic arrangement of the uplink compression-encoding-modulation chain and downlink demodulation-decoding-decompression chain is presented in Figure 6.4. The function of each of the blocks is described in previous sections of this chapter and in Chapter 5. We are concentrating here on the uplink compression elements contained within the broadcast center Earth station such as that shown in Figure 6.5 for DIRECTV. Included is the equipment that digitizes and time division multiplexes the video, audio, and data information. Systems that employ FDMA use separate carriers

Figure 6.3   DTH system operating elements; components of the broadcast center are indicated with shaded boxes (transmission and production services may be included, depending on the configuration).

for each video channel. In large networks, between 5 and 12 video channels and their associated audio and data are combined using TDM onto a single carrier that would occupy the entire transponder bandwidth. The form of TDM could be either a fixed bit-by-bit multiplex or alternatively a statistical multiplex where data rates are

Figure 6.4   Digital DTH transmission system, based on the DVB-S model.

Figure 6.5   The DIRECTV Broadcasting Center in Castle Rock, CO, is recognized as the most sophisticated television transmission facility ever built. As the heart of DIRECTV, the 55,000-ft^2 facility provides more than 150 channels of movies, sporting events, popular subscription networks, and special attractions. (Courtesy of DIRECTV.)

adjusted based on content. In either case, the carrier must be transmitted by a single Earth station as the multiplexing is done on the input side of the modulator.

The compression systems themselves fall into two categories: (1) those that comply with a standard, particularly MPEG 2 or DVB (which includes MPEG 2 as a component); or (2) those that use a proprietary algorithm and multiplexing scheme. Systems that started out in category (2) are quickly moving to MPEG 2 because of the rapidly decreasing cost of the receiving equipment. In the following sections, we briefly review the offerings and capabilities of some of the leading suppliers.

6.2.3   Suppliers of Key Elements

Compression based on the MPEG-2 standard is the core of DDTH. In the early days, only one or two viable suppliers were on the market. They produced equipment that did the job but had few features to allow growth in content and facilities. In time, the capability of these systems has increased greatly to allow more in the way of channel capacity, versatility in setting quality standards, support for HDTV and wide screen aspect ratio, and the ability to provide IP data encapsulation.

6.2.3.1   Compression and Multiplexing Systems

Included under compression are a number of elements that comprise the MPEG 2 chain. The video and audio encoder is the primary element that produces the MPEG transport stream, as discussed in Chapter 5. Another key function is that of multiplex, which allows several encoders to feed a common transport. This is how current

DDTH systems achieve up to 12 digital TV channels on a single carrier of around 36 Mbps, averaging 3 Mbps per TV channel. There are two categories of compression systems: cable and DTH. For cable, the major suppliers are Scientific Atlanta and Motorola General Instruments. Each produces an integrated system that allows cable networks and cable TV operators (e.g., the local companies) to provide what is called digital cable. On the DTH side, the major operators in the United States and Europe employ equipment from Thomson Broadcast, Harmonic Data Systems, Tandberg Television, Barco, and Scopus.

6.2.3.2   DVB-S Modulators

The modulator is a critical component for the satellite uplink. Products in this range provide all of the functions necessary to take the MPEG 2 transport steam (the information input) and apply all of the remaining functions of the DVB-S standard [see Figure 5.11(a)], including:

- Sync inversion and energy dispersal;
- Reed-Solomon outer coder;
- Convolutional interleaver;
- Inner coder puncturing and mapping;
- Baseband filtering;
- QPSK modulation.

The modulators are programmable from the front panel or remotely to configure the DVB-S carrier for the desired coding rate. Some have the ability to change the modulation format to 8PSK or 16 QAM, but neither of these is supported by IRDs on the market as of 2003. Suppliers of DVB-S modulators include Newtec, Radyne, EF Data, Tandbarg, Scientific Atlanta, Motorola, and Scopus. As discussed at the conclusion of Chapter 5, care must be taken in selection of interface standard [e.g., synchronous (SDI) versus asynchronous (ASI)].

6.2.3.3   Video and Audio Switchers

Of the many elements in the broadcast center, the one with the greatest overall impact on service is the switcher. These devices would appear to be nothing more than automated patch panels that allow video and audio signals to be transferred between systems on the contribution side to the various encoders and other transmission means. They are also essential for switching between primary and backup means such as fiber optic cable and backhaul satellite links. The common functionality in the digital domain is to provide an $N \times N$ cross-connection capability under computer and stored schedule control. The VIA32 series of routers from Leitch, for example, accommodates SDI, ASI, analog video, and analog audio for mid-sized professional applications. Terrestrial telecommunications related formats such as DS-3 (45 Mbps), E3 (34 Mbps), and DVB-ASI (270 Mbps) are supported as well. The system is modular to combine multiple VIA32 frames to create multilevel routing systems, which can expand from 4×4 to 32×32. The analog video and audio can start out at 32×16 and be expanded to 32×32 in the same frame. Using output

expansion frames, analog video and audio can be expanded to 32×48 or 32×64. Even the control circuitry is designed on a removable module to accommodate future possibilities.

6.2.3.4   Video Servers and Automation

Modern broadcast centers employ video servers in lieu of magnetic tape in its various forms. Based on the same computer storage technology that is popular in the IT field, video servers are supported by computer systems that control every facet of recording and playback, along with the switching function discussed in the previous section. There are several very good reasons for the rapid migration to this style of production and distribution [2]:

- They permit material from a single storage source to be used simultaneously by multiple users.
- They provide a migration path to the all-digital facility that is not necessarily format limited.
- They result in a reduction in lost or misplaced materials.
- They offer a reduction in the size and space requirements relative to tape.
- They have complete computer control capabilities.
- They offer near instant access and playback of video segments.

The majority of video servers are produced by Tektronix, HP, Fastforward, Doremi, and Thomson Broadcast. Makers of automation products include Harris (formerly Louth), Omnibus, Sundance, and Florical.

We have provided an introduction to the various elements and the suppliers most active in the marketplace. A complete listing of vendors of the range of broadcast center products can be found on the Web site of the National Association of Broadcasters (http://www.nab.org). The developer of a DTH system really has two options available: design the uplink facility and purchase the necessary components, or contract the entire project to a system integrator. Most projects will involve a combination of the two, as internal resources may already exist with capability to pursue at least part of the overall program. Some companies that offer integration service are Sony, Globalcomm Systems, Inc., Level 3 Communications, Raytheon, MIH Limited, and Tandberg Television. As is always the case with projects of this magnitude, it is always best to define the service requirements in sufficient detail so that either strategy can be pursued in a methodical manner.

6.3   Satellite Architecture

In this section, we discuss various design approaches for the spacecraft that would support a DTH service business. The emphasis is on how to evaluate the technology alternatives and apply them effectively. DTH is a delivery vehicle for programming, where the receiver is located with and probably owned by the end user. In the ideal case, specifics like the type of receiver, size of antenna, signal format, and satellite design are secondary. However, it is vital that the system developer understand these

alternatives so that a poor technical choice at the beginning does not become the point of business failure. The emphasis should be on making it easy and relatively inexpensive for the subscriber, since their interest is in the programming and the cost of getting access to it. The quality of satellite-delivered digital video is as good as or superior to cable or over-the-air broadcasting; hence, the picture quality will probably not be a differentiating factor. Rather, subscribers could relate more to reliability of service along with the convenience factors cited previously.

In the main, the area where quality is seen as a negative is at C-band where users have to contend with terrestrial interference. Ku-band systems, while generally free from terrestrial interference, are subject to rain fade, which produces occasional outages. Once this problem is solved through adequate link design and margin, subscribers will next be drawn by a desirable array of programming. Over the years in the U.S. market, this has come down to the range of channels, movies, and events that are delivered to cable systems and now provided over the operating digital DTH networks like DIRECTV and DISH Network. Anything less in such a competitive environment will be rated below the leaders; the question is, which will prove the winner in coming years? Innovation, while good for generating interest, introduces opportunity for consumer dissatisfaction.

In Europe, the experience is exclusively at Ku-band, and the previous points about simplicity, cost, and programming have been validated over and over again. SKY TV succeeded over its higher technology competitor, BSB, precisely because of these aspects. Other activities in France and Germany were slower at the start of the game, being hampered by government red tape and inflexibility. In many ways, Europe poses a much more exciting future for DTH simply because of the wider array of ventures and economic conditions. DDTH services operate in every major European economy and viewership is generally high. For these services, the satellites of Eutelsat and SES-Astra are primary, although national operators in Spain, Norway, Turkey, Israel, and Russia serve domestic DDTH subscriber bases as well.

The Asian environment has many opportunities because of the primitive nature of the DTH industry in the region. Cable TV is a viable business in developed countries and city-states; China, India, and Indonesia have large populations that are hungry for more and better entertainment. China now has a DDTH platform operating on Sinosat. Importantly, money flows easily into major projects and business ventures. This has fueled the creation of several new satellite operators and the development of the largest satellite market in the world. The greater rainfall in the tropical parts of the region (where a high percentage of the population lives) is a significant factor in building a technically satisfactory system at Ku-band. C-band FSS systems exist and serve a strong niche market, not unlike the early backyard dish segment in the United States. However, putting this together in a business the size of DIRECTV or BSkyB may take several years and false starts.

The story for Latin America assumes a different tone from the previous examples. In the leading countries of Mexico, Brazil, and Argentina, pay TV is accepted and very popular among the middle class. Others enjoy a reasonably wide variety of standard broadcasting services. The initial introduction of DDTH occurred with the launch of Galaxy 3R, a Ku-band FSS satellite, for DIRECTV Latin America. The subsequent start of a service on PAS 8 by a consortium of large programming interests, namely, Globo of Brazil, Televisa of Mexico, and News Corp., was not far

behind. The fact that these organizations are such powers in the area of TV content will have a big impact on the potential attractiveness of the services. It is interesting to note that both systems use FSS satellites that are owned and operated by United States–based corporations At the time of this writing, neither operator appeared to be developing a solid business. DTH would appear to be welcome in Latin America, where the viewing options tend to be limited. This is particularly the case outside major cities like Rio de Janeiro, Mexico City, and Buenos Aires. DTH still has the ability to provide a broad programming package at a fraction of the investment of cable. However, even with this benefit, building a regional DTH business may be a much more difficult challenge than once thought.

6.3.1   Medium-Power DTH Satellite Systems

Medium-power Ku-band satellites with EIRP less than or equal to approximately 50 dBW have a variety of uses, including video distribution, DTH, data broadcasting, audio distribution, point-to-point telecommunications, and VSAT interactive data networks. They are being used to create DTH businesses in the United States, Europe, Australia, Japan, and Latin America. A satellite operator who successfully creates a video hot bird from a medium-power Ku-band satellite or even a C-band satellite has a valuable asset upon which to expand a business. This is what happened with the Galaxy system in the United States, although this first involved cable TV rather than DTH. If the infrastructure is there because of the need for cable TV programming, then the DTH side can develop gradually. Building an exclusively DTH business has proven much more difficult for the pure satellite operator because of the lack of a flywheel from other revenues and easy access to programming. Another inherent problem for the medium-power DTH operator and TV programmer is that the cost of the receiving station is elevated due to lower satellite EIRP.

Medium power also means that there can be more transponders per satellite for the same total launch weight. Cost is primarily driven by weight, so medium-power DTH has an economic advantage in space. Ground costs would tend to be higher, as a result of the need for a larger dish (anywhere from 30% to 100% larger). The cost of increasing dish size is significant; but when considering the TV receive-only (TVRO) cost as a whole, the dish size has only a secondary economic effect. In terms of appearances, the larger dish could be a problem in some markets. The experience in Europe and Asia so far is that this is not the case, as consumers appear to accept the fact that to receive the programming one needs to have this type of dish. A similar story is told of niche DBS services in the United States, such as GlobeCast's WorldTV.

Medium-power DTH can be difficult to implement from the standpoint of international coordination. DTH satellites that use the WARC-77 assignments do not need to be coordinated as long as they follow the plan. On the other hand, medium-power FSS satellites must be coordinated from the beginning. This process can take years to complete and poses risk to the business. The design may have to be changed to satisfy neighboring systems that could currently be using a lower power level, or one might have to accept constraints on operating downlink power levels. A detailed discussion of the coordination process is provided in Chapter 12.

Any technical compromises resulting from coordination could require receivers to have larger antennas, better polarization and/or sidelobe isolation, or narrower bandwidth (and a different frequency plan). The incompatibilities previously mentioned will likely impact the business, unless that satellite achieves hot bird status. Ancillary services such as data broadcasting may have to be curtailed to satisfy requirements of a coordination. This is because some carriers have power densities that could cause unacceptable interference. Therefore, system developers should maintain close contact with the government regulatory people who are pursuing the coordination process. Trade-offs may have to be made along the way.

In summary, the principal advantage of the FSS approach is that existing satellite capacity can be utilized, provided that the EIRP is sufficient to allow reception by an appropriate receiver. This is an important factor for C-band because satellite EIRP is inherently lower in this band. The appropriate receive dish is in the range of 1.8m to 2.8m, which tends to reduce its attractiveness. Also, there must be enough bandwidth to support a channel capacity that will be attractive to users. In North America, the market demands that at least 100 channels be available at the same orbit position, whereas in Europe and Asia the number might be as low as 20 for a specific country market.

The existing BSS assignments might already be taken by another operator. This is precisely the situation with respect to Sky versus BSB in the United Kingdom. Sky introduced their service on the Astra FSS satellite, ahead of BSB's launch of their compliant BSS satellite.

The orbit slots and coverage patterns assigned by BSS WARC plans may not be optimum for a particular market or application. One of the big problems with the BSS coverages is that they are for a single country; the market might extend to a larger region where the same language is used. This requires changing the plan, a process described in Chapter 12. BSS assignments are only for one-way broadcasting, whereas a new FSS entrant might wish to introduce an interactive service that uses a two-way VSAT.

6.3.2   High-Power DTH Satellite Systems

The a priori plan of WARC-77 established both a planning process and a plan for two of the three ITU regions of the world. That this could be accomplished in several weeks of deliberation is incredible. Several satellites operated according to the 1977 plan but even more satellites did not. The fundamental difference was that BSS satellites used fixed orbit positions and frequency assignments, with their antenna footprints covering typically one country each or a portion of that country. The transponders were used for video transmission, although sound (program) and data broadcasting were also contemplated even in 1977. The planners established certain technical standards for the satellites based on their estimate of current and future technology. They overestimated the ability of satellites to carry high-powered transponders and underestimated the performance of future home receivers. The application of digital compression and advanced conditional access was not even considered. WARC replannings were introduced in 2000 and may be voted on in 2007.

The protection ratio (C/I) is the key interference test parameter that determines dish size and satellite spacing. The single entry value corresponds to the interference

due to one other frequency assignment. Multiple entry indicates the combined effect of all interference sources on the same channel. A single entry objective was set at 32 dB. The planners sought to hold the multiple entry C/I at 29 dB, which is a very high value. In the United States, we try to achieve 20 dB, but even here we often accept a poorer value. In accordance with the plan, the following are the assumed characteristics of each BSS satellite:

- High EIRP, typically 63 dBW;
- Fixed-frequency plan with 23-MHz transponders;
- At least one satellite with five channels per country;
- Circular polarization (left and right hand);
- Fixed orbital spacing between satellites:
  - 3° minimum separation;
  - 9° for cofrequency assignments.
- Satellites that serve different countries are colocated where possible;
- Maximum spacecraft reflector size of about ~2m;
- Elliptical beams without shaping;
- Mandated radiation levels into other countries.
- Assumed home receiver characteristics include:
  - Dish size no smaller than 0.9m (0.45 typical today);
  - Receiver noise figure no lower than 9 dB (1 dB typical today);
  - Relatively poor cross-polarization ability;
  - Relatively poor angular discrimination.

The last two bullets are addressed in Section 6.9.2.

The rules surrounding the plan have a few loopholes under which an ITU member nation can make technical changes. With regard to Region 2, any proposed change cannot raise interference by more than 0.25 dB (6%) into another planned satellite that complies with the plan. The criterion for Regions 1 and 3 is even tighter: the interference posed by the change cannot produce a ratio of carrier to interference (C/I) that is less than 30 dB. This is equivalent to an increase of 0.1%, or 0.004 dB.

There are a number of reasons for considering making a change to the plan. Some of these are:

- Change the downlink footprint;
- More constant domestic coverage using better footprint beam shaping;
- Increase service area;
- Reduce the number of satellites;
- Consolidate systems of several nations.

The key technology for altering the footprint is beam shaping onboard the satellite. It is impossible to create more energy than the power amplifier delivers to the antenna. The trick is to distribute that energy in a particularly effective way. The beam selection approaches in Section 6.5 offer more options, such as using multiple

spot beams to deliver local programming. The original plan assumed the use of very simple feed networks—that is, a single horn and an unshaped elliptical reflector. The same shaping techniques that increase the performance within the intended coverage area can be used to reduce the radiation into other countries. This can be needed if the calculated C/I into another member's system is unacceptable.

At the time of this writing, the Radio Regulatory Board was literally jammed with requests to modify BSS planned orbit assignments. This situation, as well as the general confusion over who is really proceeding, has caused the entire process to come under review.

6.4   Orbital Interference Limitations

The drive toward smaller receiving antennas brings with it the prospect of increased *adjacent satellite interference* (ASI). If the specific design of the home receiving dish and the performance of the satellite are known, then it is a simple matter to calculate the expected amount of interference into the IRD. The key measure is the ratio of received carrier power from the desired satellite to the ASI power within the same bandwidth as the received carrier (e.g., C/I). For BSS, these calculations are performed based on the assumption that the adjacent satellite is a precise number of degrees away (typically 9 for the same channel) with the same signal characteristics and the same EIRP. These are the conditions that are defined as homogeneous. In nonhomogeneous interference environments, many of the specific characteristics are not known with precision, making orbital interference calculations less precise than in the BSS domain. Homogeneity is what allowed WARC-77 to tackle the problem of creating complete plans for all nations (with the respective needs) in two of the three regions of the world. In fact, the only reason that Region 2 had its own conference was because of displeasure with the conservative approach taken in the first instance.

6.4.1   Interference Model

The basic interference model is shown in Figure 6.6, which indicates interference entering the desired system either on the uplink or downlink. In BSS interference analysis, the uplink is transmitted from large antennas with low levels of sidelobe radiation. Hence, we concern ourselves with downlink interference as this is what determines orbit spacing and receive dish size. We see on the right of Figure 6.6 that the C/I ratio, also called the protection ratio, is a simple function of the difference between transmitting EIRP from the desired and adjacent satellites along with rejection ability of our small receiving antenna. Under the principle of homogeneity, the satellites both produce the same level of EIRP at their respective beam centers. All we need to consider here is how far from beam center the particular receiver happens to be. The performance of the receive antenna, on the other hand, must comply with standards established by WARC-77. The worst case expected antenna patterns are provided in Figures 6.7 and 6.8 for the satellite transmit antenna and home (Earth) receive antenna, respectively [3]. These are typical of the types of specifications placed by the ITU on its members. These particular figures

$$C/I = [EIRP_d - EIRP_i] + [G_d - G_i] - 20 \log (R_d/R_i)$$

Figure 6.6   Basic GEO adjacent satellite interference model.

are from the WRC in 1977 that established the first BSS a priori plan, for Regions 1 and 3. The specifications are subject to revision from time to time, but the principles remain constant.

The original criterion for acceptability was set very high by the creators of the original plan. A C/I value of 30 dB is commonly used; however, this particular level is difficult to achieve in practice. Table 6.1 contains a variety of C/I values for different services. This is provided as an illustration of the complexity of trying to arrive at one number that will satisfy all.

The last line in the table footnote of the list of definitions for Table 6.1 is a unit of measure for noise in an analog telephone channel. One picowatt of noise power is equivalent to 10^{-12} watts (–120 dBW), which is 90 dB below the level of the standard 1-mW (0 dBm) test tone used to align telephone channels. Thus, 500 pW0p is the same as saying the S/N is 63 dB (e.g., –30 dBW – [–120 dBW] + 27 dB). This is about 13 dB below the background noise in the telephone receiver.

Figure 6.7   Reference patterns for copolar and cross-polar components for satellite transmitting antennas, Regions 1 and 3 (WRC-77).

Figure 6.8   Copolar and cross-polar receiving antenna reference pattern in Regions 1 and 3 (WRC-77).

Table 6.1   C/I Protection Ratios for Various Services

Wanted Service	Wanted Signal	Interfering Service	Interfering Signal	C/I, Total Acceptable	C/T, Single Entry
BSS	TV/FM	BSS, FSS, FS, BS	TV/FM	C/I = 30 dB	C/I = 35 dB
FSS	FDM/FM	BSS	TV/FM	N = 500 pW0p	N = 300 pW0p FSS
FSS	TV/FM	BSS, FSS	TV/FM	C/I = 32 dB	C/I = 37 dB
FSS	QPSK	BSS, FSS	TV/FM	C/I = 30 dB	C/I = 35 FSS
FSS	FDM/FM	FSS	FDM/FM	N = 1000 pW0p	N = 400 pW0p
FS	FDM/FM	BSS	TV/FM	N = 1000 pW0p	−125 dB (W/m²/4 kHz)
BS	TV/VSB	BSS	TV/FM	C/I = 50 dB	Not applicable

BSS: broadcasting satellite service; FSS: fixed satellite service; BS: broadcasting service; FS: fixed service; FM: frequency modulation; FDM: frequency division multiplex; QPSK: four-level phase shift keying; VSB: vestigial sideband; pW0p: picowatts of noise power, psophometrically weighted.

6.4.2   Satellite Spacing and Dish Sizing Analysis

DTH networks are designed to work with smaller receiver dishes that cannot reject adjacent satellite interference as well as the larger dishes found in video distribution. Operation in the BSS bands is made to be satisfactory because the ITU has assigned the satellites to orbit positions. This assures a satisfactory service under expected conditions. In contrast, DTH systems that employ FSS satellites are not directly protected by the Radio Regulations, except through the coordination process. Typical frequency coordination efforts for FSS satellites allow satellites to be placed as closely as 2° apart, and there are even cases where a 1° spacing has been assumed. This causes downlink interference to be perhaps the largest single contributor of degradation to the DTH service quality. It partly explains why C-band receive dishes must be 1.8m or larger (the other main reason is the relatively low EIRP afforded by currently operating C-band satellites).

The following sections analyze in more depth the mechanism for interference into the DTH receiver. Figure 6.9 shows the ways that several satellites can produce downlink interference that can enter a DTH receiver. The small-diameter antenna on the ground has a main lobe and sidelobes on either side. These are exaggerated in

Figure 6.9   Potential downlink interference into a small-diameter DTH receive antenna from two adjacent satellites operating in the same frequency band.

the figure to make the example clear. In an actual case, the 3-dB beamwidth of a 45-cm (18-inch) antenna is approximately 4° at 12.3 GHz. This means that the 3-dB point on this antenna lies 2° off the peak, in either direction. A satellite located in this direction would only have about 3 dB of isolation (by definition). If this satellite has the same EIRP as the desired satellite in the direction of the DTH receiver, then the C/I is the same 3 dB. However, if it happens to be twice as powerful and therefore has 3 dB more EIRP in this direction, then the C/I is 0 dB. Either value is, of course, unacceptable. The proper value must be determined for the specific modulation and frequency spacing, which is discussed later in this section. The second adjacent satellite in Figure 6.9 is spaced far enough away to be located in the first sidelobe of the DTH receiving antenna. If we assume that the spacing places the interfering satellite at the peak of the first sidelobe, then the isolation is in the range of 15 to 25 dB, depending on the dish design. For the example of the 45-cm antenna, this sidelobe peak occurs at an offset of 6.4°.

The orbit spacing, DTH antenna diameter, frequency, satellite power, and resulting C/I are all interrelated. Figure 6.10 presents an example of the results of a typical analysis for a DTH system operating in the BSS portion of Ku-band. The

Figure 6.10   Typical calculation of C/I values as a function of DTH receive antenna diameter, based on WRC-77 BSS plan for Regions 1 and 3.

receive antenna follows the expected main beam and sidelobe gain characteristic given in the ITU Radio Regulations, Appendix S30. The required value of C/I will depend on the makeup of the link budget for the particular application. In digital service, values in the range of 10 dB to 15 dB are typical. If we assume a value of 15 dB, then we obtain Table 6.2, which shows the relationship between orbit spacing and dish diameter.

These are the basic facts that govern the interference design of a DTH system on the receiving side. From an uplink standpoint, the broadcast center can utilize large enough antennas to reduce the uplink component to an insignificant level.

The previous discussion had as an assumption that the desired and interfering carriers are cofrequency (on the same frequency). If there is an offset between the center frequencies of the carriers, the potential interference is suppressed by receive filtering and demodulator carrier rejection. As a result, the required C/I can actually decrease for the same level of interference effect. A typical example of this relationship is shown in Figure 6.11. The consideration of orbital interference is now a critical part of designing a new DTH network. If we are applying the ITU BSS Plan as it is incorporated into the Radio Regulations, then the only concern is that the actual hardware complies with the plan's assumptions. The regulations provide for making modifications to the plan, provided that the change or addition does not increase the interference into any satellite network already in the plan by more than the threshold values previously discussed for the BSS plans. This particular rule has been followed by many DTH operators and countries since there are good commercial reasons for making changes.

Table 6.2   Relationship Between Orbit Spacing and Dish Diameter, Given a C/I Value of 15 dB

Orbit Spacing (deg)	Dish Diameter (cm)
3	120
6	60
9	40

Figure 6.11   Protection ratio template for planning in Region 2, indicating that interference is suppressed for a frequency offset from the desired carrier (lower protection ratio means that more absolute interference power can be tolerated).

The rapid introduction of DTH systems around the world is increasing the difficulty of finding adequate FSS spectrum and orbit positions. For this one reason, BSS is likely to become more popular. Progress with digital DTH technology in general and DVB in particular is making it possible for literally any country to introduce satellite-delivered services within a year or less.

6.5   Differences Among DTH Systems

The basic architecture of a DTH system is probably the simplest of all satellite application systems because it relies on a single uplink to gather and transmit the programming and literally millions of home receivers to act as individual downlinks. An example of the type of receiver is shown in Figure 5.1.

DTH systems have multiplied in number and capability over the years, resulting in a rather substantial receiving public who own their own antennas. As a result, developers of these systems and services need to pay attention to the differences among systems, resulting in potential incompatibilities and threats from competition. The degree to which these impact the viability of a DTH service depend on the business strategy. For example, a totally standalone system that does not wish to attract viewers from other operators can implement an incompatible system. The advantage of doing this is that the technical or business parameters can be optimized. One such approach is to use very high satellite EIRP and correspondingly small receiving dishes that cannot also receive from existing medium-power DTH services transmitted from FSS satellites. On the other hand, an existing base of DTH viewers must adhere to an appropriate number of these factors in order to succeed. A given incompatibility can, of course, be resolved by modifying or adding equipment on the ground—one DTH receiver at a time. This approach was pursued by EchoStar and DIRECTV as the means to add programming to relatively narrow segments such as cultural, ethnic, foreign language, and sports.

6.5.1   Downlink Frequency

The choice of downlink frequency is probably the most strategic in a technical sense. As of the time of this writing, the bands either in use or of potential interest include (frequencies indicated are for the downlink [4]):

- S-band BSS (2,520 to 2,670 MHz);
- C-band FSS (3.4 to 4.2 GHz);
- Ku-band FSS (10.70 to 12.2 GHz);
- Ku-band BSS (11.70 to 12.7 GHz);
- Ka-band BSS (21.4 to 22.0 GHz).

Systems using C-band and Ku-band FSS allocations generally use linear polarizations, while the BSS allocations are generally used with circular polarization. Most DTH receivers are designed for a specific downlink frequency and total bandwidth, with the capability to switch polarizations or receive in both polarizations simultaneously (requiring two LNBs and coaxial cables). Some offer multiple bands, using dishes with dual frequency feeds that employ separate LNBs. This approach permits reception from different satellites and service providers since the dish may be repointed using a motor drive. It is more for the enthusiast or hobbyist and not the typical consumer. For a consumer installation, the dish is fixed in position and the particular band is permanently part of the design, so access to alternative programming and satellites requires additional feeds or entire antennas.

From an uplink standpoint, there are nearly as many options from which to choose. Developers of satellites and DTH systems usually employ the paired uplink band for the particular downlink. For the Ku-band BSS allocations, the uplinks may employ FSS spectrum that is allocated for "feeder links" in the range of 17.3 to 18.1 GHz. Use of the paired band is usually a good idea because the spectrum is already allocated and probably available at the assigned orbit position. On the other hand, the 18-GHz uplink frequency is subject to increased rain attenuation, particularly in tropical regions with a lot of thunderstorm activity. The objective of the uplink design is to maintain the signal for a high percentage of the time, typically 99.95%. Above C-band, this may involve the use of some form of automatic gain control in the satellite and uplink power control in the broadcast Earth station as well.

6.5.2   Significant Differences in Satellite EIRP

Satellite EIRP has the greatest single impact on DTH dish diameter. The basic relationship imposed by free-space propagation is that every doubling of EIRP (in power, not decibels) allows the receiving dish diameter to be reduced by a factor of the square root of two, or 1.414. Thus, if we increase the satellite EIRP by 3 dB (a factor of 2 in power), a 1-m dish can be squeezed down to 1/1.414m or 71 cm. Another 3 dB (or a total of 6 dB of increase) brings it exactly to 50 cm. A DTH system designed with a high EIRP produces the smallest sized receiving dish. This minimizes installation cost and improves aesthetics; however, a smaller dish may be incapable of adequate reception from a medium-power FSS satellite and could receive unacceptable levels of adjacent satellite interference. (Adjacent satellite interference places a floor on dish size, as reviewed at the end of this chapter.)

Dish size is one of the most easily seen differentiators among services. The smaller the dish, the less obtrusive and potentially objectionable is the installation at the home. Smaller dishes are better able to withstand high winds and ice. The DIRECTV and DISH networks in the United States have benefited from a small dish size because of the lack of a perceived problem with appearance; Federal legislation also prevents local zoning and homeowner agreements from precluding their installation at residences. In contrast, there is a general feeling that 1.8-m C-band dishes are only acceptable in rural areas where normal TV reception is limited. This rule is perhaps violated in some developing countries where a rather obvious C-band receive antenna is a sign of prosperity. Developing the satellite radiated power to produce a small receiving dish is no longer the problem that it was once considered to be. Space power amplifiers, particularly the high-power traveling wave tube (TWT), are available up to 200 W/channel. These deliver very high EIRP over a wide coverage footprint reaching as much as 55 dBW at Ku-band, affording consumers the opportunity to use small receiving dishes in the broadest possible market. At C-band, EIRP values in excess of 40 dBW are produced across a hemisphere to allow very extensive coverage from a single satellite. Through the use of DVB-S with concatenated coding and the prospect of turbo codes, DTH is fast becoming a global common denominator for television broadcasting. Our discussion of business TV in Chapter 4 provides the methodology where this kind of coverage can serve a business or government communication need.

6.5.3 Polarization Selection (LP or CP)

Designers of DTH systems can select either linear polarization (LP) or circular polarization (CP) for the downlink. The standard BSS assignments at Ku band mostly employ CP to simplify the installation of home dishes because the feed does not need to be rotated after the antenna is aligned in azimuth and elevation. This is a factor in large DTH networks, where antennas number in the millions, which is why DIRECTV and DISH have gone this route (it happens to be mandated by both the FCC and ITU in accordance with the ITU orbital plan). In FSS systems, we encounter both CP and LP, depending on the satellite design. Intelsat, NewSkies, and some Russian C-band satellites were designed around CP to minimize the impact of Faraday effect (which is a day-to-night change in polarization angle if LP is used). LP is the baseline in all other C-band satellites because this tends to simplify the design of the Earth station transmit and receive feed system. It may also be possible to improve the efficiency of satellite antennas as well. In addition, depolarization caused during heavy rain is less of a factor with LP.

A linearly polarized system is incompatible with a circularly polarized one. For example, there is a 3-dB loss when a LP-to-CP or CP-to-LP connection is attempted. Even worse, the polarization isolation between CP and LP transmissions is 3 dB in the ideal case, independent of the sense of polarization of either link. With only 3 dB of isolation, most communications will be lost. As a result, the only means to control interference between adjacent CP and LP satellites is to increase the orbit separation.

Because of asymmetric field concentration on typical offset-fed dishes, CP experiences a slight squint of the beam, right or left of center; the direction of shift is opposite the sense of polarization [5]. The squint occurs when the reflector system

causes a depolarization of the incident linear vector field components, and these components are out of phase with respect to each other, as is the case with circularly polarized fields. A similar depolarization occurs for linear polarization that introduces a cross-polarized interference component; however, the beam does not squint in this case. The squint occurs in the plane transverse to the principal offset plane and is toward the "left" as an observer looks in the direction of main-beam wave propagation, when the main beam of the antenna system is right circularly polarized (RCP), and to the "right" for a left circularly polarized (LCP) main beam. The following equation provides a rough order of magnitude estimate for the squint angle:

$$\vartheta = \sin^{-1}\left[\frac{\lambda \sin\theta_{tilt}}{4\pi F}\right]$$

where λ is the wavelength, θ_{tilt} is the tilt angle from the parabola axis to angular center of the feed's illuminated region boresight, and F is the focal length in the same units as λ. Antenna diameter does not matter. Anything that can be done to reduce the antenna's depolarization will reduce the squint as well.

6.5.4   Frequency Plan Differences (Channel Spacing)

The arrangement of the frequency plan for the downlink impacts the design of the tuner in the home receiver. Examples of the most popular channel spacings for C- and Ku-band satellites are provided in Table 6.3. Channelization of the BSS frequencies impacts the maximum bit rate on each carrier and was originally supposed to be according to an international plan. Home IRDs must be capable of turning to these frequencies and separating them from each other. The original channel characteristics were selected to permit a single FM TV carrier to occupy the bandwidth. With the advent of digital compressed video, as many as 10 standard definition video channels are time division multiplexed onto a carrier that can pass within the same bandwidth and channel characteristics. FSS assignments at C-band follow a de facto standard (e.g., 40 MHz, starting at a center frequency of 3,720 MHz), while those at Ku-band generally do not. The potential use of Ka-band as a DTH medium is likewise not standardized.

Most receivers have standardized on certain transponder frequencies, bandwidths, and spacings. The most flexible design allows the user to select literally any frequency in the downlink range using a digital frequency synthesizer. However,

Table 6.3   Typical Transponder Channel Spacings in the C- and Ku-Bands

Band	Lowest Center Frequency (MHz)	Channel Spacing (MHz)	Usable Bandwidth (MHz)
S	2,114	28	20
C	3,720	40	36
C extended	3,400	40	36
Ku (FSS)	10,750	40	36
		60	54
Ku (BSS)	11,715	30	27

this type of device must be properly programmed, which can be a burden for a non-technical user (i.e., having to tune to a specific frequency, like 11.7255 GHz, as opposed to a simple channel assignment, like channel 10). Alternatively, the tuning is managed over the satellite link, making the process transparent to the user. Instead, the user selects programs using a remote control unit and on-screen channel guide. In an ideal DTH system with a uniform receiver design, there is only one channel bandwidth allowed. In the real world, IRD suppliers make their devices flexible enough to work with standard and nonstandard channel bandwidths. However, this should be verified before making a purchase.

6.5.5   Digital Transmission Format (QPSK, 8PSK, 16 QAM)

With 20 years of history, FM-TV is essentially a unified standard that satisfies the needs of the various analog TV systems used around the world. The situation with regard to digital transmission is moving in the same direction with the DVB system as the foundation. As discussed in Chapter 5, DVB-S offers a standard set of transmission parameters employing concatenated Reed-Solomon and convolutional coding, interleaving and energy dispersal, and QPSK modulation. This particular arrangement satisfies nearly all needs in the DTH domain as it is currently defined. However, the applications are evolving in new directions that demand alternative features. For example, transmitting video to aircraft imposes an even greater restriction on receive antenna design. Thus, turbo coding with 2 dB or more of reduced power requirement will be the preferred forward error correction scheme. In another application, more channels will need to be crammed into the limited bandwidth of a standard 27-MHz BSS transponder. This will require a modulation format with a greater ratio of bits per second to hertz. Two principal bandwidth efficient modulation (BEM) methods to be considered are eight phase shift keying (8PSK) and 16 quadrature and amplitude modulation (16 QAM). One can go further to a technique called trellis modulation, which combines 16 QAM with a sophisticated signal estimation scheme. The best application of trellis modulation is in the common dial-up telephone modems that stuff nearly 56 Kbps through the narrow bandwidth of a telephone connection. This works because the effective carrier power in terms of E_b/N_0 is measured in tens of decibels.

Whenever the ratio of bits per second per hertz is increased, there is generally a penalty to be paid in terms of power. In a nutshell, we are trading bandwidth efficiency for power. The move from 4 to 8 phases will reduce bandwidth occupancy to two-thirds (equivalently, a 33% reduction of bandwidth needed, allowing one-third more TV channels to be placed within the same transponder). The power penalty to do this is approximately 3 dB (e.g., a doubling of RF power in watts). If we are speaking about the satellite, then the prime dc power of the satellite increases by the same percentage. This can be directly made up by reducing the coding rate, as from 7/8 to 5/8. This change in coding rate happens to increase bandwidth by the inverse ratio, or 7/5 (e.g., 40%). This is obviously a no-win situation, so the only practical way to take advantage of this form of BEM is to find a way to increase the power by itself. An alternative is to employ turbo coding.

Similar remarks can be made about 16 QAM, which likewise requires more power to achieve a satisfactory link. In this case, the bandwidth use can be cut by a

factor of four relative to QPSK, while the power correction would be a factor of four as well.

6.5.6   Video Signal Format

In Chapter 4, we identified the three standard analog TV signal formats, namely, NTSC, PAL, and SECAM. These represent an important differentiator among DTH systems. A DTH system for North America can conveniently employ NTSC, while in China the PAL system is the appropriate vehicle to reach the mass market. Southeast Asia and South America are more of a challenge because both employ NTSC and PAL within the respective regions. Africa and Europe are each split between PAL and SECAM.

While multiple standard TV receivers exist in markets like Western Europe and East Asia, they cannot be assumed as a given in DTH service. Therefore, it may be necessary to transmit programming in more than one format. Conversion between formats is certainly possible but is not currently available within set-top boxes. A significant difference in formats is the fact that NTSC uses 30 frames per second while PAL uses 25. One can imagine that a time will come when TV sets will be more like PCs, and so the video signal format will cease to be a differentiator.

In the case of digitally encoded TV, the clear trend is toward the MPEG standards. MPEG 2 is the leader in the market at the time of this writing. We reviewed in Chapter 5 some of the specifics of the current implementation of MPEG 2 with the main profile and level. This provides a good compromise among the factors that designers can currently control. However, in time, digital processing technology will advance to the point that higher levels and more advanced profiles can be adopted into low-cost equipment. Add to this the prospect of improved or high-definition TV, which is something that satellites started delivering ahead of existing terrestrial cable and over-the-air resources.

6.5.7   Scrambling and Conditional Access

The purpose of the scrambling technique is to render the picture unwatchable unless and until the subscriber has been authorized by the service provider. The typical reasons why this is necessary or desirable are to:

- Prevent piracy of the programming by a functioning but unauthorized receiver;
- Control the customer base to assure payment for services obtained;
- Protect intellectual property rights such as copyrights that might not have been granted within all service areas;
- Control delivery of services to satisfy domestic regulations in particular countries;
- Implement PPV services to the end user so that specific programming can be turned on and off directly by the user;
- Provide levels of service to different user groups.

Because scrambling systems fall into a number of categories, they can be difficult to implement as an open system. One approach is to scramble the picture while it is in analog form, which is comparable to what is done on current cable TV systems (see Chapter 4). For a digital format, the information in the picture has been compressed using the DCT, which requires a properly designed receiver to recover it. Scrambling amounts to rearranging the bits in the sequence according to an encryption algorithm.

The strongest and most secure way to do this is through encryption using algorithms like the digital encryption standard (DES). As is usually the case, the process itself is typically not kept secret since it is likely part of a standard like DVB. The key, on the other hand, must be handled properly because with it anyone can recover the picture, sound, and data. Part of the key can be delivered over an encrypted satellite link as long as the second key used to recover it is provided by an independent path (e.g., a password or smart card). All of these aspects of scrambling represent areas of possible incompatibility. This is why scrambling systems are usually supplied either as part of a validated standard (where an agency performs type acceptance of products from various vendors) or as a proprietary package from a single vendor.

The last step of compatibility has to do with the way the subscriber's set-top box is controlled as part of service delivery. Since a geostationary satellite can broadcast its signal over a wide area to a potentially very large subscriber base, DTH operators and programmers must have an efficient and effective means to control the ultimate distribution of the programming product. The VideoCipher II system was one of the first to introduce conditional access on a wide scale. Also, DVB conditional access features reviewed at the end of Chapter 5 provide a workable baseline for modern digital DTH applications.

The satellite link offers an attractive medium for downloading control information to remote equipment, hence the preference to use it for conditional access. However, a physical connection is more secure and flexible since it provides more direct control to the DTH programmer as well as a mechanism for return data. The smart card performs many useful tasks under conditional access, such as holding the subscriber's key and securing the process of unlocking the decoder. The card also can store the viewing history for the particular installation. An alternative is to have the subscriber enter the authorization code manually (like a password) into the decoder box. A connection to the telephone network is another way to isolate the key from the satellite link.

DTH is applied to other one-way communications applications such as data broadcasting to small dishes. CD-quality audio broadcasting is another viable application. The EBU has developed standards, and DTH companies offer near CD-quality audio to subscribers. Chapter 7 considers the dedicated delivery of near CD-quality audio to vehicles and homes.

The question of piracy is a serious one for DTH services because once the CA system is broken, revenue losses grow. There are tough laws in the United States that provide DTH operators with teeth in their pursuit of pirates. Elsewhere, it may not be so straightforward. The providers of CA systems and technology continue to come up with improvements and extensions that complicate the pirate's life. In 2002, things got somewhat complicated when a supplier of CA technology was sued by some customers. The precise nature of the suit involved business as well as

technical issues, but the fact that such an action would take place provides some indication of how serious the matter may become.

6.6 Survey of DTH Systems

The history of DTH as a legitimate satellite offering provides a good introduction into the development of a modern system and service. The old adage, "those who fail to heed the lessons of history may be doomed to repeat them," would seem to apply. Therefore, we want to make sure that anyone who considers making the kind of investment that DTH requires has the benefit of this background. Readers are encouraged to conduct their own research into the fine points of whichever of these systems represent the closest parallel to what they plan to do.

At the ITU World Administrative Radio Conference in Geneva, Switzerland, in January 1977 (WARC-77), representatives of most of the nations of the world and international regulators assigned BSS orbit positions and channels to member countries except those in North America and South America (the latter, Region 2, was subsequently assigned in a Regional Administrative Radio Conference in 1983). This meant that big countries like Russia and China as well as small countries like Kuwait and Mauritius each received their dedicated assignments for domestic BSS satellites. To be consistent with expected demand, larger countries with larger populations were assigned multiple orbit positions. Characteristics of the assumptions used in the plan for ITU Regions 1 and 3 are provided in Table 6.4. The global assignment plan contained in the proceedings of both conferences can be found in Appendix S30 and S30 A to the ITU Radio Regulations, which can be obtained at http://www.itu.int for a cost.

Table 6.4 Assumed Characteristics of BSS Satellites in the WARC-BS Plan

Feature	Specifications
EIRP	Typical 63 dBW
Minimum satellites per country	1
Channels per satellite	5 (or multiples thereof)
Frequency plan	23-MHz transponders
Polarization	Circular polarization (left hand and right hand)
Orbit spacing between satellites	Fixed orbital spacing between satellites (3° minimum separation) 9° for cofrequency assignments Satellites are colocated where possible
Maximum spacecraft reflector size	2m
Beam shape	Elliptical
Sidelobe radiation pattern	Tightly specified
Assumed home receiver characteristics	
Dish size	No smaller 0.9m
Receiver noise figure	No lower than 9 dB
Cross-polarization isolation	Relatively poor
Angular discrimination	Relatively poor

This was a classical planning effort by regulatory experts and technologists who worked very hard but lacked a practical feel. They made several assumptions that may have seemed sensible at the time but failed to take account of future improvements. For example, the WARC-77 plan assumed that the maximum spacecraft antenna size would be 3.5m, the typical low-cost home receiver would have a system noise temperature of 1,000K, and the maximum transmit power would be 1 kW. Over time, it would be established that LNAs could be produced cheaply with less than 70K, and therefore transmit power could be held to 200W. Another technical assumption that has turned out to be too pessimistic was that satellites could not carry sufficient reserve battery power to sustain full capacity during an eclipse. As a result, satellites were all placed to the west of the service area so that the eclipse would occur after local midnight. This has led to recent filings for satellites to the east of the coverage area that would offer less potential interference to already assigned stations.

Due to these difficulties, the BSS assignments remained dormant for at least a decade while other approaches using the FSS allocations were pursued. Japan was the only country to actually use their assignment during the 1980s and this was for satellites with only two or three video channels. It was not until the launch of DIRECTV 1 that true BSS services were brought to the public.

In the remaining sections, we review the status and development of DTH systems around the world. The particular systems being considered are summarized in Table 6.5 according to the properties described earlier in this chapter. The

Table 6.5   Comparison of DTH Systems in Development or Operation Around the World

System	Satellite	Coverage Area	EIRP Performance	Dish Size	Frequency	Polarization	Delivered Format	House-holds
DIRECTV	DBS	United States	50 dBW	45 cm	Ku BSS	CP	DSS (MPEG+)	12 million
DISH	EchoStar	United States	50 dBW	45 cm	Ku BSS	CP	DVB-S	9 million
SES Astra	Astra	Europe (German-speaking)	53 dBW	60 cm	Ku FSS/BSS	LP	FM-PAL	25 million
SKY	Astra	United Kingdom and Ireland	53 dBW	45 cm	Ku FSS/BSS	LP	DVB-S	8 million
TDF	TDF	France	57 dBW	45 cm	Ku BSS	CP	D2MAC	<1 million
Eutelsat	Eutelsat	Europe	48 dBW	60–75 cm	Ku FSS/BSS	LP	DVB-S	15 million
Thor	Thor	Norway	53 dBW	45–60 cm	Ku BSS	LP	DVB-S	2 million
Indovision	Cakrawarta 1	Indonesia	45 dBW (S-band)	60 cm	S BSS	LO	DVB-S	<1 million
Astro	Measat	Malaysia	55 dBW	60 cm	Ku FSS	LP	DVB-S	2 million
SkyPerfectTV	JCSat	Japan	53 dBW	45 cm	Ku FSS	LP	DVB-S	3 million
STAR TV	AsiaSat	Asia	36 dBW	2m	C FSS	LP	DVB-S	5 million
SKY Latin America	PanAmSat	Latin America	48 dBW	60 cm	Ku FSS	LP	DVB-S	1 million

discussion is not intended to be all-encompassing but rather to give the reader an appreciation for the range of options as well as the strengths and weaknesses of each approach.

6.7 Digital DTH in the United States

The U.S. home TV market poses a special challenge to developers of DTH systems as witnessed by the demise of STC and USCI. Unlike the experience of Sky in the United Kingdom, systems in the United States that offered 5 to 10 channels of programming did not have a good experience. This happened principally because of competition from cable TV, with typically 50 channels, and C-band backyard systems that can access more than 100 channels. A theory evolved that a successful DTH business would also require as many as 100 channels or more and be as convenient as cable to use. Also, the equipment must be unobtrusive and have a cost comparable to that of, say, a 30-inch TV or high-quality VCR. To achieve this number of channels, the advances in the area of digital compression had to be harnessed. Thanks to the work underway by MPEG (discussed in Chapter 5), the technical base became available in 1991.

6.7.1 DIRECTV

The first company to conceive such a system and bring it to market was Hughes Electronics Corp., a subsidiary of General Motors. Later known as DIRECTV, the system pioneered the use of all digital technology for the compression, transmission and management of a DTH service. The only precursor in terms of a digital DTH was tried by a startup in Seattle called Pacific StarScan. Using off-the-shelf compression technology from Compression Laboratories, Pacific StarScan offered a trial service on a Ku-band FSS satellite. However, the project failed to reach the market because, like USCI, they ran out of money.

The digital platform of GM's DIRECTV was a success from the start and quickly reached two million subscribers in 1996, after just 2 years of operation. By 2003, DIRECTV had 12 million paying subscribers—above the objective set when the system was originally proposed to GM. Digital DTH was subsequently introduced in Latin America, Asia, and Europe in much less time than it took in the United States. As reviewed at the end of Chapter 5, the DVB standard is allowing MPEG 2–based systems to proliferate.

Studies conducted by Hughes in 1991 showed that a subscriber base of around 10 million U.S. households was achievable provided that the service meet the following goals [6]:

- Receiving dish size of approximately 45 cm to aid in installation and for aesthetics;
- A diverse programming mix offering approximately 150 channels;
- An initial cost of less than $1,000, preferably in the neighborhood of $700;
- An average monthly service charge of approximately $25.

This original target of DIRECTV, Inc. was to begin marketing to consumers in late 1994, during the peak Christmas selling season. The somewhat-proprietary DIRECTV Satellite System, as the DIRECTV consumer equipment is known, was introduced on schedule and subsequently recorded the largest initial new consumer electronics product sales ever. These first units were produced by Thomson Consumer Electronics under the RCA brand, a very familiar name in U.S. home electronics. It was critical that the DIRECTV service quality be as close to perfect as possible at the time of business launch. A delay in the roll-out would be preferable to failing to meet a high standard. This goal was also met, and consumer reaction to video and audio quality has been excellent. A second supplier, Sony, introduced their DSS system to the U.S. market in 1995. Subsequently, the Hughes Network Systems (HNS) subsidiary of Hughes Electronics began manufacturing DSS equipment.

The first three high-power Ku-band satellites for DIRECTV, Inc., built by Hughes Space and Communications (now Boeing Satellite Systems), employ the HS-601 three-axis spacecraft design. The coverage is confined to the continental United States and southern Canada, delivering EIRP in the range of 48 to 58 dBW. Initial service was provided by DBS-1, which was shared between DIRECTV, Inc., and U.S. Satellite Broadcasting (USSB), a part of the Hubbard Broadcasting family of companies. USSB was subsequently bought out by DIRECTV, facilitating one bill per customer. DBSI was launched on December 13, 1993. The DIRECTV service contained a mix of more than 150 channels at the 101° WL U.S. BSS orbit position. With the acquisition of Primestar, DIRECTV took over some channels at the 119° WL slot and has expanded services.

The DIRECTV uplink center in Castle Rock, Colorado, is shown in the photograph in Figure 6.5. Each of the 16 transponders on DTH-1 has 24 MHz of bandwidth. Two additional satellites have been launched, and their combined capacity of 32 transponders is dedicated to entertainment service. Additional satellites were added to bolster capacity and provide for the delivery of local channels. Currently, DIRECTV relays seven of the locally available TV channels to over 50 metropolitan area markets in the United States; expansion to almost 100 markets is planned. This was a massive endeavor, which required not just the transponder capacity but also the ability to backhaul seven channels from each of the 60 markets to the uplink center.

6.7.2   EchoStar DISH Network

Charlie Ergen started EchoStar Communications Corporation in 1980 under the name of Echosphere, which sold C-band satellite TV dishes to rural homes in Colorado. Under his vision and leadership, EchoStar launched DISH Network in 1996, which has become the fastest growing DTH satellite television company in the United States with more than 9 million customers. In 1987, EchoStar filed for a DBS license with the FCC and was granted access to orbital slot 119° WL in 1992. The company began its move toward providing its own DBS service on December 28, 1995, with the successful launch of EchoStar I. That same year, EchoStar established the DISH Network brand name. The eight satellites that make up the EchoStar fleet, operated by the company from their Cheyenne Uplink Center, have the capacity to provide more than 500 channels of digital video, audio, and data services via DISH Network service to homes, businesses, and schools throughout the United States.

The DISH Network has enjoyed wide acceptance and is a strong competitor to DIRECTV. Although both services offer essentially the same programming, the DISH Network has emphasized price and ease of use as differentiators. In 2001, EchoStar was selected by GM as the successful purchaser of Hughes Electronics, including in particular DIRECTV. This acquisition would result in one primary DTH supplier in the United States. As a result of political and economic objections, the FCC and Department of Justice have blocked this from proceeding.

6.7.3   Other U.S. DTH Operators

Having reached 20 million households in the United States, DTH has attracted some potential new entrants into the market. Cablevision, a leading cable system operator and provider of sports and other cable programming, is spinning off a new DTH operator. Having purchased a high-power DBS satellite from Lockheed Martin, Cablevision is in the process of launching the service.

SES Americom has announced another DTH system, which they call Americom to Home. Rather than be the programming provider, SES Americom desires to use a business model that was successful for SES Astra in Europe. This is to provide the technology platform and leave the programming and customer acquisition and management to others. This bears some resemblance to the approach in the C-band cable distribution business, where SES Americom, PanAmSat, and Loral Skynet provide the transponders to HBO, AOL Time Warner, and the other cable TV networks. Americom at Home was still in the midst of FCC proceedings to obtain permission to use BSS frequencies at the time of this writing.

6.8   European DTH Experience

The European experience with satellite TV in general and DTH in particular is almost exclusively at Ku-band. A difference from the United States is that both cable TV and DTH grew up at the same time. This is because cable TV reaches a high percentage of total TV households in only a few countries like Germany and Belgium, while in others like France and the United Kingdom DTH has been the primary market for reaching the pay TV consumer. There has also been an explosion in the use of DTH in Central and Eastern Europe, where the existing broadcasting and cable infrastructure tends to be extremely weak.

The points already made about simplicity, cost, and programming have been validated over and over again. Sky TV, the first service to be offered directly to the public, is staying ahead precisely because of these aspects. Other activities in France and Germany were slow at the start of the game, being hampered by government red tape and inflexibility. In many ways, Europe poses a much more exciting future for DTH simply because of the wider array of ventures and economic conditions. The company that continues to be successful is Société Européenne des Satellites (SES), the commercial operator based in the Grand Duchy of Luxembourg. SES-Astra, the European segment of their business, established its orbit positions as the most popular real estate with the associated large base of existing TV receive antennas. This considers both markets—cable TV systems and DTH home receivers.

It is beyond the scope of this book to go into detail for each and every one of these systems. Not shown in Table 6.5 are the systems of INTELSAT, NewSkies, and Loral Skynet. These provide some video distribution capabilities throughout Europe but are hampered to some degree by the low-elevation angles due to the western location of the satellites.

We now review some of the key DTH systems and activities in Europe, followed by a few others around the world. These break down into government-sponsored and commercial systems. Government-sponsored systems tend to be more innovated in a technology sense, pushing the state of the art. They usually have an industrial mission, that is, to promote domestic industry and gain a leg up in the international market. Commercial systems want to make money (usually), so their emphasis is on the market for the services. The information is very basic, so readers who need details for system design and service evaluations are advised to contact the appropriate satellite operator for current detailed information. Details on programming can be obtained on the associated Web sites.

6.8.1 SES-Astra

It is always good to start off with a success story. In 1988, SES was in direct conflict with Eutelsat and BSB and it was not clear at all if Astra would be successful. SES appeared to have a more marketable approach, using a medium-power satellite built in the United States with a 16-channel repeater. But Eutelsat had a big head start since its low-EIRP satellites were delivering most of the satellite TV in Europe.

SES-Astra is a part of public company, SES-Global, that is traded in Europe and with significant ownership by several public and private investors, including GE Capital, several European banks, and the Luxembourg government. SES-Global is a group management company that operates through its 100% owned companies SES-Astra and SES-Americom, and through the network of partners, in which SES-Global holds interests: Americom Asia-Pacific, AsiaSat, Nordic Satellite AB (NSAB), Nahuelsat, and Star One (formerly Embratel of Brazil).

The idea for SES-Astra, which predates SES-Global by about 15 years, came from an American, Clay T. (Tom) Whitehead, who is also credited with the successful Galaxy system established by Hughes and now part of PanAmSat. Whitehead's original Coronet enterprise failed to get started and provided the framework for the eventual SES venture (Whitehead retained a small ownership interest). Whitehead's main innovation was to determine that the spacecraft should have 16 medium-power, Ku-band transponders rather than the five high-power transponders that were being considered under the WARC plans.

By the time of launch on December 11, 1988, most of Astra 1A's transponders had been leased. The last four transponders were taken for German-language channels. The anchor customer for this hot bird was Sky TV, controlled by News Corp. For the bulk of the channels, the marketing agent was BTI, which acquired rights prior to launch. Each transponder is leased according to a private deal, much the way channels are acquired in the United States. SES-Astra was extended to a second orbit slot and the system has undergone a transformation from analog to digital television transmission. Table 6.6 presents the current programming lineup. One of the newer spacecraft, Astra 1H, includes Ka-band repeaters to be used for a return

Table 6.6   A Sample of Programming on Astra 1A, 1B, 1C, 1D, and 1E

Number	Service	Lessee	Encryption	Format
1	Screensport	WHSTV		PAL
2	RTL Plus	RTL Plus		PAL
3	TVS (Scansat)	Scansat	Eurocrypt	D2-MAC
4	Eurosport	Sky		PAL
5	Lifestyle	WHSTV		PAL
5	JSTV	JSTV		PAL
5	Childrens Ch.	Childrens Ch.		PAL
6	Sat1	Sat1		PAL
7	TV1000	Scansat	Eurocrypt	D2-MAC
8	Sky One	Sky		PAL
9	Teleclub	BetaFilm/Kirch		PAL
10	3Sat	3Sat		PAL
11	Filmnet	Filmnet	Satbox	PAL
12	Sky News	Sky		PAL
13	RTL Veronique	Sky	Irdeto	PAL
14	Pro 7	Pro 7		PAL
15	MTV Europe	MTV		PAL
16	Sky Movies	Sky	Palcrypt	PAL
16	Shopping Ch.	Shopping Ch.		PAL

channel by satellite (ARCS, the basis of the DVB-RCS standard). This arrangement allows a return channel at Ka-band to coexist at the same orbit position with the large quantity of Ku transponders that represents SES's core business.

The idea of "More Channels – More Choice" was subsequently exploited with additional spacecraft that were colocated at 19.2 EL. SES became self-taught experts at colocation of spacecraft, culminating in a world record of seven space-craft colocated for a short time in 1997 (quasi-permanently in 1999), and eight for a short time in 2001. SES has 120 RF Ku-band channels at 19.2 EL providing programming to all of the markets in Western and Eastern Europe. Since then, additional orbit slots were opened at 5.2° EL, 24.2° EL, 24.5° EL, and 28.2° EL.

The Astra spacecraft are specified with emphasis on delivery, performance, and cost. This gave SES an advantage over its European rivals. A common characteristic of ASTRA satellites is the ability to cover more channels than the spacecraft have power to serve simultaneously, so that when colocated with similarly channelized spacecraft mutual backup (for uninterrupted service) is possible. In 2002 SES had a setback with a launcher failure that resulted in a large replacement satellite, ASTRA 1K, being stuck in a nonviable low orbit from which it was then deorbited (to avoid leaving space debris). However, the robust fleet concept limited the impact to only the coverages unique to ASTRA 1K.

6.8.2   British Sky Broadcasting

BSkyB is actually a contraction of British Satellite Broadcasting (BSB) and Sky TV, brought about by the merger of the two organizations in 1992. BSkyB is a publicly

traded corporation with shares of ownership spread among several commercially oriented companies in the United Kingdom, Australia, and France, but lead by News Corp. Originally, BSB was awarded the only U.K. license to use the BSS assignments from WARC-77. This venture was completely funded though investments by owners and by bank debt. It contracted with Hughes for the manufacture and placement in orbit of two satellites of the HS-376 spinner design (a conservative approach, including use of the world's most reliable launch vehicle, the Delta). The venture had to use D-MAC video and audio encoding according to its license, which followed existing EC policy for true DBS satellites in Europe. A problem with the manufacture of D-MAC receivers delayed the roll-out of service, which severely hampered the business. In the meantime, Sky TV was delivering both set-top boxes and programming. BSB emphasized that their ground receiving antennas were much smaller than Sky's and promoted the use of flat plate arrays called "Squarials." With the merger, BSkyB dropped the use of the BSS frequencies and satellites and continued to grow using SES-Astra. It can be said that BSB did a lot to rekindle interest in DBS in the United States.

British Sky Broadcasting is a leading provider of sports, movies, entertainment, and news, and whose channels are received by more than 10 million households in the United Kingdom and Ireland. The launch of the United Kingdom's first digital television service, Sky Digital, on October 1, 1998, signaled the start of a new era in British broadcasting, and was done coincident with a change of orbital slot to 28.2° EL. It remains the fastest, most successful roll-out of any digital TV service in Europe, attracting 5.9 million customers at the end of March 2002 who yield annual average revenue per user of greater than 330 GBP with churn rates under 10%. BskyB's analog services from 19.2° EL were shut down in 2001, but 19.2° EL is still strong in non-U.K. markets.

6.8.3   Télédiffusion de France and TV-Sat

France made a strong technological entry into DBS through its Télédiffusion de France (TDF) program. The project employed the WARC-77 orbit positions and channels through a government-supported project. Several French aerospace and electronics firms contributed to ground and space elements. The French and German DBS projects were handled jointly to economize on the development of the spacecraft. These were manufactured by the Eurosatellite consortium, which was composed of the French and German companies, Aerospatiale (24%), Alcatel Espace (24%), Daimler Aerospace (24%), AEG (12%), ANT (12%), and ETCA/ACEC (4%).

From the French government side, the project was supervised and funded by TDF and CNES. In France, TDF acted as a common carrier for broadcasters by operating the broadcasting stations and the microwave network. It was natural for them to take on the operation of a DBS activity. Similarly, CNES is the national resource for space technology and operations. The project achieved many technology goals, including demonstrating transponders at 200+W each. The satellite provided coverage of France and several French-speaking areas of Europe. In addition, major French-speaking cities in North Africa could receive programming with antennas up to 2m in diameter. Both satellites were successfully put into orbit by

Ariane and became operational. Table 6.7 details the technical characteristics of the TDF spacecraft. TV-Sat was originally controlled by the Deutsche Telekom and was built to the same general specifications as TDF by the Eurosatellite consortium.

TDF and TV-Sat BSS satellites were taken over by Eutelsat, with TDF 2 and TV-Sat 2 relocated to 36 EL and 12.5 WL. Subsequently, TDF 2 was replaced by Eutelsat W-4 to provide services to Africa, Russia, and Eastern Europe. Likewise, TV-Sat 2 was replaced by Eutelsat's first international satellite, ATLANTIC BIRD 1. This brought the chapter of European governmental DBS programs to a close. Eutelsat continues as a for-profit satellite operator with quasi-governmental ownership with intentions to float ownership on the stock market, like NewSkies. The following section reviews the development and operation of Eutelsat.

6.8.4 Eutelsat

Eutelsat provides extensive European coverage in support of the full array of FSS applications. Patterned after Intelsat, the original member/owners of Eutelsat were the national telecommunications operators of western and eastern European nations. When Eutelsat was formed in 1976, the direction was clearly toward Western Europe; but after the breakup of the eastern block, countries from east to west joined. Today, leading global telecommunications operators are among Eutelsat's most active partners: Belgacom, British Telecom, Deutsche Telekom, France Telecom/Globecast, KPN, Russian Satellite Communications Company, Telecom Italia,

Table 6.7   Technical Characteristics of the TDF Spacecraft

Feature	Specifications
Launch date	
1	October 1988
2	July 1990
Launch vehicle	
1	Ariane 2 (single launch)
2	Ariane 4 (dual launch)
Start of service	1989
Projected lifetime	9 years
Orbital position	19° W
Manufacturer	Eurosatellite
Payload characteristics of TDF	
Frequency range	
Uplink	17.3 to 17.7 GHz
Downlink	11.7 to 12.1 GHz
Transponder bandwidth	23 MHz
Polarization	Circular (right hand)
Number of transponders	5
Number of TWTAs	6
Power per transponder	230W
Typical EIRP	63 dBW
Coverage area	France (Germany on TDF 2)

Telekom Polska, and many others. Eutelsat's new status as a société anonyme with a directorate and a supervisory board provide an accelerated decision-making process and give the company the full commercial freedom required to continue to grow the business and increase shareholder value.

Several of their satellites provide program distribution to cable systems, in competition with SES-Astra, and they have nearly an equal TV broadcast market share with their orbital slot at 13° EL, easily reachable with a dual feed or second 60-cm dish. Historically and unlike SES, Eutelsat's spacecraft were built by European consortia. A preferred-source policy is no longer followed by either company as they pursue a totally commercial agenda.

In the late 1980s, a study called Europesat was conducted to determine how to address the DTH requirements of several countries with a regional BSS system. One approach considered was to combine the assignments of several countries and implement a constellation of high-power BSS satellites at a single orbit slot. Originally, the administrations of many nations thought that it was good to have their own satellite positions. Some nations even thought the more satellites the better. Actually, the fewer, the better, because you can aggregate programming and create a hot bird. Eutelsat did not proceed directly with Europesat because of the perceived difficulty of combining all of the operating channels at a single orbit position.

In 1994 Eutelsat introduced the HOT BIRD series of satellites at 13° EL. These provide focus for cable TV and DTH services in the European region, primarily the West where incomes can support premium services. HOT BIRD 1, built by Matra Marconi Space, began service in March 1995 collocated with Eutelsat II at 13 EL, and has since been joined by HOT BIRDs 2 through 6, launched from 1996 through 2002 to fully populate this particular orbit position with 98 active channels covering both polarizations of the 10.70- to 12.75-GHz downlink band.

Eutelsat became a pioneer in the field of digital onboard processing with the inauguration of Skyplex on later HOT BIRD satellites. Initially tested on Eutelsat's HOT BIRD 4, and made fully operational on HOT BIRD 5, the Skyplex unit can receive low bit rate uplink signals in the range 350 Kbps to 6 Mbps. It then demodulates them for multiplexing into a single digital stream at 38 Mbps, which is then remodulated onboard the satellite for broadcast. A simplified block diagram of the Skyplex Multiplexer is provided in Figure 3.13.

The output of the Skyplex modulator, a DVB-S QPSK signal at 27.5 Msymb/sec, feeds a 33-MHz satellite transponder at full power. The resulting downlink signal is similar, although not identical, in structure to all the other digital transmissions emitted from HOT BIRD transponders that were originally multiplexed on the ground. It can be received by standard equipment, that is, by compatible MPEG 2/DVB decoders. Small uplink stations can choose SCPC or shared modes. In the shared mode, up to six stations can simultaneously access an uplink channel in TDMA mode. As a result, uplinks can be set up over a wide range of data rates, from 6 Mbps (single station accessing a 6-Mbps Skyplex channel in SCPC) down to 350 Kbps (six stations accessing a 2-Mbps Skyplex channel in TDMA mode). The satellite bandwidth leased by the broadcaster can therefore be more closely matched to the precise needs of the application.

Since Skyplex enables DVB transmission direct-to-home, the broadcaster can decide whether to encrypt the uplink streams—with a choice of conditional access

systems—or to simply leave them in the clear, as with a standard ground-multiplexed DVB transmission.

Due to the consolidation of global satellite operators, Eutelsat has aggressively added orbit positions and satellites since the late 1990s. This involves direct investment in satellites such as ATLANTIC BIRD 1 and 2, and purchases of shares of satellite operators such as Hispasat and Express.

6.8.5   Thor

The Thor series of DTH satellites operate in the BSS frequencies and provide service to several Nordic countries, including Norway, Sweden, Denmark, and Finland. The first satellite, Thor 1, was purchased from BSB in orbit as it was already functioning as a part of the defunct U.K. service (replaced by Sky aboard the Astra satellites). Telenor Satellite Services AS of Oslo, Norway, ordered an HS 376 satellite from Hughes Space and Communications International, Inc., to provide DTH programming to Scandinavia and northern Europe.

Telenor is also party to the Nordic Satellite Distribution (NSD) joint venture company, which employs other satellites such as Intelsat 702. The Thor series have adequate power to transmit to 50-cm dishes. The Nordic market is moderate in size but hungry for programming, witnessed by the significant number of satellites and transponders serving the subregion.

6.9   Expansion of DTH in Asia

The first introduction of TV transmission to a small receiving antenna was conducted in 1977 in Indonesia using the Palapa A1 satellite. A cooperative effort of Perumtel (now PT Telkom) and Hughes Space and Communications, it demonstrated the viability of a form of direct video transmission. While the antenna was 4.5m in diameter, which is large by current standards, it nevertheless proved that low-cost installations could provide a quality service. The next major step was Aussat A, first launched in 1985, with its medium-powered Ku-band FSS satellites, which achieved the same result with an antenna of around 1m in diameter. However, it was not until the launch of AsiaSat 1 on April 7, 1990, that a true commercial DTH service appeared in the region.

The Asia-Pacific region experienced rapid expansion in commercial satellite terms, fueled by the anticipated demand for video transmission capacity. A summary of the satellite operators and the size of their respective constellations are provided in Table 6.8. The largest countries, China and India, sparked a lot of interest and several satellites went into operation to provide respective capacity. It is interesting to note the early experiment with S-band television transmission to India (the SITE project) from NASA's ATS-6 satellite, which was relocated to the Indian Ocean region in 1977. In spite of this early success with the technology, the Indian market has remained a difficult challenge for would-be DTH operators. Likewise, China has not produced the hoped-for level of business. In other countries like Thailand and Malaysia, local companies arose to satisfy domestic demand and pursue

Table 6.8 Satellite Operations in the Asia-Pacific Region

Satellite/Operator	Location	Number of Satellites	Coverage	Primary Services
Palapa B/PT Telkom	Indonesia	2 (all C-band)	Indonesia, ASEAN region	Video, including DTH, other FSS
Optus and Aussat/Optus Communications	Australia	4 (all Ku-band)	Australia and New Zealand	Video, including DTH, other FSS
ChinaSat/MPT of PRC	Peoples Republic of China (Beijing)	4 (2 all C-band, 1 C/Ku)	People's Republic of China	Video, other FSS
Insat/ISRO	India	3 (C/S-bands)	India and India Ocean	Video, including DTH, other FSS
JCSat/Japan Satellite Systems	Japan	3 (all Ku-band)	Japan, East Asia	Video, including DTH, other FSS
AsiaSat/Asia Satellite Telecom Co. Ltd.	Hong Kong	2 (1 all C-band, 1 C/Ku-band)	Asia and Middle East	Video, including DTH, other FSS
Superbird/Space Communications Corp.	Japan	3 (2 Ku/Ku-bands, 1 all Ku-band)	Japan	Video, including DTH, other FSS
Palapa C/Satelindo	Indonesia	2 (C/Ku-bands)	Indonesia, ASEAN region, Asia	Video, including DTH, other FSS
APStar/APT Satellite	Hong Kong	2 (all C-bands)	China, Asia	Video, including DTH, other FSS
PSN/Pasifik Satelit Nusantara	Indonesia	1 (all C-band)	Southeast Asia	Video
BSat/NHK-TAO	Japan	2 (all Ku-band)	Japan	DTH
Taicom/Shinawatra	Thailand	2 (C/Ku-bands)	Thailand, East Asia	Video, including DTH, other FSS
Agila/Mabuhay	Philippines	1 (C/Ku-bands)	Asia and Pacific	Video

the DTH opportunity. Japan has evolved into a mature satellite TV market, and a principal operator, JSAT, now offers capacity throughout the Asia-Pacific region.

6.9.1 Indovision (Indonesia)

The Indonesia satellite TV coverage is the most extensive in the region as it delivers national video signal across a country of over 10,000 islands. As indicated in Table 6.8, there are three active satellite operators and five satellites in operation. The bulk of the capacity is at C-band, which is consistent with two important factors. First, these satellites were originally intended to provide an effective telecommunications infrastructure for the country. The availability of regional C-band service for the expanding video market was more or less a fortunate accident. Second, this is a tropical region with some of the greatest thunderstorm activity in the world (see Figure 2.5). The result is that high rainfall rates will not impair the signal as much as they would have in a Ku-band system design.

Palapa A1 and A2 were launched successfully into service in 1976 and 1977, respectively, followed by the Palapa B series in the mid-1980s. PT Telkom has replaced the B series with a high-capacity C- and Ku-band satellite named Telkom 1, constructed by Lockheed Martin. Reverting to C-band alone, a smaller satellite named Telkom 2 was procured from Orbital Sciences.

A second satellite operator called Pasifik Satelit Nusantara (PSN) appeared on the scene in 1991, gaining access to the orbit by purchasing a Palapa B satellite that was nearing end of life. The satellite operates in an inclined orbit to extend life to serve programmers in Indonesia and Taiwan. An uplink is installed on Batam, a small Indonesian island about 50 km south of Singapore. PSN has more recently begun an ambitious MSS project called ACeS, which is discussed in Chapter 11. The third satellite operator in Indonesia, PT. Satelit Telekomunikasi Indonesia (Satelindo), has three government telecommunications licenses: an international PSTN gateway in Jakarta, a GSM cellular license for Indonesia, and the owner and operator of the Palapa C series of satellites. One of these satellites is operating, providing video transmission at C- and Ku-bands.

6.9.2   ASTRO/MEASAT (Malaysia)

Binariang Satellite Systems Sdn. Bhd. (BSS) is the owner and operator of the MEASAT satellite system. The Malaysia East Asia Satellite (MEASAT) system currently comprises two HS 376 spacecraft built by Boeing Satellite Systems, Inc. Both spacecraft provide high-EIRP C-band and Ku-band capacity to address the increasing Internet, telecommunications, and broadcasting needs in the Asia-Pacific region. In 2003, BSS purchased MEASAT3 from Boeing Satellite Systems.

The system has the capacity to provide digital video and audio broadcasting, international and domestic VSAT and telecommunications services, as well as high-speed Internet access. MEASAT was a technology breakthrough with its pioneering Direct-To-User (DTU) system transmitted via its high-powered Ku-band transponders specifically designed to cut through the Asian region's heavy tropical rainfall. MEASAT-1 is used by the ASTRO DTH operator in Malaysia to provide digital transmission using 60-cm receive dishes and represents a new era in broadcasting services in the region.

MEASAT-1 was successfully launched to the orbital slot of 91.5 EL on January 13, 1996, from Kourou, French Guiana. It has 12 C-band and 5 Ku-band transponders. MEASAT-2, which was successfully placed at the orbital slot of 148 EL on November 14, 1996, has nine Ku-band transponders that cover Malaysia, Indonesia, Eastern Australia, Vietnam, Taiwan, and the Philippines. MEASAT-2 also has six C-band transponders covering the Asian region as well as Australia and Hawaii. The MEASAT Satellite Control Center (MSCC), the nerve center of the MEASAT system, is located in Pulau Langkawi, an island off the northwest coast of Peninsular Malaysia. The MSCC is manned by a team of highly trained technical specialists comprising spacecraft controllers, ground engineers, and orbit analysts.

MEASAT Broadcast Network Systems Sdn. Bhd. (MEASAT Broadcast) is an integrated electronic media enterprise offering wide-ranging multimedia broadcasting services to Malaysia and the region. The Asian Broadcast Center in Kuala Lumpur is equipped with the latest in digital broadcasting technology and positioned strategically within the Multimedia Super Corridor. MEASAT Broadcast sets the stage for Malaysia's advent into twenty-first century broadcast, information, and interactive technologies. The ASTRO DTU service is subscription based and presently offers more than 30 television channels and 16 radio channels in digital format. The DTU service will be expanded to include a range of interactive

applications, such as distance learning, home shopping, home banking, and software download capabilities. This service is delivered via the high-powered Ku-band spot beams of the MEASAT system that are over Malaysia and other countries in the South and East Asia region, including India, the Philippines, Vietnam, Indonesia, Taiwan, Singapore, and Brunei.

6.9.3   SKY PerfecTV (Japan)

Japan was an early developer of DTH service with the BS series of satellites. The governmental broadcaster, NHK, has been delivering a DTH service to the Japanese audience since 1980, and it became very popular. There were estimated to be more than 7 million home dishes in Japan by 1997, and it clearly represented the most successful service of its kind in the world. Viewers pay an annual charge to NHK more or less in the form of a tax, akin to the policy of the BBC in the United Kingdom.

NHK provided the following DTH channels using the BS series of spacecraft:

- DBS-1: a combined 24-hour satellite feed of news, sports, and general interest programming coming from Japan, the United States, and Europe;
- DBS-2: a cultural channel available about 23 hours per day, consisting mostly of Japanese material.

NHK has been very successful with DBS-1, primarily because of the attractiveness of the programming. They have obtained the rights to rebroadcast CNN, ESPN, ABC, and other popular U.S. channels to the Japanese market.

The DTH situation changed dramatically after the Japanese government allowed the entry of commercial satellite broadcasting using the FSS satellites of JSAT and SCC. The NHK project went through a transformation to a commercial joint venture under the category of BSat. On the other hand, the FSS services became known as CSat (indicating its origins in communications). A subscription DTH service is available over the BSat satellites that use Japan's BSS assignments at 110 EL. It consists of analog channels, including the old DBS 1 and 2 of NHK, and 10 digital commercial channels comprising HDTV, news, movies, and special programs. BSAT purchased an HS 376 spacecraft from Hughes, which was placed into the 110 EL orbit location in 1997. It replaced the aging BS series operated by the Telecommunications Advancement Organization of Japan (TAO), the government-supported satellite operator. Additional satellites were purchased from Boeing and Orbital Sciences.

The newest and now largest DTH operator in Japan is SKY PerfecTV, which inaugurated service on JCSAT-3 in April 1996. The original investors included Itochu, Fuji TV, JSAT, Nippon TV, Mitsui, Tokyo Broadcasting, and others. During the ensuing years, News Corp. established a joint venture called JSkyB with Softbank and later Sony. However, this company merged with PerfecTV to form SKY PerfectTV in 1998. Hughes established DIRECTV Japan and went into operation on Superbird; however, they terminated service in September 2000, leaving SKY PerfecTV as the sole CS digital broadcasting platform service in Japan. Boosted by the migration of former DIRECTV Japan subscribers to its service, SKY PerfecTV's subscriber base topped 2 million subscribers in June 2000. The company listed its shares

on the MOTHERS market of the Tokyo Stock Exchange in October 2000. At 3.3 million subscribers by September 2002, SKY PerfecTV is Japan's largest, most powerful digital satellite broadcasting service, with the country's largest offering of channels (including TV, digital radio, and data). Using JCSAT-3 and JCSAT-4A, the service contains 38 basic channels, 100 radio channels, and another 24 channels of premium programming. Special packages are offered for movies, professional sports, live events, and PPV.

SKY Perfect Communications, by use of a multiplatform business model, is working to provide broadcasting and other advanced services via the N-SAT-110 FSS satellite (launched in October 2000 and placed in at the same orbit position as BSat). This satellite, also known as JSAT-110, was manufactured by Lockheed Martin and provides a minimum of 57-dBW EIRP throughout Japan. Moreover, the company will be developing next generation set-top boxes for use in the existing service provided via the JCSAT-3 and JCSAT-4A satellites (see Figure 6.12 for an illustration of JCSAT-4A), and through these set-top boxes will provide new services linked with Internet, mobile communications equipment, and other devices.

The SKY PerfecTV platform provides a common transmission and management system, using an IC card, which allows for signal receptions from two satellites (orbiting at 124/128 EL) with one set comprising set-top box and dish. The operation of the service benefits from the following:

- A comprehensive system that ensures video and audio reception with clarity and quality;
- Stable broadcast operations via distributed, automated systems. Distributed uplink sites in Meguro, Aoyama, Tennozu, Ariake, and Osaka, plus backup satellite systems;
- High-speed unlocking/decoding and high-speed EPG;
- Upgrading service via satellite;
- New service developments, including high function interactive services, data communications services, broadband content distribution, and CS digital broadcast services via the communications satellite orbiting at 110 EL.

Figure 6.12 JCSAT-4A satellite and footprint coverage of Japan. This satellite is used to deliver the SKY PerfecTV DTH service to Japan.

6.9.4   STAR TV/AsiaSat (Hong Kong, SAR)

Satellite Television for the Asia Region (Star TV) was formed by Hutchison Whampo, a leading "Hong" conglomerate corporation in Hong Kong. The brainchild of Richard Lee, son of Hong Kong's most notable businessman, Li Ka Shing, Star TV remains the best model of a video distribution system created by business interests in Asia. AsiaSat, the satellite operator serving Star TV, is based in Hong Kong as well. It was formed by three partners: Cable and Wireless PLC, Hutchison Wampoa Ltd. (also 100% owner of Star TV at the time), and the Chinese International Trust Investment Corporation (CITIC). CITIC is the leading overseas investment corporation in the People's Republic of China. AsiaSat is a very unique combination of western, Hong Kong, and Chinese interests.

The first satellite, AsiaSat 1, was obtained from the insurance underwriters who were left with a previously launched but nonetheless nonfunctioning spacecraft. Originally launched as Westar 6, the spacecraft had been marooned in the wrong orbit by a failed perigee boost rocket after deployment from the Space Shuttle. Hughes, the manufacturer, was paid by the underwriters to retrofit the spacecraft and modify the antenna for Asia coverage. Acquired by AsiaSat and renamed AsiaSat 1, it was relaunched by the Long March 3C rocket from China on April 17, 1990. The satellite has allowed Star TV to begin service from 105.5° EL and thereby become the leading Asia-based satellite TV service. AsiaSat 2, built by Lockheed-Martin, was launched in November 1995, on the Long March 2E, and went into service in 1996 at 100.5° EL. Initially an analog family of channels, STAR went digital in 2000–2001. STAR is actually a collection of country- and language-directed services. For example, the service to Hong Kong consists of the following channels:

- Star Movies;
- Star World;
- Star Sports;
- Channel [V] music;
- National Geographic;
- Phoenix Chinese Channel;
- Phoenix Info News;
- ESPN.

The package directed to mainland China adds a Mandarin language channel called Xing Kong Wei Shi. With the broad footprint and a total of 39 channels, STAR claims to reach 80 million homes with 300 million viewers in 53 countries.

The coverage of AsiaSat 1 is split into two beams: the northern beam that focuses on China, Korea, and Japan; and the southern beam that extends through southern Asia and the Middle East. Consequently, the STAR services are repeated on the two beams. With the advent of AsiaSat 3S, which had higher power transponders, a single-beam C-band coverage was possible.

The exact number of DTH antennas currently pointed at AsiaSat is impossible to know. This is because few of the services are encrypted and controlled. Users simply purchase the equipment as cheaply as possible (usually from domestic sources) and put them up. It is estimated that there are over 1 million antennas in China

alone, which is an amazing fact when you consider that it is illegal for most individuals to have a dish. Service is widely used in India—owing to the popularity of ZEE TV. In this case, individuals subscribe to cable in their community. This is an unregulated business, where entrepreneurs invest the few dollars needed to string cables from residence to residence. Subscribers pay very little for access and nothing for the programming itself.

AsiaSat purchased a third satellite, AsiaSat 3, from Hughes, which failed to achieve orbit and was recovered for another purpose by Hughes Global Services. Its replacement, AsiaSat 3S, was successful. The replacement for AsiaSat 1, called Asia-Sat 4, was placed into service in 2003. In this case, STAR TV is no longer an exclusive occupant from a video standpoint. In 1998, Hutchison and C&W sold their shares, which changes ownership and allowed SES-Global to become the strategic investor. The company maintains its CITIC connection and is traded on the New York and Hong Kong stock exchanges.

6.10   Expansion of DTH in Latin America

Latin America has followed the United States in the development of terrestrial and satellite TV. Because U.S. cable TV satellite transmissions do not extend well below about the middle of Mexico, many groups and individuals in South America have still taken the trouble of installing 13-m dishes just to be able to receive HBO and CNN. More recently, these programmers have introduced Latin American versions of these channels and offer them through PanAmSat, INTELSAT, Loral Skynet, and SES-Global.

Homegrown operators in Mexico, Brazil, and Argentina have been acquired by the global operators. The fact that the United States is so closely tied to Latin America means that, effectively, U.S. operators are Latin American operators. The best example is PanAmSat, which was partly funded with Mexican backing, and as a public company it is a major provider of capacity to Latin American markets.

From a DTH standpoint, the new additions to Latin America are Galaxy 8i and PAS 8. Both satellites are owned and operated by PanAmSat; however, the DTH service is provided by DIRECTV Latin America and SKY Latin America, respectively. DLA was originally developed by a joint venture of Hughes, Organisacion Diego Cisneros (ODC) of Venezuela, TV Abril of Brazil, and Multivision of Mexico. The company has since been totally taken over by Hughes Electronics and is headquartered in Ft. Lauderdale, Florida. The services are uplinked from a broadcast center in Long Beach, California. Additional broadcast centers will be operated from Mexico, Brazil, Argentina, and Venezuela. The frequencies are within the FSS portion of the band, but the system employs circular polarization. Sufficient EIRP is provided to allow the use of antennas typically 60 cm in diameter.

Sky Multi Country operates 24-hour digital DTH television service in Argentina, Chile, and Colombia. With more than 150,000 subscribers, Sky Multi Country is a leading DTH satellite television broadcaster in Latin America. Sky is offered by the strategic alliance of Organizações Globo (Brazil's leading entertainment group), Mexico's Grupo Televisa, S.A. (the largest production and media company in the Spanish-speaking world), The News Corporation, Ltd. (one of the world's largest

media companies), and Liberty Media International Inc. (one of the world's largest communications and multimedia companies).

References

[1]   Elbert, B. R., *The Satellite Communication Ground Segment and Earth Station Handbook*, Norwood, MA: Artech House, 2001.

[2]   Whitaker, J., (ed.), "Video Server Storage Systems," in *National Association of Broadcasters Engineering Handbook*, 9th ed., Washington, D.C.: National Association of Broadcasters, 1999.

[3]   International Telecommunication Union, *Radio Regulations – Volume 2*, Appendix APS30, Annex 5, Geneva: ITU, 2001.

[4]   International Telecommunication Union, *Radio Regulations – Volume 1*, Geneva: ITU, 2001.

[5]   Cwik, T., and V. Jamnejad, *Beam Squint Due to Circular Polarization in a Beam-Waveguide Antenna*, TDA Progress Report 42-128, February 15, 1997, http://tmo.jpl.nasa.gov/tmo/progress_report/42-128/128M.pdf.

[6]   Elbert, B. R., "Digital Direct to Home TV Broadcasting Via Satellite," *Cable and Satellite China*, Beijing, China: International Institute for Research, 1995.

Satellite Digital Audio Radio Service

Since the printing of the first edition of this book, another direct broadcasting application has reached the market: satellite-delivered digital audio radio service. What the systems that provide these services have in common is that they deliver a multiplexed combination of several audio program channels transmitted directly to automobile receivers, portable radios, and homes using special frequency allocations in the region of L- and S-bands. S-DARS overcomes the range limitation of terrestrial FM radio broadcasting and provides quality of sound comparable to other digital formats such as MP-3 and possibly CD. Another advantage is that the digital multiplex approach permits narrowband services like talk radio and sports broadcasting to be combined with wider bandwidth high-fidelity music. Another term used for the service is digital audio broadcasting, although this is generally reserved for digital audio over terrestrial broadcasts—under development in Europe.

Satellites have been used to deliver audio programming for decades, but these systems were directed toward fixed installations at radio stations and commercial buildings. Additionally, DTH systems typically include a package of music channels that can be played through the TV set. What sets the new S-DARS apart is that it provides coverage to automobiles and portable radios, and offers some unique program formats not popular enough to sustain themselves as commercial operations. Thus, we have a new generation of direct broadcasting services to provide universal coverage of radio-like services to the general public. Begun in Africa as a free, advertiser-supported service by WorldSpace, S-DARS has been propelled into a potentially major satellite business for subscribers who are willing to pay a monthly fee (e.g., pay radio) akin to what is already standard for cable and DTH TV (e.g., pay TV).

As mentioned, WorldSpace pioneered the concept of broadcasting a radio service from space and launched Afristar, their first operating satellite, in October 1998. As the pioneer, they demonstrated technical feasibility and the ability to reach a dispersed ground audience. The technology to produce a subscription S-DARS has been available for about a decade, but the pieces first came together with the launch of XM Satellite Radio in the United States in March 2001. Sirius Satellite Radio, which operates in the same spectrum with XM and is their sole competitor in the United States, began service shortly thereafter. With its head start, XM has reportedly been able to meet its subscriber growth expectations, and in 2002 their receivers became available preinstalled in automobiles from General Motors, one of their key investors. The technically more interesting system, Sirius, had a delayed start due to late delivery of receiver chip sets, but their system nevertheless works as it was designed to work. This author compared both radios and services side by side in

a local Best Buy store and found them nearly identical in terms of the variety of services and the audio quality.

Development of S-DARS in Europe and Asia has begun but along a later timescale as compared to that in the United States. One reason for this is that the European community has chosen a path based on terrestrial DAB. A system called Eureka is available in many countries and could eventually pose a threat to S-DARS if and when it is introduced in the United States. However, this is not a forgone conclusion and at least one S-DARS startup is developing on the Continent.

7.1 Satellite Radio Broadcast Concept

Satellite radio broadcasting is not so different from TV program distribution and in fact shares many of the same principles and components. First use of dedicated satellite audio was by Muzak, a company that delivered elevator music in the 1970s and moved from tape to satellite. A strictly satellite form of radio broadcasting was introduced by Jones Intercable in Colorado, consisting of a compilation of advertisement-free music channels in multiple formats (easy listening, jazz, classical, rock, and country). Subsequently, all DTH TV operators included digital versions of the Jones service as part of their programming content through the facilities of Music Choice. Another form was the private radio broadcast to chains of retail stores, pioneered by Supermarket Radio Network. But it was not until Noah Samara created WorldSpace that S-DARS really got its start. The concept is indicated in Figure 7.1, where a broadcast center obtains audio content from a variety of sources: tape, local studio, audio CD, and existing radio stations and networks.

The trick here is to be sure that the programming flows in the same manner as listeners are accustomed to hearing. The actual broadcast transmission is fairly standard, as in TV, using analog-to-digital conversion, compression appropriate to the

Figure 7.1   Basic architecture of a satellite digital audio radio service.

content, forward error correction, modulation, and RF amplification. The satellite to be used may be a bent-pipe design with sufficient EIRP to allow the use of portable or vehicular receivers. The service operates like a mobile application system, discussed in Chapter 11, which means that it must consider the mobile fading environment—multipath, shadowing by terrain and buildings, and absorption by foliage and nonmetallic walls. In the absence of countermeasures, any break in data delivery will cause total silence of the receiver. This is similar to the effect on the sound portion of a digital DTH service during a rain outage. In contrast, AM radio signals transmitted around 1 MHz follow the Earth using ground wave propagation and signals pass through walls with relative ease. FM carriers around 100-MHz experience more disruption as it is now mostly line of sight, but it has better consistency of sound delivery than S-DARS (above 1-GHz carrier frequency) in the presence of fading mechanisms. Partially, the reason for this is the high carrier power of FM broadcasting.

Countering mobile fading requires a multiprong strategy that makes the same content available to the receiver via diverse paths. The basic link margin should exceed 6 dB to deal with signal cancellation caused by reflection off the road, buildings, or hills. As a further step, simultaneous broadcasts are provided by at least two satellites spaced sufficiently apart to provide different angles of arrival to the user antenna. In urban and crowded suburban environments, blockage of both paths can be anticipated, mandating the use of terrestrial repeaters that attempt to fill in the holes. These approaches, as well as time and code diversity, are reviewed in the case studies to follow.

7.1.1   S-DARS Spectrum Allocations

S-DARS uses the lower frequency bands, L and S, to provide a signal more resilient to mobile fading. Allocations were made by the ITU in Resolution 528 (WARC-92) "Introduction of the broadcasting-satellite service (sound) systems and complementary terrestrial broadcasting in the bands allocated to these services within the range 1–3 GHz" [1]. These international allocations have very favorable propagation characteristics, as there is effectively no rain attenuation. However, the low frequency brings with it the issue of reduced bandwidth, amounting to approximately 25 MHz total per band. The latter is less of a problem for audio services, which individually requires only a small fraction of the bandwidth of full motion video (e.g., ~100 Kbps versus ~5 Mbps, ignoring the benefit of statistical multiplexing).

The specific allocation at L-band is the range of 1,452 to 1,492 MHz on a global basis—with the exception of the United States where S-band is mandated. The following footnote is indicated in the Table of Frequency Allocations: "Use of the band 1,452–1,492 MHz by the broadcasting-satellite service, and by the broadcasting service, is limited to digital audio broadcasting and is subject to the provisions of Resolution 528 (WARC-92)." Resolution 528 allows countries to use this spectrum subject to a number of conditions, one being that a future WRC should be held to provide a more detailed plan. In the meantime, WorldSpace has made good use of this band for its three-satellite global coverage S-DARS service.

The S-band allocations cover 2,310 to 2,360 GHz. In general, S-band reception on the ground must consider the use of the adjacent spectrum for industrial,

scientific, and medical (ISM) applications, which are typically unlicensed. One use of ISM that is growing in application is for wireless local area networks (W-LANs) built on the 802.11b standard. Likewise, the Bluetooth personal LAN standard butts up against S-DARS. A recent study by the FCC identifies a potential for significant interference from W-LANs into S-DARS as newer, lower cost devices are inserted into mobile phones, computers, and a multitude of other appliances and systems. Microwave ovens are another popular use. Any of these can produce harmful interference to reception, within the following specific range: 2,400 to 2,500 MHz (center frequency 2,450 MHz).

The band that is employed in the United States by XM and Sirius comes by virtue of the following footnote in the Table of Frequency Allocations: "5.393. Additional allocation: in the United States, India and Mexico, the band 2,310–2,360 MHz is also allocated to the broadcasting-satellite service (sound) and complementary terrestrial sound broadcasting service on a primary basis. Such use is limited to digital audio broadcasting and is subject to the provisions of Resolution 528 (WARC-92), with the exception of resolves 3 in regard to the limitation on broadcasting-satellite systems in the upper 25 MHz. (WRC-2000)." This nice provision seems to go counter to the old adage among WRC participants, "What the Table of Frequency Allocations provides, the footnotes take away."

The following footnote makes additional S-band spectrum available for S-DARS: "5.418 Additional allocation: in Bangladesh, Belarus, Korea (Rep. of), India, Japan, Pakistan, Singapore, Sri Lanka and Thailand, the band 2,535–2,655 MHz is also allocated to the broadcasting-satellite service (sound) and complementary terrestrial broadcasting service on a primary basis. Such use is limited to digital audio broadcasting and is subject to the provisions of Resolution 528 (WARC-92). The provisions of No. 5.416 and Table 21-4 of Article 21, do not apply to this additional allocation. Use of non-geostationary-satellite systems in the broadcasting satellite service (sound) is subject to Resolution 539 (WRC-2000)." This footnote was updated at WRC-2000.

Antennas for S-DARS receivers likely have very broad beamwidths, which cannot discriminate among satellites transmitting on the same frequency even if they are separated by tens of degrees. Thus, the opportunity for new entrants in S-DARS is limited. Furthermore, the Table of Frequency Allocations will allow non-GEO S-DARS within the same spectrum. This is always a complicated matter because sharing in reality involves splitting up the spectrum between the systems (referred to as *band segmentation*). Currently, the only non-GEO system is Sirius, which, as discussed later, uses a highly inclined elliptical geosynchronous (24-hour) orbit. The FCC had auctioned off the S-DARS spectrum in two segments, which were won by Sirius and XM.

7.1.2 Propagation for Mobile Broadcasting

Mobile broadcasting has a lot in common with two-way interactive mobile satellite communications. The following discussion reviews propagation aspects for mobile broadcasting (please refer to Chapter 11 for the general treatment on mobile satellite propagation issues). There are two principal differences between mobile broadcasting and two-way voice or data services:

- The signal power time series seen below is a time expanded version of a portion of plot shown in part (b). The van was driving from a clear area into an area with foliage shadowing.

(a)

- Signal blockage causes deep signal fades.
- Satellite power margin is insufficient to overcome deep fades.
- Signal diversity is a reasonable solution.

1-minute record of TDRS 2-GHz signal with roadside tree blockage, Rose Bowl area residential 27 mph

(b)

Figure 7.2   Measured S-band carrier strength from the TDRS satellite: (a) signal fading due to tree shadowing; and (b) severe foliage shadowing.

1. Broadcasting must rely on one direction of information transfer (i.e., there is no possibility of requesting retransmission). As a result, when the signal finally fades out, the receiver has nothing to play except silence or what might have been prestored.
2. The programming must be delivered on a continuous basis, regardless of the location of the user. This is why both XM and Sirius include terrestrial relay transmitters to fill in dead spots in dense urban environments.

We are talking about a digital transmission system that consistently delivers bits to the receiver. The receiver, in turn, must be able to reassemble the entire stream in order to decompress and convert from digital to analog. Loss of bits causes dropouts at best and a complete loss of signal at worse. There is no graceful degradation aspect that we are accustomed to as FM listeners. The FM receiver only completely breaks lock if there is interference from a stronger signal on the same or adjacent channel frequency. In S-DARS there is only one signal, but the fades can be deep and either rapid (due to multipath) or prolonged (due to blockage). Blockage in conventional FM is limited to when one drives through a long tunnel or enters a parking structure. There are other reasons why FM is more resilient: signal strength tends to be quite high, and the frequency range centered on 100 MHz is more able to bend and reflect in and around obstacles. Absorption by foliage is also much less than at L- and S-bands.

An example of what S-DARS propagation would look like is graphed in Figure 7.2, which is the actual measured digital audio signal transmitted by the TDRS satellite during an S-band test conducted by the NASA Jet Propulsion Laboratory [2]. Several aspects are evident in this measurement taken during 1 minute of driving past trees:

- Fluctuations between 0 and 10 dB are rapid, mostly caused by signal cancellation from a ground-reflected path (e.g., Ricean fading).
- Some fades are deep when the signal is blocked by tree trunks. These cannot be countered by an increase in satellite EIRP.
- Any chance of providing a continuous service under these conditions is small, thus the need for diverse paths containing the same broadcast.

In the case of a fixed or portable type of receiver (e.g., one that does not move after locating the signal), these factors do not matter. In fact, with L- or S-band transmission, the signal into a fixed receiver with clear line of sight to the satellite will generally remain solid. In tropical regions around the geomagnetic equator, ionospheric scintillation can be a problem. Fades from this phenomenon can be 6 dB or more in seasons near the equinox. Fortunately, the S-DARS systems generally have margins of 8 dB or more, which should provide a commercially acceptable service. The Faraday effect is avoided through the use of CP, which thereby demands an acceptable axial ratio.

7.2 First Introduction—WorldSpace

WorldSpace was founded in 1990 and represents an interesting startup venture in the commercial satellite industry. Initial financial support made it possible for WorldSpace to build and launch satellites under contracts placed with Alcatel and other major companies. The WorldSpace system was the first S-DARS system and therefore was the innovator in applying L-band spectrum to audio broadcasting [3]. Of critical importance are the size of the coverage areas in relation to the cost of the satellites, advanced low bit rate audio coding, and simple satellite uplinking arrangements. However, WorldSpace is less suitable for mobile reception than XM or Sirius

because of low elevation angles in some areas served. Without any form of diversity, signal fades and dropouts make reception extremely problematic in moving vehicles. Instead, the service lends itself to reception for fixed and portable operation. As stated by Andrew Hope of Satellite Solutions–Australia, having a WorldSpace receiver on site in Central Africa is the difference between abject boredom and sanity.

The key provider of technology for the delivery system was Alcatel Space, of Toulouse, France [4]. While several technologies for spacecraft and receivers have made this possible, the key here is the first commercial satellite use of the FDMA/TDM principle for broadcasting, which in turn permits operation of the spacecraft transponders close to their saturation points [3, 4]. Alcatel was the primary contractor responsible for the complete end-to-end system and was directly responsible for the design and construction of the satellite communications payload, satellite ground control system, WorldSpace broadcast services control system, business control system, and intercommunications network. Audio coding developed by the Fraunhofer Institute (FHG) of Germany for the project is based on the MPEG Layer 3 algorithm with customization to suit the WorldSpace project. The coding rate for each service is available in simple multiples of a basic 16-Kbps channel, up to a maximum of 128 Kbps. The system offers four audio subjective quality standards and associated capacities shown in Table 7.1. Each carrier has an information rate of 1,532 Kbps and employs concatenated FEC using (255,233) Reed-Solomon block coding and rate 1/2 convolutional coding. The modulation system employed is QPSK, which was selected in favor of the ITU standard for DAB terrestrial broadcasting, coded orthogonal frequency division multiplexing (COFDM).

7.2.1   Transmission and Network Design for WorldSpace

The WorldSpace system is built on the premise of serving the needy populations of the world, as divided into three regions: Africa and the Middle East (AfriStar), Asia and the Pacific Rim (AsiaStar), and Central and South America (CaribStar). The system configuration in each region, indicated in Figure 7.3, comprises the following:

- *The space segment:* the satellite and its associated TT&C facilities needed to control the satellite;
- *The broadcast segment:* the studios and feeder link systems to uplink the programming to the space segment;
- *The radio segment:* the individual receivers used by the public;

Table 7.1   Audio Bit Rates, Quality, and Satellite Channel Capacity for WorldSpace Regional

Service Type	Quality	Bit Rate, (Kbps)	Channel Capacity of 1,536-Kbps Carrier
Audio	Near AM radio	16	96
Improved audio	Better than AM radio	32	48
Stereo audio	FM-like	64	24 stereo
Stereo improved audio	CD-like quality	128	12 stereo

Figure 7.3   Regional coverage offered by the three WorldSpace satellites.

- *The mission segment:* to control and monitor the broadcast segment and the satellite payload. According to Alcatel, the developers of WorldSpace, it is composed of Communication System Monitoring which receives all signals transmitted by the satellite, controls the quality of the link in terms of bit error rate, and controls the mapping of the programs with what is expected. This is validated in the Mission Control Center, which has remote monitoring and control on the broadcast segment and on the satellite payload.

To date, AfriStar and AsiaStar went into operation and are serving the related regions. The focus on lower latitudes allows each satellite to provide service at elevations angles generally above 50°. As discussed later in this chapter, this has the advantage of limiting the shadowing and multipath effects. Unlike Sirius and XM, WorldSpace networks rely 100% on satellite broadcast and have no terrestrial fill-in repeaters.

7.2.2   WorldSpace GEO Satellite Design

To meet the system requirements, a specific payload design has been performed by Alcatel onto a standard Eurostar 2000+ platform from Matra Marconi Space (MMS) to form the WorldSpace satellites. These satellites are three-axis-stabilized satellites, to be operated on the geostationary orbit with a 15-year maneuver lifetime. The two solar arrays generate more than 6 kW of power. Arianespace has been selected for the launch of the three satellites. The launch mass of the satellite is 2.75 tons. The satellite is designed for full eclipse operation, and is capable of full 24-hour-per-day operation. The three satellites have the same design, offering flexibility for in-orbit delivery with respect to the risk management and business development access. An additional space satellite would be able to replace any of the three satellites. The difference between the three satellites is the downlink coverage; the satellite commonality is achieved using the capability to modify the coverage by satellite biasing, beam pointing, and antenna feed switching.

The basic block diagram of the communications payload in Figure 7.4 indicates both a bent-pipe mode and a processor mode. Regardless of which mode is used, the downlink carrier is amplified to 300W in parallel 150-W TWTAs. The resulting TDM signals are assigned frequencies within the 1,467- to 1,492-MHz band, allocated for the BSS. The transmitting antennas offer a total of three spot beams to increase the EIRP into the service areas. Uplink operation is at X-band through a global beam pattern, allowing the originating Earth station to exist anywhere on the visible Earth. The onboard digital processor includes ASICs developed by Alcatel based on technology from nonspace programs.

7.2.3   WorldSpace Receivers

Receivers that are compatible with the WorldSpace transmission format are manufactured by leading consumer electronics companies and distributed worldwide. They employ the basic block diagram shown in Figure 7.5. Also shown is a photograph of a receiver and antenna produced by Sony. The unit receives the L-band signal, demodulates the full TDM, and extracts the useful prime rate channels from the TDM stream, which is FEC decoded into a broadcast channel.

7.3   Sirius Satellite Radio

Sirius Satellite Radio is a commercial radio broadcasting company, publicly traded and headquartered in the heart of New York City. Both Sirius and XM provide a programming package within a total of about 5 Mbps comprising 100 total audio channels, half of which are music formats and half of which are talk radio. Using advanced digital recording systems, the music may be assembled off-line for later playback and without advertising. The talk formats include standard services like Fox News Channel and CNN along with a variety of shows to appeal across a spectrum of interests. Talk channels that are taken from existing program sources may include advertising. Sirius has assembled at its headquarters in New York a large suite of studios and editing facilities to allow them to originate a substantial number of the audio channels.

The satellite and network control functions are also provided in New York by engineering staff using automation onboard the spacecraft as well as within computers at the headquarters. Tracking, telemetry, and command stations are located near the equator in Quito, Ecuador, and Utive, Panama; these locations see the entire orbit and each have two antennas to allow for one spare across the system to maintain coverage.

7.3.1   The Use of the Inclined Elliptical Orbit

In the beginning, Sirius adopted the standard geostationary satellite model using half of the S-band spectrum allocated in the United States for S-DARS. This approach, nearly identical to that of XM, would only have required one satellite to enter service but would need to rely on a large quantity of ground repeaters to fill gaps in coverage. During the development of Sirius, the technical organization led

Figure 7.4 General block diagram of the WorldSpace communications payload, providing both bent-pipe and digital processor channels.

Figure 7.5 Radio receiver block diagram and example of WorldSpace receiver from Sony.

by Robert Briskman considered all of the aspects of providing a commercially acceptable service via satellite. Through a detailed review of the mobile fading environment (discussed previously), they concluded that the approach they were taking could not be counted upon to address the demands of U.S. consumers. This led them to consider and ultimately select a system more akin to the old Molniya satellites of the former Soviet Union. To better serve the northern latitudes of Russia, Molniya was maintained in a highly inclined orbit with an apogee higher that GEO (e.g., greater than 36,000 km).

The final configuration (Figure 7.6), which is patented by Sirius, consists of three 24-hour orbits that are staggered around the earth at 120° increments [5]. The

Figure 7.6 Sirius orbital configuration.

satellites are likewise staggered in their revolution, so that they individually reach the apogee peak 8 hours after each other. The satellites follow the same ground track (centered at 96º WL), shown in Figure 7.7, differing only by residual orbital errors. This approach assures that at least one satellite will be above 60° elevation angle from any point within the 48 contiguous states. Figure 7.8 demonstrates how the three satellites work together to assure the 60° elevation angle criterion at northern locations in the contiguous United States (CONUS); a GEO satellite would lie at about 30° for a similarly located user. In addition, a second satellite can be counted on to add a diverse path to the user at an elevation angle of greater than about 25°. This along with a frequency offset between the transmissions and a 4-second time delay for one satellite versus the other provide several measures to assure delivery of a continuous bit stream to the decoder on a moving vehicle. The techniques cited provide spatial, frequency, and time diversity. As illustrated in Figure 7.9, near-100% continuity of reception is assured to a moving vehicle in suburban environments (e.g., trees and low buildings). Service in cities is generally good because of the high elevation angle, providing clearance over many obstacles. Since this cannot be assured, Sirius rebroadcasts through a hundred terrestrial transmitters that get their signals over a VSAT network originating in New York. The spectrum for the terrestrial signal is contained in a guard band between the carriers transmitted by the two operating Sirius satellites.

Another important factor in propagation is attenuation caused by foliage. As presented in Figure 7.10, this loss factor can be predicted based on the path elevation angle and probability of outage. The graph indicates that for a 30° elevation angle, a worst case for a single GEO satellite, there would need to be more than 20 dB of fade

Figure 7.7   The Sirius orbital ground track allows the three satellites to be visible above North America—at least two satellites are visible at any one time. TT&C stations are located near the equator to provide continuous view of the three satellites.

Figure 7.8 Satellite elevation angles at Seattle, WA (47 N), for the Sirius three-satellite constellation.

margin to maintain an availability of 99% (e.g., 100% minus 1%). In comparison, the orbit used by Sirius can yield a worst case elevation angle (assuming three working satellites) of 60°, resulting in a loss factor of about 8 dB at the same 99% availability. This data is for L-band; at S-band, the corresponding attenuation increases to about 14 dB.

7.3.2 Satellite Design for Sirius

The spacecraft for the Sirius program, shown in Figure 7.11, was developed and manufactured by Space Systems/Loral. As mentioned, the project began under the assumption that two GEO satellites were to be located at 80° and 110°. The switch

Figure 7.9 Path blockage probability in suburban environment for a highly inclined elliptical orbit. N = number of satellites.

Figure 7.10   Foliage attenuation and probability of outage for a single GEO satellite at various elevation angles from the ground.

to inclined elliptical orbit did impact the design of the spacecraft in the following areas:

- Sun angles more oblique, causing reduced solar flux at times for panels with conventional single-axis rotation;
- Increased radiation resulting from spacecraft dropping below GEO into closer proximity to the Van Allen belt;
- Nonuniform eclipse seasons;
- Different Sun angles causing a new environment for thermal control;
- Antenna beam steering needed to maintain the footprint on the coverage area;
- Somewhat greater impact of gravitational forces of Sun and Moon due to nonequatorial orbit;
- Variation in slant range as the satellite appears, passes through apogee and then disappears;
- Variation in orbital rates and apparent Earth size (coupled with antenna beam steering);
- Launch and orbit raising differences to deploy three satellites into highly inclined orbit.

Based on the SS/L 1300 bus, the Sirius spacecraft is capable of transmitting at 60.3 dBW across CONUS through the elliptical beam shown in Figure 7.12. This

Figure 7.11   Sirius satellite fully deployed.

footprint is produced by a 2.4-m dual-axis steered main reflector and a rotating shaped subreflector. Using offset Gregorian optics, the antenna permits rotation of the spacecraft about its Earth-pointing (yaw) axis so as to improve alignment of the solar panels with the sun. The latter is necessitated by the high degree of orbit inclination, which could cause the sun's rays to be nearly parallel with the panels at times during the year. By rotating the spacecraft about the yaw axis and using the standard solar panel single-axis rotation motor, the panels can provide nearly 100% power output at all times. This is significantly different from a GEO satellite, where the incident sun angle on the solar panels is never more than 23.5° from

Transmit pattern @
ascending node

Figure 7.12   Sirius transmit antenna coverage.

perpendicular, which produces a tolerable 8% power loss (compensated with extra cells). Based upon a broadcast satellite design, the Sirius spacecraft provide 8,500W of prime power at 15 years end of life and a pointing accuracy of 0.38°. The uplink to the satellite employs X-band (7.1 GHz) into a circular beam with minimum G/T of 0 dB/K that is dual-axis steered. Both the uplink and downlink are circularly polarized.

The satellite provides effectively one channel of RF transmission (e.g., a single carrier) using an extremely high power bent-pipe transponder with an output power of nearly 4,000W. As shown in Figure 7.13, a three-for-one redundancy scheme is used for the wideband receiver. From there, the low-level carrier is split four ways to drive four quadrants so that the high output level can be built up using individual TWTAs with 120-W RF output each. Two amplifiers are phase combined to create a paired unit called the dual TWTA (DTWTA). Each quadrant contains six DTWTAs, four of which are operating at a given time (the remaining two are backup). The output of the four operating DTWTAs is phase combined to a level of 950W and applied to one of four ports on the antenna feed. With an appropriate phase shift between adjacent quadrants of 90°, the feed will radiate circular polarization toward the reflector system with full power of approximately 3,800W. This approach reduces RF losses as well as the maximum power that would be applied to a single component (except the horn itself). To make this work correctly, the design must manage and control phase errors between amplifiers and quadrants; the resulting loss due to phase imbalance is less than 1 dB. The actual output section of the repeater, indicating the 32 TWTAs, is shown in Figure 7.14.

One would expect that satellites of this type in inclined elliptical orbit would require a great deal attention from ground operations personnel and supporting systems. However, Sirius and SS/L have made the spacecraft as easy to operate and maintain as their GEO counterparts. The orbits selected have exactly the same ground track, but are separated by 8 hours. In addition, station-keeping maneuvers and power management activities during eclipse are staggered throughout the year. This allows the ground system to work in sequence, moving from one satellite to the next in a methodical manner. The onboard computers have special software to make the pointing and alignment of solar panels and antennas autonomous as well. Through appropriate ground automation, the entire constellation can be controlled by a single operator. Other spacecraft and orbital analysis staff are required, but are not actively in control of the satellites in orbit.

TT&C for these satellites is performed using a doubly redundant ground network composed of the previously mentioned equatorial Earth stations in Ecuador and Panama, along with U.S.-based TT&C sites in Hawley, Pennsylvania, and Three Peaks, California. The TT&C subsystem is controlled from the satellite control center (SCC) in New York City, Hawley, and Three Peaks. It is connected to Ecuador and Panama by redundant data lines routed through diverse paths. The SCC receives telemetry from these antennas and sends commands back for transmission to the satellites; ranging operations, used for orbit determination and control, are likewise performed through these links. There are four 4.5-m tracking antennas at Vernon, New Jersey, to provide the uplink for the broadcast and for carrier monitoring purposes in the downlink. The antennas are remotely controlled from the NYC National Broadcast Center, discussed in the next section.

Figure 7.13   Sirius communications payload block diagram.

Figure 7.14   Repeater high power output section for the Sirius satellite.

7.3.3   Network Technical Design

The Sirius Satellite Radio delivery system is an integrated architecture of space segment and ground segment components, as illustrated in Figure 7.15. TT&C and satellite control, indicated at the left of the figure, were discussed in the previous section. Note that TT&C operations are conducted at C-band using 11-m antennas at the equatorial sites. The broadcast carriers—designated TDM 1 and TDM 2 at 2.3304 and 2.3221 GHz, respectively—are uplinked at X-band from Vernon, New Jersey, with a 4-second time delay of the second relative to the first. The latter, along with the physical separation of the satellites and the use of terrestrial repeaters in urban areas, provides significant protection for signal reception on moving vehicles. The X-band uplink at Vernon has four tracking antennas of 4.5m, three being required during normal operation. This allows for antenna maintenance and troubleshooting on one antenna without disruption of service.

Service to urban canyons is enhanced using terrestrial DAB remote repeaters. Operating at a frequency midway between TMD 1 and TDM 2, the DAB signal provides fill-in where satellite reception is difficult or impossible. Interestingly, this also represents the first entry of DAB into the United States. The multiplexed programming is uplinked separately from the roof of the New York studio for delivery using a GEO Ku-band satellite in a point-to-multipoint distribution network. Thus, there are 100 downlink receivers located at the DAB remote repeaters around CONUS.

The flow of music programs from source to receiver is shown in Figure 7.16. Radio channels are created by producers using standard production bays; output is taken in uncompressed digital form and delivered to a common server computer where it may be assembled with other material into the broadcast channel. This is a highly organized and automated system that employs people to schedule the content and verify signal form and quality. The on-air programs are then digitally compressed followed by concatenated Reed-Solomon and convolutional coding for

Figure 7.15 Sirius S-DARS delivery system architecture.

forward error correction. The stream is further protected from burst errors by an interleaver before using statistical multiplexing of the 100 channels. This scheme provides a 100-channel multiplex with an average information rate per channel of approximate 60 Kbps. In reality, the listening experience is subjectively better due to the performance of statistical multiplexing, which grants more bits to those channels with greater audio content and takes bits from those with silence or low content

Figure 7.16 Flow of music from source to delivery within the Sirius network.

levels (e.g., talking). Following this, the audio stream is combined with control data and program information using time division multiplexing. This digital baseband is transferred over terrestrial redundant paths to Vernon, where two replicas are formed with one being delayed by 4 seconds (e.g., TDM 1 and TDM 2). These are uplinked by the RF electronics and 4.5-m antennas mentioned previously.

The satellites relay the respective carriers (one per satellite) into the downlink CONUS footprint and thereby are available for reception by vehicles and in homes. The receiver can receive two carriers, demodulate and decode them, and can select the best of both using internal processing logic. A basic link budget for Sirius is provided in Table 7.2. The X-band uplink is sufficiently robust at $E_b/N_0 = 27.2$, with rain margin included; at this level, it does not introduce significant noise into the service. The downlink at S-band must survive the mobile environment as well as propagation through the ionosphere, for which a single-satellite margin of approximately 6 dB is provided. This is further enhanced with spatial and time diversity to achieve another 12 dB of effective link margin.

7.3.4   Receiver Equipment and User Experience

The user's Sirius receiver is either an adaptor that plugs into an existing FM radio or a new generation of three-band AM/FM/Sirius radio. It plays the music and displays

Table 7.2   Basic Link Budget for Sirius Satellite Radio

Studio-to-Satellite	7,060.3	MHz
Uplink power (after losses)	23.0	dBW
Antenna gain (4.5m at 62% aperture efficiency)	48.4	dBi
EIRP	71.4	dBW
Path loss and pointing loss	203.6	dB
Rain loss (99.99%)	2.5	dB
Satellite antenna edge gain (including pointing loss of satellite)	30.0	dBi
Received power at satellite	−104.7	dBW
Satellite noise power ($G/T = -0.5$ dB; 4.2-MHz noise bandwidth; QPSK)	−131.9	dBW
E_b/N_0	27.2	dB
Satellite-to-Mobile	2,330.3	MHz
Satellite-corrected beam-edge EIRP	61.0	dBW
Path loss	191.8	dB
Mobile receiver antenna gain (includes pointing and ohmic loss)	−140.4	dBW
E_b/N_0	12.6	dB
E_b/N_0 for 10^{-5} BER	5.6	dB
Calculated margin:	7.0	dB
E_b/N_0 loss (impairment/ uplink/propagation)	1.0	dB
Single satellite margin	6.0	dB
Diversity path advantage (estimate)	12.0	dB

content data, such as channel name and number, song selection, and artist. As is the case with all S-DARS systems, there is no retuning of frequency as the driver leaves one town for the open road. The National Broadcast Studio located on Avenue of the Americas in midtown New York City is the point of origination for the broadcast carrier containing 100 channels of audio programming (discussed in the next section).

The block diagram of the typical receiver is shown in Figure 7.17, indicating an integrated AM/FM/Sirius configuration. Based on a chip set designed and manufactured by Lucent Technologies, the elements provide a logical flow of signals and data from antenna to speakers. The AM/FM portion employs standard components consisting of a receiver/tuner and demodulator that feed a common stereo amplifier and speakers. For S-band reception, a hemispherical coverage vehicular antenna feeds an LNA and downconverter. This provides the carrier for recovery of the data stream containing up to 100 audio channels. The arrangement receives three carriers simultaneously, from each of two Sirius satellites in view and the terrestrial DAB transmitter if available, and using a comparator, selects the best data for processing. Downstream from the comparator, the demultiplexer selects the desired audio channel for decompression and conversion from digital to analog. The receiver also supports a display to indicate the program and provides front panel controls.

Sirius programming is arguably much richer than that available in North America over terrestrial radio. Broken down into music and talk, there are enough formats to satisfy almost any taste. It is the task of the producer at Sirius to select music content, add his or her voice-over, and create the program for airing. The variety and level of specific interest can be seen in the following listing of music channels available as of the time of this writing:

Figure 7.17   Typical Sirius receiver block diagram.

- *Pop:* US-1 (Top 40 hits), The Pulse (adult contemporary), The Trend (alt pop mix), StarLite (love songs), Sirius Gold (1950s and early 1960s oldies), '60s Vibrations (1960s and early 1970s hits), I-70 (the best of the 1970s), I-80 (the best of the 1980s);
- *Rock:* The Bridge (soft rock), E-1-7 (eclectic rock), Sirius Rock Hits (rock hits), Octane (modern rock), Big Rock (mainstream rock), Classic Rock (classic rock I), The Vault (classic rock II), First Wave (classic alternative), Alt Nation (alternative I), Left Of Center (alternative II), Hard Attack (hard rock);
- *Country:* Sirius Country Hits (country hits), New Country (today's country), Big Country (country mix), Classic Country (classic country), Alt Country (alt country), Bluegrass (bluegrass);
- *R&B/Urban*: Sirius R&B Hits (R&B hits), Hot Jamz (today's R&B), Slow Jamz (soul ballads), The Express (classic soul), Soul Revue (R&B oldies), Sirius Rap Hits (rap hits), Hip Hop (today's rap), BackSpin (classic rap);
- *Dance:* 50: Sirius Dance Hits (dance hits), Planet Dance (mainstream dance), The Vortex (electronica), The Strobe (disco);
- *Jazz and Standards:* Pure Jazz (classic jazz), Jazz En Clave (Latin jazz), Planet Jazz (contemporary jazz), Jazz Cafe (smooth jazz), Standard Time (standards), Swing Street (swing), Broadway's Best (Broadway's best);
- *Latin:* Tropical (Latin hits), Romantica (Latin pop mix), Alt Ñ (rock en Español), Mexicana (Mexicana), Tejano (Tejano);
- *Classical:* Symphony Hall (symphonic), Vista (chamber works), Classical Voices (classical voices)
- *Variety:* Sirius Blues (blues), Sirius Reggae (reggae), Praise (gospel), Spirit (Christian hits), Horizons (world music), Soundscapes (new age), Sirius Kids, The Galaxy (specialty showcase).

The nonmusical program lineup is as impressive as what was just reviewed. These include formats as found on AM radio in the United States—the well-recognized talk-radio familiar to many Americans who drive to and from work, as well as news and entertainment channels normally associated with cable TV. The following is a sampling of the formats on Sirius that add value to the broad selection of music:

- *Sirius News:* CNBC, Fox News Channel, CNN Headline News, Bloomberg, NPR, World Radio Network, BBC World, C-Span Radio, The Weather Channel, Sirius Talk, and ABC News and Talk.
- *Sirius Sports:* ESPN, Sports Byline USA, The Speed Channel, OLN Adventure Radio;
- *Sirius Hispanic Talk:* BBC Mundo, La Red Hispana, Radio Diportivo, Radio Mujer, Radio Amigo;
- *Sirius Entertainment:* Radio Disney, Discovery Radio, E! Radio, A&E Satellite Radio, Radio Classics, Sci Fi, Sirius Entertainment, Sirius Comedy, Sirius Arts, Personal Achievement, Wisdom Radio, African American Talk, The Scandal Channel, Women's Talk, Guy Talk, Trucker Channel, Preview Channel.

7.4   **XM Satellite Radio**

XM Satellite Radio was first to market in the United States, providing a comparable programming package to Sirius but using the standard GEO satellite approach. The two high-power Boeing 702 S-DARS satellites, named Rock and Roll, were launched on March 18, 2001, and May 8, 2001, by Boeing Sea Launch. Positioned at 115º WL and 85º WL, respectively, Rock and Roll each transmit two carriers (total of four for the system) that contain half of the channels each. Due to elevation angle constraints of using GEO, there are a multitude of terrestrial repeaters throughout the United States. The satellites transmit using left-hand circular polarization while the terrestrial repeaters are linearly polarized (linear is superior in the terrestrial multipath environment). The frequency plan is shown in Figure 7.18. Unlike Sirius, delivery of the broadcast channels to the repeaters is accomplished using the X-band downlink from Rock and Roll. The exact same content is transmitted three times in three different signals: once on each of two satellites and a third time by repeaters.

As illustrated in Figure 7.19, the overall XM system integrates many elements to provide a commercial subscription service. The system and all of its elements are described in detail in an excellent paper by Richard Michalski, chief systems engineer of XM Satellite Radio [6]. Programming is created in a broadcast operations complex located in Washington D.C., where there are contained production studios and management facilities that operate in much the same manner as the comparable facilities of Sirius. This will be addressed later in this section. Transmission between studio and subscriber uses an uplink complex, again located in Washington, D.C., the space segment consisting of two bent-pipe GEO satellites (the aforementioned Rock and Roll), and a terrestrial repeater segment to supplement the signal delivery into urban areas. The subscriber radios themselves are provided by the technology segment, consisting of custom chip sets that are integrated into receivers that may be installed in vehicles and operated in homes. The enter complex is supported by the enterprise IT complex, with associated operation support and business support software and databases, and the network management complex. Figure 7.20 shows the control center where all XM network operations are managed.

Figure 7.18   XM frequency plan.

Figure 7.19 XM system block diagram.

Figure 7.20 XM operations control center.

7.4.1   Satellite Design for XM

The XM spacecraft provides 15 kW of prime power at end of life, which supports a repeater with a total RF output of 6.4 kW using two sets of 16 paralleled 215-W TWTAs. The bus and integration were provided by Boeing Satellite Systems (formerly Hughes Space and Communications Company, with whom the original contract was placed in 1998) [7]. Shown in Figure 7.21, this was the largest class of spacecraft at the time, necessary to deliver sufficient power to support S-DARS services to CONUS. The high-power mission was facilitated by use of dual-junction GaAs solar cells and xenon ion propulsion to conserve fuel mass. Boeing also provided launch services for the first two satellites by Sea Launch, using the Zenit-3SL rocket provided by Yuzhnoye/Yuzhmash and the Block DM-SL upper stage from Energia. While among the most powerful GEO satellites in service, Rock and Roll are nevertheless not particularly different in terms of the design and operation. A third satellite was contracted in 2002 as the original pair experienced more rapid solar array performance degradation than predicted.

The communications payload was provided by Alcatel Space, which was the same source for WorldSpace. As illustrated in Figure 7.22, the payload consists of two halves, each providing two banks of TWTAs which have their outputs combined in phase and fed to 5-m transmit reflectors. The flux density coverage across CONUS thus obtained is indicated in Figure 7.23. The reflectors deploy from a folded configuration (like bus doors) and are shaped to focus power to regions of greatest population density and to help compensate for range and local foliage conditions. This produces near-uniform availability for all locations within CONUS by adjusting the link margin for the two satellites in accordance with the predictions made by the Extended Empirical Roadside Shadowing (EERS) shadowing model, developed by the Applied Physics Laboratory of Johns Hopkins University and by W. J. Vogel of the University of Texas. The study is available through the Internet [8]. As a result, the delivery would achieve an acceptable level of availability to mobile and fixed users. This, or course, is dependent on the integration of two orbiting satellites, providing angle and frequency diversity, along with terrestrial repeaters to fill in dark areas that result from terrain blockage. The latter is more pronounced than for Sirius due to the lower range of elevation angles afforded by the two GEO positions of Rock and Roll.

Figure 7.21   XM satellites: Rock and Roll.

Figure 7.22 XM satellite communications payload block diagram.

Rock at 115

Figure 7.23   Flux density radiated by XM satellite on coverage area.

7.4.2   Transmission and Network Design for XM

The fundamental approach to creating the multiplex of audio channels is not much different from that of Sirius. In fact, the system is simpler in terms of the number of elements and their placement. This is a natural consequence and benefit of using the GEO approach, where the satellites—once placed on station by Boeing—can be accessed through fixed antennas within CONUS. The signal flow is according to the following steps:

1. Application layer processing (audio compression);
2. Service layer processing (encryption and introduction of other data for the network);
3. Payload channel transport layer (not related to the satellite payload, but where the content is packaged into uniform chunks of data);
4. Satellite multiplexer transport layer (FEC coding and various short- and long-term data interleaving takes place);
5. Physical layer (QPSK modulation on the X-band uplink).

While these layers are needed to support satellite service delivery, layers 4 and 5 are undone in the terrestrial repeaters and new coding is applied before being modulated in a different physical layer implementation. Regarding these layers, the bandwidth of the satellite transponders is 1.886 MHz and each slot contains one of the QPSK carriers with its multiplexed TDM signal with data and symbol rates of 3.28 Mbps and 1.64 Msps, respectively. Block error correction increases the transmitted rate to 2.048 Mbps. Rate 3/4 convolutional inner coding is provided to further

improve link performance. The two satellites use different coding and interleaving signals so that when combined in the mobile receiver, they have an effective FEC coding rate of 3/8. This is different from Sirius and its time diversity, where the same signal is sent with a fixed 4-second offset over two different transmission paths. The terrestrial repeaters are linearly polarized to better deal with ground multipath produced by reflections. Each repeater has a satellite dish to receive the carriers transmitted by Roll on an outer channel and retransmitted in the central band, thus avoiding self-interference. Multicarrier modulation (MCM) with rate 5/9 convolutional coding is used, along with other proprietary techniques to produce a signal that can better tolerate frequency selection fading. Fading on the satellite link caused by shadowing and absorption by foliage is not frequency selective and is referred to as "flat" fading (i.e., flat versus frequency). Methods such as MCM, also used in Sirius terrestrial repeaters, and orthogonal frequency division multiplex (OFDM), used in the European DAB system, are ineffective against flat fading.

For greater convenience and operational control, the uplink delivery system at 7 GHz uses 7-m antennas on the roof of the broadcast center in Washington, D.C. They are fed by 3-kW klystron amplifiers and provide EIRP of 70 dBW per carrier. The baseband subsystem has 143 encoders that compress the channels using the AES-EBU format. Other encoder cards can insert up to 10 data streams from a LAN connection. These are encrypted, and various other pertinent information, such as channel names, song titles, and artist names, are inserted into the data and displayed on subscriber receivers. There is an uplink management system in use to control and monitor the baseband and RF equipment, and to provide alarms to Network Management. Another interface allows Broadcast Operations to control the routing of audio content and text information associated with each channel.

The studios are collocated with the technical facilities in Washington, D.C. Numbering 82, these rooms range from small booths that allow a producer to review material that is planned for airing, all the way to what could support a concert. These support talk channels, music studios, multifunction rooms that adapt to almost any format, and the aforementioned performance studio. All of these facilities are digital in nature, as is the Ethernet-based LAN that interconnects studios with baseband equipment. Ample routing and switching permits content, studios, and channels to be rearranged as required, including much provision for failover and backup. A large control room houses operations staff that respond to alarms and other calls to attention. All of the music is transferred from CD and tape to the computer servers, from which it can be recalled on a random access basis during the actual broadcasts. This allows voice tracks and commercials to be developed off-line in rapid sequence, with the computer patching things together at air time.

7.4.3 Radio Equipment Development

The project long considered how it would make chips available to consumer electronics manufacturers on a timely basis. Rather than subcontract this critical aspect of the project, XM brought it inside to an internal technology group in Florida. The design allows the radio to receive, decrypt, and decompress the XM broadcast; the actual chips were produced by ST Microelectronics. The first receivers were manufactured by Sony, Alpine, and Pioneer. An example of the vehicle antenna including

satellite and terrestrial elements is shown in Figure 7.24. However, in 2002, XM turned to GM's Delphi Electronics affiliate to produce the compact and versatile SkyFi receiver shown Figure 7.25. The new radio is extremely compact, features a large scrolling display, and can move between car and home using the various devices shown in the figure.

The S-DARS market in the United States is just starting to gel as the two operators achieve greater than 1 million subscribers between them. Prospects for the service improve as these companies experiment with programming formats in order to find the mix that will cause drivers to tune in satellite channels and pay a monthly subscription fee of around $10. In addition, radios are being installed at the automobile factory, reducing the barrier to entry.

7.5   Expansion of S-DARS into Other Regions of the World

A few commercial S-DARS projects are moving from the design stage into development at the time of this writing. The following examples provide a glimpse of how S-DARS is budding in Japan and Europe, building on the satisfactory introduction of this technology in the United States. As witnessed by D. K. Sachdev, one cannot assume that success in the United States will translate to other markets in other countries. Note in the discussion of Japan, for example, how Mobile Broadcasting Corporation is planning to include video in their program mix.

7.5.1   Mobile Broadcasting Corporation of Japan

The Japanese market is of considerable interest as many families own automobiles that they enjoy driving on weekends and vacation. In addition, new electronic devices tend to be adopted quickly in Japan, particularly in the home where all family members may enjoy them. Mobile Broadcasting Corporation (MBC) of Japan contracted in 2001 with Space Systems/Loral for MBSAT. This system will deliver digital multimedia information services such as CD-quality audio, MPEG 4 video,

Figure 7.24   Combination XM satellite and terrestrial S-band receive antenna.

In vehicle

Home receiver

Radio module

Boombox and antenna

Figure 7.25   SkyFi radio unit and installation options, by Delphi Electronics.

and data to mobile users throughout Japan. On-orbit delivery of the spacecraft was scheduled for the fourth quarter of 2003 with service expected to begin in early 2004. MBC's services will deliver high-quality music, video, and data to mobile users through various kinds of mobile receiver terminals, including those in cars, ships, trains, handheld terminals, PDAs, mobile phones, and home portables. A small antenna will allow reception of MBC broadcasting signals even inside office buildings and in vehicles moving at high speed.

MBC's new broadcasting system was authorized by the Japanese government and registered with the ITU. System capabilities and performance quality have been successfully verified in dense urban locations by various field demonstrations in the Shinbashi and Ginza areas of Tokyo. MBC added SK Telecom of South Korea to its partnership and as a result will use a shared format for receivers in Japan and South Korea.

MBSAT will provide 2,400W RF power over 25 MHz of S-band spectrum to run more than 50 channels of audio and video from 16 S-band transmitters operating at 120W [9]. The new spacecraft will be a version of SS/L's three-axis, body-stabilized 1300 bus, tailored to meet the specific requirements of MBC which include a 12-m antenna reflector deployed in orbit to transmit the MBC programming. A system of high-efficiency solar arrays and lightweight batteries provide uninterrupted electrical power. It will also provide a 25-MHz Ku-band service link to transmit the broadcast signal to terrestrial repeaters. The satellite will generate more than 7,400W of dc power continuously throughout its 12-year life. MBSAT's S-band payload will deliver data using code division multiplexing (CDM) MPEG 4 for video, and advanced audio coding (AAC) for audio. The system will be able to broadcast more than 50 programs simultaneously.

Toshiba Corporation, Toyota Motor Corporation, Fujitsu Ltd., Nippon Television Network Corporation, and other partners established MBC in 1998. In

November 2001, SK Telecom, South Korea's largest cellular phone company, took a stake in MBC, becoming the second largest shareholder after Toshiba. Only recently, Hitachi Ltd. took a stake in MBC, thereby shelving plans for its own service.

7.5.2   European Digital Audio Broadcasting

The European model for digital radio is terrestrial—using advanced coding and modulation to produce a resilient signal that can withstand rapid fading due to multipath. Dubbed Eureka 147 Digital Audio Broadcast, the system is already providing services in Europe and could appear in Canada as well. Like S-DARS in the United States, DAB provides a multiplex of channels for a greater range of programming. The modulation system is OFDM, which transmits multiple carriers to provide a signal that can be received even as some of the carriers are canceled out. It is claimed by the World DAB Forum that more than 285 million people around the world have access to DAB signals and could receive them with appropriate equipment. These new DAB receivers, much like the units produced for XM and Sirius, are manufactured by leading European and Asian consumer electronics companies.

Eureka 147 uses MPEG 1, Layer 2 and MPEG 2, Layer 2 digital audio compression. Controlled coding redundancy applied to the signal gives good error protection and high power efficiency. According to the ITU, government-operated national broadcast networks are particularly well suited to Eureka 147, which is also in use for local radio as well. However, regions where broadcasting is primarily local, or community based and privately owned and operated, or regions with limited spectrum available for new services, may find it more difficult to adopt the Eureka approach [10]. The first in-service date for DAB is 1995, established in the United Kingdom, Norway, Denmark, and Sweden. The development of the Eureka 147 digital broadcasting system was started in 1987. It is a multiservice system that can be operated at any transmission frequency up to 3 GHz. It can deliver a robust signal to fixed, mobile, and portable receivers; all with simple nondirectional antennas. Some broadcasters throughout the world are now operating Eureka 147 terrestrial networks on extended pilot tests and trials or as regular broadcasts. However, the system is also suitable for delivering services by satellite only, hybrid systems (satellite with terrestrial cochannel fillers), mixed systems (satellite with terrestrial rebroadcasting), or via cable networks. By spreading the transmitted information in both frequency and time, the effects of channel distortion and fades at the receiver are avoided, even under severe multipath conditions. This applies to frequency-selective fading and not flat fading, as discussed in a previous section.

DARS in Europe would probably have a different reception than in the United States or Japan. According to Tim Farrar, President of Telecom, Media and Finance Associates, the main issues affecting European S-DARS are the fundamental differences in travel patterns—that is, Europeans spend much less time in cars—and public service broadcasters provide good (often advertising-free) terrestrial radio services with excellent national coverage. Both of these limit the opportunity for any satellite S-DARS.

7.6   Issues and Opportunities Relative to S-DARS

Satellite-based S-DARS has many advantages in the market, some of which are apparent and some of which are underlying. From a practical perspective, services like XM and Sirius offer more audio programming options than one can possibly receive by AM/FM radio at any given time. Coupled with this is the added feature that the same channel complement is available throughout the country according to a constant name and number assignment. Audio quality is comparable to clear FM reception and the radios have the added feature of displaying the channel number and specific piece being played. The latter features are provided by Eureka 147, which cannot so easily deliver consistency in terms of program lineup. S-DARS can then compete effectively with terrestrial and digital radio in the same manner that DTH competes with conventional TV broadcasting and cable. Perhaps the biggest issue in front of S-DARS is that a subscription fee seems to be needed to offset the costs of operation and programming.

The fact that S-DARS operators are themselves radio programmers makes them more like SKY in the United Kingdom that either DIRECTV or DISH Network in the United States. SKY delivers its programming by satellite as well as cable. The U.S. counterpart of SKY, Fox Television, does not own its own DTH platform but offers its programming to all delivery systems. It will be interesting to see if XM and Sirius are able to develop other outlets for their evolving program offering, thus providing additional sources of revenue. At the time of this writing, the biggest issue facing these companies is their cash situation. According to Armand Musey, noted satellite industry analyst, these operators are at the stage where they need to accelerate the sign-up of new subscribers [11]. To do this, they need to reduce the financial hurdle to the early majority of potential customers by subsidizing the sale and installation of receivers and possibly subscription charges as well. This has worked well in the U.S. DTH market, with the result that the subscriber count exceeds 20 million. Accomplishing this in a matter of 8 years has cost DIRECTV and EchoStar dearly. It is the kind of business that requires deep pockets.

There are risks in this market, as demonstrated by failed projects discussed elsewhere. Becoming a mainstream service taken by millions of paying subscribers is still only an expectation at best or dream at worst. This author recalls hearing from a senior executive that if a new DTH service ever reached 10 million subscribers, he would be overcome with joy. Today, we are at 20 million in the United States, with the prospect of reaching even greater heights. This is the promise of a wildly successful new service introduction; it could ultimately take much more effort to reach a fraction of this kind of number for S-DARS.

The true potential of S-DARS lies in the need to change the attitude of the radio listener, who is now accustomed to free service (supported through a continuous stream of advertising or requests for donations). Being able to get the specific channels that interest you, possibly without commercial interruption, represents a new kind of luxury for a market that at times craves luxury. Consider, for example, that people in the United States are willing to pay more for bottled water than for gasoline. The nominal $10 per month charge for S-DARS would not set many people back and in fact is well below the threshold of $50 one associates with "new" services like DSL and the pricey subscription packages on DTH. If equipment cost were

to be reduced to around $100 (or provided nearly free as an already installed feature in new automobiles), Musey suggests that subscriber take-up would accelerate.

References

[1]   International Telecommunication Union, *Radio Regulations – Volume 3*, Geneva: ITU, 1998, p. 273.

[2]   Presentation by David Bell, Jet Propulsion Laboratory, at UCLA Extension, January 2002.

[3]   Sachdev, D. K., "The WorldSpace System: Architecture, Plans, and Technologies," http://www.sadiba.co.za/PDFfiles/worldspace.pdf.

[4]   Courseille, O., P. Fournié, and J. F. Gambart, "On-Air with the WorldSpace Satellite System," *IAF-97 - M.5.05, Satellite Communications Symposium*, International Astronautical Federation, 1997.

[5]   Briskman, R. D., and R. J. Prevaux, "S-DARS Broadcast from Inclined, Elliptical Orbits," *52nd International Astronautical Congress*, Toulouse, France, October 1–5, 2001.

[6]   Michalski, R. A., "An Overview of the XM Satellite Radio System," *AIAA 20th International Communications Satellite Systems Conference*, Montreal, Canada, paper AIAA 2002-1844, May 12–15, 2002.

[7]   "Cruising to Orbit – XM-2," http://www.sea-launch.com/special/sea-launch/past_xm2.html.

[8]   Goldhirsh, J., and W. J. Vogel, *Handbook of Propagation Effects for Vehicular and Personal Mobile Satellite Systems*, Chapter 3.3, http://www.host.cc.utexas.edu/research/mopro.

[9]   "Space Systems/Loral to Build Digital Satellite for Mobile Broadcasting Corporation," in *The News – Loral Space and Communications*, New York, August 15, 2001, http://www.loral.com/inthenews/010815a.html.

[10] ITU Telecommunication Development Bureau, *Digital Radio Guide*, Document 1/014-E, Geneva: ITU, August 31, 1998.

[11] Musey, A., "Satellite Radio Subscriber Economics—Nothing to Worry about Yet," *Industry Note – Satellite and Communications Towers*, SolomonSmithBarney, New York, July 25, 2002.

PART III
Two-Way Interactive Applications for Fixed and Mobile Users

VSAT Networks for Interactive Applications

VSAT networks are composed of low-cost Earth stations for use in a wide variety of telecommunications applications. Unlike the point-to-multipoint systems discussed in Chapters 4 through 7, VSATs are two-way communications installations designed to achieve interactivity over the satellite; interconnection with various terrestrial networks is also a feature. This chapter provides a framework for the use and architecture of VSAT networks, while Chapter 9 is dedicated to technology and design issues.

Since the first edition of this work, the Internet has taken over the role of the common structure for integrating data communications for the majority of applications in information technology (IT). This has rationalized the field to the point that a single protocol and interface standard provide almost all of what an organization needs. The same approach works equally well for individuals and the *small office/home office* (SOHO) environment. Satellite communications technology has adapted to this new world as well. Oddly, it was not until the early 1980s that satellite systems found a direct place in this expanding field. The overriding principle of the VSAT is that it is a small bidirectional Earth station that delivers integrated data, voice, and video services within a package that is often cost justified when compared to terrestrial alternatives.

IT networks can serve basic administrative needs, like payroll processing and e-mail, or strategic needs, like a customer reservation system in the automobile rental business or a just-in-time inventory control system that ties a major customer to its network of suppliers. Today, terrestrial copper and fiber lines and data routing and switching in conjunction with VSATs provide a fast and effective mix to advance the competitive strategy of many medium to large businesses. VSAT networks also address the needs of small businesses and individuals, although the market is still developing at the time of this writing. The three classic architectures for IT networks are host-based processing (utilizing centralized large-scale computers like mainframes), peer-to-peer networks (usually employing minicomputers or large servers that are deployed at different locations to serve local requirements), and client/server networks (which tie together personal computers, servers, and peripherals using LANs and WANs).

8.1 Interactive Data Networks

Data networks usually require a duplex connection for information to be requested, delivered, or exchanged. There is a wide variety of data communications

applications, leading to a very significant difference in the specific requirements for the type and amount of interactivity. Traditional host-based computer networks are perhaps the easiest to manage, while peer-to-peer networks and client/server systems have replaced the host/mainframe approach in many organizations (this dichotomy has its ebb and flow, as organizations attempt to convert from one to the other to improve operational performance and deal with changes in business needs). On top of this, the specific nature of the data varies greatly, and this variety is itself a changing landscape from one year, or even month, to the next. Thus, what was an effective IT network architecture today may become a burden in times of change. These factors make it impossible to generalize on the ideal architecture, data communications structure, or application mix. Instead, organizations must select the network architecture that best satisfies the needs of users and customers. For this reason it is useful to comprehend how VSAT networks and other forms of satellite communications can potentially solve both planned and unexpected needs.

8.1.1   Principle of Protocol Layering

Modern data communications theory and practice is literally built upon the concept of protocol layering, where the most basic transmission requirement is at the bottom and more complex and sophisticated features are added one on top of another. While this concept is abstract, it is important to understanding how the data in a network is assembled, processed, and reliably transferred between sender and receiver. It has evolved over decades of telecommunications development, beginning with the most simple voice radiotelephone network, through networks that support national air defense, applied in business for large-scale data processing, and evolved into the pervasive structure of the Internet. The layering concept is embodied in the Open Systems Interconnection (OSI) model shown in Figure 8.1 and contained in relevant standards of the International Standards Organization (ISO) and the ITU-Telecommunication Sector (ITU-T). It applies in general to all protocol systems, particularly the Internet Protocol suite on which most data communications are provided, yet is concrete enough to allow more general analysis. As we move up the stack, each layer above provides a standardized service, defined in the relevant protocol, to the layer immediately below. In this way, the details within the layer can be optimized for performance and isolated from the other layers. What is specified is the details of how data is transferred to a lower layer for processing and how that layer sends data to its counterpart at the other end of the medium. This used to be called a handshake in reference to how the two sides of a physical connection hand data and acknowledgment across to each other. Without the proper acknowledgment, the delivering side does not know if the receiving side got it. With a satellite network in the middle, we are required to understand how the process works and how to render the result at least as good as what terrestrial networks (which form the basis of the OSI model) can provide.

The standard structure of the OSI model is presented in Figure 8.1, where each box represents a module of functions that are performed at that particular layer by hardware and software. At the very top of the structure is the actual information processing application that requires the network in order to do its function. A detailed discussion of the layers can be found in numerous references on data

Figure 8.1   The layers of the OSI model and the TCP/IP protocol.

communications, such as the familiar book by William Stallings [1] and the informative Web site of Cisco Systems (http://www.cisco.com).

- *Layer 1, physical:* provides the mechanism for transmitting raw bits over the communication medium (e.g., fiber, wireless, and satellite). It specifies the functional, electrical, and procedural characteristics such as signal timing, voltage levels, connector type, and use of pins. The familiar RS-232 connector definition is a good example of the physical layer. A way to look at this is that the physical layer takes the raw bit stream at the sending end and introduces it to the network. All together, most of the investment in a satellite network is at the physical layer.
- *Layer 2, data link:* provides for the transfer of data between adjacent nodes or connection points either by a dedicated point-to-point line (e.g., a T1 private line or a satellite duplex link) or a medium capable of shared bandwidth (e.g., an Ethernet cable or satellite TDMA channel). The link layer can offer a one-to-one connection (the most common approach) or one-to-many delivery (associated with broadcast or multicast).
- *Layer 3, network:* responsible for routing information from end to end within the network, which would consist of multiple data link paths. This may involve decisions about the most effective route through the point-to-point links that comprise the network. A VSAT network may serve as one of these links and hence would have to interface properly with the network layer.

Popular examples of the network layer are the IP that is employed in the majority of router-based private networks and ATM.

- *Layer 4, transport:* provides another level of assurance that the information will properly traverse the network, from end user to end user. Two services are commonly available: connectionless, which transfers packets of data, one at a time; and connection oriented, where a virtual circuit is first established before sending multiple packets that make up the entire conversation. The familiar TCP layer of TCP/IP provides a connection-oriented service to computer applications.

- *Layer 5, session:* somewhat more complicated than layers 3 and 4 but provided to instill yet greater degrees of reliability and convenience of interface to applications. It manages the data exchange between computer systems in an orderly fashion to provide full-duplex or half-duplex conversations. One important service is that of reestablishing the connection in the event that the transport layer is interrupted for some reason.

- *Layer 6, presentation:* provides syntactic and semantic services to the application layer above. What this is saying is that the presentation layer is inserted to resolve the complexities between transport/network layers and the more simplistic needs of the actual application that employs the network in the first place. Some specialized services like encryption and data structure definition are considered to be part of the presentation layer. Interaction of the presentation layer with elements of a satellite network may cause incompatibility, requiring additional processing to be performed in Earth station equipment or user terminals.

- *Layer 7, application:* includes the actual data communication applications that are common in open systems, such as file transfer, virtual terminal, e-mail, and remote database access. We refer to these as applications because they include not only the protocol elements that support specific types of information but also features and facilities that ultimately interact with the end user. Most nonexpert users will not use the application layer directly, instead relying on specialized software within the computer to improve the interface and functionality. For example, most subscribers to on-line information services use the e-mail package supplied by the provider. This package, in turn, will engage layer 7 e-mail services to do the actual function of sending and receiving message traffic.

This was a brief introduction to the structure of modern data communications networks. A clear and useful summary of the characteristics and operation of the Internet can be found in the guides by Matthew Naugle [2] and Floyd Wilder [3]. The VSAT network is ideal for centralized computer networks—that is, those that employ a host computer. The majority of such installations are assembled from standard computer components supplied in the United States by Dell, HP, IBM, and Sun Microsystems; major European and Japanese suppliers like Bull, Siemens, Fujitsu, and Hitachi are in this market as well. Following the layering concept, each computing or terminal device in the network has a unique address that identifies that device at the specified layer. Some examples of addressing schemes are given in Table 8.1. This is very important to the proper interfacing of VSATs with both the host and

Table 8.1 Layers, Protocols, and Their Associated Addressing Schemes Used in Private and Public Networks

Layer	Protocol	Address Scheme	Practical Consideration for VSAT Networks
1 – Physical	Constant bit rate with framing	None—dedicated connection	Use proper connector, bit definition and timing.
2 – Link	Ethernet (IEEE 802.3 series of specifications)	Media Access Control (MAC) address	Widely available at 10 and 100 Mbps and 1 Gbps (10 Gbps offered).
3 – Network	Internet Protocol	IP address—32 bits segmented into 4 bytes	May use fixed (dedicated IP address) or dynamic addressing (dynamic host control protocol—DHCP).
4 – Transport	Transport Control Protocol	Port number	Assigned to particular applications on the end computer systems. VSAT networks employ this layer in TCP acceleration.
5 to 7 – Session, Presentation, and Application	Hypertext Transfer Protocol (HTTP)	Universal Resource Locator (URL)	Assigned to a particular page and object within the page. VSAT networks may actually employ this layer to improve end-to-end performance.

client ends of the network. It is even more critical when the satellite network interconnects with a much larger and more established public network such as the Internet or PSTN.

One of the objectives of OSI—and why the U.S. and European governments pursued it so strongly in the 1990s—was its prospect of creating an open, nonproprietary world for data communications. Efforts to generate economy of scale in nonproprietary systems (i.e., the multivendor environment) produced the OSI open systems architecture, sponsored both by ISO and the ITU-T. While the total picture never came to being, it did produce one success: the X.25 network layer protocol. X.25 is found in many open systems and proprietary systems, including IBM and Digital Equipment Corporation (now part of HP). Both the terrestrial networks and VSAT developers embraced X.25 because of its functionality and universal availability. There are still many countries that employ X.25 in commercial and government networks; however, it is largely being overcome by the Internet.

8.1.2 Protocols Supported by VSAT Networks

A summary of the protocols in general use and their support over typical VSAT networks is provided in Table 8.2. When first introduced in the 1980s, VSATs played heavily on the traditional IBM proprietary protocol, Systems Network Architecture (SNA), which followed the same centralized approach as the VSAT star network. While still in existence in some legacy environments, it has been replaced with the more open Internet Protocol suite (TCP/IP). As many readers are aware, TCP/IP has its shortcomings, which are being addressed by standards bodies and major vendors like Cisco. Employing TCP/IP in a private network is very straightforward and is well within the means of any organization or individual. However, the complexity

Table 8.2 Network Protocols and Applications in Common Use for IT Networks and Their Availability over VSATs

Protocol	Applications	Availability on VSATs
Internet (TCP/IP)	Web, e-mail, file transfer, VoIP, streaming video, videoconferencing	Supported since 1995, now becoming the standard for access and data handing
Frame Relay (ISDN)	Wide area network, private voice networks	Limited (may be substituted by TCP/IP)
Ethernet (MAC layer)	Virtual LANs	Supported since 1992
Novell NetWare (IPX/SPX)	Wide area network	Supported in early VSAT implementations; being replaced by TCP/IP which is provided by NetWare 6

comes when an organization wishes to interconnect with the global Internet and with other organizations. This is due to the somewhat complex nature of routing protocols like the Border Gateway Protocol (BGP) and a new scheme called Multi-Protocol Label Switching (MPLS). The role of BGP is to allow large intranets (which are private networks based on TCP/IP) to interconnect with one another without introducing undesirable interactions and the potential for instability and confusion with IP addresses. MPLS, on the other hand, is intended to improve the performance of the Internet to serve up-and-coming applications that require guarantees as to delivery time or accuracy. Thus, it would provide levels of QoS that can support real-time applications like voice and video transfer along with standard data communications applications over the same infrastructure.

Frame Relay has been popular in WANs for more than a decade, thanks to its ease of interface at the router and availability in (and between) major countries. It is capable of near-real-time transfer and can support voice services. However, the access speeds generally available are at 2 Mbps or less. The specific protocol structure is actually part of the Integrated Services Digital Networks (ISDN) standard. It is used to provide links between intranets and to connect hosts to remotes in a traditional mainframe environment based on IBM and other large systems. Satellite provision of Frame Relay has been limited to point-to-point circuits as the protocol is not directly supported in VSATs currently on the market. The best approach would be to use TCP/IP in lieu of Frame Relay when VSAT links are interfaced at the router.

Ethernet (IEEE 802.3 standard), the most popular access protocol, supports the range of higher-level protocols and networks that are common in the LAN environment. It is very effective as an access medium to the Internet, making bandwidths in the range of 10 Mbps to 1 Gbps widely available at low cost. Being a multiple access technology, Ethernet may not provide the full rated data transmission speed; however, this is only an issue when several devices are connected to the same cable segment (something that was part of the original standard but which is largely overcome through the use of Ethernet switching). Ethernet is also becoming available as a metropolitan area and even national service, where the user can use the MAC addressing approach (discussed in Section 5.7.3). These are hardware addresses that are built into devices like PCs, servers, and peripherals. Currently available VSATs now employ the Ethernet standard to connect to a local PC or LAN.

Table 8.2 also includes Novell NetWare as a networking scheme that is popular in small to medium-sized businesses and government organizations. It has employed a proprietary protocol called Internetwork Packet Exchange/Sequence Packet Exchange (IPX/SPX), where the former moves traffic across a network and the latter operates at the transport layer to work on top of IPX and provides extra features [4]. NetWare is a cost effective means of delivering a basic set of office and business applications across a LAN and WAN of moderate size. Novell has upgraded their offering in version 6 by employing TCP/IP in lieu of IPX/SPX.

Satellite links can support interactive data applications through two fundamentally different architectures: point-to-point connectivity (also called mesh topology) and point-to-multipoint connectivity (also called star topology). Point-to-point connectivity can be either temporary or dedicated links between pairs of Earth stations. Temporary connections emulate dial-up service over the telephone network as well as connection-oriented data communications services. This type of service is common to the fixed telephony networks discussed in Chapter 10. TCP is effectively used on a point-to-point basis for common Internet services like FTP, SMTP, and access to the World Wide Web. Dedicated point-to-point links are useful for telephone trunks and broadband data lines that connect LAN segments together. Alternatively, point-to-multipoint connectivity includes a hub as the center of the star through which all communication passes and is controlled. Each is discussed in detail in the following sections.

8.1.3   Point-to-Point Connectivity

The first satellite networks to be implemented were employed for point-to-point connectivity to complement the cross-country microwave and undersea cable links of the time. This topology remains an effective means of transferring information with minimum delay between pairs of points. As illustrated in Figure 8.2, node 1 in a point-to-point service conducts a full-duplex conversation with node 2 (shown with heavy arrows), and node 3 does likewise with node 4 (shown with broken arrows). As will be considered in Chapter 10 on fixed telephony satellite services,

Figure 8.2   Satellite network topology for point-to-point connectivity using a full mesh.

point-to-point connectivity between node 1 and node 3 can be changed on demand. The maximum number of duplex connections is equal to the permutation $N(N - 1)/2$, where N is the number of nodes. This type of arrangement is called a full mesh, potentially requiring many links.

It has become common practice to refer to the use of separate carriers (e.g., FDMA) of this type as single channel per carrier (SCPC). This may be a misnomer because the aggregate data rate on the carrier may use time division multiplex (TDM) from several independent sources, such as television signals or ISDN circuits. For this to work, the satellite provides separate bandwidth and power for each direction of transmission, either by assigning spectrum (FDMA) or time (TDMA) for the duration of the particular interaction. The link must support transmissions between any pair of terminals; hence, the design must be balanced in terms of critical parameters like Earth station G/T and EIRP. The duration of a connection could be limited, as in the case of a dial-up call or a virtual circuit, or a fixed duration, such as a permanently assigned channel or dedicated circuit. A mesh network is potentially the most versatile because it allows any-to-any communications. On the other hand, it must be managed since users may need to initiate connections on their own without central coordination.

Another consideration is that satellite capacity in a full mesh FDMA network is consumed according to a geometrical relationship: $N*(N - 1)$. The number of carriers and potential capacity can be reduced through the technique of multidestination carriers (MDCs). Pioneered in the early Intelsat and Indonesian Palapa A networks, each Earth station transmits one MDC that contains traffic for all other stations. It is comparable to the outbound hub carrier in a star VSAT network, discussed later in the chapter; however, all stations transmit MDCs as there is no single hub. As illustrated in Figure 8.3 for a network of four Earth stations ($N = 4$), each station has one modulator and transmit chain, and three ($N - 1$) receive chains. In this manner, each station can select from the TDM baseband of each carrier the traffic destined for it [5].

Interactive mesh networks are implemented in almost any frequency band, depending on the availability, pricing, and technical requirements of the particular application. Link capacity is usually symmetrical, meaning that the same speed is used in both directions. This ranges from about 64 Kbps on the low end (for medium-speed data, voice, and low-quality VTC) through T1 or E1 speed (1.544 and 2.048 Mbps, respectively), DS-3 (45 Mbps), topping out at the OC-3 or STM-1 capacity of 155.52 Mbps. Using the link budget principles given in Chapter 2, it is possible to determine the proper antenna and HPA sizes on the ground as well as the required amount of satellite bandwidth and EIRP.

The RF components for point-to-point transmission consist of the upconverter to take the carrier from IF to RF, the HPA to develop the requisite output power level, RF switching to connect devices and allow for redundancy, and waveguide and the transmitting Earth station antenna. Critical to performance is the digital modem that takes the input data stream, provides the desired type and degree of forward error correction, introduces energy dispersal, and modulates the data onto an IF carrier using QPSK, 8PSK, or possibly 16 QAM. It is common practice to purchase a unit that performs both the modulation and demodulation function in one box (hence the term modem). Under the aegis of Intelsat and Eutelsat, these devices have become nearly standardized according to input/output data rate (as discussed

Figure 8.3 Application of FDMA and multidestination carriers in a mesh satellite network.

previously), coding method, and modulation. Coding methods include convolutional coding with Viterbi decoding, which is often concatenated with Reed-Solomon block coding (the turbo product code is available as well), and trellis coding. These standards have been identified under the class of Intelsat Business Service (IBS), IDR, and Eutelsat Satellite Multiservice System (SMS). For example, the Eutelsat offering for point-to-point high-speed digital links has the following basic specifications:

Eutelsat provides network interconnectivity and interoperability in a "Standard Network" in which Earth station characteristics and the satellite transmission parameters are specified. The capacity used is operated in SCPC/ FDMA. Transmission parameters are:

- Information bit rates: 2.4 Kbps to 2 Mbps (up to 8 Mbps on W and SESAT satellites);
- FEC rates: 1/2 or 3/4;
- Modulation: QPSK.
- Earth station type S-1, S-2 or S-3; all operate at 14-GHz uplink and 12-GHz downlink with G/T of 30.0, 27.0 and 23.0 dB/K, respectively [5];
- SMS is suitable for the transmission of the following services (this list is not exhaustive) [6]:

- Data;
- Text and images;
- Voice.

Specific applications for which the system is particularly suitable are as follows:

- Videoconferencing;
- Audioconferencing;
- Computer-to-computer transfer;
- Remote printing;
- Packet switched data carrier;
- Fast facsimile;
- Teletex;
- Slow scan TV;
- E-mail.

SMS carriers use coherent QPSK modulation and rate 3/4 or rate 1/2 convolutional coding with Viterbi decoding. The rate 3/4 is a puncture type of convolutional code and is constructed from a rate 1/2 encoder by periodically deleting specific bits from the rate 1/2 output bit sequence. (This is the same technique used in DVB-S.) The SMS specification goes into considerable detail as to data frame format, coding, and modulation, and is best left to the manufacturer to insure compliance. However, the benefit of these specifications is that equipment from different vendors should interoperate. Eutelsat services and standards were selected only as examples and are indicative of what is offered by major operators like SES-Global, PanAmSat, Intelsat, Loral Skynet, and JSAT.

There are a number of specialized suppliers of compliant modem equipment, including Comtech EF Data, Radyne Comstream, Hughes Network Systems, ND Satcom, and NEC. As an example of this type of device, the Comtech EF Data model CDM-600 is a bidirectional satellite data modem capable of rates between 2.4 and 20 Mbps. It can satisfy requirements of the majority of satellite operator standards for SCPC/FDMA applications, providing QPSK, 8PSK, and 16 QAM modulation along with convolutional, Reed-Solomon, and turbo product codes. The modem can be purchased with the appropriate features to meet the requirements of the link. Feature enhancements (such as turbo product code) can be purchased at a later date for activation in the modem.

8.1.4   Point-to-Multipoint Connectivity (Star Topology with VSATs)

The point-to-multipoint connectivity is illustrated in Figure 8.4. The thick, shaded arrows represent the digital broadcast "outroute" from the hub to the remote nodes (other acceptable terms for the hub transmitted signal are "outbound," "forward," and "downstream"). It contains all hub-originated data to be delivered to the VSATs throughout the network. This transmission is received by all remotes within the satellite footprint; however, it would typically contain address information that allows only the desired remotes to select the information destined for them. The thin lines represent the "inroutes" from the individual remote nodes (likewise, acceptable terms include "inbound," "return," and "upstream"). Several VSATs may share each inroute frequency and so must be separated in time (TDMA), frequency

Figure 8.4 Satellite network topology for multipoint-to-point connectivity using a star with hub.

(FDMA), or code (CDMA). Combinations and variations of these multiple access techniques are employed to further optimize data transfer and capacity usage. All information transfer is between remote node and the hub; direct single-hop remote-to-remote data transfer is not possible with this topology. This approach has the following benefits:

- The hub maintains tight control of transmissions to and from remotes.
- Information transfer is optimized for star topology information networks, such as those maintained by centralized corporations and service/content providers.
- The RF design of the return link can be arranged to minimize dish size and transmit power at the remote VSAT. The resulting weak carrier signals can be received by the larger diameter hub station antenna (this assumes a bent-pipe satellite transponder—without an onboard processor that demodulates the bit stream).
- A common network management system is provided through the hub station.

Unlike the point-to-point arrangement, the link design is imbalanced to simplify the VSAT and thereby reduce its cost. An example of a typical Ku-band link budget for the star network is presented in Table 8.3, which is a summary of a detailed analysis of outbound and inbound transmissions through the SatMex 5 satellite, located at 116.8° WL. The antenna diameters of the VSAT and hub are assumed to be 1.2m and 4.7m, respectively. This results in a gain difference of 12 dB in favor of the hub, an imbalance that allows small VSAT antennas to be used in the first place. The overall link C/N is comparable for both the outroute and the inroute, assuring more or less balanced performance in the presence of rain attenuation. The link

Table 8.3 Example of a Ku-Band Link Budget for a VSAT and Hub Star Network

Link Component	Outbound[1]	Inbound[2]	Units
Information data rate	12.0	0.512	Mbps
Uplink			
Frequency	14.25	14.25	GHz
Antenna diameter	4.7	1.2	Meters
Antenna gain at peak	55.5	43.2	dBi
HPA power	4.6	1.8	Watts
	6.6	2.6	dBW
Waveguide loss	2.0	0.5	dB
Mispointing loss	0.5	0.0	dB
Uplink EIRP	59.6	45.3	dBW
Path losses (no rain, including satellite mispointing)	207.7	207.3	dB
Spacecraft G/T	4.0	3.0	dB/K
C/T_{up}	−144.1	−159.0	dBW/K
Boltzmann's constant	−228.6	−228.6	dBW/Hz/K
C/N_0	84.5	69.6	dB-Hz
Bandwidth	11.7	0.492	MHz
	70.7	56.9	dB(Hz)
C/N_{up}	13.8	12.7	dB
Downlink			
Frequency	11.95	11.95	GHz
Satellite saturated EIRP	51.0	49.0	dBW
Output backoff per carrier	10.5	25.5	dB
EIRP per carrier	40.5	23.5	dBW
Earth station mispointing loss	0.0	0.5	dB
Path losses (no rain, including satellite mispointing)	205.7	206.1	dB
Receive antenna diameter	1.2	4.7	Meters
Antenna gain at peak	41.7	53.5	dB
LNB noise temperature	70	70	Kelvin
Antenna noise temperature	50	50	Kelvin
Feed coupling loss	0.25	0.25	dB
System noise temperature	123.4	123.4	Kelvin
	20.9	20.9	dB(K)
Earth station G/T (at peak of beam)	20.8	32.6	dB/K
C/T_{dn}	−144.4	−150.0	dBW/K
C/N_0	84.2	78.6	dB-Hz
C/N_{dn}	13.5	21.7	dB
Total link thermal C/N	10.6	12.2	dB

Note: satellite: SatMex 5; hub: Mexico City; VSAT: Hermosillo.
[1] 12 Mbps, QPSK,=0.35; FEC: R = 3/4 convolutional, (204,188) Reed-Solomon.
[2] 512 Kbps, QPSK,=0.28; FEC: R = 2/3 convolutional, no Reed-Solomon.

budget is simplified in that it does not consider interference sources, such as cross-polarization, adjacent satellite and terrestrial interference. These would need to be included through the noise-combining technique given in Section 2.2.4. Also, the budget is stated for the clear, no-rain condition. Link margin would therefore be

necessary to accommodate these factors. For a detailed analysis, we recommend the use of a consistent link budget software tool such as SatMaster Pro from Arrow Technical Services of the United Kingdom (which was used to verify the data in Table 8.3).

As a result of the star architecture, all communication passes through the hub, whether it is destined there or not. It is the equivalent of a star topology used in many host-based data processing environments. Networks of this type are highly coordinated and can operate very efficiently. The disadvantage (which may actually not apply in many important cases) is that direct remote-to-remote communication is on a double-hop basis through the hub. As will be discussed in the next chapter, the link budget provides the means to size the network in terms of transponder capacity requirements (bandwidth and power), antenna sizes of remote and hub, and HPA powers.

Frequency bands in use for VSAT networks include Ku-, C-, and X-bands. In principle, any of these bands may be applied as the baseband elements of the hub, and remotes are what establish the network (these interface at IF frequencies such as 70 MHz and L-band). Ku-band is generally preferred for large networks as the link typically can support smaller dishes at the remotes. The reasons for the Ku-band capability include:

- In the Americas and certain other regions, a portion of the Ku-band FSS allocation is not shared with terrestrial microwave radio systems (e.g., the fixed service); thus, neither terrestrial coordination nor terrestrial interference would typically be a concern.
- For a fixed dish size, the antenna beamwidth decreases as frequency increases; the C/Ku ratio of 6/14 for transmit means the Ku beamwidth is slightly less than one-half that of C-band. This factor tends to control adjacent satellite interference in a 2° spacing environment
- Ku-band satellites are designed to transmit at a higher EIRP level, supporting smaller dish diameters or greater bit rate for the same diameter. This is offset by the greater rain attenuation experienced in areas of high thunderstorm activity, such as the southeastern United States and tropical regions of South America, Africa, and Asia.

The case for C- and X-band can still be made when suitable Ku bandwidth may not be available or the rain environment may be unsuitable in the region of interest. For these reasons, C-band, while demanding dish sizes effectively twice that of Ku-band, is the only practically solution for VSATs in the Pacific Islands, equatorial South America, Southeast Asia, Central Africa, and India. With regard to X-band, this allocation is typically assigned to government services and is appropriate for fixed, transportable, and mobile applications.

The point-to-multipoint architecture is very common in modern satellite data networks and is the basis of the success of VSATs in may commercial and government application environments. A VSAT is a complete Earth station that can be installed on the user's premises and can provide business communication services. For the purposes of this book, we regard a VSAT as the remote Earth station in the star network architecture. An example of a typical VSAT with a 1.2-m antenna is

illustrated in Figure 8.5. The antenna feed has an RF outdoor unit (ODU) attached to it that contains the transmit-receive portion of the terminal. Cables connect between the ODU and the indoor unit (IDU), which is a complete baseband system contained in a cabinet about the size of a PC or compact router. The market at the time of this writing was dominated by three specialized companies: HNS, Gilat, and ViaSat; however, a variety of other smaller companies deliver products that are very worthy of consideration (as will be discussed later). One current consideration is that unlike the SCPC equipment discussed in Section 8.1.3, VSAT networks are normally proprietary and do not interoperate. There is a movement in Europe to create a standard VSAT platform under the DVB group, called the DVB-Remote Channel by Satellite (DVB-RCS). The status and benefits of this approach will be reviewed later in this chapter.

8.2   VSAT Star Networks

Organizations employ VSATs primarily as replacements for terrestrial data networks using private lines in a variety of applications, including retailing, postal and package delivery, automobile sales and service, banking and finance, travel and lodging, and government administration and security. Perhaps the first major installation was for Wal-Mart, the leading U.S. retailer with stores throughout the United States and other locations around the world. Today, there are more than 250,000 two-way VSATs installed in the United States and over 600,000 worldwide. Not included is the consumer VSAT designed to provide Internet access, an application that will be covered in Chapter 9. VSAT technology should only be used as a supplement to high-quality digital fiber optic and wireless networks of the world. In fact, the best strategy is often to complement the terrestrial network infrastructure with VSATs so as to achieve an optimum and reliable mix. For example, a European company needing to connect only five domestic locations to a data center would find that conventional VSATs may not be cost-effective. Likewise, a large industrial organization that needs high-capacity links between major sites is not a candidate

Figure 8.5   Typical installation of a 1.2-m VSAT. (Courtesy of Hughes Network Systems.)

for existing VSATs. This would clearly be a better application for fiber optic links, if that were feasible, or point-to-point satellite links discussed in Section 8.1.3.

8.2.1   Applications of Star Networks

Many centralized companies build their IT systems around the host computer that is located at the headquarters or outsourced hosting facility. This is an ideal starting point for VSAT network adoption since it is centralized. Table 8.4 provides a listing a popular IT applications now provided over enterprise VSAT networks. We offer

Table 8.4   IT Applications Provided by Star VSAT Networks

User Application	Network	Technology
Internet access (one user; small group; remote site)	High-speed access to Internet backbone; TCP/IP	One way over satellite; terrestrial return
		Two way over satellite; broadcast outbound with multiple access inbound
Remote access to corporate Intranet (LAN extension)	High-speed access to private network infrastructure; Web-based applications; TCP/IP	One way over satellite; terrestrial return
		Two way over satellite; broadcast outbound with multiple access inbound
Remote access to corporate business applications	Medium- to high-speed access to private network infrastructure; applications employ client/server or mainframe style; may employ proprietary protocol	Two way over satellite; broadcast outbound with multiple access inbound
		Two way over satellite; point-to-point circuit, either preassigned or demand assigned
Content distribution	Multicast uplink for wide area distribution to PCs and content caching servers; UDP/IP and Multicast Transport Protocol (MTP)	One way over satellite; verification of 100% reception via terrestrial or satellite return
Video teleconferencing	High-speed access to private network infrastructure or public ISDN; H.320 or H.323 standards	Two way over satellite; broadcast outbound with multiple access inbound
		Two way over satellite; point-to-point circuit, either preassigned or demand assigned
Telephone	Low- to medium-speed access to private network infrastructure or PSTN; POTS or VoI standards	Two way over satellite; broadcast outbound with multiple access inbound; echo cancellation
		Two way over satellite; point-to-point circuit, either preassigned or demand assigned; echo cancellation
Leased line	Medium- to high-speed connection; T1/E1	Two way over satellite; point-to-point circuit, preassigned

the following examples from the varied experience with VSAT star networks in developed and developing countries.

8.2.1.1   Retail Marketing—Wal-Mart and JD Group

Wal-Mart was an early adopter of the VSAT star network, extending the reach of its IT resources at the HQ in Bentonville, Arkansas, to thousands of remote towns in rural America. Without VSATs, Wal-Mart would have had great difficulty integrating their systems of distribution, credit verification, training, and financial management. With credit card verification, price and inventory management, video-based training, and telephone services, the VSAT network proved its worth in financial and operational terms. One could argue that without the VSAT network, Wal-Mart might not have grown as rapidly and become the number one corporation in America. The network supported the most advanced just-in-time automated replenishment system in the industry, which uses regional distribution centers and links to suppliers who deliver direct to Wal-Mart stores. This strategy was so effective that it was imitated by Kmart, Target, Home Depot, and Staples.

JD Group is South Africa's leading furniture retailer operating through six chains that cover a spectrum of consumer needs. While each chain has its own identity, merchandise range, and market profile, they all concentrate on offering customers a wide range of value for the money on products like quality furniture, appliances, home entertainment, and consumer finance products, supported by a high level of personal service. JD Group serves the mass market through 665 (2000: 671) stores in urban and rural areas across southern Africa. They service 1.3 million customers and employ 40 distribution centers that deliver products directly to customers' homes. Customers predominantly pay in cash (not credit card or check as is common in developed countries) and are afforded extended terms. As a result, customers travel to stores once a month to make payments to their accounts. The result is more than a million cash transactions a month for JD Group to contend with.

With inaccessible terrain, the company decided in 1995 to move from messenger to data communications using terrestrial telephone lines. However, the prevailing situation gave a new meaning to the concept of "cable theft" (normally associated with stealing pay TV signals using illegal devices). It was reported that cut and stolen copper cable in a city near their headquarters resulted in the local community being without telephone service for more than 2 weeks. Rather than rely on terrestrial copper telephone lines, JD Group selected a star VSAT network provided through a shared hub operated by Telkom of South Africa. More than 600 stores in South Africa were fitted with 1.2-m bidirectional dishes supplied by Hughes Network Systems. Installation in stores in neighboring Namibia, Swaziland, and Lesotho are pending negotiations with the local telecom providers [7].

8.2.1.2   Automotive—Daimler-Chrysler and Toyota

Daimler-Chrysler Corporation, one of the Big Three U.S. auto makers, and regarded as an innovator in design and engineering, has provided every one of their 6,000 U.S. dealers with a VSAT to be used for order entry, financing, parts inventory management, distance education and training, and other applications that get added as time

goes on. The VSAT network allows Daimler-Chrysler to treat every one of its deal-ers equally, which is important because they are usually individually owned and operated by local business people. Following Chrysler, General Motors and Ford Motor Company introduced VSATs as both a two-way and a one-way medium. As discussed later in the chapter, GM University and FORDSTAR use VSATs to pro-vide all sales and maintenance training to U.S. dealers.

Toyota of America's Lexus Automobiles, the most successful new luxury car line in the United States, uses their Lexus VSAT network as part of a strategy to give the customer the impression that there is one dealer and service organization operating in North America. A Lexus purchased in Torrance, California, can be taken for service in Falls Church, Virginia, where all of the maintenance records are immediately avail-able over the satellite link. All customer information is right in front of the service rep-resentative, just as if the car had been purchased in Falls Church instead of Torrance.

8.2.1.3   Postal Services

The U.S. Postal Service, a quasi-government agency, has the largest VSAT network in the United States. With approximately 11,000 Skystar Advantage VSATs from Gilat, the USPS can reach even the remotest places and provide enterprise applica-tions on a seamless basis. The network is used for the following applications [8]:

- Point of sale;
- Data polling;
- Credit authorization;
- Delivery confirmation;
- Software distribution;
- Recovery control (backup) of the terrestrial Frame Relay network.

Using VSATs as a backup for terrestrial systems is a proven and reliable solu-tion. Nowhere was this truer than the September 11 attack in New York. As a result of the attack on the World Trade Center, the Frame Relay network of USPS's New York branches became ineffective. Yet once the USPS redirected its VSAT system in Virginia to New York, the states' post offices were back on-line. Generally, the switchover is automatic: When Frame Relay goes down, a satellite connection takes over [9].

8.2.1.4   Hotels and Travel—Holiday Inn

Holiday Inn, part of the global hospitality chain of Six Continent Hotels, uses VSATs in North America to connect between headquarters and the larger properties for access to the reservation system and other applications for business manage-ment. The original VSAT network was installed by HNS and demonstrated the value of satellite communications in the large hotel chain environment. Subse-quently in the second generation, Holiday Inn went with Gilat and their Spacenet subsidiary. The VSAT network supports the IP-based automated central reservation system in 2,100 Holiday Inn hotels in the United States [10]. This always-on, Web-based application provides the following services:

- Guest reservations;
- Credit card authorization;
- Polling services.

The comprehensive solution, from Gilat subsidiary, Spacenet includes turnkey installation of all remote site equipment, network operations, and field maintenance.

8.2.1.5   Retail Banking—Banamex

Banco Nacional de Mexico (Banamex), the largest financial organization in Mexico, relies on its VSAT network to reliably communicate with its branch offices. Tellers are supported with terminals that can connect back to the headquarters' computers and databases. The system allows Banamex to provide consistent services to its clients whether they are in Mexico City or a small outlying town.

8.2.1.6   Recent VSAT Trends

VSAT networks arose in the mid- to late 1980s as a result of electronic and software innovations that permitted all of the necessary features to be contained into an affordable package about the size of a personal computer. By affordable, we mean that the user pays a total network cost that is competitive with many terrestrial data network alternatives. This includes the cost of ownership of the VSATs and possibly the hub, along with the rental of adequate space segment capacity. The price of a Ku-band VSAT itself started around $15,000 in the mid-1980s, dropping over the years to a level of around $6,000 in 1995, settling at approximately $1,500 at the time of this writing (C-band VSATs may cost slightly higher due to a larger dish needed to receive from a satellite with lower EIRP than at Ku-band, and a somewhat more expensive SSPA due to the higher cost of this class of device). This represents a 90% discount; however, the specific price of a VSAT depends on its configuration and feature set, along with the size of RF amplifier and dish. In recent years, greater integration and large-scale production have brought to price to the same level as a high-end PC and thus makes more applications justifiable from a financial perspective. This permits large-scale application of two-way VSAT technology, including small businesses and SOHO use.

8.2.2   VSAT Network Architecture

The VSAT antenna is typically in the range of 1.2m to 2.8m; the size depends on the satellite coverage performance (G/T and EIRP), the capacity requirements, and the frequency band. The following discussion will address the most popular case: Ku-band VSATs with antennas between 1m and 1.8m. As reviewed previously, their rapid rise in the United States was based on Ku-band because Ku-band is not shared with terrestrial microwave. Hence, VSATs can be placed where needed to provide service. No frequency coordination is needed, and licenses can be granted on a network basis—this is termed *blanket licensing* because the national regulator (the FCC in the United States) authorizes the network based on equipment characteristics

rather than antenna location (critical at C-band). Ku-band FSS satellites usually have higher EIRP than C-band (50 versus 36 dBW) to allow the smallest antennas on the ground and to overcome the increased rain attenuation at this higher frequency. Adjacent satellite interference is less for the same Earth station antenna size because main-lobe gain increases while sidelobe gain does not (for the same satellite spacing). This is taken advantage of by recognizing that a smaller Ku-band antenna has about the same interference potential as the larger antenna used at C-band. A smaller Ku-band antenna can be placed in an inconspicuous place such as on the roof of the building or behind a low screen wall. The transmit power level, being below 4W into the antenna, would not be a hazard to humans for these types of installations (provided that one does not stand close to the antenna and directly in the RF path). C-band VSATs are required in areas without adequate Ku-band satellite capacity; therefore, the user must usually carry out the necessary frequency coordination prior to use in order to protect existing terrestrial microwave stations. These aspects are covered in Chapter 12.

Because potential customers are really users of telecommunications rather than technology managers, VSAT marketers must make a complete service offering that includes equipment, installation, maintenance, and on occasion, hub, satellite capacity, and network operations. This requires that entrants in this market be prepared to become more vertically integrated than previously required for satellite or teleport operations and equipment sales. Going into this business requires both a significant investment and a commitment to a specific technology platform in terms of VSAT supplier, satellite transponder bandwidth (which must be taken for an extend term), and skilled human resources.

The architecture of the typical VSAT star network is provided in Figure 8.6. This emphasizes the configuration of the ground equipment and ignores the

Figure 8.6 Typical hub Earth station and remote VSAT in a star network for corporate data communications. (Courtesy of JSAT International Inc.)

techniques used for modulation and multiple access (referred to as the air interface). It depicts how the user connects computers, PCs and other terminals, PBX and telephone systems, and video equipment used in private broadcasting. The hub of the star is shown on the right in the form of a complete Earth station facility with a relatively large antenna (typically 4.7m at Ku-band and 9m at C-band). The most common implementations of the star network use TDM on the outroute and TDMA as well as a derivative called ALOHA (discussed in Chapter 9) on the inroute. The hub and the entire network of VSATs can be managed from the single network management console attached to the hub (in fact, many implementations allow the console to reach the hub over the Internet).

8.2.2.1   Hub Equipment

The hub is the origination point for the outroute carrier that is received by all VSATs in the same network. It has a fixed RF capacity and transmits information as a constant bit rate stream of data using time division multiplexing. The detailed structure of the bits and bytes would comply with a protocol standard, such as IP encapsulation into MPEG 2, discussed in Chapter 4. In addition to user data, the outroute also includes control information that allows the hub to manage each VSAT and thus control all network resources. Outroute data rates are in the range of 10 to 80 Mbps, where the rate depends on the total network capacity transponder bandwidth and the architecture of the particular vendor's equipment. As discussed in Chapter 9, the data rate dictates the RF power and bandwidth, and hence the cost of satellite capacity. One must employ sufficient bandwidth to provide a quality service, yet it should not cost more than the network can justify. Purchase of transponder capacity is covered in Chapter 13.

The RF equipment is fairly conventional, usually consisting of a redundant set of LNAs attached to the antenna feed and TWT power amplifiers in the range of 80W to 400W, depending on satellite G/T, capacity requirements, and a need for extra uplink rain margin. Up and downconverters provide a 70-MHz or L-band IF interface to the indoor baseband equipment. High powers and voltages exist at this equipment, so it is important that it is protected from casual contact and the elements. Also, it should be possible to work on the equipment even during heavy rain (Murphy's Law would indicate that the time when you need to service the equipment also happens to be during a rainstorm). An optional feature in some Ku-band installations is uplink power control, which increases power during heavy rain to maintain the link without consuming excess satellite capacity during clear weather.

The heart of the hub is the baseband equipment, which performs all of the protocol processing, encapsulation and multiplexing, forward error compression, modulation, and multiple-access functions needed to manage traffic over the entire network. The customer's host computer connects directly to a data port in the same way as it would to a terrestrial data line operating at 64 Kbps or some multiple thereof. Alternatively, the interface may employ Ethernet at any of the available transfer rates (e.g., 10, 100, or 1,000 Mbps). Hubs can also support telephone service within the confines of the star network. In this case, human speech is compressed from 64 Kbps down to the range of 8 to 32 Kbps, depending on the requirements of

the customer. Generally, the higher speed provides a more natural-sounding service comparable to what is available on the public telephone network. Still, purchasers of VSAT networks prefer the 8-Kbps rate as it reduces space segment costs. One caution is that in-band data using V.90 modems and the like is only possible at 32- and 64-Kbps speeds. Voice services can be transmitted at constant bit rate or variable bit rate using VoIP. The benefit of the latter is that voice can be more easily integrated with other forms of data in an IP-based VSAT network. The hub provides demand assignment features, as will be discussed later, so that calls are set up and taken down as needed.

The third capability offered by the hub is video broadcasting using either live or recorded material. There are two means for adding video capability to the VSAT outroute transmissions:

- Use a separate RF carrier with the video and associated audio modulated onto it. The modulation method may be analog or digital—companies may employ their own video uplink or rent time on video uplinks since they generally use this capability only on occasion. The carrier must be on the same satellite and preferably the same polarization to permit reception by each VSAT antenna and LNB.
- Integrate video and data using IP encapsulation into MPEG 2. In actuality, the video and associated audio provide the base MPEG 2 stream into which IP data may be encapsulated. This has the double benefit of using one carrier, and allows the data to burst to higher rates by grabbing unused MPEG 2 null packets in the video stream.

It is highly likely that the second process, which employs a common transport mechanism, will use considerably less satellite capacity, particularly if we are talking about more than half of a transponder. The first case requires two carriers instead of one, which represents an advantage for the second. However, the fact that the video may come from a different source and not be full time could give the first case enough advantage to warrant its use.

The hub provides the central coordination and management point for the entire network. Control of the VSAT network is extremely important to the overall user or network operator. The basic approach is to use the hub as the coordinating point and to provide specialized network management functions over the same links between the hub and each remote VSAT. The inbound link provides the path for status information from the VSAT back to the management system at the hub. A network management workstation is connected to the hub equipment to allow the network operator to configure the network for service and to monitor its performance over time. This allows the operator to be proactive, meaning that the network is always under surveillance and every remote is monitored for its activity and potential for malfunction. If a problem arises, then it is possible to isolate and resolve it in a relatively short time. The remote VSATs themselves do not require local operator support as long as the link is functioning. One caution is that a malfunctioning VSAT could break frequency lock and cause radio frequency interference (RFI) across the transponder or the 500-MHz band. Without hub control in this case, the errant dish must be shut down manually by someone at the remote site.

8.2.2.2   Remote VSAT

A full-capability remote VSAT is shown at the left of Figure 8.6. The primary function is that of a communications network access device for the local PCs on a LAN, legacy data terminals, and phones (analog, ISDN, and VoIP). In the most common configuration, the VSAT offers an RJ-45 connection to the LAN, for which there are three possibilities:

- A direct LAN connection (making the VSAT appear to the LAN as either a switch or router);
- Connection through a PC that has been loaded with custom software for the particular VSAT network architecture (e.g., proprietary software that is either downloaded through the Internet or from a CD-ROM);
- A router that provides the interface to the LAN (the router also allows local connection to the Internet for alternate routing of traffic).

The baseband equipment performs the necessary protocol conversion or adaptation, an important part of which is to fool the end computer systems into thinking that it is passing data to the distant computer over a minimum delay connection. Various approaches to this have been tried in the past and some have remained effective in allowing the delay of a GEO satellite link to be accommodated. The key is to understand the differences between satellite and terrestrial protocols, which are summarized in Table 8.5. Basically, terrestrial protocols (TCP/IP, ATM, and Frame Relay) were designed for use on terrestrial networks with their low propagation delay and low error performance. Satellite protocols, on the other hand, must deal effectively with the delay and error properties of the path length and noise. Data networks typically use terrestrial links provided by several organizations; as a result, the user cannot control nor even know what different links and service providers will be involved. In contrast, satellites networks are offered by a single organization, which can be the user group or an outsourced service from one provider.

For terrestrial protocols to work properly over satellite links, the concept of protocol adaptation (also called spoofing) can be applied. Figure 8.7 indicates that the application on the left end of the connection employs the standard terrestrial protocol labeled P_1. This would be one of the common network protocols, such as TCP/IP. An important property of P_1 is that it automatically recovers from errors on the link through automatic retransmission of bad or lost packets. With the satellite propagation delay of approximately one-quarter second from end to end (one way), such retransmissions will introduce a very significant reduction of throughput and even

Table 8.5   Key Differences Between the Popular Terrestrial Data Protocols and Those over GEO Satellite Links

Terrestrial Protocols	Satellite Protocols
Highly variable comm. bandwidth and quality	Wide bandwidth and stable comm. quality
Relatively short propagation delay	Fixed, long propagation delay (for GEO)
Many intermediate nodes and connections	Very few intermediate nodes (for single hop)
Wide variety of alternatives	Common network architecture
Many different service and equipment providers	Limited number of providers, but must interface properly

Key:
P₁   Terrestrial protocol (e.g., TCP/IP)
Pₛᵢ   Satellite protocol, inbound direction
Pₛₒ Satellite protocol, outbound direction
R   Router and/or firewall

Figure 8.7   Adaptation of terrestrial protocol into a satellite protocol.

cause the application to hang. We help reduce this possibility through forward error correction, which will take the nominal BER of the satellite link and reduce it several orders of magnitude (e.g., from 10^{-4} to 10^{-9}). For example, if the information rate were 1 Mbps, then 10^{-4} represents an average of 100 errors per second. However, at 10^{-9}, the average interval between errors would be 1,000 seconds (about 18 minutes). As a wireless/radio medium, the satellite link cannot reduce errors to zero; hence, protocol P_1 must at times engage in automatic retransmission. Our objective is to reduce these instances to once in a great while. In addition, the spoofing technique provides the locally attached computer with immediate response to eliminate the delay during retransmission over the satellite link.

Reliable data transfer across the satellite link is the responsibility of the pair of protocols indicated as P_{so} and P_{si} (representing outbound satellite protocol and inbound satellite protocol, respectively). The outbound transfer of data is generally done using a continuous flow of packets on a broadband satellite link, using MPEG 2 or another transport format. This broadcast is received at all stations, which can select the data addressed to them within the packet structure. DVB-S is currently popular for this as it includes concatenated FEC and can deliver error rates of less than 10^{-9}. Protocol P_{si} is transmitted by the VSAT and usually employs one or a combination of multiple access methods. This provides the physical and link layers for inbound data; higher layers may be implemented within the VSAT or provided by software loaded in the PC and server. The operation of the satellite protocols is illustrated in Figure 8.8 and typically operates as follows:

- User terminal device initiates session using TCP/IP protocol.
- VSAT indoor unit initiates satellite link connection with hub.
- Hub initiates session with server using TCP/IP protocol.
- IP packets are buffered at VSAT and host, and converted into satellite link packets.
- Satellite errors are detected and corrected over the VSAT network.
- Host IP packets are likewise spoofed at hub.

| TCP/IP | Custom protocol | Custom protocol | TCP/IP |
| (DTE-DCE) | (VSAT-hub) | (Hub-VSAT) | (DCE-DTE) |

Figure 8.8   End-to-end connection example.

• End-to-end TCP/IP session proceeds and terminates upon PC command.

After the session is established, the protocol over the satellite link takes care of reliable data transfer, delivering the actual information from source to destination. The delay encountered is principally that of a single hop (about one-quarter second) or at most a double hop (about a half of a second). Spoofing along with reliable data transfer are the techniques that can make a satellite network perform as well or better than a terrestrial telephone network.

If telephone service is offered, then the VSAT can connect to a PBX or key system to allow placing calls over the satellite to the PBX at the hub. VSAT-to-VSAT calls are generally avoided as they experience a double hop. Historically, voice calls were provided with a constant bit rate of 8 or 16 Kbps using a standard compression algorithm. More recently, VoIP has been added as a means to integrate voice on a packet basis directly into the IP environment of the terrestrial and satellite network. Another important application is business TV and videoconferencing, using either MPEG 2 or IP-based video. For MPEG 2, the video receiver would be included if the site is to be part of the private broadcasting network. This is a TV receive-only system, so any response would use the narrow bandwidth of either the data or telephone facilities of the network. Two-way video would most likely use an IP standard, such as H.323. More detail, including an example of an actual business TV network using VSAT technology, is provided in Section 8.3.3.

The typical VSAT is designed to be unattended, requiring only limited attention by customer personnel, who only need to worry about maintaining power and occasionally inspecting the minimal display on the indoor unit. Problems can arise within either the indoor unit or the RF head, which can render either the satellite link or the local connection insufficient to maintain service. These problems are relatively infrequent because of the simplicity of the electronic design and the structure of the satellite network. This situation is usually detected by the hub Earth station and corrected by a competent field technician who would be sent to the site.

8.2.3   Integrator of PCs, LANs, and Internets

Data communication networks have grown far beyond the realm of the host computer, propelled by the popularity of three facilities that are now commonplace: the PC, the LAN, and the Internet suite of protocols. The latter is actually more than a set of protocols, since it now comprises many public and private networks

worldwide. VSATs have worked to keep pace with these developments, motivated by the extensive use of IT facilities in literally every organization in every developed and developing country. We consider here how the three facilities can be supported by VSATs.

8.2.3.1   Personal Computer Integration with the VSAT

The PC is the ideal direct user interface with the VSAT in applications where on-line information delivery is required. Typical telephone networks have a real throughput of about 40 Kbps; lower rates are common in areas where line quality is poor. This was once adequate for applications such as on-line service connection, dial-up terminal access to e-mail, and fax. With the growth of the World Wide Web and the increasing demand for the transfer of large files for graphics, database, and engineering applications, the analog telephone network ceases to be adequate. The marketplace is provided with VSAT networks that have typical inbound throughputs in the range of 128 Kbps to 2 Mbps. This bandwidth can be delivered to a single user, being able to transfer a 1-MB file in seconds rather than minutes.

A historic example of using the point-to-multipoint feature and the PC is shown in Figure 8.9. The data files or streams are uplinked from a hub Earth station at the left. A public ISP or content delivery network service would own and operate the hub. Information is delivered to the hub over backhaul circuits from one or more servers or other information sources (e.g., a stock market ticker). Subscribers purchase and install a receive-only VSAT, which need only receive the high-speed forward link broadcast from the hub. Some existing and potential applications of this medium are listed in Table 8.6. These were selected based on a predominately outbound data transfer requirement.

The interface for most data applications between the PC and the VSAT is through the Ethernet connection standard. (RS-232 serial data, while once very popular, is disappearing from common usage.) Client software in the PC can select information from the stream based on a variety of criteria chosen by the user and displayed in real time. If it is a file transfer, then the block of data is loaded into RAM and possibly saved to the hard drive. This type of application has actually

Figure 8.9   Architecture for data broadcasting with a terrestrial request and return channel.

Table 8.6 Typical Internet and Information Services That May Be Provided by a Satellite Data Broadcast Network

Application	Use	Protocol
Web site access	Can provide access to corporate intranet Web site or public Web sites normally accessed through the Internet	Hypertext Transfer Protocol (HTTP)
E-mail (text)	Standard electronic mail; additional features like voice and video mail may be provided	Simple Mail Transport Protocol (SMTP); custom applications
Streaming audio and video	Broadcast of real-time events (one way)	Real-Time Protocol (RTP)

been around since 1983 but lacked both the throughput afforded by a megabit-per-second data rate and the demand for the service.

The quality of transmission can be very high because of the simplicity of the radio link from the satellite. The error rate can be maintained at less than 10^{-9} using forward error correction, which means that there would be one bit error per gigabit of data received, on average. The basic satellite delivery system operates much in the same manner as the digital DTH systems discussed in Chapters 5 and 6. This suggests that DTH operators are in a position to offer advance data delivery services to PCs at any time they care to enter the market.

Readers are probably wondering how some interactivity can be added. Figure 8.9 offers the simplest and least costly solution—the PSTN. The way we use the PSTN is as a request channel, which means that the amount of information is relatively small. Also, most of the time is taken up with the user either thinking or slowly typing. As a result, there are several good reasons for using the PSTN where it is available:

- Low cost of access, including service charges and the cost of a standard modem (most PC owners have a modem for fax and on-line access);
- Low cost of return data transfer (assuming short requests);
- Relatively low delay or latency for short messages;
- Wide availability in developed countries and major cities in developing countries;
- No transmit license required for the remote site. This could be a big advantage.

Perhaps the most familiar service still available in the United States at the time of this writing is DirecPC, an offering of hardware, software, and satellite service delivery from HNS. As the originator of VSAT technology, HNS was in the best position to bring a high-performance data broadcasting service into being. It implements a 24-Mbps outroute transmission from a Ku-band satellite.

The process to access the public Internet is as follows:

- Dial your Internet access provider (HNS provides this as an option, but there are literally hundreds of possible providers in the United States alone).
- All keyboard and mouse actions are transferred over the terrestrial line.
- All download requests are fulfilled over the DirecPC satellite link instead of the telephone line.

After 2000, DirecPC was subsumed within DIRECWAY (discussed later), the public broadband offering of HNS. The PCI card was replaced by an external modem style of device to allow a broader range of PCs, including laptops, to be used. From this discussion it is easy to see that PC integration is possible and attractive on a standalone basis. The terrestrial return connection provides a request path to assure 100% delivery accuracy of content.

8.2.3.2   Integrating LANs and VSAT Networks

The LAN represents the foundation of nearly every IT strategy because of its versatility and cost-effectiveness. LANs are found in small companies comprising only a few employees and all the way up to the largest organizations engaged in global business. In fact, the LAN environment is now common in many homes. The challenge that IT managers continue to face is how to interconnect LANs into an enterprise data communication system to accommodate a full suite of IT applications mainframes from industry leader IBM. The door is open to VSAT networks, provided they can meet very stringent usability, performance, reliability, and economic objectives.

LAN and WAN systems are heavy users of the Internet series of protocols, that is, TCP/IP. In contrast to proprietary architectures and protocols, like IBM's SNA and Digital Equipment Corporation's DECnet, TCP/IP is completely open and vendor independent. The protocols themselves are in the public domain, and much of the networking software is available free of charge over the Internet or by license for a very low cost. Most users do not employ TCP/IP directly; rather, it is integrated into the actual applications that are sold by leading vendors like Microsoft, IBM/Lotus Development, Novell, Oracle, SAP, PeopleSoft, and Sun Microsystems.

As a review, the fundamental pieces of networking equipment that are used by all organizations to interconnect LANs are the Ethernet switch and the IP router. According to Cisco, the definitions are as follows:

- *Switch:* This is a device that improves network performance by segmenting the network and reducing competition for bandwidth. When a switch port receives data packets, it forwards those packets only to the appropriate port for the intended recipient. This capability further reduces competition for bandwidth between the clients, servers, or workgroups connected to each switch port. The switch operates at the data-link layer of the OSI model.
- *Router:* This is a device that moves data between different network segments and can look into a packet header to determine the best path for the packet to travel. Routers can connect network segments that use different protocols. They also allow all users in a network to share a single connection to the Internet or a WAN. Working fundamentally at the network layer (layer 3), the router reads the destination address from each IP packet and chooses the "best" path among the possible choices available to it (e.g., from the links the it terminates at the particular location). Making this decision depends on the routing table it contains in memory as well as the stored logic it uses. There are a variety of routing protocols (the rules for routing packets) employed in industry today, including Routing Information Protocol (RIP) and the more

advanced Open Shortest Path First (OSPF). Another vital function of a router is to act as the gateway between an intranet and the Internet, and between the Internets operated by different tier-1 ISPs. This involves what is termed a Border-Gateway Protocol, necessary for the proper handing off of traffic across these high-capacity interfaces.

- *Firewall:* This is a computer that stands between the internal LAN and the exterior world, such as the Internet, to prevent undesirable intrusion by third parties and hackers. It may operate at a number of levels of the OSI model, such as the link (MAC address filtering), network (IP address filtering), and application (Web, e-mail, and file control). Firewalls range in compatability, complexity, and strength.

These devices are produced by several suppliers in the United States, particularly Cisco Systems, Nortel Networks, 3Com, Hewlett Packard, and Alcatel. Smaller companies, such as Foundry Networks, Redback Networks, and Linksys produce routers for specialized markets. Users of IT capabilities implement client/server architectures using these devices to build a private WAN or to access much larger networks. Through the use of stored-program processors, the switch and router intelligently transfer information between LANs on more or less a seamless basis. Management of a WAN is provided through a series of standards for network management, including the SNMP. Organizations purchase network management workstations like Hewlett Packard's OpenView and IBM's NetView that use SNMP to gather relevant information and control many of the devices and systems in the WAN environment. Support for SNMP is growing with respect to the mainstream of satellite communication networks.

We first review the application of the switch and router before going into detail on how VSATs support a WAN strategy. The LAN and switch, shown in a typical application in Figure 8.10, provide the most basic connection between Ethernet

Figure 8.10   Ethernet LAN arrangement using a switching hub.

segments at the link layer (layer 2) of the OSI model. Each computer and other intelligent device on the LAN are able to originate Ethernet frames (the term for packets at this layer) that are transferred on a point-to-point basis. The intelligence in the switch reads the link layer address (called the MAC address) to determine the destination. If it is not in a local segment, the frame is switched internally to the appropriate segment. Modern switches are able to filter IP packets (a layer 3 function normally associated with routers) to limit transfer and route IP packets to external networks. Modern LAN switches operate at 10, 100, and 1,000 Mbps.

A typical application of the router is shown in Figure 8.11. While looking a lot like the switch application in Figure 8.10, the important feature of the router is that it can create a versatile WAN environment and interface with the public Internet as well. The router operates at layer 3 of the OSI model, which means that it transfers packets based on the network layer protocol. The selection, configuration, and programming of routers is complex, requiring considerable knowledge of the associated routing protocols that comprise the Internet. The remote LANs and their associated indoor VSAT units fulfill the role of a WAN with a star topology; that is, the satellite network emulates a star with dedicated connections between the switches at the respective LANs.

Transfer of packets from a remote LAN across the VSAT network to a distant LAN is controlled through protocol filtering, a function previously identified for the Ethernet switch. The firewall also functions in this regard and has more intelligence to better protect internal LAN resources. The router can also implement a firewall that prevents local users from making outside connections and conversely protects a given location from intrusion over the VSAT network. Protocol filtering may be enabled or disabled on each direction, independently. Filtering can be based on learned addresses or addresses that are configured (i.e., downloaded by the operator). The question may arise about how this star network might support applications that need communication between pairs of remote LANs, without passing through the host computer. An approach is for the indoor unit at the hub to redirect such packets to the single-destination LAN (not all remote LANs simultaneously). In the typical WAN environment, this is performed by a core router. Hence, the star VSAT network emulates the operation of the router even though this particular equipment need not be provided.

In cases where remote-to-remote traffic is heavy and where the double-hop delay is unacceptable, it is better to implement a mesh topology. This type of

Figure 8.11   Use of a router to connect into a WAN and the Internet.

arrangement, as discussed at the beginning of this chapter, implements direct point-to-point links that do not pass through a hub. Most mesh networks over a satellite employ circuit switching, so the connection must be first requested before it is available for data transfer. The time needed to make the connection is typically of the order of several seconds due to the existence of multiple satellite delays and the necessary processing time to arrange the connection. Once the circuit is established, data can be exchanged with the delay of a single hop.

8.2.3.3   DVB Return Channel by Satellite

While the majority of VSAT networks installed to date have been of a proprietary nature, there is a process underway towards an open standard. The objectives include reducing the manufacturing costs of components through greater production quantities, and achieving a form of interoperability where products from different suppliers can function within the same network. The effort attempts to replace the dominance of proprietary VSAT products and is known as Digital Video Broadcast–Return Channel by Satellite. Now part of the DVB Forum activities, DVB-RCS adds the inbound transmission over the satellite to the popular and effective DVB-S outbound transmission system (reviewed in Chapter 5). This would appear to be a good marriage owing to the success and attractive pricing of DVB-S components and systems.

A number of VSAT manufacturers and component suppliers support DVB-RCS and offer products to the commercial market. EMS is selling systems primarily in North America, while Newtec is concentrating on the European market. The functionality of these networks is very similar to those of the main suppliers: HNS, Gilat, and ViaSat. Chip sets for the VSAT IDU are available from SpaceBridge Semiconductor, based in Gatineau, Quebec, Canada. DVB-RCS relies on the same fundamental air interface concepts as proprietary VSATs. For example, the return channel options include TDMA and multifrequency TDMA (MF-TDMA). The packet structure within the time frame is based on ATM.

8.3   VSATs in Business TV

VSAT networks offer a special capability to introduce business TV into the environment at a relatively low cost. In Chapter 4, we reviewed the role of business TV as a means to tie organizations together. A common application is distance education for internal training and corporate communication. Video teleconferencing requires a two-way interactive link, which is feasible with VSATs that provide sufficient bandwidth in both directions (e.g., greater than 64 bps). Otherwise, only the audio is bidirectional along with the outbound video broadcast.

Most business TV installations have been additions to VSAT data networks, making use of existing antenna installations. However, there have been a few networks installed for business TV activity alone. Typically, the demand for it has come from executive management to improve control. This does not mean that all business video applications are not economically justified. It will depend on the manner in which the network is implemented and operated. Once the network is installed,

however, users become very cost conscious. This has given rise to the business TV service provider—one of the largest is Convergent Media of Atlanta, Georgia.

8.3.1   Video Teleconferencing

The medium provides duplex video transmission, so the groups get the impression of a face-to-face meeting in real time. The duplex link is a digitally compressed video channel using multiples of 64 Kbps (termed P*64) or switched 64-Kbps service. The first VTC meeting rooms in the United States were specially designed and cost upwards of $1 million. By 1995, equipment was housed in roll-around cabinets that allow any conference room to support a conference. VTC products available at the time of this writing are of the monitor-top variety, housing the camera and codec. With the work going on with the MPEG 4 standard, it is entirely possible that VSAT star networks might become capable of allowing two-way VTCs to be conducted without reverting to a mesh topology. An example of a VTC linkage is shown in Figure 4.8.

An adaptation of VTC is being applied to the field of medicine as a means to assist physicians in the field who need expert advice from specialists. This field of telemedicine employs a variety of telecommunication technologies to assist in the proper diagnosis and treatment of victims of accidents and rare sicknesses who otherwise would have had to be transported to a major city where the specialists reside. Paramedics in the United States may already practice telemedicine to the extent that they transmit EKGs from the scene of the accident to a hospital. This has been extended to allow a patient in Saudi Arabia to employ the expertise of a specialist of Columbia Presbyterian Hospital in New York.

Adding VTC allows the specialist to actually see the patient and speak to the on-site physician. The reverse direction is important so that the physician can better interact with the expert and also view medical image data after evaluation or processing. Hughes Electronics and Telecomm Mexico cooperated in a telemedicine trial using VTC between a clinic in Chiapas and a hospital in Mexico City. This allowed the local physician to work with the specialist of his own nationality who was located at the best medical facility in their country. While telemedicine always seems to impress those who experiment with it, the lack of adequate telecommunication facilities remains a roadblock to extensive use. This is an area where VSATs can fill the gap once the applications are refined to the point where a cost-effective architecture can be introduced to the medical community.

8.3.2   Private Broadcasting

This is a private satellite video distribution network, constructed much like that used in entertainment TV. Many of these networks were created to allow the CEO to talk to organization members and to hold nationwide press conferences. More important for the business, some companies have found ways to use their VSATs to deliver specialized broadcasts that contribute to corporate effectiveness. Examples include retailers (JC Penney), fast food servers (Domino's), computer manufacturers (HP), and automobile manufacturers (GM and Ford).

An added feature is audio return using the satellite inbound link. This works similar to audience call-ins for radio and TV talk shows, but the line may stay up for the duration of a teleconference. This author participated in a private nationwide panel discussion with four other industry executives. We assembled in the studio of the Prudential Insurance Company in New York and were uplinked to Westar IV. As many as 12 sites around the country received the signal for viewing at coordinated meetings of the Society of Satellite Professionals International (SSPI). After each speaker made his or her presentation, the floor was opened up for questions. People at the remote locations were in contact through the telephone network (the satellite network did not include return links).

One of the most ambitious and extensive private broadcasting networks is that operated by GM University. The background of GMU is summarized as follows [11]:

- Formed in 1997, GMU is a global network of education and training resources designed to help GM's executive, management, technical, and professional employees continuously improve their competitive performance to conduct and grow the business of General Motors. It is one of the world's largest universities for professional, supervisory, and salaried employees.
- GMU currently has 15 functional colleges charged with developing curricula tailored to the professional needs and unique challenges facing employees from a business sector, divisional, or regional perspective.
- The colleges are as follows: Communications, Engineering, Finance, Health and Safety, Human Resources, Information Systems and Services, John G. Smale Brand Management, Leadership, Legal, Manufacturing, Planning, Public Policy, Quality, Worldwide Purchasing, and Sales, Service and Parts.
- GMU focuses on providing real-world learning that enhances performance and drives business results.
- Course delivery is through classrooms, learning laboratories, satellite, electronic, and Web-based applications.
- GMU offers approximately 1,300 courses to GM's 88,000 managerial, executive, professional, and technical employees.
- In 2000, GMU provided approximately 230,000 student days of learning.
- There are nearly 1 million Web hits monthly to GMU On-line that result in nearly 70,000 Web enrollments per year. The Web site also allows users to create Individual Development Plans, track individual training histories, and use Web-based tools to align their training with functional development strategies.
- GMU's Interactive Distance Learning System has 10 broadcast studios that support a satellite delivery infrastructure across more than 7,300 dealer and corporate sites. The system provides more "live" broadcast hours of programming than any private broadcast network.
- By concentrating on e-learning and expanding its global offerings, GMU intends to make learning available on demand, anytime and anywhere.

GM's principal U.S. competitor, Ford Motor Company, developed a similar network: FORDSTAR. In the early 1990s, Ford realized that it needed to upgrade

its workforce to enable it to deal with the increasingly computerized nature of its cars [12]. The workforce of 350,000 is scattered throughout 200 countries. The company found that the literacy skills of the workforce were not up to the requirements of print-based training, and its independent dealership network was averse to training. Ford decided to overcome these barriers by video training that would piggyback on an existing but under-used satellite network that was used for data transmission to the dealerships. Ford invested $100 million to create FORDSTAR, its educational arm; this did not include the cost of curriculum development or the $7,000 each dealer had to invest in a satellite receive-dish and recording equipment. Unlike most universities, Ford was in a position to compel training of its employees and dealers. Today, FORDSTAR operates on $440 million per year (2.5% of its payroll) to train its engineers, repair technicians, managers, and sales staff in the dealerships. Ford leveraged the satellite investment to provide a value-added service. Developers of such networks might consider the following questions:

- What resources exist on which a virtual education program might piggyback?
- Do any multinational companies have unused delivery capacity that they might contribute to the cause?
- Are there training resources that might be expanded to become a backbone of a virtual education program?
- Are there other existing resources that could be leveraged to give the virtual education program a boost?

Determining those needs is greatly facilitated by the primary distance learning response mechanism, the One Touch keypad on each student's desk [13]. While each classroom learning station has a microphone, it is the keypad that prevents cacophony, making it practical to conduct classes that often number hundreds of students at dozens of sites. Instructors can ask true/false and multiple-choice questions and instantly see the results. Because the keypads are addressable, the instructor can pinpoint students who are having trouble, as well as gauge the class's overall level of comprehension. At the push of a button, students can also "raise their hands" to ask or answer a question over their microphones. Even better for some students, another button allows them to anonymously signal the instructor that they're confused. This multifaceted interactivity is potentially more effective than face-to-face classes at assuring that the instructional goals are met by all students.

These applications are only a sample of what is possible, based on what is actually going on in industry. One or more of these can be combined and implemented to create a unique capability that could mean success in a business context. This involves experimentation by the organizations, perhaps starting out with one or two applications that justify the value of installing the VSAT network. Then, more applications can be introduced as the original applications are modified or replaced. The idea is to continue to innovate.

References

[1]   Stallings, W., *Business Data Communications*, 4th ed., Upper Saddle River, NJ: Prentice Hall, 2001.

[2]   Naugle, M., *Illustrated TCP/IP—A Graphic Guide to the Protocol Suite*, New York: John Wiley & Sons, 1999.

[3]   Wilder, F., *A Guide to the TCP/IP Protocol Suite*, 2nd ed., Norwood, MA: Artech House, 2002.

[4]   Newton, H., *Newton's Telecom Dictionary*, 18th ed., New York: Telecom Books, 2002.

[5]   Elbert, B. R., *Private Telecommunication Networks*, Norwood, MA: Artech House, 1989, p. 235.

[6]   SMS QPSK/FDMA System Specification, EESS 501, Issue 2 - REV. 0, Eutelsat, Paris, 1996.

[7]   "Case Study: Real-Time Information Gives Real Competitive Edge to South Africa's JD Group," *Global VSAT Business – Emerging Markets*, Global VSAT Forum and Satellite Business International, p. 36.

[8]   "Case Study: United States Postal Service (USPS)," http://www.gilat.com/Solutions_CaseStudies_Usps.asp.

[9]   Cope, J., "For Disaster Recovery, Put Your IT Eggs in Different Baskets," *Computerworld*, July 15, 2002, http://www.computerworld.com/securitytopics/security/story/0,10801, 72638,00.html.

[10] "Case Study: Six Continents Hotels/Holiday Inn," http://www.gilat.com/Solutions_CaseStudies_BassHotels.asp.

[11] Life@GM, GM University, http://www.gm.com/company/careers/life/lif_gmu.html.

[12] Dirr, P. J., "The Development of New Organizational Arrangements in Virtual Learning," *The Changing Faces of Virtual Education,* The Commonwealth of Learning, 2001.

[13] Gach, G., "Fordstar: A Guide to the Future for Ford," *Envision*, Convergent Media Systems Corporation, Summer 1995.

Technical Aspects of VSAT Networks

This chapter focuses on the design and operation of VSAT networks that provide digital services, particularly data communications, through a star topology. While the star is the most popular (there being approximately 600,000 of this type of VSAT in operation worldwide at the time of this writing), another segment of the market provides mesh capability. Design principles for VSAT networks build upon those of telecommunications systems in general and data networks in particular. Unlike telephone networks, which tend to follow traditional models like the Erlang B formula reviewed in Chapter 10, data networks are extremely variable as to the specific types of information that are being processed and transferred. As a result, the analysis only approximates what is going to happen within the network once it is turned over to a user community. A good example of this is where VSATs are used to provide Internet access to small offices and remote offices (SOHO) because no two users will employ their transmission path in the same manner. Some types of information access and transfer include:

- Using a Web search engine such as Google to find Web sites of interest and value to the user;
- Clicking on the URL link within a Web site to download a Web page in HTML;
- Downloading and uploading e-mail messages using the Web browser or a separate e-mail package such as Microsoft Outlook or Eudora from Qualcomm. E-mail messages often include attachments, such as text documents, photo images, and compressed files like programs, that themselves can be 1 MB or more;
- Meetings conducted with a presentation tool such as PlaceWare (now part of Microsoft) that allows participants to see the same PowerPoint slides and exchange text and notes in near real time;
- Access to documents on a server to facilitate remote collaboration by members of a team;
- Use of a real-time Internet application like videoconferencing or VoIP to engage in a discussion or meeting. The latency of the satellite link combined with that of the Internet can result in some user dissatisfaction; therefore, it is wise to conduct tests to assure that the service is usable.

From the above list it should be clear that Internet access is a very demanding application and generally pushes the bandwidth required. The dynamic nature of these usage patterns also requires that the VSAT network have a fast response to

make users happy. This is in contrast to VSAT networks from the 1980s pioneered by HNS that were designed for a common service, such as credit card verification or automated teller machine transactions. In these, the network would service a standard message format and it was only a question of allowing for the maximum loading—such as Christmas season in North America for credit card usage or the last day of the month for banking.

Other applications of VSAT networks are in the media field, to deliver and exchange content in various forms. This is expanding because terrestrial broadcasters in television and radio expand markets and respond to competing systems such as DTH and DARS. With the satellite point-to-multipoint broadcast connectivity common in television and radio, the inbound return channel is a straightforward way to add interactivity that is independent of terrestrial means. The reason for doing this is that many cities and towns, which are the broadcast outlet for the network, do not have broadband services available to everyone. For example, this author was contacted by a U.S.-based radio broadcaster who had purchased radio stations in American Samoa. How, then, would they deliver their program out in the islands and obtain a dependable link back for station management purposes? A VSAT network was their best option.

Regarding the mesh alternative, there are pure TDMA VSAT products on the market that allow any station to have a single-hop connection to any other. Unlike the star, there is no hub station with a large antenna to offset the small dish size and transmit power of low-end VSAT terminals. Instead, each terminal has the capability to reach any other at the full data rate, which is in the range of 256 Kbps to 10 Mbps. This impressive capability is sufficient to support a full channel of commercial quality MPEG 2 video along with audio and other data. This scheme was chosen by NBC in the United States using SkyWAN equipment from ND SatCom to update the news contribution service network, where stations can send content either to an NBC studio on the East or West Coast, or among each other.

This chapter provides the basics of network design and an introduction to the difficult task of network sizing. Readers should keep in mind that this discussion provides the conceptual framework and that VSAT network sizing is more art than science. In the end, the designer makes a best estimate of transmission rates, capacity, and other aspects; then, the network is started up so that operating data may be collected. From this, the parameters can be adjusted to match what is actually going on. The challenge comes when usage patterns are altered—or new applications are added—producing changes that cannot easily be anticipated. Hopefully, the VSAT network has excess capacity or the ability to grow modularly so that these dynamics can be accommodated with only minor pain and inconvenience.

9.1 Capacity Planning and Sizing

The greatest challenge to developers of VSATs was to build versatile and reliable data communication systems that could compete with existing terrestrial technology. Consumer broadband Internet access (what is generally meant by "broadband") is a reality for more than 40 million homes in developed countries, but pushing this out in a global context has proven to be a challenge. This has been

referred to as the Digital Divide, a popular phrase to indicate the challenge of making broadband Internet access reach people and places that would benefit greatly from it [1]. Implementers of VSAT networks have their own challenges to face, that of properly sizing the network and maintaining its performance as requirements change. The purpose of this section is to provide some guidelines and examples of how to approach the problem of capacity planning and network sizing.

9.1.1   Collecting Requirements for the VSAT Network

We are taking the systems approach to this problem, which consists of determining the requirements, sizing the network, and determining overall performance against the requirements. The typical requirements for a data communication network fall into the following categories.

1. The applications that the network is to support, in specific terms such as software, user devices, message content, and standards to be employed.
2. The number of users and locations that are to be serviced, which defines the geographic properties (physical distribution) and the associations (linkages) of the requirements. The output of this part of the exercise is the topology of the network.
3. The traffic or information volume that is offered to the network by each user or by an expected volume of users considering the particular amount of information as well as the timing of its occurrence (also referred to as its temporal nature).
4. Throughput is the amount of useful data that is transferred per unit of time, measured in bits per second, packets per second, or the like. This provides an estimate of the aggregate bandwidth required from the VSATs, hub, and satellite.
5. Time delay (also called latency) is the amount of time required to transfer a specific amount of data. Applications should be subdivided as to their fundamental needs in this area—real-time applications versus non-real-time applications. Latency is composed of line transfer time, access protocol time, propagation time, and node processing time. Interestingly, satellite system users focus on propagation time, which is only one of the factors. If this is a circuit-switched type of network, then the call setup time must also be considered. All of these times are the sum of a number of contributors, some of which are constant and some of which are variable or random in nature.
6. Response time is measured from when a user initiates a request to when the response is displayed on the user's terminal device. This applies mostly to data communication networks where users employ PCs or other types of display terminals, phones, and video systems.
7. Other service demands that are unique to the user community, such as mobility and transportability, growth, and ability to support new and evolving applications.
8. Service management aspects, such as reliability, availability, mean time to repair (MTTR), and help-desk and accounting support.

Each requirement has a definite impact on the capacity, complexity, and cost of the network, VSAT or otherwise. Requirements 1, 2, and 3 are the basic inputs to the network design, defining the structure, distribution, and timing of the information to be carried and/or distributed among users. Believe it or not, this is the hardest thing to determine because in most situations it is simply not known with any precision. Another real but less tractable reason is the internal structure of the organization where functional groups tend to not want to share this kind of information. Breaking down these barriers can be extremely time consuming and this often is a reason why the network fails to satisfy users who were unable to have their particular needs reflected in the design of the network.

Requirements 4, 5, and 6 are performance measures that can be obtained from an operating network. User satisfaction is often driven by these measures, which is why it is important that they be learned before a major commitment is made to a particular technology. In the absence of a working network, tests with a pilot system can gain insight into what users like or can live with. It turns out that the user experience and level of satisfaction is based on what they already know from using other networks and services. If their expectation is low, then any improvement will be viewed favorably. The worst situation occurs when the user comes in having experienced something a lot better, either because of their former environment or because they had absolutely no basis of comparison. Any effort to shed light on this area is going to be well worth it.

Throughput (requirement 4) measures the actual quantity of data flowing between users over the network. What we are observing here is the user data that reaches the end of the link for the particular computer application. In the process of carrying data, the network introduces overhead information that is needed to control and manage network operation and is not part of throughput. Another nonthroughput contributor is data that must be retransmitted due to congestion or errors on particular links. For these reasons, the throughput rate offered by a vendor's VSAT product may in fact be overstated by a considerable margin, as much as 100%. Since it is difficult to sort out overhead, the best way to know the actual throughput is to perform a live test.

The logical way to design a data network is to begin with the requirements for data transfer among the various users and points of operation. If you are starting with an existing star network, there would already be a working host computer and many remote users. Typically, this would use terrestrial leased lines that connect from the remote sites back to the host. The data requirements can be determined at the host where all of the data can be observed for each remote and each leased line connection. Statistics on the information packets, their delays, and the overall response times are collected using monitoring software that is loaded into the host.

In a peer-to-peer or client/server environment, the network is far less structured and there is no single point of concentration where data characteristics can be gathered. Remote hosts and servers allow some of this information to be collected, as can monitoring devices on LAN segments and routers used to interconnect LANs. This information may be available at a network management station that employs SNMP. If necessary, technicians can make manual measurements using a protocol analyzer such as the Sniffer by Network General or the Lanalyzer by Novell. Information such as quantity of packets and their lengths, throughput, retransmissions,

number of virtual circuits, and transfer delays are definitely of value in sizing the VSAT network. It is likely, however, that this information will only address part of the requirement because there will be new applications and computer systems. This brings us to the rather rough estimating process outlined in the next section.

Service demands and management (requirements 7 and 8) may represent the difference between success and failure in a real implementation. Engineers who focus on technical details can satisfy themselves that the network will, in theory, meet the need. Once deployed, the network will be subjected to a variety of stresses in the environment:

- The users themselves, who cause the demand to rise and fall in unpredictable patterns; their actions can cause hiccups and failures due to bugs in software or situations that the developer did not expect;
- The physical environment, in terms of location, weather, and distance to be covered;
- Availability of support for remote locations too far for repair people to visit;
- Lack of extra components and test facilities that allow the operator to identify and correct problems quickly and with minimum disruption; this is the result of the extended distances and remoteness that itself is justification for using VSATs in the first place.

Developers of VSAT networks must assess and track all of these factors if they wish to succeed beyond seeing the system enter service. Experience is the best teacher when it comes to VSAT network operation and extension. Our advice is to begin the design effort with a research project where requirements are collected, categorized, and refined. Section 9.2 provides further details and examples of the requirements gathering process. Three months dedicated to this on the front end is the best insurance policy we know.

9.1.2 Estimating Delay and Response Time

Delay and response time (requirements 5 and 6) are quality measures that can have a direct impact on user acceptance. Values of acceptable time delay and response time must be established ahead of time so that the network components can be sized properly. Figure 9.1 shows the major elements of the link between user terminal and server computer over a VSAT star network. This is at a high level and should not be regarded as a specific design (we provide such an example later in this chapter). We can see that there are several contributors and each adds to the delay as experienced by the user of the PC client workstation. At the start of an exchange between client and server, the user enters data and hits the return key. This causes a block or frame of data to be applied to the local access line between the PC and the VSAT indoor unit. The speed of this line determines the time delay for transfer of the entire packet (we can ignore propagation time since the distance is usually less than 1 km). Hence, the data is repackaged within the indoor unit and formatted for the satellite return link. VSATs may initiate an exchange of data using a fast access technique like ALOHA or CDMA; subsequent transfers of either large blocks or files or a real-time information transfer could be accomplished with continuous transmission (SCPC)

Figure 9.1   Major contributors to delay and response time in a VSAT star network.

or TDMA. The amount of additional delay will depend on the access technique, channel loading, and need for retransmission (a consequence of using either ALOHA or CDMA). Access protocol performance is covered in detail in the next section.

The uplink and downlink introduce nearly constant propagation delay, amounting to approximately 260 ms total (for a single hop). The delay of the satellite repeater is typically less than 1 ms and therefore can be ignored. An exception would be an advanced processing type of repeater that requires a certain amount of time to route the packets. This is specified by the satellite manufacturer and would be expected to be a constant number much less than the propagation delay. The next significant contributor to delay is the processing within the hub baseband equipment, resulting in a restored block of data at the local access line. Most hub installations multiplex several streams of such data together, a process that can add a small but measurable delay. Finally, there is the transfer and propagation time associated with any backhaul circuit between the hub and the server. Additional delay is introduced by the server (including its operating system, file management system, and application processing).

Processing time within the host is not supposed to be counted as part of the response time of the network, but it is often added since it will, of course, be experienced by the user at the PC client. A well-engineered VSAT network will fail if the server cannot process the required number of user actions. For that reason, it is essential that the performance of the server by itself be budgeted and verified before connecting to the satellite network. This establishes the baseline with which to certify the overall service on an end-to-end basis. The same can be said of additional network connections that data must be negotiated before reaching either the end user or the server.

An example of a typical time budget for this end-to-end connection is provided in Table 9.1. We have made the assumption that the user wishes to transfer a block of 10 KB, which is 80,000 bits. Assumed line transmission speeds and processing delays are indicated in the table. In this example, we looked at one direction only; also, there are other contributors, such as an end-device queuing delay, that add to the total. There would be a separate estimate for the forward direction compiled in the same manner. This estimate shows that the access-line delay contribution (item 2) occurs only once and is determined by the speed of the link (256 Kbps at the original access point, producing a delay equal to $80,000/256,000 = 312.5$ ms). The uplink/downlink

Table 9.1   An Example of a Delay and Response Time Budget for a PC-to-Server Application over a VSAT Network, Assuming Transfer of a Block Containing 10 KB (Including Overhead)

Item	Element	Line Speed (Kbps)	Intrinsic Delay to Transfer Block (ms)	Propagation Delay (ms)	Actual Contribution (ms)
1	PC client workstation	—	0	—	0
2	Access line	256	312.5	—	312.5
3	VSAT indoor unit, inbound direction	—	10.0	—	10.0
4	Multiple access protocol (TDMA)	—	25.0	—	25.0
5	Uplink propagation	—	—	135.0	135.0
6	Satellite (bent pipe)	—	—	0.0001	0.0001
7	Downlink propagation	—	—	130.0	130.0
8	Hub baseband equipment	—	50.0	—	50.0
9	Backhaul data circuit (hub to server)	2,048	39.1	3.0	42.1
10	Server processing	—	100	—	100
	TOTAL – one direction	—	—	—	804.6

contributes a total of 265 ms of propagation delay, to which is added only 3 ms for the backhaul circuit (assumed to be 750 km in length). Processing delays for the equipment in the system are estimates but can be measured on real devices to improve accuracy. The multiple access delay of 25 ms is based on a TDMA frame length of 50 ms, dividing it in half to represent the average holding time before transmitting a burst. Finally, we have assumed a host processing time of 100 ms, which could be an underestimate in the case of a heavily loaded computer system.

The estimation process shows that there are many contributors to total delay of about 0.8 second, of which satellite path propagation is 33%. If we merely doubled data communications delay to account for the round-trip situation, then the response time is 1.6 seconds. This would satisfy a requirement of 2 seconds, which is a typical value used in industry.

9.1.3   VSAT Access Protocols

Capacity estimation in a VSAT network is a complex problem driven by the fact that many of the key variables are not well known ahead of time. Also, even if you have a good estimate of the required number of users, their locations, and the data throughput needs, there is still the problem of considering how these users will interact with each other and with the system at any given time. We typically work with an average of some sort and allow for a peak during the busy hour or seasonal peak of activity (like just before Christmas in the retailing industry or summer in air travel and hotels). But this does not consider the statistical peaking that occurs from an instantaneous heavy load after some unexpected event or if some particular users activate an application that produces a surge in the network. As illustrated in Figure 9.2, the network loading has statistical peaks and an overall average during any particular period (1 week is shown). A network sized for the average load will work during off-

Figure 9.2   Illustration of data traffic load on the VSAT network, assuming random activity of users and applications.

peak periods. However, the noise-like peaks could overload any of a number of potential choke points—some within the control of the VSAT network designer and some outside of it—resulting in greatly increased latency and even dropped data communication sessions. As in any computer system, this can only be dealt with by allowing some headroom in terms of bandwidth or processing capacity.

What we have are some basic methods of estimating the load, and from that, the capacity. This should be taken as a starting point for sizing and not used as if it was an accurate prediction. Fortunately, VSAT networks have a number of built-in features that tend to forgive estimating errors. For example, the overall capacity of the network is determined by the amount of satellite bandwidth that is both allocated to the network and that can be accessed by the Earth stations. The fact that this pool of capacity is available to all means that any given user has potentially a large reserve that can be utilized under a critical load. This is constrained by the equipment that is available to the user, such as the bandwidth of the particular uplink chain through which the data must flow to reach the hub. The hub, on the other hand, has a much larger reserve of bandwidth, but this too is not unlimited, as we shall see.

The basic technology of the VSAT star network exists on a foundation of many years of experience with digital satellite communication. We use a large hub Earth station to receive all direct transmissions from VSATs, which are sent in packets. The hub transmits one or more continuous streams as a data broadcast to the entire network—the DVB-S standard is favored for this portion. Individual VSATs select packets addressed to them from the broadcast. A group of VSATs transmits information to the satellite on the return link. The use of burst transmission in the inbound channel and broadcast TDM for the outbound channel is illustrated in Figure 9.3; the outbound channel is shown as a continuous data stream, and the inbound channel is discontinuous, allowing individual VSATs to timeshare the particular narrowband channel. Two other inbound channel approaches, CDMA and FDMA, are discussed in Sections 9.1.3.3 and 9.1.3.4, respectively. TDMA return channel provides maximum throughput for a constant bit rate (e.g., voice and video). FDMA likewise supports constant bit rate services but requires that inbound channels operate on separate frequencies. CDMA and ALOHA, on the other hand

Figure 9.3 Typical arrangements of outbound and inbound transmissions in a star VSAT network. The outbound carrier is broadcast from the hub to all remote VSATs; inbound transmissions from VSATs employ one of a number of multiple access techniques: ALOHA, TDMA, FDMA, or CDMA.

are designed for short and infrequent data transmissions used to request constant bit rate service (reservation mode) or to transfer a single packet.

VSAT and other satellite data networks make use of ARQ protocols that operate on top of the multiple access system. As discussed in Chapter 8, this technique is central to reliable data transfer in standard protocols like TCP/IP. In the case of satellite networks, the parameters are adjusted so that the round-trip propagation delay of about 0.5 second does not cause the throughput to reduce to zero. We employ a block-oriented protocol that allows the receiving end to detect errors and limit the requests for retransmission to only those blocks that contain errors. Another name for this technique is "look back N" because the receiving end checks for bad blocks and tells the sending end to look back in the frame and only resend the incorrectly received block. Detection of such a bad block is accomplished using an error-checking technique like CRC [2], where the originating end appends a calculated binary number called the frame check sequence to each block. This sequence is computed using a special algorithm that assures a high degree of uniqueness relative to the original block. The receiving end likewise makes the computation on the block and checks that the appended frame check sequence is the same. A difference indicates that the block has errors within it, resulting in the aforementioned request for retransmission. The power of the CRC approach has made it the standard operating procedure for computer memory and storage systems, data communication networks, and VSAT systems as well.

9.1.3.1 Operation of the TDMA Access Protocol

The inbound channel is shared by multiple VSATs that transmit their data in bursts. TDMA is one of two basic multiple access methods used for this purpose. An

example of a TDMA burst time frame lasting about 45 ms is provided in Figure 9.4. As applied to the inbound channel, the transmissions from the VSATs are coordinated and highly synchronized so as to prevent overlap and a resulting loss of information. The master frame in this example is 45 ms, which is a multiple of 360 times the frame rate requirement for telephone service. Each station (numbered 1 through 10) is allotted a fixed interval of time in which to transmit data. The frame repeats every 45 ms, producing an average delay per inbound channel burst due to multiple access of 45/2 = 22.5 ms. Obviously, the shorter the frame, the less the average delay. This also reduces the amount of storage memory needed in the VSAT and hub to accumulate data and voice samples for inclusion in the next burst. The start of the frame may be identified by a sync burst that is transmitted by a master station for reference. In the actual case of the VSAT inbound channel, no sync burst is needed because the hub controls transmission via the outbound channel. Additional delay and loss of throughput results from the addition of the preamble, as shown at the bottom of Figure 9.4. However, this overhead is needed to permit the hub to lock onto the incoming burst, restore the bit stream, identify the source VSAT, and subsequently recover the original data packets. The only loss of capacity relative to a single continuous carrier is due to guard times and overhead.

9.1.3.2 Operation of the ALOHA Access Protocol

Another approach for separating the inbound channel transmissions in time is the ALOHA protocol. The scheme is simpler in that the transmissions are uncoordinated; however, the complexity occurs because there are occasional overlaps that result in lost communication. This is overcome by retransmissions from the affected

Figure 9.4 An example of a TDMA time frame format, indicating burst transmissions from 10 different VSATs on the same frequency. Stations synchronize their bursts to a reference transmitted by the hub (which may in fact be sent over the outbound broadcast carrier rather than within the TDMA time frame).

VSATs. For example, a slotted ALOHA channel with three users is shown in Figure 9.5. Slotting refers to requiring that the ALOHA packets fall within timed periods, indicated by the vertical lines. The upper three horizontal lines represent three VSAT uplinks; the bottom timeline depicts the downlink showing how the ALOHA packets appear after passing through the satellite repeater. Each VSAT remains in an idle state until there is data to be transmitted. In this example, VSAT 1 is the first to need the channel and so transmits the block of data without waiting. VSAT 2 transmits next, independently of what happens at users 1 and 3. From the downlink timeline, we see that VSAT 1 and VSAT 2 do not overlap and hence get through in the clear.

The next packets from VSATs 1 and 3 have reached the satellite at approximately the same time and so have produced a collision. In the event of such a time overlap, the signals jam each other and the information is lost (indicated by the presence of a dark block in the downlink). Neither packet is received at the hub—a condition that is inferred by these VSATs because of nonacknowledgment by the hub over the outbound channel. The way that packets are ultimately transferred is through automatic retransmissions, as shown at the ends of the curved arrows in Figure 9.5. The delay between the original and retransmitted packets is selected randomly by each VSAT to reduce the possibility of a second collision. The result of this protocol is that the delay is as small as it can possibly be for a packet that does not experience a collision. For one that does, the delay is lengthy since it includes at least two round-trip delays plus the delays of the random offset as well as from processing within the hub and VSAT. In an acceptable operating situation, only 1 in 10 ALOHA packets will experience a collision.

The vertical dashed lines in Figure 9.5 represent a timing reference provided by the hub and used by the VSATs to align the bursts. Referred to as slotted ALOHA,

Figure 9.5   The operation of the slotted ALOHA channel with three VSATs. The collision occurs when packets from two users overlap at the satellite. The vertical dashed lines represent reference times to reduce the potential collision to times of 100% coincidence of VSAT transmission.

inbound transmissions can only collide if they happened to select the same precise slot. Unslotted ALOHA does not employ this timing reference, as stations transmit at any time in a completely random manner. As a result, any amount of overlap—no matter how small—will corrupt both packets. Slotted ALOHA, therefore, has a significantly higher (factor of two) throughput than the unslotted flavor. We will see later that a CDMA technique (called Spread ALOHA by Aloha Networks, a provider) counters this limitation on unslotted ALOHA due to the added isolation provided by the use of the spreading code.

The bandwidth of the inbound channel provides the most fundamental limit of capacity at the VSAT level. If we could perfectly allocate capacity to a group of VSATs, then a 512-Kbps inbound channel could serve any number of VSATs as long as the total throughput does not exceed 512 Kbps. TDMA offers the best prospect for high throughput and predictable delay because the bursts are ordered in time. Still, there must be guard time between bursts to allow stations to keep their burst separate. This fixed guard time causes the throughput to decrease more or less proportionally to the number of bursts (stations) on the same frequency. In TDMA, we do not expect to exceed about 90% utilization, which gives a throughput of 460 Kbps. If we estimate that each VSAT needs to transmit about 56 Kbps on the average, then this inbound channel could support eight VSATs. The fact that we have eight independent users on the same frequency means that we can probably allow a significant amount of occasional peaking by individual stations since the total bandwidth can be shared dynamically (e.g., when one user is peaking some others are either idle or at least operating below their average throughput).

The situation with slotted ALOHA is quite different because the bursts are not completely coordinated. In any multiple access scheme where there is no such coordination, the throughput will be substantially reduced. The reason for this is that conflicts will occur on a statistical basis, as we see in Figure 9.5. Since the originating VSAT users are operating independently of each other, combining their activity into the same channel will follow what is called a Poisson arrival process. As defined in queuing theory, a Poisson process is one in which the occurrence of an event (e.g., the initiation of an inbound packet at the VSAT) is defined by a probability density function which is exponential; for example,

$$f_\tau(\tau) = \lambda e^{-\lambda \tau}$$

where τ is the time interval between arrival events and λ is the average arrival rate. The general statement behind this simple probabilistic model is that the longer the instantaneous interval, the higher the probability a packet will arrive (the longer the interpacket wait, the more likely one will come). Figure 9.6 illustrates some of the key aspects of this simple queuing model. The rate at which packets are cleared by the VSAT in the inbound direction is a vital factor. If the rate of forwarding packets, μ, is much greater than the arrival rate, λ, things work smoothly and congestion is avoided. On the other hand, as arrival rate approaches the forwarding rate, the VSAT becomes a point of congestion as packets line up (queue) for action. At the point where they are equal, the queue becomes unbounded and transmission effectively comes to a halt. Two different curves are shown to demonstrate that performance is degraded further if packet lengths and arrivals are random as opposed to fixed (deterministic).

Figure 9.6 Performance of a simple queuing model, based on the Poisson probability density function.

9.1.3.3 Operation of the CDMA Access Protocol

As an access protocol, CDMA is a relative newcomer to the VSAT scene. The strong play made by QualComm with its CDMA cellular standard and evolving 3G technology has greatly enhanced the position CDMA holds on a global basis. However, the benefits of CDMA for GEO satellite applications are considerably less and hence it has not played a major role to date. This may change to some extent due to some special features of CDMA in a random access scenario like that described for ALOHA in the previous section. As reviewed in Section 2.3.3, CDMA allows multiple stations to transmit at the same time and on the same frequency. The bandwidth of the CDMA carrier is equal to N times the information bandwidth, where N is the ratio of chip (spreading sequence) rate to input data rate. In very robust CDMA systems, N would normally be in the range of 100 to 1,000 to gain maximum interference rejection and the greatest capacity. However, the version of CDMA to be applied in VSATs, where bandwidth is at a premium, is typically less than 100.

An illustration of how CDMA would improve inbound access as compared to ALOHA is shown in Figure 9.7. This is based on information provided by Aloha Networks, a company founded by Dr. Norman Abramson, inventor of ALOHA. The upper illustration shows what happens with conventional ALOHA (unslotted in this example) when attempting a throughput exceeding 50%. Note that every packet experiences a collision, with a result that none actually get through. The actual throughput is therefore 0%; along with this, the access delay goes to infinity (because of the need for continuous retransmissions and subsequent collisions). This is what mathematicians call a pathological case. Contrast this with CDMA (called Spread ALOHA by the supplier): The bursts overlap as in the previous ALOHA case but there are very few collisions (none in this illustration). The inbound channel transmissions use the same repeating chip sequence, but they are not aligned in time. A collision is still possible, and will in fact occur, when two VSATs happen to have aligned chips (because their transmissions reached the transponder at the same instant or the chip sequences are aligned even though one started before the other). If, for example, the chip sequence is 50 bits long, the access channel could support

Figure 9.7 Application of CDMA in spread ALOHA to inbound multiple access. Narrowband ALOHA cannot support high throughput, as illustrated in the upper timeline. Overlap in time, as shown for Spread ALOHA does not result in collision unless chip sequence aligns in time. (Courtesy of Aloha Networks.)

50 stations. In this instance, true collisions, requiring retransmission, will only occur 2% of the time (1 out of 50 attempts).

For this illustration to be realistic, it is necessary that the bursts be short as in the previous ALOHA example. This type of service works for short packet applications using a protocol such as UDP where a connection is not established and a constant bit rate is not involved. It can be used to request a synchronous channel of transmission that subsequently uses FDMA or TDMA, a feature that could be introduced in the near future.

Another supplier of this technology is ViaSat in their ArcLight product line. The term used for access is code reuse multiple access (CRMA), which identifies the scheme as using a common chip for all stations. ArcLight is also intended for networks where the inbound channel supports brief transmissions for request purposes or to transfer short packets. ArcLight also makes use of a bandwidth-efficient technology for the outbound broadcast carrier called paired carrier multiple access (PCMA). The object of PCMA is to allow the outbound carrier to occupy the same bandwidth as the inbound CRMA carriers. The hub performs an RF cancellation on the receive side to remove its own carrier from the downlink spectrum. The cancellation technique is similar to what is used within an echo canceler, and it only works at the hub, which can sample its uplink and then subtract itself in the downlink. Some residual carrier and distortion remain, reducing the power efficiency to some extent. ViaSat claims this can be countered through better forward error correction.

9.1.3.4   FDMA as an Access Protocol

FDMA, and more specifically SCPC, have long been used in preassigned and demand-assigned networks of multiple Earth stations in point-to-point and mesh topologies. At least one company, Shiron Satellite Communications, offers an SCPC star system for broadband IP services. Unlike the other access techniques, the Shiron FDMA approach provides a continuous data transfer at the full bandwidth of the carrier. In addition, the carrier is automatically activated and adjusted in data rate in

response to the needs of the locally connected devices. The range of bandwidth available is between 64 Kbps and 2 Mbps, with the power of the carrier automatically adjusted to match the bandwidth and link fade conditions. These features pertain to the inbound channel only; the outbound channel uses the common standard of DVB-S and MPEG 2 multiplex formatting.

The other way to apply FDMA is that of conventional constant bit rate carriers that provide symmetrical transmission between pairs of Earth stations. As discussed at the end of Chapter 8, this is very popular for providing backhaul of IP traffic from ISPs in remote areas that need connectivity to the worldwide backbone. PanAmSat offers a specialized service for ISPs called SpotBytes, wherein the FDMA carriers are provided in standard increments. Hawaii Pacific Teleport, a niche teleport on the island of Oahu, provides Asian and Pacific ISPs with such access to the U.S. Internet backbone. The SCPC channels can also be arranged in a star network, hubbed on a common teleport that is connected to a corporate headquarters. This forces the user to pay for satellite capacity whether or not it is being used, which is in contrast to the more dynamic approach taken by Shiron with their VSAT DAMA solution.

9.1.3.5 Operation of the ATM Protocol over Satellite

Many of the original applicants for Ka-band satellite licenses proposed to use ATM as the access protocol of choice. The reasons for doing this might be summarized as follows:

- ATM was originally developed for broadband networks—its short, constant length cell structure promotes fast action from routing and switching equipment.
- ATM is adapted to many telecommunications applications—toll quality voice, video, high-speed data.
- ATM is inherently a mesh network structure—onboard routing between beams is easily accommodated.
- ATM hardware has been available and proven for more than 10 years—development timelines could be reduced with the possibility of using off-the-shelf designs.

To date, two have progressed in development of the repeater and VSAT hardware and software for ATM-based implementation: Astrolink and Spaceway. Their respective developers, Lockheed-Martin/TRW and Hughes Electronics/Boeing, have invested a great deal of time and resources to produce what is reported to be working flight hardware. At the time of this writing, Astrolink has been put on hold and Spaceway continues in its development.

According to a published account [3], the Astrolink system employs TDMA and FDMA for the uplink air interface from the user terminal. Adaptive concatenated coding with a Reed-Solomon [236,212] outer code and a biorthogonal [8,4,4] inner code provides strong forward error correction. The inner code is adaptive to link conditions used to counter rain attenuation. Uplink power control is provided to the Earth terminal transmitter to compensate for rain fade without forcing an increase in power during clear weather; it is used to help comply with frequency

coordination requirements as well. The satellite onboard digital processor demodulates and decodes the data stream, and recovers the ATM cells for routing to the appropriate downlink beam. This uses TDMA with configurable beam hopping. Downlink coding is an adaptive concatenated code with a Reed-Solomon [236,212] outer code and a rate 3/4 or 3/8 convolutional inner code—again adaptive to compensate for rain fade. Bandwidth and connectivity is dynamically assigned through a central management system, providing an efficient DAMA scheme.

The operation of ATM is based on the principle of the virtual circuit, which is a point-to-point logical connection established in software. Unlike the Internet Protocol, which relies on a destination address to route packets, ATM uses a VC number to define the end points. Like IP, the information is fragmented into individual packets, which, in the case of ATM, are a fixed length of 53 bytes. As a connection-oriented protocol, ATM would be used to transfer voice, MPEG, or IP data from a specific sender to a specific receiver. The satellite repeater can participate in this connection by switching ATM cells between uplink and downlink streams. This discussion would indicate that ATM is more of a transfer mechanism than a multiple access method. For example, Astrolink actually uses TDMA on the inbound channel but ATM for cross-connection within the satellite processor.

9.1.4   Comparison of Access Protocol Performance

TDMA and ALOHA are very established as VSAT access protocols for the inbound channel. Recently, CDMA has been developed as another inbound channel access method, and ATM has been selected for some of the Ka-band systems. As can be seen from the descriptions in the previous section, these techniques differ in their operation and most certainly in their performance. The consequence of this is that users should select the protocol based on the requirements of the application. As we indicated previously, ALOHA and CDMA have the benefit of offering the transfer of a data packet with the minimum possible delay. This requires that the channel be lightly loaded. TDMA is best applied to an application that requires near-continuous data transmission. Lastly, FDMA is a simple approach that guarantees that bandwidth is available and where additional time is introduced for frequency assignment purposes.

It is the function of the VSAT indoor equipment to take a continuous data stream, such as for a telephone channel or file transfer, and break it up into a sequence of bursts. A time-division or SCPC channel can be provided on a preassigned basis to be maintained until the network is reprogrammed. Alternatively, the connection can be established on demand, through a process called bandwidth reservation. This would employ the ALOHA or CDMA protocol to request the bandwidth, which is subsequently provided via the TDMA or FDMA protocol. At the conclusion of the call, a release request is transmitted and the bandwidth is made available to others.

9.1.4.1   Use of an Analytical Model

The analysis of throughput and time delay on the VSAT inbound channel is complicated by the fact that user data requirements are usually unknown or at least highly

variable. There are some theoretical relationships that give us a good understanding of how these protocols tend to perform. For example, Figure 9.8 indicates the average delay experienced by each of three types of links, namely, an unslotted ALOHA channel, a slotted ALOHA channel, and a TDMA channel used in reservation service. As described previously, slotted ALOHA requires that the packets are transmitted within defined slots, according to timing provided by the hub Earth station. This means that a collision only occurs if the packets are actually transmitted in the same slot, as adjacent slots are nonconflicting. The upshot of this is that, on average, the time delay of slotted ALOHA is one-half that of standard ALOHA. The average packet delay is plotted against the throughput fraction, which is the ratio of data that arrives at the destination divided by the maximum data communication capacity of the channel. The point of intersection on the y-axis is the single packet transmission delay, which is the time to transfer the packet unimpeded. Say that the channel operates at 512 Kbps; then a throughput fraction of 0.10 means that the actual usable data transfer is 51.2 Kbps. This particular value (e.g., 10% throughput) introduces only a small additional delay due to collisions and subsequent retransmissions. The graph indicates the average delay; in fact, about 90% of attempts go through without collision, and the remaining 10% are delayed while the packet is retransmitted and confirmed. Figure 9.8 displays how delay increases as one attempts to increase throughput toward the ideal ratio of 1.

The importance of the throughput fraction is that it allows the network designer to determine how many inbound channels are needed to satisfy the data transfer requirements. The ALOHA channel has the lowest delay at zero throughput, a condition under which there are no collisions (that is, our VSAT has the channel to itself and wishes only to transmit one packet). At this point, there is a fixed delay that results from the basic transmission speed of the channel. Based on simple Poisson

Figure 9.8   Comparison of the contention delays for three VSAT access protocols: unslotted ALOHA, slotted ALOHA, and TDMA with transaction reservation.

statistics, the standard ALOHA channel experiences infinite delay at a throughput of 0.18 due to a continuous procession of collisions and the resulting loss of data transfer. The channel effectively breaks down. For the slotted ALOHA channel, this point of instability occurs at 0.37, which is about twice the throughput of unslotted ALOHA. Operation with a satisfactory frequency of collisions and average delay would probably correspond to throughput values of 0.1 and 0.2 for the standard ALOHA and slotted ALOHA channels, respectively. While not clear in the figure, the standard ALOHA curve starts at a value of delay below that of the slotted ALOHA channel and crosses above it well before a throughput of 0.18 is reached. For this reason, VSAT networks employ the slotted ALOHA scheme in recognition of its greater throughput potential.

A practical network design must operate sufficiently below the appropriate vertical broken line to provide satisfactory performance. This should be below the point of inflection of the curve (below the knee). We see that operating slotted ALOHA at 18% throughput is practical, but not so for unslotted, which must be reduced to less than 10%.

These results are well known; as a consequence, the ALOHA technique is recognized as effective for light traffic. Similar remarks can be made with respect to CDMA on the inbound channel. The performance of CDMA on the inbound channel side should be significantly better because the probably of collision is substantially reduced through reuse of the chip code. Aloha Networks claims that Spread ALOHA will deliver between two and four times the bandwidth efficiency of slotted ALOHA. This, or course, depends on your assumption for allowable collision rate. Furthermore, the individual messages themselves should be uniform in length and relatively short. This further reduces the potential for collision. Slotted ALOHA seems to be preferred because it has roughly twice the throughput for the same average delay.

For the case where the throughput is less than 10%, the relative number of retransmissions is essentially equal to the throughput. An example of the application of this rule of thumb, derived from the Poisson statistics of the packets, is that if the throughput is 6%, then, on average, 6% of the packets will experience a collision. Above 10%, the percentage increases rapidly and reaches 100% at a throughput of 37%. Any variation in packet length will further degrade this result. The performance of the channel in this case is dominated by the largest packet length, which means that capacity will be lost as compared to using a standard shorter packet.

The frequent occurrence of retransmissions is a strong indication that the throughput is pushing too high. The following approaches can be applied, either individually or as a combination:

- Reduce the quantity of packets per second from the VSAT with the greatest demand. This form of load shaving could provide a substantial improvement in overall network performance in the short term; however, one of the following steps would need to be considered if traffic loading must be restored for the affected VSAT.
- Reduce the length of the longest packets or force the use of a single packet size. The statistical behavior of a queue for constant length packets is decidedly better because the overall randomness of the channel is reduced.

- Reduce the number of VSATs sharing the same inbound channel—that is, add another inbound channel frequency to reduce the average load. This point is what could give CDMA an advantage in terms of the number of VSATs sharing the same bandwidth. Exactly how much better is subject to debate and can only be established for a particular network implementation and user community.

The performance of TDMA is also indicated in Figure 9.8 for the case of a bandwidth reservation type of service. There is a fixed delay corresponding to the time required to establish the connection. Delay increases slowly with throughput to a knee in the curve at a throughput of approximately 0.8, which is considerably higher than for the ALOHA channels (but comparable for CDMA). Increasing throughput further produces added delay, in this case because the bandwidth reservation process employs ALOHA as the request mechanism. A mesh-type TDMA network would have delay increase linearly to a point somewhere just below 1.0, which accounts for the guard times allowed in Figure 9.4.

9.1.4.2   Use of Computer Simulation

Analytical curves like those in Figure 9.8 provide some clarity into the operation of the protocols employed in VSAT networks. The difficulty comes in using the results for network design purposes. A more accurate way of proceeding is to build a discrete-time simulation model in a digital computer and then to simulate the operation of the network under the expected packet loading. During the 1990s, off-the-shelf software tools like COMNET III from Compuware were made available to help address the need for assessing network performance before committing to hardware. These tools were also popular on campuses for graduate education and research, and this trend continues even as the tools have lost their attraction among user organizations. Currently, most of the simulation research is being conducted using custom written simulation programs that precisely model the network technology under consideration. These are written in procedural languages like C and Ada, and have also moved to object-oriented programming using C++.

An example of the results of such a simulation, taken from our first edition because of its clarity and relevance, is shown in Figure 9.9 for a typical VSAT star network employing the slotted ALOHA protocol on the inbound channel. The graph displays the round-trip delay from VSAT to hub to VSAT and therefore represents the response time minus the host processing time. To add complexity, it is assumed that both short and long messages are to be transferred. The graphs are cumulative distributions of user message delay in milliseconds, shown for the short message, the long message, and the combined average.

The curves indicate that message delay is not a constant but in fact varies statistically across the total quantity of packets carried over the link. We see that the minimum delay is about 750 and 1,200 ms for short and long messages, respectively. On the other hand, the maximum delay experienced is 3,500 ms (e.g., 3.5 seconds). Between these limits, we can choose the 50% cumulative percentage as a point of reference. At this mean value, half of all the packets take less time and half take more. The nominal values of delay at the 50% point are 1,200, 1,400, and 2,000 ms for the short, combined, and long message conditions, respectively. The

Figure 9.9 An example of the results of a computer simulation of the performance of a VSAT network using the ALOHA access method. (Courtesy of TSCI, Berkeley, California.)

useful result is that this network would meet an objective of a 2-second response time on an average basis for the worst case type of message.

More recently, a comprehensive simulation was developed by a team of engineers with Georgia Institute of Technology, Broadcom Corp., and others to quantify the performance of an interactive VSAT network architecture for providing Web access services [4]. The model itself was for a two-way system that uses a different protocol, based on the MAC facility used on hybrid-fiber-coax cable TV systems. In it, the remote uses a hierarchy of packet sizes, based on what it is trying to accomplish. The hub interacts frequently with the remote to allow it to optimally make those adjustments. The block diagram of the simulation is indicated in Figure 9.10,

Figure 9.10 Software implementation of the satellite MAC channel.

where the Web browser at the upper left employs TCP/IP for its communication purposes, and the satellite splits data transfer between queued forward and return links with the hub and server on the lower right. To make this realistic, the action of the Web browser is driven from a probability distribution that was developed using measurements of real Internet subscribers. These were, in fact, students in dormitories at Georgia Tech and amounted to more than 1,900 participants browsing about 24,000 Web pages. A Sun workstation on this network captured and recorded the packet traffic to allow the statistical model to be developed. Even with this type of real information, the simulation was so massive that approximations had to be made so that simulation computations could be performed in a reasonable time.

Results from a series of simulations are shown in Figure 9.11. The curves demonstrate something different from the ALOHA characteristics previously described. Rather than delay going exponential around 36% throughput for slotted ALOHA, the performance seems to scale almost linearly. This may be deceiving because the delay values along the y-axis amount to many tens of seconds as channel utilization (throughput) exceeds about 30%. Utilization can increase toward 80% (impossible in the case of ALOHA) as long as one is willing to accept queuing delay of Web pages of the order of 100 seconds. These results allowed the authors to come up with an alternate scheme called piggybacking that cuts page delay in half. In piggybacking, the first request goes forward, and the next one is piggybacked on the channel before the first is fulfilled. Simulations were also performed to compare the performance of the MAC protocol with slotted ALOHA. This showed that MAC, which manages transmission by remote terminals, results in substantially reduced page delivery times under heavy loading.

These results are interesting in that they report on the dynamic behavior of a network. We have to keep in mind, however, that they were obtained for specific network configurations for a given traffic input. Change the configuration or traffic, and the results will change, possibly dramatically. The ideal situation is one where we closely model the network and traffic and then test the sensitivity of the

Figure 9.11   Whole page delay versus number of stations and channel utilization versus number of stations. After 50 stations, the curve becomes linear. (*From:* [4]. © 2001 IEEE. Reprinted with permission.)

delay to changes in assumptions. One useful type of study would be to look at how delay increases as we add more VSATs to the network or if we were to increase traffic loading by 50% in the original network.

9.1.4.3 Integration of Several Multiple Access Techniques

As we have discussed, the ALOHA protocol has certain benefits in applications that require short packet delivery with minimum delay. CDMA may prove to be superior to unslotted ALOHA as it reduces the incidents of collisions and thus increases the number of simultaneous users on the channel. On the other hand, TDMA is better suited for the transmission of synchronous and longer blocks of information, such as voice conversation and file transfer. The bandwidth reservation mode, which uses ALOHA or CDMA to request bandwidth and TDMA to provide stream transmission, combines both. Finally, the MAC-based access technique discussed in the previous section may provide a good balance between all of these alternatives.

Since a given inbound channel can only operate at a data rate consistent with the dish size and amplifier power, carrier bandwidth is generally limited to 2 Mbps, with 512 Kbps being more typical. This bandwidth must be shared among the remote VSATs and thus limits the overall capacity of the network. Introducing FDMA (which allows operation on multiple frequencies) adds bandwidth and provides a number of valuable benefits in dynamic VSAT networks:

- Adding another dimension to provide random access to satellite capacity. This can increase the throughput and occupancy, and potentially reduce the incidence of collisions in ALOHA.
- The network can be made much larger than what a single inbound channel can accommodate. This is because more frequency channels can be included in the capacity bundle; extension to additional transponders is also possible.
- Subdividing capacity into relatively small channel bandwidths without loss of total network capacity will reduce SSPA power requirements for the inbound channel.

For ALOHA and TDMA, the benefits are powerful enough to cause FDMA to be an integral part of modern VSAT networks. The two general terms for this approach have been called frequency-hopping TDMA (F-TDMA) and multifrequency TDMA (MF-TDMA). While fundamentally the same, the approaches vary among the suppliers, and there are subtle yet potentially beneficial attributes thereby obtained in particular applications. MF-TDMA has been incorporated in the DVB-RCS standard introduced in the previous chapter; all major suppliers of VSATs currently employ some form of MF-TDMA.

The above remarks would not necessarily apply to CDMA because it depends on the code separation property as opposed to that of time or frequency. Network capacity in CDMA is a very complex subject and continues to be debated even though this access method is so popular in the mobile telephone arena. At the root is the multiplication factor, N, which is the ratio of the chip rate to the input information rate; however, there are other factors that will diminish the quantity of simultaneous CDMA users—the ability to control power level being principal among them

[5]. In normal operation, the remote transmitter power must be adjusted so that the signal-to-interference ratio received at the hub is acceptable. This should also produce the maximum system capacity. However, any increase in remote power level raises the interference experienced by all users, and capacity is compromised. A highly complex and dynamic scheme can be used to control the received power and thus maximize channel capacity. If the uplink experiences a fade due to rain attenuation or other propagation effects, the power level must increase to compensate. This compensation must be tightly controlled since removal of the attenuation will initially allow the power at the satellite to exceed the nominal level. There will therefore be a residual error in power control, measured in decibels, which will directly reduce capacity by a like amount.

ALOHA and TDMA were first integrated together by HNS in the early 1990s. This was in response to application trends in the commercial U.S. marketplace, where several types of user traffic were to be served on a common network. Other suppliers of VSAT products have recognized the value of offering a hybrid solution of this type. The basic structure of the combined frame, which works with a combined purpose TDMA method, is illustrated in Figure 9.12. The frame period of 45 ms in this example is 360 times the minimum time required for speech transmission; therefore, each speech burst must contain 360 voice samples. The figure indicates how this frame is allocated among the protocols, with the first segment available for pure TDMA, referred to as stream (continuous real time). Bursts that appear in this segment will repeat indefinitely or until the network is reconfigured on a long-term basis. A second TDMA segment is allocated for demand-assigned bursts, that is, for bandwidth reservations. Here, the bursts have been allocated through a dynamic process in response to requests from remote VSATs.

ALOHA transmissions are confined to the third block of time. Within this segment, VSATs can transmit packets using the slotted ALOHA protocol. This means that there will be instances of collisions since VSATs transmit in an uncoordinated way. The only means that we have to control the frequency of collisions is to limit the number of VSATs that share this particular block of time (which limits the maximum possible throughput).

One important outcome of using a common time frame for all services is that every user will experience the same average time delay of one-half of the time frame (22.5 ms in this example). This is only about 10% of the one-way satellite hop delay and therefore might be considered acceptable. An exception, of course, is for ALOHA users who are subject to an additional delay due to collisions and retransmissions. The benefit of the common time frame is that a single inbound channel can be shared very effectively and the bandwidth allocated as needed over the life of the network. It is possible to move applications from one protocol to another as a

Figure 9.12   The time frame for an integrated VSAT inbound channel, providing ALOHA, TDMA, and demand assignment/transaction reservation services. These allocations may be static or dynamic.

way to better meet user requirements. Also, the number of VSATs assigned to a particular segment can be adjusted up and down. These are functions for the network control center that is associated with the hub Earth station.

The fact that the access protocols need to be tailored to the application has caused two additional modes to be introduced in the market. The first is the dynamic selection of access method based on the traffic demand. The activity level of the data channel is monitored at the remote VSAT and at the hub. If the situation requires a change, then the hub automatically causes a switch in access protocol. For example, if the particular user is employing slotted ALOHA and the requirements shift so that more collisions are experienced, then the network can automatically transfer to a bandwidth reservation mode of operation. The user application would only experience an improvement in performance, recognizable as a drop in response time.

The switching between access protocols would be based on a criterion to be selected by the operator. Two such criteria are: (1) switching based on the rate of traffic flow and (2) switching based on the size of the message being applied to the remote. In traffic rate switching, the remote VSAT senses the amount of data in the queue buffer that holds information that is to be placed into packets for transmission to the hub. When the amount of that data exceeds a threshold, the switch is made from ALOHA to bandwidth reservation. Later, when the amount of data drops below a second threshold, the protocol is switched back to ALOHA. In message size switching, the VSAT uses the measure of the size of the message to make a determination on the protocol. As long as short messages are coming into the remote, the ALOHA protocol is employed. However, when long messages begin to appear, the protocol is switched to bandwidth reservation or continuous TDMA, as appropriate. The selection criterion is a programmable parameter.

The other technique is to make a bandwidth reservation on the fly without having to employ the ALOHA protocol. This reduces the setup time and therefore improves the operation of the particular application. Referred to as piggybacking, the request for bandwidth is attached to another user packet that is already going from the remote to the hub. The hub recognizes the request and then returns the reservation information to the remote. This is invisible to the user application.

9.1.4.4   Structure of the Outroute from the Hub

The outbound channel is typically transmitted as a continuous stream of data using a TDM format, with MPEG 2/DVB-S being the most common structure. It is not TDMA because only one station is providing this uplink and there is no need to employ burst transmission. This also simplifies the design of the remote modems in the VSATs since the carrier is continuously on, reducing how often receivers must resynchronize. As we indicated earlier in this chapter, the outbound channel speed is typically some multiple of the inbound channel speed to take advantage of the larger antenna size and greater potential transmit power of the hub. For example, in one design the outbound channel operates at 10 Mbps while each inbound channel operates at 512 Kbps. There would then be multiple inbound channels to balance each outbound channel, determined by the specific data transfer requirements of the network. The transmission would employ strong forward error correction and QPSK

or higher order modulation such as 8PSK, which would associate nicely with the DVB-S standard.

By using TDM on the outbound channel, the network is forced into a star topology, which is consistent with centralized services such as Internet access or provision of corporate IT computing functions. Alternatively, an outbound channel based on TDMA could allow mesh services when direct connection between remote sites is required. VSAT networks of this type are offered by ND Satcom, VIPERSAT, and iDirect, which employ TDMA for all transmissions.

There could be a situation when one outbound channel is insufficient to provide the forward link for the expected quantity of VSATs and their expected data throughput requirements. In this case, additional outbound channels would have to be provided. The general architecture of the system usually requires that there be one set of equipment at the hub for each outbound channel. This set of equipment is referred to as a "network," not to be confused with the overall network itself. A network of this type could support one outbound channel at, say, 10 Mbps, and a quantity of 20 inbound channels at 512 Kbps each, all within the same transponder. There might be instances where the traffic is imbalanced—that is, where the total inbound channel data rate is greater than that of the outbound channel. Also, the number of VSATs that can share an inbound channel will depend on the aggregate data throughput requirements, including the extent to which the ALOHA, CDMA, and TDMA access protocols are used. This can change on a dynamic basis and therefore would have to be adjusted as the needs vary.

There are architectures where the hub can be downsized to reduce cost and therefore the price of entry for the user. (Another way to accomplish the same result is to share the hub among several networks, as discussed later in this chapter.) Reducing the outbound channel speed has a direct effect on the bandwidth and power that must be obtained from the satellite provider. Keeping the hub and VSAT antennas the same size will allow the commitment for satellite capacity to be reduced by the ratio of outbound channel speed. It also means that the hub antenna might be smaller in size, say, 2.8m instead of 6m. For some applications, a smaller antenna might be required by constraints at the hub site. Also, having a lower capacity on the outbound channel, such as 512 Kbps, would allow a smaller high-power amplifier to be used. This will depend on the size of the hub antenna and the need to maintain service in the local rain environment. Reducing the outbound channel speed also decreases the number of VSATs that can be supported in the network. Therefore, for a small network with perhaps only 25 remotes, the smaller hub and outbound channel speed could be a good match.

9.2   Sizing of VSAT Networks

The typical VSAT can be configured to satisfy a range of data transfer requirements and thereby achieve an acceptable response time or whatever the appropriate quality of service measure may be. Whether a satisfactory result can be obtained depends on this match; not every application will fit the particular architecture and throughput capabilities of the typical VSAT network. Cases where there is a good fit usually adhere to the following:

- The network is clearly a star topology, where there is a central host computer that supports a large quantity of remote locations. This depends on having a centralized data processing structure, where remotes depend on the host for information access and network management. Alternatively, the hub is connected to the Internet through a suitable ISP.
- The data rates on the forward and return links are sized adequately for peak demand.
- The throughput requirements are consistent with the basic data transfer capabilities of the VSATs and the hub. This must be verified through the process of sizing.

The focus of this section is on the proper sizing of star VSAT networks. Before this activity can begin, all of the requirements for the network must be listed and quantified. These are broken down as to form of information and application that processes it, locations and the associated number of users, and the measures of quality of service to be addressed. Table 9.2 provides a format for this type of analysis, based on a typical organization's needs. From this, the requirements for the hub, remote VSAT, and transponder can be surmised.

9.2.1   Hub Sizing

The sizing of the hub equipment is particularly important to VSAT network design because it relates to the most costly single element. Quantities of line interface cards,

Table 9.2   Example of a Summary of VSAT Network Requirements

User Application	Network Requirements	Number of Locations	Bandwidth Requirements	Quality of Service
Internet access (one user; small group; remote site)	High-speed access to Internet backbone; TCP/IP	250	Better than dial-up	Page download within 30 seconds
Remote access to corporate intranet (LAN extension)	High-speed access to private network infrastructure; Web-based applications; TCP/IP	250	Better than dial-up	Page download within 30 seconds
Remote access to corporate business applications	Medium- to high-speed access to private network	250	32 Kbps per user	Response time less than 5 seconds
Content distribution	Multicast uplink for wide area distribution to PCs and content caching servers; UDP/IP and Multicast Transport Protocol	15	256 Kbps	100% accuracy of delivery within 1 hour
Video teleconferencing	High-speed access to private network infrastructure or public ISDN; H.320 or H.323 standards	25	256 Kbps	1.5-second one-way delay
Telephone	Low- to medium-speed access to private network infrastructure or PSTN; POTS or VoIP standards	250	8 Kbps	0.5-second one-way delay
Leased line	Medium- to high-speed connection; Frame Relay	15	256 Kbps	300-ms one-way delay

modulators and demodulators, and baseband processing equipment must be determined and placed on order with the supplier. However, this is a complex problem that warrants careful consideration and significant involvement by the supplier since they understand the unique architecture and capabilities of their products. The hub is often modular in nature, being composed of a variety of racks of equipment that are configured for the particular network requirement. It is relatively simple to add, remove, or rearrange equipment within the hub, thus allowing the capacity to be managed and new capabilities introduced from time to time. In the case of star networks using MPEG 2 and DVB-S, several outbound carrier components are standard items that are supplied to the broadcasting industry. These include the MPEG multiplexer, IP encapsulator, and DVB-S modulator. Specialized components would be added by the VSAT manufacturers for inbound-to-outbound synchronization and restoration of proper MPEG PCR timing.

The RF terminal is typically of standard design as applied in teleports and either located in an external shelter or outdoor cabinet. Once the antenna, LNBs, and RF power equipment are installed, it is unlikely that many changes will be required to accommodate the hub baseband equipment. The critical step, discussed later, is the sizing of the HPA, which could be one of three types: SSPA, TWTA, or klystron power amplifier (KPA). Adding equipment or increasing RF power may require a major modification of this part of the hub, and even a complete replacement could be a consequence.

It is usually best to allow the VSAT hub supplier to participate in the sizing of the network. This can be done using the power of competition to obtain the best estimate, both in the terms of the quantity of equipment and the ultimate price. A low estimate could result in under-capacity; additional equipment could be brought after the network is put into service to resolve cases where response times are excessive. Backup, spare items, and any specialized test equipment and software should be purchased at the same time and tested to verify proper operation.

The critical elements of the hub related to sizing are shown in the simplified block diagram given in Figure 9.13. In its role as the central point of the star network, the hub coordinates all transmissions over the satellite and provides the connection to the host computer. These elements must be sized to support the maximum capacity of the hub, which also determines the capacity of the network. In this example, the applications are serviced by a central host or server computer that may provide access to the external world as well (e.g., the Internet). From a network standpoint, the server establishes all of the sessions with remote user devices and assures reliable

Figure 9.13   The elements in the hub that determine network size and performance.

transmission. The backhaul circuit is either a short direct connection in the case of a locally connected hub or a dedicated circuit using private or public transmission facilities. Communication between the port and the server follows the protocol used by the host in its normal operation. Most corporate applications operate through the TCP/IP suite of protocols, and the physical and link layers are likely to be provided with Ethernet (IEEE 802.3). While it is customary to use a synchronous physical layer such as a T1 or E1 for the backhaul, there are instances where an asynchronous metropolitan area Ethernet service can be employed (discussed in Section 5.7).

The capacity of the backhaul circuit is chosen so as not to restrict the flow of data on the outbound channel and inbound channels. Thus, it must support every application that the host delivers to the network. Because congestion here will directly impact all users, it is a good practice to overdesign the backhaul circuit in terms of its bandwidth (for example, by providing double the expected throughput). This allows for peaking of traffic during expected and unexpected increases in demand and for growth. Another consideration is the provision of backup service in case of failure. A typical strategy is to provide an alternate path using possibly a lower speed data line or even a point-to-point satellite link. The link can be used for additional capacity and thus provides for unexpected peaking of demand.

The hub baseband equipment is difficult to size because it is really a system composed of several functional elements. Port speeds and quantities are selected based on the needs of the backhaul circuits and host. Typical interface options include RS-232 (for speeds under 56 Kbps), RS-449 (for 64 Kbps), V.35 (for 56 Kbps up to approximately 256 Kbps), T1 and E1, and LAN interfaces such as IEEE 802.3 (Ethernet). A DVB-S type of hub, illustrated in Figure 9.14, employs the DVB- ASI where the MPEG digital video and encapsulated data are introduced.

Port cards and the supporting buffering equipment make the digital traffic available to the outbound channel equipment and accept traffic from the inbound channel equipment. The buffering is necessary to allow the packetized information to be placed into the time division frame that comprises the outbound channel transmission. This introduces an amount of delay that consists of:

Figure 9.14   Specific components for the outbound DVB-S carrier of a VSAT.

  - Processing delay;
  - Buffering delay;
  - Outroute access delay (nominally half the frame time).

The outbound channel capacity is dictated by the speed of transmission. If we assume this to be 4.096 Mbps, then each 45-ms frame contains $4,096,000 \times 0.045 = 184,320$ bits or 23,040 bytes. A typical frame, illustrated in Figure 9.15, is divided among the VSATs that receive this outbound channel. The header indicates the start of the frame and provides a timing reference for all remote VSATs. Real-time packets are assembled into the first segment of the frame; these correspond to the TDMA stream inbound channels that repeat periodically. Nonreal-time packets in the second segment are dynamic and respond to ALOHA and bandwidth reservation allocations to the remotes. The breakdown need not be constant as services can be quite dynamic in nature. The data can be packed as tightly as the accessing traffic allows. Obviously, maximum throughput occurs when the traffic is preassigned and not subject to change. If we are responding to requests from remote VSATs who are using ALOHA on a totally random basis, then the makeup of the outbound channel frame will change constantly. It is assumed that the user data can be contained within the variable portion of the packet. The architecture of the network is constrained so that there is a maximum length, as indicated in the figure. If the user data input from the host exceeds the maximum, then a segmentation process in the hub breaks the user data packet down into smaller pieces that fit within the maximum size allowed by the network.

The complexity of the architecture and data structure of the VSAT network along with the wide range of possible user configurations at the remote and hub make sizing a very difficult process. The VSAT manufacturer may provide a detailed analysis of traffic loading for use in sizing their hub equipment. The dynamic and flexible nature of the VSAT and hub allow user demands to change as they may (something that is expected). This means that after the hub and VSATs are installed, there are many options to be explored. This includes letting the network itself choose the best multiple access technique in response to variation in user demand. If the capability of the installed equipment is exceeded, then it is still possible to add ports and hub baseband equipment elements to bring the limiting element up to sufficient capability.

Figure 9.15 Outbound packetized TDM time frame (typical).

Sizing of the RF terminal, indicated to the left in Figures 9.13 and 9.14, is based on the standard principles for Earth station design. An RF engineer would perform a basic link calculation as outlined in Table 8.3 to determine the uplink EIRP for the outbound channel carrier. From the uplink EIRP, the RF engineer works backward through the hub antenna gain and waveguide loss to determine the power output from the HPA. The design might have to allow for multiple outbound channels, either for the initial service or for future growth, with acceptable intermodulation distortion. The HPA would be selected to have greater power than the requirement, calculated as the product of the power per outbound channel carrier times the number of carriers times a backoff factor. A typical value of the backoff factor is in the range of 4 to 6 (e.g., 6 to 8 dB). More power margin should be provided if the hub is to use UPC. An example of an RF terminal sizing is given in Table 9.3 with four outbound channel carriers. The last item in the table is the power of an off-the-shelf HPA that can be used for this station. It provides an additional margin since the minimum requirement is 631W.

9.2.2　VSAT Remote Sizing

The remote site where a VSAT would provide service is typically a branch office, retail store, government service center, or even a home. Historically, the type of data communication to such locations was using either a dial-up telephone connection or an analog leased line, a problem that the VSATs were originally designed to address. With the advent of the Internet, broadband DSL and cable modems, and real-time applications like VoIP and videoconferencing, expectations for data speeds have increased to more than 128 Kbps. For a single user, this speed can provide a reasonable approximation of a broadband service. More users on the same line will demand greater access speed—the precise amount will depend on the activity these users engage in. The ability of the VSAT to meet the requirements of a remote location is constrained by the maximum throughput of the inbound channel data rate, which typically is not greater than about 2 Mbps, and more likely is between 300 and 600 Kbps. Outroute speeds, however, can reach tens of megabits per second with the maximum currently supported by the DVB-S standard being in the neighborhood of 100 Mbps.

Table 9.3　HPA Sizing for the VSAT Hub (No Uplink Power Control)

Factor	Value
EIRP per carrier	68.9 dBW
Antenna gain (6m)	56.9 dBi
Waveguide loss (100 ft)	3.0 dB
HPA power per carrier	15.0 dBW, 31.6W
Multicarrier output backoff	7.0 dB
Number of carriers (4)	6.0 dB
Total carrier power	28.0 dBW
HPA requirement	631W
HPA size selection	800W

The elements that must be sized in a remote VSAT are indicated in Figure 9.16. Normally this is done by the supplier, according to the features the customer requires. The customer premise equipment, including computers and other peripheral devices, are attached at the left. In substituting the VSAT for a terrestrial data circuit, we need to consider the bandwidths and performance of each critical element. Depending on the requirement and the type of VSAT equipment being considered, the access port could employ one of a variety of interface standards. These include RJ-45 [Ethernet, using Category 5 wiring with four twisted pairs (eight conductors)], RS-232, RS-422, and V.35 specifications. The port speed depends on the connection at the user device such as a PC or router. Usually, the VSAT is within the same building as the customer premise equipment; hence, we can use a much higher data rate than we could normally afford on a long-distance basis. For example, if the throughput requirements are estimated to be an average of 100 Kbps, then we can employ a LAN connection at 10 Mbps to prevent this part of the system from ever impacting performance. This is also consistent with the much higher-speed outroute channel. The cost of doing this is relatively minor since all that we are purchasing in this case is Category 5 cable.

If the VSAT serves a single user, there is no contention for local access, and therefore, the user will not experience any delay for his portion of the system. Otherwise, this is resolved either by a remote router, PC, or the VSAT indoor equipment. This is consistent with the protocol used by the hub to determine which remote devices are in need of service. As we discussed previously, the host computer is spoofed by the hub baseband equipment into thinking that it is exchanging protocol information with each remote device. Instead, protocol operation is re-created at the remote and information is transferred over the satellite on an optimized basis using the appropriate access technique. Communication between the VSAT indoor equipment and multiple PCs on the access line is accomplished on a more or less seamless basis as long as the bandwidth at the port is sufficient for satisfactory service. Utilization in a contention scheme can exceed 90%, as in the case of TDMA if

Figure 9.16   VSAT remote elements to be sized for service.

the VSAT controller is able to force an orderly entry of information from the user devices. It may be just a matter of ensuring that the total access speed is sufficient for the expected demand from the particular number of users who share the port.

The delay that the port card introduces involves two components: the data transfer time (determined only by the line speed) and the queuing delay (determined by the number of users sharing the line and the contention control method). A dedicated port involves no contention, while a shared port must be evaluated carefully and perhaps should not exceed 60% utilization to avoid significant additional delay.

The RF equipment, shown with the antenna in Figure 9.16, must support the transmission between the VSAT and the hub. The capacity of the link is determined by the access protocol and the associated throughput versus delay characteristic discussed earlier in this chapter. The link itself can be evaluated using standard link budget calculations, telling us the required EIRP and G/T of the station (see Table 8.3). From this, we can select an appropriate antenna size and corresponding HPA. The typical approach is to make the antenna as small as possible, consistent with the available solid-state power amplifier that can come with it. These are usually sold as a system. Readers should note that the RF head tends to determine the overall reliability of the VSAT, primarily because of the rather hostile environment to which it is exposed (including such factors as rain, snow, salt air, sand, and sun—including that focused by the reflector). For this reason, a well-proven design with many similar units already in use is the best choice.

Sizing of the antenna and SSPA is sensitive to air interface parameters, including data rate, type of modulation, forward error correction employed, and margins to cover baseband impairments. These characteristics vary among equipment designs and are impacted by the nonlinearity of the amplifiers on the ground and in space. Typical information about the air interface is provided in Figure 9.17, which shows the BER versus E_b/N_0 performance for six common error correction techniques. This data is then used to perform link budgets in parametric form, from which the curves in Figure 9.18 were created. If we were to assume a fixed satellite EIRP of 35 dBW and a threshold BER of 10^{-8}, the figure indicates that the receive dish size would be

Figure 9.17 Examples of modem and coding performance for commercially available air interface characteristics.

Figure 9.18   Impact of forward error correction on dish size for C-band.

3m for convolutional coding with $R = 3/4$, or about 2.2m for Turbo Product Code. The benefit of improved FEC is clearly evident. The curves have been adjusted for the adjacent satellite interference, which is greater for the smaller dish size. A specific site location in a very rainy region was assumed for this example. Figure 9.19 adds the inbound channel and provides an indication of the required SSPA power. We have assumed AsiaSat 2 serving three locations: Suva, Fiji; Manila, Philippines; and Palembang, Sumatra, Indonesia. The antenna size is fixed at 2.4m and the inbound channel data rate set at 256 Kbps. The data indicates that Turbo Product

Paramter	Units	
Data rate	Kbps	256
E_b/N_0	dB	
R=1/2 (without R-S)		6.4
Turbo		4.5
Transponder capacity	channels	
R=1/2 (without R-S)		161
Turbo		225

		Suva, Fiji	Manila, Phil	Palembang, Indon
SFD	dBW/m2	−80.0	−90.0	−88.9
G/T	dB/K	−12.0	−1.6	−2.9
R=1/2	Txpower	30.7	3.6	3.0
Turbo	Txpower	22.0	2.6	1.9

Figure 9.19   Illustration of sensitivity of SSPA sizing and transponder capacity for C-band VSAT to air interface and satellite characteristics (AsiaSat 2 footprint). (Courtesy of AsiaSat.)

Code will reduce the uplink power required for a satisfactory link by the ratio of about 3 to 2. This type of study would need to be performed for any network of VSATs, for the different satellites and air interface characteristics being evaluated. One also sees that with the 10-dB satellite G/T difference it might be useful to invest in a bigger dish at Suva.

9.2.3   Transponder Capacity Sizing

A given VSAT network may only require a fraction of a satellite transponder. This is shown in Figure 9.20, where a transponder supports several outbound channel carriers and their associated inbound channel carriers within its usable bandwidth. This packing efficiency permits a given network to go into operation at a relatively modest cost because transponder capacity can usually be purchased based on the percentage of the transponder that is needed. As in the case of a multicarrier Earth station HPA, sharing the transponder power must allow for multicarrier output backoff in the range of 3 to 5 dB to control intermodulation distortion. Typical TWT input/output and intermodulation characteristics are shown in Figure 9.21.

The basis for the selected backoff is illustrated in Figure 9.22, which is conceptual in nature. On the x-axis is plotted the input backoff to the transponder; this varies linearly with the power transmitted by the uplink Earth station. Five separate variables are indicated on the y-axis in terms of ratios of carrier to noise. The uplink C/N tracks exactly the uplink power and is therefore linear with input backoff. In response to the uplink drive, the downlink C/N increases until the transponder amplifier begins to saturate at the right extreme of the curve. This is the typical RF input/output characteristic for a multicarrier transponder (Figure 9.21). The dark saturation curve shows the combined uplink and downlink C/N. The ratios, as well as others to follow, are combined using the formula in Section 2.2.4. Working in the downward direction is the intermodulation C/IM, which decreases as the transponder is driven toward saturation. The final measure of the overall link is presented in the total link C/N, which is the humped curve shown at the bottom. The action of adding together intermodulation with thermal noise produces a maximum point that occurs at what is called the optimum backoff. Note that the top of the curve is relatively flat so that some imprecision in backoff "operating point" is permissible.

Figure 9.20   Allocation of transponder bandwidth for inbound and outbound carriers, and multiple VSAT networks.

Figure 9.21   Typical linearized TWTA power and intermodulation characteristics for a C-band transponder.

The optimum values of total input backoff (*x*-axis) and output backoff (*y*-axis) are displayed in Figure 9.22.

Extra transponder bandwidth can be used to advantage by spreading the carriers out and thereby reducing the effective intermodulation noise (which increases the C/IM). This improvement is approximately equal to the ratio of the total RF

Figure 9.22   Optimization of transponder backoff for multiple carriers, assuming equal carrier power and bandwidth (graph is notional and does not represent specific data).

bandwidth in the transponder divided by the occupied bandwidth, converted to decibels. Thus, carriers that only occupy about half of the transponder will experience 3 dB higher C/IM, provided the carriers are spread out equally. As a further refinement, the outbound channel carriers in Figure 9.21 can be unequally spaced to avoid some of the stronger intermodulation products. This particular technique is also practiced in constant-carrier FDMA networks.

The amount of satellite transponder capacity that is required for a VSAT network is determined by the combined bandwidth and power for the outbound channel and inbound channel carriers. It is a very straightforward process to sum up the bandwidth of each type of carrier. Because ground receivers cannot separate carriers that are too close to each other, it is appropriate to allow a guard band of 10% to 15%. A typical example is provided in Table 9.4. The outbound channel data rate is 2 Mbps and employs four-phase PSK, while the inbound channel data rate is 512 Kbps and employs MSK. The power allocation includes backoff and is based on the percentage of the total transponder power. We are assuming the transponder bandwidth is 54 MHz. The amount of power used is approximately 40% of the total available, while the amount of bandwidth is 10%. Under this condition, the satellite operator would charge for service based on 40%, which is the bigger fraction. The power of the transponder would therefore be used up before the bandwidth, which is a typical result in multicarrier operation.

9.3 Hub Implementations

That a VSAT can provide a range of services through a relatively inconspicuous antenna is well established. The star topology incorporates the hub to coordinate transmissions and provide access to a central host computer. However, the hub represents a major investment as well as a commitment to operate a complex telecommunication facility for the duration of the network lifetime.

Users of VSAT star networks have two avenues open to them: the dedicated hub and the shared hub. The dedicated hub is owned by the network user and operated solely for the benefit of the user. The shared hub is owned and operated by a service provider who makes the investment and operates the hub for the benefit of multiple user networks. The following sections compare these alternatives in general terms. Anyone considering the choice of dedicated versus shared should conduct a detailed

Table 9.4 Example of an Estimate of Total Capacity Required for a VSAT Network

Component	Carrier Bandwidth (kHz)	Percent Power	Quantity	Total Bandwidth (kHz)	Total Power (%)
Outroute (2 Mbps)	2,000	15	1	2,000	15
Inroute (512 Kbps)	700	4	6	4,200	24
Guard band allocation	250		1	250	
Total network				5,250	39

examination that considers the operational, business, and financial aspects of the question. Before committing to a shared hub, be sure that the VSAT technology they employ is right for your applications. This means that the same kind of evaluation that would be done for a dedicated hub is needed in the shared hub situation as well.

9.3.1   Use of a Dedicated Hub

Early adopters of VSAT networks were large companies who were willing and able to risk being pioneers. Not only did they have to overcome the early barriers to integrating satellite communication into the conventional terrestrial networks of the time, but they also had to make the full financial commitment to hub ownership and operation. Companies like Wal-Mart, the world's largest retail chain, and Edward Jones, a regional stock brokerage in the United States, built their businesses on VSAT technology. Taking responsibility for the hub was an expected commitment along the way to achieving a competitive advantage over their rivals. Interestingly, the feature of VSATs that attracted them the most was their ability to deliver a consistent set of telecommunications applications independent of location. Wal-Mart placed their stores in rural towns where there was no competition. Likewise, Edward Jones served stock brokerages in the Midwestern United States where the latest telecommunications technology was being installed at a slower pace than in major cities like New York and Seattle.

The basic arrangement of a dedicated hub serves a single data center, as that shown in Figure 9.13. In the ideal case, the host computer and hub occupy the same building and are within a short cable distance of each other. This removes the backhaul circuit that could otherwise impact throughput and response time. In addition, the service is not susceptible to failure of a backhaul circuit provided by an outside telecommunication operator. The hub Earth station site may have to be located away from the host computer due to practical considerations such as difficulty with placement of the antenna or availability of adequate room for the equipment. Also, the organization may maintain two host locations for redundancy reasons. The backhaul circuit would be required in these instances.

The following are some general guidelines in favor of a dedicated hub.

- *The total number of VSATs exceeds 400.* This is not a hard and fast rule but rather is based on economics. The actual breakpoint number is determined by a variety of parameters, in particular, the total investment in the hub and the cost of obtaining competing data communication services over the terrestrial network. There are instances where a dedicated hub was installed for a network of only 60 to 100 remote sites. Likewise, networks much larger than 400 remotes have been placed on shared hubs. If an organization wants to consider this question on an economic basis, then it should run a thorough evaluation to see which direction offers the lowest cost over the expected period of operation.
- *Applications are very strategic.* A strategic application is one that gives a company a significant competitive advantage in its market. Telecommunications can be an ingredient in achieving this aim, as we discuss in our other work [6].

The VSAT network is truly a private network infrastructure that is independent of the public network. Furthermore, it can deliver features and resources in ways different from what is available to the general public and, in particular, competitors (unless they implement their own VSAT networks). This strategy is employed by Wal-Mart to grow rapidly in rural communities.

- *The highest level of reliability and availability are required.* A dedicated hub can be made as reliable as resources and money will allow. The facility is not shared with other users, and hence there is no conflict over how resources are allocated and what steps are taken to assure the availability of services. For example, all of the equipment is available to one organization, which can decide which particular service or end user will be maintained in service even during a partial interruption. Also, there will tend to be less potential conflict between the services that run through the common equipment within the hub.

- *Adequate technical staff is available.* Design and installation of a VSAT network, including the hub, can be contracted out to the supplier. After the network is up and running, it is time for the user organization to take over the operation and maintenance and to evolve the network as applications are modified or added. This type of activity can be very technical and complex in nature. You are, after all, a service provider to your internal and external users. Some organizations prefer to stay out of this line of activity, as they do not consider it to be a core competence. On the other hand, through recruiting and training, the organization can raise its standard to become a competent VSAT network operator. Outsourcing of the operation to the manufacturer is another option.

- *The backhaul circuit is unacceptable.* The dedicated hub configuration shown in Figure 9.13 uses a backhaul circuit to connect the hub to the host computer. While this is the general case that allows for separation of the two elements, there are instances where backhaul circuits of adequate quality and reliability are simply not available. This could be in a developing country or in a remote area of an advanced economy. The dedicated hub allows the host and hub to be physically adjacent to each other. In this case, the backhaul circuit is a piece of cable that would be provided along with the installation of the hub.

- *Video transmission is frequently required.* We discussed earlier in this chapter how business TV can be added to the network simply by installing a video receiver with each VSAT and arranging for an appropriate uplink. The latter need not be part of the hub but could be obtained on an occasional basis from a video service provider (e.g., a teleport). For large VSAT networks that can justify their own video uplink, the hub should probably be of the dedicated type. This gives the user complete control over all aspects of its operation, removing the possibility of conflict over resources. With IP encapsulation, data can be effectively merged with MPEG video to gain capacity in an opportunistic way.

Clearly, the cost and organizational and strategic implications should be examined thoroughly before making this kind of commitment. If good shared hub facilities are available, then this option must be examined as well.

9.3.2   Use of a Shared Hub

A shared hub is a telecommunication facility that is operated as a business for the benefit of customers who wish to reduce their costs and operational commitments. Generally speaking, the cost of using the shared hub would be less than that of implementing a dedicated hub and operating it for the benefit of one organization. The idea is that there is an economy of scale in loading a hub with several independent networks, each with their own respective community of remote VSATs. Since the hub would have one RF terminal and antenna, the remotes would all have to be within the coverage area of the same satellite. Access by the hub to terrestrial networks is another consideration.

Figure 9.23 indicates how a single hub can be shared by three customers, each with their own host computer. From the customer's point of view, their network is an independent entity supported by their host, shown at the right of the figure. Not shown are the remote VSATs themselves, which would be located around the satellite coverage area. Each customer's host is connected to the hub via a dedicated backhaul circuit (one per customer). This circuit would be maintained by a terrestrial communications service provider (probably the public telephone operator) or could be a private point-to-point link using cable or microwave. There have been instances where the backhaul circuit is provided over a GEO satellite as well, although this introduces an additional propagation delay of one-quarter second.

The hub operator has the responsibility of satisfying the special data communication and application needs of each individual user. Their respective networks may employ different types of hosts, protocols, and computer applications. This increases complexity and might introduce some conflict within the operation or performance of the hub. For these reasons, the hub operator will have to maintain an experienced staff and be on call to resolve difficulties as they arise.

The following are the general guidelines for determining if the shared hub approach is of potential value to a customer.

- *There are fewer than 200 VSATs.* As with the case of the dedicated hub, there is no hard and fast rule based on the maximum quantity of VSATs. However, if there are fewer than 200, then it is often difficult to justify the investment in

Figure 9.23   The arrangement of a shared hub allows several customers to operate independent networks, either logically or physically, through individual backhaul circuits and an allocation of hub equipment and satellite capacity.

a dedicated hub. The reason for this is that the hub is a major telecommunication facility, costing up to $1 million. To make this effective, you should divide this cost over a sufficient quantity of remotes to make it pay. The breakpoint can be reduced if there is an existing RF terminal and telecommunications access already in place.

- *Technical staff are unavailable.* This is the converse to the situation posed for the dedicated hub. Organizations that do not wish to extend their activities into the highly technical operation of a hub should best leave the effort to an experienced provider. Then, they would not be concerned with the recruitment, training, and management of hub staff. The cost of doing this can be significant but is usually recovered in large networks. On the other hand, some owners of dedicated hubs prefer to hire outside contractors to run the hub (a process called outsourcing the hub maintenance).

- *Reliable backhaul services are available.* Shared hub service only makes sense where reliable backhual services are available. This is because of the criticality of the backhaul circuit, which must have near 100% availability and high-circuit quality in terms of error rate. Fiber optic or coaxial cable links are preferred.

- *Technology is changing rapidly.* The investment in the hub covers the cost of hardware and software. The software part can be modified and upgraded over time to match the needs of users and to correct for problems and bugs that are detected in this and other hubs. The hardware portion is less flexible and can become obsolete when new equipment arrangements and options are introduced by the vendor. If the customer believes that technology is changing rapidly, then it would be prudent to allow the shared hub operator to make the investment and take the associated risk. They are also in a position to move equipment around among users as their specific needs change.

- *Capital commitment is not possible.* Organizations sometimes find themselves in a position where they cannot make investments outside of the core business. Money might more appropriately be spent on adding a new production line or retail outlet as opposed to investing in a telecommunications facility.

9.3.3 Network Management and Control

VSAT networks with a hub have the benefit of centralized management and control, provided through the same outbound channel and inbound channel transmissions that support user traffic. Consequently, all suppliers of this equipment offer their particular computer software and workstation arrangement to facilitate good operation of the entire network on a consistent basis. Without introducing a significant amount of overhead, these network management arrangements allow an NOC to know the status of every VSAT in the network and can perform a variety of tests over the satellite itself. Some permit the remote location of the NOC, using a leased line or Internet connection to transfer monitor and control information.

The previous remarks relate to the baseband and protocol processing portion of the network. Hub RF equipment is almost always dealt with separately through proprietary hardware and software supplied either by a systems integrator or manufacturer of monitor and control systems. As a result, there will be a bifurcated style of

management, where one part of the NOC supports the hub baseband and remotes, and the other deals with the hub RF equipment. On top of this, there would normally be a separate NOC function for the routers, switches, and multiplexers in the network, which provides standard processing of network users and traffic. This adds yet another layer of complexity, but is something that most IT and broadcast organizations already understand quite well.

9.4   VSAT Networks at Ka-Band

Moving from C- and Ku- to Ka-bands is simple in some ways (it is just another piece of microwave spectrum) and difficult in others (commercial experience is less and rain attenuation poses a greater challenge). There are a variety of reasons why operators and manufacturers are pursuing systems above 18 GHz:

- *More bandwidth available:* as much as a total of 2,000 MHz, in a relatively unoccupied section of the spectrum.
- *Smaller antennas:* for the same diameter, beamwidth reduces by a factor of 2 to 3 relative to Ku-band, and adjacent satellite interference is decreased by approximately 6 to 10 dB. Designers take advantage of this by reducing antenna diameter rather than reducing satellite spacing.
- *Satellite antennas are capable of generating small spot beams:* allows greater frequency reuse and increases the EIRP for the same HPA power. Designers take advantage of the latter by using lower power TWTs or SSPAs.

Getting to these benefits does not come for free, as there is a fairly high price to be paid in the near term:

- *Atmospheric absorption is greater:* up to 1 dB in clear weather as opposed to less than 0.5 dB below 15 GHz;
- *Extreme environment of rain attenuation:* total degradation is three to four times (in decibels) that at Ku-band. This requires a combination of strategies, including the following:
  - Using higher satellite EIRP;
  - Not reducing VSAT dish size below about 75 cm;
  - Employing greater forward error correction, which uses bandwidth less efficiently but saves on power;
  - Providing an adaptive link, which can change the data rate and coding rate during rain;
  - Allowing more frequent outage: that is, designing for 99.5% (rather than 99.9%) availability at the VSAT. The hub side can be helped by using site diversity (e.g., a pair of receive locations, spaced at least 10 km apart).
- *Lower antenna efficiency for diameters greater than about 5m:* hub antennas will generally be very costly and require tracking;
- *Hardware that is more delicate and generally performs less effectively (lower dc-to-RF efficiency and higher noise) than Ku-band counterparts:* prices are

quite high in early markets due to cost of development and manufacture. This may well improve as Ka-band moves to a mass market.

These factors neither guarantee success nor undermine it completely. What one must do is design a network that meets the needs within the financial and operational constraints of the market. Illustrated in Figure 9.24 are Ka-band hub and remote antennas with diameters of 3.4m and 75 cm, respectively. These devices are designed and manufactured to maintain the necessary mechanical tolerances to minimize loss due to surface accuracy. The VSAT includes a typical RF head that operates at Ka-band with translation to L-band over the interfacility link.

The majority of the VSAT data networks to date employ the star architecture to provide effective networks through area coverage bent-pipe transponders. In moving to the Ka-band, designers are working with more complex satellite repeaters including those with digital onboard processing. As discussed in Section 9.1.3.5, a specialized access protocol like ATM would probably be used to facilitate beam-to-beam traffic routing. Therefore, the baseband systems at the hub and remote will be different in many respects from what was discussed in this chapter. Currently, no such network is in operation (except for the demonstrations by NASA during the ACTS program). At the time of this writing, three systems were in some stage of development and could reach orbit before 2010. One could be launched by 2005 by Telesat Canada on Anik F-2. As shown in Figure 9.25, the satellite covers North America with 45 beams intended for VSAT broadband access. This will demonstrate how the combination of new network technology, advanced satellite repeater design, and the challenge of Ka-band systems can compete in the VSAT market.

9.5 Suppliers of VSAT Networks

The market for VSAT equipment is diverse in terms of the number and types of suppliers. Most networks are based on proprietary technology from leading companies, but there are two standards—DVB-RCS and DOCSIS—that could impact the

Figure 9.24 Typical 3.4-m hub antenna and 75-cm remote antenna designed for Ka-band service. (Courtesy of NASA Glenn Research Center.)

Figure 9.25   Ka-band spot beam coverage for Anik F2. Gateways (GW) are provided in selected beams. (Courtesy of Telesat Canada.)

business in coming years. A key issue for the prospective buyer is the long-term viability of the chosen supplier. However, if the supplier can deliver the desired capability now and the user is technically able to maintain the system on a continued basis, then long-term supplier viability is less of an issue. Perhaps more important in this case is the performance and reliability of the equipment coupled with the available support from the supplier during the time that the network is being brought into use.

To give the reader an idea of the scope of resources, Table 9.5 provides a selected listing of VSAT suppliers that were addressing this market at the time of this writing. The companies on the list have delivered products to the market and are viable in terms of their technology.

The process of selecting a supplier is as complex as one wants to make it. A good strategy is to make sure that you understand your requirements very well before looking into suppliers. The next step is to gather data on the suppliers and to learn as much as possible about what can and cannot be done with currently available equipment, software, and services. This can take several months, particularly if the network is to be installed outside of a major developed country. A consideration also is the quality of service and support available for the supplier. Since the VSATs and hub are an integrated system that must handle a range of services and computing facilities, a safe approach is to rely on the larger companies like Hughes, Gilat, and ViaSat. They have addressed more customer situations than the other companies and hence have the software and expertise that can get your network into service as quickly as possible. On the other hand, a specialized supplier like ND Satcom, Shiron, iDirect, and Aloha Networks may have the precise product or software that

Table 9.5 Selection of VSAT Networks Suppliers in the Global Market

Supplier	Capability	Product Lines	Standard Support
Hughes Network Systems	Full range of VSAT and wireless products	PES, TES	Proprietary, TDMA
Gilat Satellite	Full range of VSAT and wireless products	SkyStar Advantage, SkyBlaster	Proprietary and DVB, TDMA
ViaSat	VSAT products	LinkStar, ArcLight, WildBlue	Proprietary and DVB, TDMA and CDMA, DOCSIS
Tachyon	VSAT network	VSAT	Proprietary IP, TDMA
Shiron	Interactive VSAT product line	InterSKY	Proprietary and DVB, FDMA
iDirect	Interactive VSAT product line	Netmodem II	Proprietary, IP-based
ND SatCom	Mesh VSAT	SkyWAN	Proprietary, TDMA
EMS Technologies	Wide range of satellite communications products; leading supplier of DVB-RCS products	VSAT	DVB-RCS TDMA
NewTec	Leading supplier of DVB-RCS products	VSAT	DVB-RCS TDMA
Aloha Networks	New supplier	VSAT	Spread ALOHA
VIPERSAT	Standard mesh VSAT	VSAT	IP and DVB, TDMA

your network requires and may make the most attractive offer in financial terms. For example, Shiron employs an FDMA return channel that provides bandwidth on demand. This enables not only the assignment of fixed-size satellite bandwidth segments to a user, but also the allocation of bandwidth and data rate to match the application need [7]. Bandwidth on demand is automatically configured based on the immediate requirement and presets imposed by the network operator (e.g., QoS specifications).

The hardware and software of the VSAT can be improved and made less expensive over time. However, much of the overall future advance in VSAT network architecture will come as the satellite repeater moves from a simple bent-pipe to an onboard processor like what is carried on modern MSS satellites. The processor can adjust gains and data rates between the hub and the remote. For example, instead of having a high data rate on the outbound channel and a low data rate on the inbound channel, the roles could be reversed with a remodulating type of repeater. With greater bandwidth on the inbound channel, the network could support greater information transfer for data collection and video. There is also the potential for full mesh connectivity, using the satellite as the switching point in the sky.

Consumer VSATs appeared on the U.S. market around 2000 to compete in the broadband Internet access market. The real cost of these units was estimated to be $1,000; yet they sell for around $500 when taken with a 1-year commitment to the service. The other aspect is the performance of the satellite itself. We see many organizations advancing plans for satellite-based digital communication networks capable of delivering megabit-per-second speeds directly to users. Hub Earth stations may or may not be required, depending on the application. It would be possible to implement such networks at Ka-band, relying on multibeam coverage and

high power to overcome rain fading and to provide adequate signals to antennas as small as 60 cm in diameter. The technology base is there to do these things today; all we need is the business base to make it a success.

References

[1]   Elbert, B., (guest ed.), "Satellites Address the Digital Divide," *Online Journal of Space Communication,* Issue 5, July 2003, http://satjournal.tcom.ohiou.edu/issue5/main.html.

[2]   Stallings, W., *Business Data Communications*, 4th ed., Upper Saddle River, NJ: Prentice Hall, 2001, p. 214.

[3]   Rosener, D., et al., "A Robust, High Performance Network Provided by Astrolink," *AIAA International Communications Satellite Systems Conference*, Montreal, Canada, 2002.

[4]   Choi, H.-K., et al., "Interactive Web Service Via Satellite to the Home," *IEEE Communications Magazine*, March 2001, p. 182.

[5]   Lee, J. S., and L. E. Miller, *CDMA Systems Engineering Handbook*, Norwood, MA: Artech House, 1998, p. 368.

[6]   Elbert, B. R., *Networking Strategies for Information Technology*, Norwood, MA: Artech House, 1992.

[7]   Barda, A., Z. Ner, and S. Laufer, "Closing the Digital Divide Through Resource Allocation Techniques," *Online Journal of Space Communication*, Issue 5, July 2003, http://satjournal.tcom.ohiou.edu/home.html.

Fixed Telephony Satellite Networks

Satellite links have long been used to carry telephone calls to extend the reach of public and private networks. This role has changed over the years due to the adoption of digital transmission and fiber optic systems throughout developed countries. During the 1990s, undersea fiber was introduced on a global basis and so the original demand for satellite telephony has diminished. Another trend that has been underway is the rapid economic expansion of many underdeveloped countries throughout Asia, Africa, and Latin America, as well as increased inhabitancy of rural areas of developed countries. Owing to the Internet, the telephony sphere is undergoing a transformation in the form of VoIP technology for connecting calls through this pervasive medium.

The cost and time required to extend the terrestrial network to these areas amount to a large barrier to development. Cellular and mobile satellite networks cover some of this gap, but they are expensive in relationship to income levels in developing areas. There is now a growing awareness of and market for satellite-based telephony networks to fixed locations such as homes, small businesses, and branch offices. These have proven themselves in years past, and the technology used in low-cost VSATs as well as third generation cellular systems is being applied to greatly improve the flexibility and cost-effectiveness of these fixed telephony satellite networks.

The focus of this chapter is on the application of satellite networks to public and private telephone service, emphasizing its attractiveness in rural and underdeveloped areas. By fixed telephony we mean that the Earth stations are located on the ground and serve stationary users. The idea is to use a GEO satellite and ground segment implemented in the classical manner to replace local loops, switches, and trunks that serve subscribers and end users. The intriguing feature of FTS is that the switching function is provided by the satellite, either using SCPC or TDMA. Most FTS networks have employed constant bit rate speech transmission; however, the rise of VoIP is significant because variable bit rate speech transmission becomes the norm in most organizations. The technologies and methodologies have been covered previously, so we concentrate here on the special situations and requirements for the FTS application. In addition, the following chapter on mobile satellite service covers a closely related area and should be considered to be part of the solution set for extending telephone services to literally every corner of the planet.

10.1   Role of Satellites in Telephone Services

One of the most important roles of FTS is to fit into the conventional PSTN architecture. Later, we will review how FTS can supplement or supplant a private voice network used to provide internal telephone services. In either case, the FTS must interface and interoperate with the PSTN. As shown in Figure 10.1, the PSTN is arranged as a hierarchy of telephone exchanges or central offices, which we simply refer to as switches. The local exchange is the first point of entry into the PSTN, serving each subscriber on a separate local loop. The subscriber local loop (A-1 connected with line 5 in the figure) would normally be implemented with twisted pair wire and extend a distance of a few kilometers. A radio-based technology called wireless local loop (WLL) is an alternative in rural areas and overbuilt cities to extend local telephone service quickly and cheaply. FTS addresses this situation by extending the local loop hundreds or thousands of kilometers.

Many businesses employ privately owned telephone switches called PBXs to concentrate telephone usage within a building or campus environment. This improves the effectiveness of telephones and reduces the number of required local loop lines, as indicated for subscriber A-2 in Figure 10.1. A FTS capability could be used to interconnect the PBX with the local exchange or to connect several PBXs together into a private network. The benefit of doing this is that FTS links can be implemented for efficient calling (e.g., using abbreviated dialing) even if the PSTN is

Figure 10.1   A typical PSTN hierarchy, showing local subscribers, a PBX, local offices, toll offices, and international gateways. Satellite links and FTS can be applied for any links, both permanently assigned and demand assigned.

not in existence within the area of interest. Also, such lines can bring a remote location into an existing private network and thereby improve integration.

FTS links can be implemented within the PSTN itself, particularly in regions where the construction of terrestrial lines is impractical or too expensive. Examples would be in highly mountainous areas, over barren deserts, or to remote islands. These multiple voice channel trunks extend between a local office and a domestic toll office (4) and between toll offices (3), and would be dedicated T1 or E1 links to support traffic carrying requirements. Access to the international gateway (2) is also a candidate for a dedicated satellite link. Figure 10.1 shows PSTN connections between two countries (1) that can use the satellite capacity on a global basis. Traditional telephone switches and transmission systems have employed time division multiplex at a constant bit rate. With the rise of packet switching, many long distance networks have moved to a combination of ATM and IP to offer greater flexibility and better integration with other data networks and applications.

10.1.1   Domestic, Regional, and International Services

That satellite links continue to be used in telephone networks should be no surprise. Telephone communication is a common denominator throughout all communities of the world, yet many locations are still lacking in quality service. The complimentary nature of the telephone hierarchy and FTS provides an excellent match of need and capability. Potential connections and their applicability to FTS are indicated in Table 10.1. The number in the left column corresponds to the link shown in Figure 10.1 and is comparable to the class of exchange where the link terminates. Domestic FTS was introduced in the United States around 1980 so new competitors like Satellite Business Systems (SBS) could enter the long distance telephone market. This died rather quickly after fiber optic systems were installed between major cities. This is because the cost per telephone channel of a high-capacity fiber cable is considerably less than a comparable satellite link, provided that the fiber capacity is fully utilized. However, FTS is still important in private networks and for addressing the telephone needs in Alaska and U.S. territories in the Pacific Ocean. Indonesia, China, and Russia are highly dependent on FTS to tie together the far-flung reaches of their respective countries. While cable and microwave are effective transmission means, FTS proved from the beginning that it was more cost-effective and practical for long thin routes.

We see that two modes are available to achieve connectivity in domestic FTS systems: demand assigned and preassigned. In Chapter 8, we introduced the concept of demand-assigned bandwidth in VSAT networks as a very useful technique for

Table 10.1   Application of Satellite Links for Circuits Used in Fixed Telephony Networks

Class	Purpose	Technology
5	Connect subscriber (or PBX) to local exchange	Demand assigned, SCPC or TDM/TDMA/
4	Connect local exchange to toll (long distance) office	Demand or preassigned
3	Connect toll office to toll office (interexchange)	Demand assigned or preassigned
2	Connect toll office to international gateway	Preassigned
1	Connect international gateway to international gateway	Preassigned

improving the loading of the satellite network. This was done using two technologies: TDM/TDMA with bandwidth reservation and FDMA (e.g., SCPC with DAMA). Both approaches are available for FTS as well. In SCPC transmissions, each voice channel is assigned to a particular frequency for the duration of the call. Frequencies are maintained in a pool of transponder bandwidth and managed by a central network control system. This is the same function as a telephone exchange, except that it is performed at a national level (e.g., as if every user were connected directly to the country's national switch). FTS can therefore bypass all of the exchanges in Figure 10.1, except the international gateway. A classical demand-assigned FTS network is illustrated in Figure 10.2. It includes two types of individual terminals, a multichannel terminal connected locally to a PBX, and a gateway to the national PSTN. The master station at the lower right manages the shared satellite bandwidth and responds to call requests from all of the traffic stations.

There is a caution with regard to Figure 10.1 and Table 10.1, which is that not more than one satellite hop should be included in any end-to-end connection. This insures that the total delay remains below 400 ms. Otherwise, the quality of telephone service will be degraded to the point that a significant fraction of subscribers will be dissatisfied. Such a double hop is avoided through what is called "class marking" of individual calls: any call that includes a satellite hop for any of the connections indicated in Table 10.1 will be ineligible to be placed on another satellite connection. Of course, there are circumstances where a double hop cannot be avoided, such as when an individual is using a fixed or mobile satellite link from an isolated location (e.g., a double-hop is better than none).

Figure 10.2 Typical arrangement of a demand-assigned FTS network with central network control and management.

10.1.2   Estimating Telephone Traffic

Satellite links used in FTS fall into the category of either trunking or thin route. As suggested in Figure 10.1, a trunking application provides links between telephone exchanges at the higher levels of the hierarchy. They tend to support multiple channels of telephone transmission, using T1, E1, or greater bandwidth. Individual voice channels within the trunk group may use fixed time slots (TDM) or asynchronous packet transfer (IP or ATM). Consequently, thin-route telephony responds to the need for only a few voice channels of transmission. With demand assignment, these channels can be used to connect to multiple destinations so that specific links do not have to be arranged ahead of time. The actual manner in which telephone channels might appear on the satellite is indicated in Figure 10.3. Here, we have a collection of SCPC carriers occupying most of the bandwidth of a transponder. These channels can be assigned by network control in response to requests for calls by individual FTS stations. The larger carriers on the right represent the control frequency channels, which are used by the stations to request channels and for network control to transmit the corresponding frequency assignments to the calling and called stations.

The previous discussion implies the existence of a signaling function within the overall PSTN to perform the following functions:

1. Respond to an off-hook condition when a calling subscriber wishes a dial tone;
2. Receiving and processing the dialing digits that identify the called subscriber (who is remote to the caller and possibly at a location connected through the PSTN);
3. Passing back the indication that the called party has answered or is occupied (busy);
4. Passing back the indication that no channels are available over the network (fast busy);
5. Collecting call information for billing and management purposes;
6. Network state and performance monitoring and control.

The FTS network must accommodate and even use the existing signaling systems; otherwise, calls will not be properly processed and subscribers will be left high and dry. An "in-band" signaling system employs the same voice frequency channel

Figure 10.3   Demand assignment FTS scheme, using transponder bandwidth divided into single-channel frequency slots. Power is allocated to each SCPC channel, allowing for transponder intermodulation distortion. A master station assigns channels form the "pool" to conversations.

as the subscriber to carry the majority of the six functions listed above. The other approach, "out-of-band" signaling, passes the associated signaling data through an independent network not employed for telephone traffic.

The satellite transponder in Figure 10.3 represents a bundling point for all communications in an FTS network of the type illustrated in Figure 10.2. This is unique in telecommunications because the transponder provides one trunk group that is shared by all, regardless of the calling and called station locations. The principle of bundling at the transponder level is shown in Figure 10.4. The collection of SCPC channels provides such a bundle since any FTS station may employ any channel to connect to another station for call connection. This provides the highest degree of efficiency in terms of traffic as will be discussed later. There is a similar bundling of traffic at the individual FTS station that can access multiple satellite channels. For example, it does not require 20 transmission channels to serve a local community of 20 individual subscribers because of the infrequent nature long distance calling.

The telephone traffic unit of measure is the Erlang (E), which is equal to an average load of one channel during the busy hour [1]. Calculating Erlang quantity involves a process of measuring calling activity on each telephone circuit in a trunk group. In a busy hour, if we measure that the circuit is busy 30 minutes, then the load is precisely 0.5 E. If the trunk group carries 10 such channels, then the group is carrying 5 E. Development of this methodology and the mathematical models to follow is credited to A. K. Erlang, a Danish telephone engineer who lived in the early twentieth century. The traffic requirement between a pair of thin-route service locations can be a small fraction of 1 Erlang, but the efficiency of handling traffic in terms of Erlang will improve as more circuits are added to the trunk.

Figure 10.5 provides the basic model of a telephone switch connected to a trunk group. On the left a large number of subscribers have access through the switch; however, the actual quantity of channels within the trunk group is considerably less. In mathematical terms, the number of Erlangs of load is $A = \lambda/\mu$, where $1/\lambda$ is the average idle time between calls and $1/\mu$ is the average call holding time. (The variables λ and μ are the arrival rate and the call completion rate, respectively.) Another way of looking at A is that it is the average utilization in terms of the number of channels. Because the calls are coming from outside sources (presumably subscribers), we are concerned that there be enough channels to serve the peak load as well as the average. To determine the peak load, we must first specify the grade of service, which is measured by the probability of blockage, P_b. A standard value of P_b in telephone networks is 1% or 0.01. In words, this states that there will be one blocked call out of 100 attempts (e.g., 99 successes). Cellular networks are designed for two

Figure 10.4 FTS networks resemble telephone switching networks, where conversations occupy voice channels. Bundling of the trunks allows more efficient channel usage and the application of standard traffic engineering principles.

times this value, in recognition of the higher infrastructure cost per subscriber and value of the radio spectrum.

The following formula relates the grade of service to the average utilization and number of available channels. Any demand that exceeds the peak cannot be served, resulting in blocked calls. The probability of this occurring is

$$P_b = \frac{A^N/N!}{\sum_{n=0}^{N} A^n/N!} \qquad (10.1)$$

where N is the number of available channels in the same group. In standard telephone networks, a group would consist of a bundle of channels that connect the same pair of telephone exchanges together. The more channels in the group, the greater the relative efficiency of handling calls. When we look at a demand-assigned SCPC network, the group consists of all of the available frequency channels on the satellite. For TDMA, the channel bundle consists of time slots, which can achieve the same class of bundling efficiency.

This formula is highly nonlinear; Figure 10.6 provides a solution curve for a range of 1 to 100 channels in the trunk group and with a grade of service of 0.01 (1 out of 100 calls blocked) [2]. A separate curve for channel groups between 1 and 10 is presented so we can examine how a satellite can provide bundling efficiency. Take the following simple example:

- Consider an FTS network of 20 Earth stations equipped with five telephone channels each.
- We are to determine the required number of satellite channels to deliver a grade of service of 0.01.
- To make this determination, find the value of A, in Erlang, which five channels will support:
  - This is approximately 1.4, according to the curve at the upper left.
  - Twenty stations can generate 20 times 1.4 Erlang, which is equal to 28 E.
- The solution to the problem amounts to determining the number of satellite channels that would support 28 E at the required grade of service:
  - Use the curve at the lower right of Figure 10.6.

Figure 10.5   The basic model of how telephone calls are switched between calling and called parties. The call generation and completion rates are λ and μ, respectively. The inputs to the switch (or collection of shared telephone channels) are from i individual subscribers, but there are only N available channels to service calls.

Figure 10.6   Performance curves for use in traffic engineering. This is based on the Erlang B formula (blocked calls dropped), with a grade of service of 0.01, or 1 out of 100 calls blocked.

- Enter the curve along the *y*-axis at 28 E.
- Find the number of channels that can satisfy this traffic load, which is 40.

What this little example shows is that 20 stations with five channels each (a total of 100 channels on the ground) can be served effectively with only 40 channels on the satellite. The ratio of 100 to 40 (i.e., 2.5) is the bundling efficiency. A few observations need to be made about this result:

- Both the Earth station and the transponder will introduce call blockage of 1 per 100; thus the overall call blockage will be 2 per 100 call attempts.
- Most calls will be handled without experiencing a busy due to all channels previously being in use.
- This analysis is for the busy hour; all other times will provide even better grade of service.

This concept also applies to cellular telephone networks, but the irony is that satellite-demand-assigned FTS existed long before modern cellular telephony was even conceived. In cellular networks, the bundling occurs at the base station for a specific cell. Users in cell A, for example, who experience a lack of channels cannot use available channels in cell B because the associated coverage is on different frequencies or out of range. This is not true for a bent-pipe GEO satellite, since all channels are available to all users in the same footprint.

The topology of FTS networks is almost always full mesh in nature, as compared to the star topology of the VSAT. By avoiding a double hop, FTS is able to deliver service at a quality that is comparable to the terrestrial PSTN. When employing GEO, the ITU has recommended that the double hop be avoided. This particular rule can be broken in MEO and LEO satellite networks (covered in Chapter 11) because their lower altitude allows a double hop to meet typical delay standards. To avoid a double hop in GEO systems, a domestic FTS that is destined for another country should be connected over an international terrestrial link.

Another important but often overlooked aspect of telephone quality is the post-dialing delay (also called the connect delay). This is the time interval from when the caller finishes entering the phone number and the phone on the other end begins to ring. There are several contributors to the postdialing delay, some of which are fixed in duration and others that are variable. Table 10.2 provides a hypothetical example of an end-to-end connection over an FTS link. The specific values are rough estimates of what it might take in an actual system, yielding a total of nearly 5 seconds for a call placed between parties that are directly connected to the FTS network. In reality, the postdialing delay will be determined by the nature of the particular connection. If, for example, one user is directly connected to one node while the other must be reached through an international gateway, several seconds would be added to account for the greater complexity. User expectations for postdialing delay vary according to their individual situations. The estimate is a sum of constant and random variables; if one of the users is reached over the PSTN, the time for this portion of call setup adds to the total as well.

Table 10.2   Hypothetical Example of an Estimate of Postdialing Delay

Contributor	Symbol	Example Value (seconds)
Deliver dial tone to caller connected to node A	T_s	0.50
Processing at node A	$E(T_p)$	0.25
Waiting time	$E(W)$	0.35
Connection request	T_1	2.00
Processing at node B	$E(T_p)$	0.25
Ringing message	T_s	0.50
Processing at node A	$E(T_p)$	0.25
Answer message from called party connected to node B	T_s	0.50
TOTAL time $T_c = 3T_s + T_1 + 3E(T_p) + E(W)$	T_c	4.6

Note: values indicated as T_x are constants, while $E(T_y)$ indicates the expected value (average value) of a random variable.

10.1.3   VoIP

VoIP is a very broad application category designed fundamentally to transmit voice conversations over a packet data network using the Internet Protocol. The specific network infrastructure could be either the public Internet or a corporate Intranet. According to *Newton's Telecom Dictionary,* there are several potential benefits to moving voice over a packet data network using IP:

1. You may save some money.
2. You may achieve benefits of managing a voice and data network as one network.
3. If you have IP phones, "adds, moves, and changes" will be easier and cheaper.
4. You have simplified call setup, tear down, and transfer of calls.
5. The key attraction is added (and integrated) new services, including integrated messaging, bandwidth on demand, voice e-mails, and the development of "voice portals" on the Web.

Form one point of view, VoIP is nothing more than another use of the Internet, along with e-mail, Web site services, file transfer, and virtual private networks. However, like videoconferencing using the H.323 standard, VoIP represents a departure for the Internet since it requires end-to-end connections with minimum delay. It is not a given that VoIP services will satisfy all or most of the five points listed above, simply because the Internet was not designed as a real-time medium like the traditional PSTN. Nevertheless, corporations, nonprofit organizations, and government agencies want a greater return on investment for their Ethernet LANs, Frame Relay and ATM WANs, and high-speed connections to the Internet. The same can be said of VSAT and other satellite networks and the transponder capacity committed to broadcast and data communications. For these reasons, it has become imperative that FTS systems now address VoIP in a consistent manner.

Our context for VoIP will be that of an integrated solution for providing voice telephony services in the same manner described previously for FTS. This means that the technical features of VoIP must address the following:

- Speech compression to reduce the bit rate below 32 Kbps without impacting voice quality;
- Use of RTP to provide packet transfer within acceptable delay constraints;
- Requirement for a VoIP gateway to connect between the private network, a PBX, and the PSTN;
- Complex management of telephony service akin to the PBX;
- Interface properly within a VSAT network to support the protocol, compression, and call processing.

Important to realizing VoIP in an FTS network are the standards that address the above features. Table 10.3 provides a summary of the two areas where standards are useful: the application layer protocols and voice compression. Currently, there are three application layer protocols competing for market share. Their maturity and extent of deployment vary inversely to the order presented (e.g., H.323 is the

Table 10.3   Application Layer and Voice Compression Standards Employed in VoIP Networks

Application Layer Standards	Voice Compression Standards
H.323 – ITU-T	G.711 – PCM at 64 Kbps, no compression, toll quality (MOS > 4)
V1 and V2 (faster setup) Best current end-to-end solution Should work with a single vendor	G.726 – ADPCM at 32 Kbps, sounds exactly like G.711 (MOS > 4)
SIP–IETF	G.729 – Conjugate Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP) – 8 Kbps, MOS ~ 4 (near toll quality)
Coming up fast, could be better than H.323	G.723.1 – Maximum Likelihood Quantization (MLQ) – 5.3 to 6.3 Kbps, MOS < 4 (not toll quality)
Multivendor approach used by telecommunication equipment suppliers and Internet specialists	Latency/delay – ITU "carrier class" 150 ms (compression adds ~50 ms and packet process adds ~100 ms; greater than 400 ms is deemed unacceptable for a commercial service)
MGCP–Megaco/H.248	
Also an Internet RFC	
H.248 to be an ITU-T standard	

most common, by virtue of extensive experience gotten with videoconferencing service). Cisco and other vendors are supporting Session Initiation Protocol (SIP) and Media Gateway Control Protocol (MGCP) and it is likely that one or the other will win out as the basis for VoIP through 2010. On the voice compression side, the user-perceived quality decreases with the bit rate. As will be discussed in Chapter 11, quality is assessed in terms of a subjective measure called the *mean opinion score* (MOS). Toll quality, as delivered at rates in the range of 64 to 16 Kbps, is what we expect of the terrestrial PSTN. Compression at 8 Kbps provides what is termed near-toll quality as its MOS is only slightly less than what is obtained above 12 Kbps. Dropping below 8 Kbps almost always brings with it the risk of user dissatisfaction as to what is expected in a commercial telephone service. The same can be said of end-to-end latency (one-way delay) that exceeds about 400 ms, considering all sources.

Another important issue to consider is the overhead required to adapt constant bit rate digital telephone transmission to the packet structure of the IP. Figure 10.7 provides the family of protocols used for the standards indicated in the above table. It can be seen that using IP is by no means a simple matter, which is why careful analysis and verification testing are vital for proof of the final product. We see from Figure 10.7 that a telephone call involves the use of two session layer protocols on top of IP: TCP for signaling and UDP for gateway control and compression/voice transfer. Once the connection is established, voice packets are transferred using the RTP over UDP. This adds overhead in the range of 100% to 200%. As a result, an 8-Kbps basic speech compression rate is expanded to 16 to 24 Kbps over the satellite link. The positive side of this trade-off is that these packets can be intermingled with other traffic to achieve efficient usage of an IP transmission path. The compression at 8 Kbps will hardly be noticeable and the delay will, hopefully, be below the suggested limit of 400 ms.

An example of a hypothetical mesh FTS that provides VoIP telephony is provided in Figure 10.8. The satellite portion employs the LinkWay mesh TDMA

Figure 10.7   The protocol stacks involved in providing a complete VoIP service using the IP suite. (Courtesy of Shiron Satellite Communications.)

MRT – master reference terminal
NMS – network management station
NCC – network control center

Figure 10.8   An example of the application of VoIP in a VSAT network. Equipment in the example consists of a ViaSat LinkWay and Ericsson WebSwitch.

system provided by ViaSat and the VoIP gateways are Ericsson WebSwitches. A master reference TDMA terminal is provided for LAN1, on the right, which delivers VoIP packets to the WebSwitch 2000 gateway. Individual telephones are directly connected to the WebSwitch. The other end of the satellite connection at LAN2 employs a LinkWay traffic terminal and WebSwitch 100; this same configuration would be found at the other remote sites in the overall network.

10.1.4   Interfacing to the Terrestrial Telephone Network

The satellite network can be designed for efficient operation and good commercial quality. However, the most critical problem encountered in the application of FTS is

the proper interfacing of the service to the terrestrial PSTN. Telephones worldwide are nearly the same, as far as the twisted wire connection (the local loop, 1, in Figure 10.1). The problems come when using FTS for the higher level connections (2, 3, 4, and 5) where information is coded and signaling features are included. With a variety of implementations of ITU Common Channel Signaling System No. 7 (SS7) and analog signaling systems like R-1, R-2, and SS5, many countries have multiple interfaces in operation within their domestic borders. This makes the interface into a potential bottleneck and quagmire.

An understanding of the importance and complexity can be viewed in Figure 10.9, which indicates the range of interface options encountered in FTS networks. Support of thin-route service and private networks is indicated at the bottom in terms of the interface to end-user equipment. This is comparable to what is possible with VSAT star networks (covered in Chapter 9). The fact that FTS also supports direct connections to telephones, computers, modems, fax machines, and the PBX reflects the importance of satellite communications in business and government. In many cases, the satellite network provides the primary means of communication and therefore must be at least as flexible as the telephone network.

The use of FTS in the PSTN is indicated at the top of Figure 10.9. Each type of telephone exchange corresponds to a level of the PSTN hierarchy shown in Figure 10.1, covering local, toll, and international gateway interfaces. A brief summary of some of the appropriate technical interface standards that apply in each case is provided in Table 10.4. This is a summary and not a design guide because the details vary significantly from situation to situation.

The telephone interface brings with it unique electrical, bandwidth, and signaling characteristics that must be reflected in the FTS Earth station, gateway, and network control. The interface to the local exchange is like that to the PBX and could

Figure 10.9   The range of interfaces to be employed in FTS applications, considering the connection to user devices as well as various options for the interface to the PSTN.

Table 10.4 Typical Interface Standards for Connection from an FTS Network to the PSTN or Private Network

Facility	Type of Access	Standard	Requirement
Private: end user	Telephone	Two-wire, touch tone (DTMF)	Response to off hook/on hook and tone
	PC data	RS-232	Typically 50 Kbps
	ISDN	Basic rate interface (BRI)	Total 144 Kbps; terminal adaptor for PC
	Modem/fax	Same as telephone	Bypass speech processing and compression
	PBX	Four-wire and E&M; alternatively T1 or E1	Private trunk
PSTN facilities	Local exchange	Two-wire touch-tone (DTMF); alternatively four-wire and E&M, T1 or E1	Range of options considers a wide variety of local conditions
	Toll exchange	Four-wire and E&M, T1 or E1	PSTN trunks must comply with national standard(s)
	International gateway	Domestic access, T1 or E1	Must comply with international network

be considered to be the same in many cases. However, there are many options here, based on the local telephone network standard and the types of switches in existence in the particular country. One may even find manual switches (e.g., switchboards and operators) in some underdeveloped areas. Going forward, all trunk interfaces support SS7, a highly robust and feature-rich out-of-band signaling system.

The toll exchange interface is almost always digital in nature and follows a national standard. The exchanges themselves are supplied by world-class manufacturers, and many were installed recently as part of major network expansions. Therefore, this is perhaps the easiest interface for the FTS implementer. Because this is at a rather high level of the hierarchy, we can expect the capacity requirement to be in multiples of, say, 10 channels.

Interconnection to an international gateway can occur either on the domestic side or on the international side. On the domestic side, the FTS interface is basically the same as that for the toll exchange, following the domestic standard for transmission and signaling. Interfacing on the international side means that the FTS link is part of the global network such as that of Intelsat or PanAmSat. This implies that the international transmission plan and signaling system is in use. Operators of these types of Earth stations and networks comply with the appropriate ITU-T standards. By following these rules and standards, the operator of the FTS facility will have many of its questions answered and should be able to avoid the risk of incompatibility. An exception exists when an independent operator wishes to compete with existing FTS operators and hence must find a means of entry into the market. The consequence of this is that the only standard that exists is that required to make the connection at each end. VoIP and the SIP and MGCP standards identified in Table 10.3 will play an ever-increasing role here.

To clarify some of the abbreviations of Table 10.4, the two-wire type of interface is that normally found on subscriber telephones. The common standard today is the RJ-11 modular jack, which actually has provision for six wires. Voice frequency

information passes over one pair in both directions at the same time. The Earth station detects when the phone goes off-hook or on-hook and provides ringing as well. Within the FTS baseband equipment, a device called a hybrid separates the send and receive energy so that it can be connected to the corresponding parts of the Earth station (e.g., send must be routed to the uplink and receive to the downlink). The hybrid is a transformer device that can introduce an undesired return path for audio signals. This results in an echo that can be heard at the distant end over the satellite link. Because of the quarter-second delay, even a low level of echo on a satellite link is very objectionable. The advent of the digital echo canceler has made it possible for satellite telephony to become competitive with terrestrial phone service. The ITU-T has specified the digital echo canceler in its Recommendation G.165 for use in all long distance connections (terrestrial and satellite). Every FTS installation will include echo cancelers either as part of the Earth station or within the associated telephone exchange where it can service long distance terrestrial circuits as well. In integrated digital networks, the echo cancellation function is often performed on a bulk basis on T1/E1 circuits.

The usable bandwidth within the analog voice channel extends between 300 and 3,400 Hz, which is an international standard. It is capable of transferring speech, low- and medium-speed computer data with the use of V.90 and V.92 modems, and subscriber signaling through the touch-tone pad. The touch-tone signaling approach is officially referred to as the discrete tone/multifrequency (DTMF) code. In this scheme, the telephone number digits can be sent over the active voice channel using pairs of audible frequencies.

In the case of a four-wire connection, the send and receive are separated from the origination point, eliminating the hybrid. In a totally private network, it might be possible to use four-wire telephones and other devices to provide the best possible transmission quality and eliminate the potential for echo. The other aspect of a four-wire transmission to the exchange is that on-hook and off-hook indications may be conveyed over a separate pair of wires call E (for ear) and M (for mouth). The M lead interface to the Earth station accepts line and address signaling by detecting dc signaling state changes. The complementary function is performed by the E lead, which generates output line and address signaling. This also improves network performance and simplifies interface design. It still is possible to use DTMF to transfer dialing digits and other network information.

The signaling system allows the network to transfer dialing digits, supervisory information, and customer care information. The R1 system was generally employed in North America and South America, while the R2 system is the analog standard for Europe and much of the rest of the world. Both depend on DTMF to transfer dialing digits and the E and M leads to service the circuit. However, they perform DTMF signaling differently in that the standard version of R2 employs the compelled mode. According to this protocol, the sending point waits to end its transmission until the receiving end has responded. This introduces long time delays for each signal sent. Semicompelled R2 signaling was first implemented in Indonesia's Palapa A FTS network as a cure to this compelling problem. This uses pulses of supervisory signals that are not held for response from the distant end. Whenever DTMF signaling is used in the R2 signaling system, it should adhere to the semicompelled scheme.

With the rapid digitization of the PSTN worldwide, SS7 is the norm. This is a natural outgrowth of the popularity of the intelligent network, a forerunner to the more comprehensive ISDN standard. Intelligent networks offer services like the use of calling cards or credit cards, virtual private networks for voice and data, and flat hierarchies that need not follow the traditional structure shown in Figure 10.1. The nature of SS7 is that it is transmitted outside of the voice channel, a technique referred to as out-of-band signaling. Furthermore, the signaling information for several voice circuits is combined onto one packet-switched SS7 channel, which runs at a multiple of 64 Kbps. It would be possible to add SS7 as a signaling interface to a demand-assigned network. On the other hand, the out-of-band or common channel signaling is standard in a demand-assigned SCPC network, as will be discussed later in this chapter. The basic rate interface (BRI), indicated in Table 10.4, provides two digital "B" dial-up circuits at 64 Kbps and one 16-Kbps "D" channel for signaling.

The table covers a very wide range of options, and it is doubtful that one type of FTS interface equipment will satisfy all them without modification or reconfiguration. The approach taken by suppliers of FTS Earth stations is to make the interface portion as modular as possible. In addition, there is a high degree of software control needed to accommodate the protocols involved with the interface to the PSTN. For the international gateway situation, there are fewer standards to contend with, so the interface can be satisfied in a straightforward manner.

10.2 Demand Assignment SCPC Network Architecture

The capability of FTS is greatly enhanced through the demand assignment feature that has been adopted in thin-route networks around the world. Rather than identify specific frequency channels or time slots ahead of actual demand, the required bandwidth is made available only for the duration of the call. Channel pairs are assigned in response to call demand. This is precisely the same technique that is applied in terrestrial analog cellular systems like AMPS and TACS. In Chapter 7, we discussed the use of demand assignment in conjunction with the bandwidth reservation mode of TDMA. This is applied to FTS networks as well, but the more popular approach is to use FDMA in the form of SCPC.

10.2.1 Demand-Assigned Network Topology

Most demand-assigned SCPC networks consist of various types of Earth stations that are capable of mesh connectivity. An example of such an arrangement is shown in Figure 10.2. In response to the variety of interface situations identified in Figure 10.9, there are two classes of Earth stations, namely, the remote terminal and the gateway. The remote terminal serves private or public end users who require one or more telephone access lines into the FTS network. The gateway is a much larger Earth station that serves all remotes by providing access to the PSTN. The network is controlled from a central point, indicated by the network control system (NCS) Earth station. This could be located at a gateway or operated as an independent facility.

The use of the term "gateway" here should not be confused with an international gateway, which is a telephone exchange that connects calls to other countries. Rather, it is a gateway Earth station (GES) used to complete calls from remote terminals to terrestrial subscribers and for allowing terrestrial subscribers to call subscribers that can only be reached via the FTS network. There is also the potential for placing calls between gateway Earth stations so that the FTS network can serve at higher levels of the telephone network hierarchy (Figure 10.1).

The demand-assigned FTS network allows any station to establish an SCPC circuit to any other station. Thus, we have a fully interconnected mesh network activated on demand. Reflecting the compact design of the VSAT, a typical remote terminal is composed of an outdoor antenna and RF head and an indoor unit that contains the electronics and interface equipment. The following telephone interface options are suggested in Figure 10.9:

- Two-wire telephone;
- Data communication device (PC or terminal);
- Fax machine (not shown);
- PBX, whether two-wire or four-wire;
- Host computer.

The GES typically uses a larger antenna and RF amplifier to be able to give a high availability to the DAMA control links. One normally takes advantage of this by placing it at a busy node that will service more simultaneous calls than average. Sufficient SCPC channel units would be provided to support the maximum calling demand of the particular GES. The channel units would interface to the local or toll exchange on a four-wire or T1/E1 basis. The latter requires a multiplexer to take the individual analog outputs of the channels units and combine them into a high-speed digital stream. Alternatively, the digital outputs of the modems within the channel units can be transformed into the T1/E1 directly using a device called a transmultiplexer (also referred to as a transmux).

The third element of the topology is the NCS, which implements the demand assignment feature. It has become a common practice to put the intelligence of network control into a central facility that is operated by a service provider. In this way, one entity can manage all aspects of the network and collect the necessary call accounting information for billing purposes. There is an Earth station associated with the NCS since all control and information gathering is performed over the same satellite transponder that contains the traffic channels. The actual control channel frequencies are shown at the right of Figure 10.3. Control channel units at the NCS are available at all times to support the calling demands of the various Earth stations of the network. Also, software and computers at the NCS process the data transferred over these links and make frequency assignments to the remote stations and gateway Earth stations. Call setup time, reviewed at the end of Section 10.1.2, is mainly determined by the efficiency of this function. Attached to the NCS is a network management station that allows operations personnel to configure the network for service and deal with various problems that arise from time to time.

10.2.2 Fixed Telephony Earth Station Design

The typical FTS remote Earth station, shown in Figure 10.10, incorporates the VSAT RF terminal with SCPC transmission features to provide mesh connectivity. In this example, the Earth station interfaces with an analog telephone on a two-wire basis. This analog interface emulates the local exchange with such functions as:

- Provides dc current for the phone;
- Detects on-hook/off-hook conditions;
- Responds to DTMF signaling;
- Provides ringing functions;
- Provides two-wire to four-wire conversion with a hybrid and echo canceler.

The speech processing function within the channel unit is critical to the operation and quality of performance of the FTS service. Within it, analog speech is converted into digital data through the A/D process. This bit stream is compressed through one of the techniques in Table 10.3 to produce a lower data rate to conserve bandwidth and uplink power. In addition, error correction is introduced to improve overall performance. The output of the speech processor would be in the range of 5.6 to 16 Kbps, depending on the requirements of the service. Both ends of the connection must use the same processing. We consider various speech compression techniques in the context of mobile communications in Chapter 11.

The channel unit, as illustrated in Figure 10.11, is the core of the FTS Earth station. There is one channel unit for each active conversation. These devices are fairly complicated and, as a consequence, represent a significant part of the expense of the Earth station. A station with a large capacity will, of necessity, have a large investment in channel units. From the telephone interface, analog voice band information is digitized, compressed, and modulated for transmission. The reverse direction is

Figure 10.10 Typical equipment arrangement for an SCPC Earth station.

provided so that the channel unit is a balanced device. Within the channel unit, the speech processor provides the following important functions:

- Echo cancellation to control the echo produced in the two-wire-to-four-wire conversion process, which should satisfy the requirements of ITU-T Recommendation G.165;
- Signaling tone generation and detection to properly support the calling process;
- Speech activity detector to turn the transmitter on and off in response to the local user's voice, which reduces the average uplink power usage by a factor of 0.4 to enhance transponder capacity;
- Speech compression, using an accepted algorithm such as residual excited linear predictive (RELP), code excited linear predictive (CELP), or a variant.

The greatest data throughput is achieved when the computer device is connected directly to the satellite modem within the channel unit. Speech processing substantially reduces the information bandwidth and thus will degrade the audio signal from a conventional voice-band modem or Group 3 fax machine. For these services, the speech processing is bypassed and the data is applied directly to the SCPC modem within the channel unit.

The following is a brief and somewhat simplified explanation of how the fax type of call is handled between a communicating pair of SCPC channel units. The presence of a fax signal (instead of human speech) is detected in the fax interface unit. The speech processing function is then disabled and a data channel is established in its place through the data interface. Then, the fax interface on the sending end conducts a handshaking process with the fax interface at the distant receiving end. This includes the determination of the best data rate to be used for the fax transfer (a function normally conducted by standard Group 3 fax machines). At this point, the channel unit at the sending end engages the local fax machine to begin its transfer of the page. The two fax machines can then communicate in the normal manner as if connected over an analog telephone line.

A direct digital interface is provided when the efficient transfer of computer data is desired. The data interface must satisfy the requirements of a common standard such as RS-232 or V.35. In the case of the latter, the data interface provides bit

Figure 10.11   A digital voice channel unit that employs speech compression. Dial-up data is accommodated by first detecting the modem carrier, demodulating the bits, and transferring those bits directly to the satellite modem. The fax interface converts modem tones into bits.

scrambling to randomize data so that there are enough bit transitions for synchronization at the receiving end. The satellite modem is another important element of the Earth station as it must achieve a high-quality link from end to end. The following are the functions of the modem:

- Differential coding of the data coming from the speech processor or data interface to set up the data for phase shift modulation by removing phase ambiguity over the satellite link;
- FEC coding to improve bit error rate performance over the satellite; options can be provided to allow various rates to be used, such as rate 1/2 and rate 3/4 convolutional coding or turbo coding;
- Modulation, which is a series of functions including baseband filtering, modulation, channel frequency selection, carrier on/off control, and carrier power control;
- Demodulation, which is a series of functions including downconversion of the selected channel, demodulation, automatic gain control (leveling), automatic frequency control, carrier recovery, symbol timing recovery, and preamble detection; modems may employ QPSK, MSK, or GMSK.

There are a variety of potentially useful configurations for the remote FTS Earth station. The most basic has been called the jungle telephone, so named because it can provide a usable line in a jungle or literally anywhere else. Remote site calls are connected through a GES to the national PSTN. Connections between jungle telephones provide basic telephone services totally independent of the terrestrial infrastructure. This would be useful at industrial sites or for emergency communications. In many ways, the jungle telephone resembles the mobile user terminal discussed in Chapter 11.

The multichannel Earth station configuration would supplement the PSTN by providing both local access to a village or industrial site or as an added means of trunking calls between toll exchanges. A typical configuration of a multiple-channel FTS Earth station is provided in Figure 10.12. This could be used as a higher capacity remote Earth station or, with enough equipment, as a GES in a major city. A number of channel units are indicated, each of which contains all of the elements discussed previously. The outputs of the units are added in power and applied to the RF terminal, shown at the center of the figure. It interfaces with the indoor equipment at

Figure 10.12   Interfaces and major elements of a multichannel FTS Earth station.

an IF frequency such as 70 MHz or L-band to provide access for the quantity of channel units. A twisted-pair signal cable is often included with the IF cable between indoor equipment and the RF terminal to transfer monitor and control (M&C) signals.

The RF terminal can be designed for operation in any of a number of frequency bands, particularly C-, X-, Ku-, and Ka-bands. C- and Ku-bands are the most popular at the time of this writing simply due to the much greater availability of satellite capacity around the world. FTS at Ku band provides good support to SNG vehicles for initial coordination at the site and backup communications if the video link fails. On the uplink side, the IF input to the RF terminal is translated as a group of channels to the particular transponder frequency within the upconverter. The output is on the right frequency but at too low a power level for the desired transponder operating point. Power is increased substantially by the HPA at the end of the uplink equipment chain.

The types of HPAs that are found in FTS service are either the solid-state power amplifier (SSPA), which is constructed from high-power field effect transistors, and the TWTA. Typical power levels that are available on the market at the time of this writing are listed in Table 10.5. SSPAs under 10W have been in use in the RF head of the standard VSAT. The same type of amplifier and RF head can be used in an FTS remote station with a capacity of one or two channels. For higher capacity requirements, the medium-power SSPA is available. It consists of multiple low-power stages that have been paralleled to achieve an output of 400W. The table indicates that Ku-band power levels are lower than those at C-band, which is a result of the difficulty of handling the power within the smaller dimensions of Ku-band FETs and associated amplifier circuitry.

For an FTS station with a high-channel capacity requirement, such as that used for a GES, the TWTA becomes necessary. This is the standard way to provide this level of power, but it brings with it higher cost and complexity in operation. The TWTA requires relatively high voltages and currents and so is subject to a number of potential failure modes. Untrained people should not attempt to open or operate high-power equipment. In addition, it is absolutely critical that moisture not enter the high-power section of the outdoor unit. While an SSPA should have an indefinite lifetime, a TWT will wear out due to the limited lifetime of its cathode. It will normally be necessary to replace a TWT after 2 to 4 years of service. Earth stations with TWTAs usually have an active backup amplifier so that one is available in the event of failure or the need to perform routine or emergency maintenance. Redundancy,

Table 10.5   Typical HPA Systems in Use for FTS Stations

Type of Amplifier	Maximum Power (Watts)		Comments
	C-Band	Ku-Band	
SSPA, low power	5	4	VSAT type
SSPA, medium power	40	20	Mounted on antenna
SSPA, high power	400	400	Ground mounted
TWTA, medium power	200	150	Laboratory type
TWTA, high power	800	400	Video type
Klystron, high power	3,000	2,000	Video type

backup power, and other support facilities are not shown in Figure 10.12. More details on the design and operation of such Earth stations can be found in our other work [3].

On the receive side, the full downlink frequency range is amplified and translated to L band (1 GHz) within the low-noise block converter (LNB). The spectrum of SCPC carriers is then applied to the interfacility link (IFL). A downconverter ahead of the indoor circuit selects the desired transponder and translates the carriers to the 70-MHz IF. The individual channel units within the indoor equipment can then select the assigned channels from within the IF range. Not shown in the figure are elements of the indoor unit that control the demand assignment function and control of the RF terminal from the indoor equipment.

10.2.3 Use of Satellite Capacity

We have indicated that the majority of demand-assigned FTS networks employ mesh topology with the SCPC mode of satellite transmission. The SCPC approach has all carriers operating at the same power level at the satellite, producing a uniform spectrum across the transponder bandwidth. The NCS assigns channels in pairs to allow two Earth stations to make a duplex connection for the duration of a call. In addition to these traffic channels, there are the control channels that are used to pass call setup and administrative messages between the NCS and the Earth stations (remote and gateway). The architecture of the control channel network is similar to the VSAT star, with outbound control channels between the NCS and the Earth stations and inbound control channels between the remote Earth stations and the NCS. As with the bandwidth reservation mode of the VSAT, the outbound channels use TDM and the inbound channels use ALOHA.

The manner in which spectrum is provided and used in a demand-assigned SCPC system is shown in Figure 10.3. In a typical network, there might be a total of 2,500 channels that are available to service calls and for network control (the figure shows a lesser number for illustrative purposes). Total transponder power must be divided among the carriers and an allowance made for output backoff to control intermodulation distortion. The precise number of channels is determined by the required EIRP per carrier (from the link budget) and the associated bandwidth per channel. A smaller allocation of bandwidth is made for the control channels between the NCS and the Earth stations.

There are a number of factors that the network designer can control in the use of the bandwidth and power of the transponder. The required power per carrier is an important characteristic because it is the basis of determining the total amount of transponder capacity and the resulting cost of the space segment. In Table 10.6, the total number of traffic and control channels is estimated for a typical network at Ku-band. The transponder in this example has sufficient power to support the equivalent of 1,000 full-time SCPC channels. The carriers are assumed to be voice activated, which means that the individual transmitters are turned on and off in response to speech. An average duty cycle of 40% is accepted as an industry standard. Therefore, the transponder can support 1/0.4 times the number of active channels, which amounts to 2,500. There must be sufficient bandwidth for this number of channels at full loading. Of these, an allocation of 10% is made for control

Table 10.6   Capacity Calculation for a Typical
Demand-Assigned FTS Network

Factor	Value
EIRP per carrier (from link budget)	10.0 dBW
Transponder saturated EIRP	45.0 dBW
Required output backoff	5.0 dBW
Number of channels	1,000 = 30 dB
Voice activity	40%
Channel capacity	2,500
Control channels (10%)	250
Traffic channels	2,250

channels, leaving 2,250 for revenue-bearing telephone traffic. This assumes that each control channel can support approximately 112 call requests during the same time interval.

This example is somewhat simplified in that we assume that all carriers are operated in the same manner and that the channels are uniformly loaded across the transponder bandwidth. In a real network, carrier powers and bandwidths vary between types of Earth stations and channel capacity requirements. Also, there are variations in power due to link fading and equipment power output misadjustment. All of these factors tend to reduce the usable capacity. A thorough analysis for the type of network should be performed in cooperation with the network supplier. They most likely know the detailed assumptions and operating parameters for their system and can provide specific examples for the planned network.

10.3   Preassigned Point-to-Point Link

Satellite transmission has long been applied to FTS service with preassigned links that operate at the higher levels of the hierarchy in Figure 10.1. These are preassigned point-to-point links that carry several telephone channels through the concept of multiplexing. The level 3 links are between pairs of domestic toll exchanges, while level 2 links serve the international gateway. International links themselves are provided over the Intelsat system and are definitely candidates for FTS service.

The preassigned links first implemented over satellites employed FDM with capacities ranging from 12 voice channels at the low end all the way up to the 1,800 channels on the highest capacity links. However, analog FDM has been replaced by digital TDM, which is more efficient on the basis of satellite power use. This has raised the quality of satellite voice communication to be comparable to what the digital terrestrial network can deliver. Intelsat's IDR carriers are standard in the international system. Digital services like medium-speed data and video teleconferencing can be introduced and integrated with voice using multiplexers and bandwidth managers of various kinds. A private satellite service called Intelsat Business Service became popular in the late 1980s, joined by Eutelsat's version called SMS.

10.3.1 Multiple-Channel Per Carrier Transmission

The capacity per link is often composed of a T1 or E1 channel, although higher bandwidths have been used for time to time. For example, digital carriers with 6 or 8 Mbps can share the bandwidth and power of a common transponder through the multiple-channel per carrier (MCPC) technique. A graphic spectrum example of how MCPC carriers would be placed in a satellite transponder is provided in Figure 10.13. The large carrier would come from an uplinking Earth station that needs to transmit a relatively wide bandwidth, while the other two sizes match Earth stations with smaller traffic requirements. The carriers do not occupy all of the bandwidth because of the loss of capacity from providing sufficient output backoff. Still, a properly loaded 36-MHz transponder can support up to about 30 Mbps of useful transmission in the multiple-carrier mode. The precise capacity will depend on the saturated EIRP of the satellite and the G/T of the Earth stations.

The approach to transponder loading is the same as for SCPC, with the exception that a variety of carrier capacities occupy the available bandwidth. The power level of each carrier is adjusted to meet the needs of the particular link. Intermodulation distortion is controlled by allowing an appropriate amount of output backoff, typically in the range of 3 to 5 dB. The precise amount can be determined once the specific carrier arrangement and power amplifier characteristics are known. Fortunately, the answer is not particularly sensitive to the precise value of backoff, and it is often the practice to plan to use a round number of 5 dB for the case of a standard TWTA or 3 dB for an SSPA or linearized TWTA. This means that a Ku-band TWTA with a saturated EIRP of 46 dBW (single carrier) would be capable of delivering 41 dBW of total EIRP when transmitting a group of MCPC carriers such as that shown in Figure 10.13.

A permanently assigned FTS needs a good system of multiple access to keep the individual transmissions separate. Most FTS systems in operation today use either FDMA or TDMA; however, there is a decided interest in exploiting the interference rejection characteristics of CDMA. The Earth station equipment in the case of FDMA is perhaps the simplest because separation is maintained by permanently assigning different carrier frequencies. For TDMA to work properly, stations must transmit their information as synchronized bursts, as discussed in Chapter 9. The CDMA mode of operation does not require any coordination between Earth stations, which can transmit at the same time on the same frequency. However, the complications come in from the need to properly despread the desired carrier and synchronize to its timing. Also, the determination of the number of Earth stations that can share the same frequency channel or transponder is a complex matter, as we

Transponder usable bandwidth

Figure 10.13 An example of an transponder that is fully loaded with MCPC transmissions.

will discuss in Chapter 11. In the following sections, we describe two topologies used in preassigned MCPC links.

10.3.1.1   Single-Destination Carriers

A single-destination carrier is an Earth station transmission that is intended for one other Earth station, representing half of a full-duplex link. The bandwidth is equal in both directions, as in a microwave radio path or fiber optic cable that connects two end users. The flow of information for such an MCPC link pair is shown in Figure 10.14. A simple TDM format is used so that the information transmitted by A is perfectly balanced by that from B. This provides full-duplex service and does not require any preprocessing within the Earth stations other than what is normally associated with the modems on each end of the satellite link. In the figure, we assume the first block of time is used for data transmission, the second for fax, the third for digital video teleconferencing, and the final block for telephony. The sequence repeats according to the frame format of the link.

10.3.1.2   Multidestination Carriers

The first generation of the Intelsat system introduced full mesh connectivity using analog ground equipment that was available in the late 1960s. Up to that time, the single destination carrier using FDMA was heavily applied to terrestrial microwave networks and was a natural choice. However, to apply it, every Earth station would have had to transmit a separate carrier for each destination. This would have resulted in a proliferation of transmission chains at the ground stations and a somewhat inefficient loading of the transponder. As a solution, the multidestination carrier was adopted to allow each station to transmit only one carrier that would contain the multiplexed traffic for all destinations. To connect N stations one needs $N(N-1)$ half-duplex links. With multidestination carriers, only N links are needed.

   An example with four Earth stations transmitting multidestination carriers is shown in Figure 10.15. See also Figure 8.3 for a network illustration. Each station uses TDM to transmit blocks of information to the three other stations, achieving full mesh connectivity. The carriers are the same bandwidth and capacity, but blocks are unequal in size, in recognition of the fact that balanced information

Transmission from:

Figure 10.14   Transponder timelines for single destination TDM carriers for a pair of communicating Earth stations.

transfer is only required between pairs of stations. For example, station A transmits its half of a duplex link to station C within its second block. In return, station C uses its first block to transmit the complementary side of the link to station A. The remaining connectivities can be derived from the figure in the same manner. Note that there are two unused blocks at the ends of stations C's and D's time frames. This was not assigned to traffic and hence goes unused. It would have been possible to increase the size of the transfer between stations C and D only, since stations A and B already fully utilize their time frames. Alternatively, the carrier data rates of stations C and D could be adjusted downward and the blocks extended in time to fill out the unused time slots.

One step at each station that is not shown in Figure 10.15 is the assembly of the three received downlink data streams into a single time frame. This is necessary to meet the terrestrial interface requirements to the telephone exchange or user-owned PBX. Station B, for example, must pull off its traffic blocks from the downlinks from stations A, C, and D and plug them into a time frame that corresponds to the exact format of its uplink. This bidirectional TDM link is then in proper format for use in standard digital exchanges.

10.3.2   Bandwidth Managers and Multiplexers

With the relatively high capacity on these point-to-points, it is useful to employ intelligent multiplexers on each end, which can combine streams of data from different sources in a seamless manner. Popularized in the 1980s as replacements for less-flexible TDM muxes and digital cross-connect systems (DCS), the new breed seemed to offer a solution for all of the telecommunication manager's problems. However, this did not turn out to be the case because of the rise of the Internet in the 1990s. Today, there is a role of bandwidth managers in satellite communications when one needs to integrate IP services along with constant bit rate applications like standard telephone and videoconferencing. This is not a long-term solution because of the rapid rise of H.323 for videoconferencing and VoIP for telephony.

Figure 10.15   Transmission timelines for multidestination TDM carriers for three communicating Earth stations in a full mesh.

10.4   Application of FTS

FTS has increased in popularity due to the rise in demand for good quality telephone service in the developing regions of the world, for transportable applications and for emergencies. The techniques discussed previously make FTS a very versatile problem solver in remote areas where it is too expensive or difficult to extend the terrestrial PSTN. One big advantage of the satellite approach is that once satellite capacity is available in the particular country, the network can be installed in a month or even a few days. The network then services smaller cities, villages, and even individual locations with all of the capabilities of the modern digital PSTN. A well-designed FTS can enhance the quality of life and economic opportunity in these regions as many of the more advanced features of the intelligent public network become available to almost anyone. Improvements in technology and reduction of cost will make FTS a good alternative to cellular telephone for fixed uses.

10.4.1   SCPC FTS Example

An example of a complete FTS network implementation is shown in Figure 10.16. The network provides basic telephone, fax, and voice-band data service to villages and remote users with dedicated Earth stations. Connection to the PSTN is through a gateway Earth station at the capital city (this may or may not be associated with the international gateway for the country in question). The heavy arrows represent permanently assigned trunk channels that support predictable traffic volumes between villages and the capital. Villages 3 and 4 have enough traffic between them to require such a permanently assigned trunk. Any of the permanently assigned links can be implemented either with single-destination or multidestination carriers.

The real power of the rural network is achieved with demand-assigned links, indicated with the shaded arrows. These are established only when a telephone call is placed, allowing a station of any type to communicate with any other. Remote stations can call villages, the capital, or even other remote stations. By providing connectivity on demand and without restriction, the FTS network creates a nationwide telephone infrastructure that does not depend on the degree of sophistication of the existing PSTN.

Satellite capacity for an FTS network can be drawn from a domestic C- or Ku-band satellite. Once Ka-band satellites are launched, they, too, can supply capacity for rural FTS networks. The capacity of a typical C-band transponder is in the range of 500 to 3,000 channels, depending on the voice channel data rate and the G/T of the Earth stations. In a typical network, a transponder could supply 1,000 simultaneous calls involving 500 pairs of subscribers. These could put rural callers in contact with people in cities who are connected through the PSTN to a gateway. Alternatively, they all could be in rural areas, using the satellite network as their only means of communication. Since individual subscribers do not use their telephones continuously during the busy hour, the same 1,000 channels would support a user community of between 5,000 and 50,000. Multiply the number of transponders, and the size of the user population can likewise be multiplied. One satellite with 24 transponders can therefore provide rural telephone service to approximately 1 million subscribers.

Figure 10.16   An example of a rural FTS network using permanently assigned and demand-assigned connections. The capital gateway provides access to the PSTN.

There is a natural extension of FTS from fixed communications to the area of providing mobility to users. Chapter 11 reviews the progress made in the introduction of mobile telephony using satellites that operate in GEO and non-GEO orbits. This is another exciting development in the application of satellites to commercial communications.

References

[1]   Schwartz, M., *Telecommunication Networks*, Reading, MA: Addison-Wesley, 1987, p. 522.

[2]   Freeman, R. L., *Reference Manual for Telecommunications Engineering*, New York: John Wiley & Sons, 1985, pp. 13–14.

[3]   Elbert, B. R., *The Satellite Communication Ground Segment and Earth Station Handbook*, Norwood, MA: Artech House, 2001.

Mobile Satellite Service (GEO and Non-GEO)

Satellite communications is a natural facility for serving users while they travel by various means. These include ships on oceans, rivers, and lakes; commercial and private aircraft; land-based vehicles of various types; and individuals using portable and handheld devices. The Mobile Satellite Service is an established user of spectrum at frequencies between 1 and 3 GHz, where simple antennas provide flexible access to the space segment. This regime is preferred because of its greater ability to penetrate foliage and nonmetallic structures and bend around obstacles. (The low end of the usable spectrum is probably 100 MHz, which is able to penetrate the ionosphere under all conditions.) Frequencies above 3 GHz, while more readily available, are easily blocked by natural and manmade obstacles and introduce practical difficulties when it comes to generating transmit power. Above 10 GHz, rain attenuation must be factored into the equation; as a result, there is an optimum window for MSS in the 1- to 3-GHz range. The MSS portion of the spectrum is limited to segments of around 50 MHz each due to the following factors:

- A general lack of bandwidth to begin with, due to the lower frequency as compared to C-, Ku-, and Ka-bands;
- Competition with land-based services such as cellular telephone, mobile data communications, and a wide variety of unlicensed services, the popular 802.11 wireless LAN being a notable example;
- Reduced ability to achieve frequency reuse because user antennas have little or no gain and cannot easily discriminate among satellites within view.

Operations at 1.5 GHz (L-band) began in the 1970s with the launch of Marisat and subsequent creation of the most successful maritime mobile satellite operator: Inmarsat. However, 1990 was a turning point for MSS, when Motorola introduced their concept of a low Earth orbit satellite system capable of directly serving handheld terminals. When the idea first appeared, the belief within the established GEO satellite industry was that even the most advanced satellites of the day were not capable of supporting direct links to handheld phones. However, by 1994, several companies had devised schemes to provide mobile service to handheld phones, employing literally all altitudes ranging from Iridium's 800 km all the way up to 36,000 km at GEO. Iridium went into service around 1999, followed closely by Globalstar; however, neither service gained enough early subscribers to overcome the skepticism in the financial markets and both went into bankruptcy. Iridium restructured into a new company that provides basic voice and data service to

government and industrial users. Globalstar was acquired by Craig McCaw and continues to seek its future as a private company, focusing instead on aeronautical and fixed telephone services. Two GEO MSS systems began to offer similar service to handheld phones around 2002—AceS and Thuraya—and both have continued as operating businesses at the time of the writing. All of these systems are usable for fixed telephone and temporary and emergency communications, which will have significant impact in areas and times of great need for rapid deployment and high mobility where the terrestrial infrastructure is either poor or nonexistent. Inmarsat's fourth generation satellites could offer near-broadband services using a satellite similar in design to Thuraya.

11.1   Foundation of the Mobile Satellite Service

The role of satellites in mobile communications has expanded since the first MSS satellite, Marisat, in 1976. However, the basic connectivity has not changed even though a mobile terminal is now often a handheld instead of mounted on a ship. As shown in Figure 11.1, a full-duplex link is provided so that the user can conduct a normal telephone conversation through the satellite and GES. In this example, the user terminal (UT) is in the form of a specialized handheld mobile telephone that can transmit and receive within the L-band part of the spectrum at 1.5 GHz downlink and 1.6 GHz uplink. This in itself is not a major accomplishment since there are existing terrestrial mobile telephone services (PCN and PCS) that use the 1.8- and 1.9-GHz bands, respectively.

Figure 11.1   The mobile-to-fixed duplex link, using L band to communicate with the mobile subscriber and Ku-band feeder link to the gateway Earth station. The mobile subscriber can make and accept calls from the PSTN provided by the space segment; direct user-to-user connections may be possible on a single-hop basis.

The antenna of the illustrated UT provides essentially unity gain (0 dBi) because of the need to maintain small size and to provide a nearly omnidirectional pattern to cover the full sky. The GEO satellite in the figure, on the other hand, has a much larger antenna system than found in FSS systems in order to close the link to the user. In most of the modern MSS systems, the spacecraft antenna produces an array of small spot beams that cover the ground with a grid of cells that emulate a terrestrial cellular network. This concept will be reviewed in more detail later in the chapter.

The link between the satellite and the GES is called the MSS feeder link and can be implemented within a standard FSS allocation. The Ku-band is illustrated in the example in Figure 11.1; in contrast, the Inmarsat and Thuraya systems employ C-band for this purpose. In the case of the Iridium non-GEO MSS system, the feeder and intersatellite links operate at Ka-band. The GES, representing a large fixed asset of the service provider, interfaces with the PSTN and has a large antenna and high uplink power capability. The feeder link is engineered for adequate availability to minimize its impact on the overall quality of service as rendered to mobile users. Within the GES is the RF terminal and associated baseband equipment, which performs similar functions to the counterpart in FTS applications (reviewed in Chapter 10). Figure 11.1 indicates that the Earth station includes or is directly connected to a MSO to service telephone calls that are routed over the satellite. The MSO is a standard local exchange or cellular MSO design that has been modified for satellite mobile telephone service. Through the MSO, the satellite network delivers a range of telephone, data, and fax services to mobile users who can be literally thousands of kilometers away.

The overall architecture for MSS telephony, with its four operating levels, is illustrated in Figure 11.2. Each of these levels contributes heavily to the functionality and investment of the total system:

- Satellite constellation, consisting of a quantity of operational satellites that deliver the service over the coverage area. These can employ any of the possible orbit constellation arrangements, with intersatellite links being optional.
- User terminals of various types: vehicular, handheld, transportable, ship and aircraft, and fixed terminals.
- Gateway Earth stations that allow traffic to pass between users and the public networks, and to manage the service on a consistent basis. Also considered are TT&C facilities to control and monitor the satellites.
- Terrestrial networks to address the service needs of the users. These include the PSTN, the Internet, and other networks, both public and private.

A state-of-the-art MSS system can provide other capabilities besides connection of calls to the PSTN. The most striking is direct mobile-to-mobile calling that allows subscribers to talk to each other regardless of their location and situation. The quality of the terrestrial telephone network in different countries served by the system will also play heavily into the attractiveness of direct mobile-to-mobile calling. Some systems will address this by connecting these calls through a common gateway, introducing the delay and degradation of a double hop. This is not a concern in LEO and MEO systems, where the propagation delay is relatively low from the

Figure 11.2 Overall architecture of an MSS system, showing the four primary levels: the satellite constellation, the user terminals, gateway Earth stations and control systems, and the terrestrial networks.

start. On the other hand, this type of call ties up double the capacity of the space segment and GES resources.

Satellite communication among moving Earth stations is very different from fixed and broadcasting services discussed previously in this book. The situation of a user moving relative to the satellite (or, in the case of non-GEO satellite systems, a satellite moving relative to the user) causes dynamic behavior in the link. This introduces new forms of fading that cannot be predicted in the same way as rain attenuation and ionospheric scintillation. Users who are accustomed to making cell phone calls inside buildings could be frustrated by MSS services, since it is likely they will need to go outdoors or stand by an appropriate window. The need to take the properties of satellite visibility into account is referred to as *user cooperation.*

Terrestrial cellular systems also undergo dynamic fading as the subscriber drives past buildings and trees, under overpasses, and into isolated locations where the base station simply cannot reach. This fading is very rapid in time because it is produced primarily by multipath propagation, where the strongest signal that is received may at times be a reflected signal off of one or more buildings. The various reflections add and subtract at the receive antenna in random ways, with the instantaneous sum varying very rapidly. What tends to overcome the problem in cellular is the very high link margins provided (in excess of 30 dB) along with the availability of multiple signal sources from several base stations that could potentially reach the subscriber at any one time. Receiving the same signal through dual antennas, called antenna or space diversity, is another approach for overcoming multipath. This requires a scheme for selecting the stronger signal or constructively combining signals within the receiver.

The propagation model for MSS depends on the situation within which the mobile user is found. Ideally, there should be a direct line-of-sight path between the

user and the satellite. The service quality in this case, which is suggested in Figure 11.1, is ideal because the link can be engineered for no outage due to the mobile-to-satellite path. If the user or satellite is moving, then the link will experience periods of blockage when the user is "shadowed" from the satellite transmission. The MSS network would either (1) allow the inevitable dropouts in the data transfer, which would intermittently halt the conversation or information flow, or (2) attempt to eliminate or reduce the dropouts through path diversity. The latter is very expensive because it requires that the number of satellites be increased by perhaps a factor of two. In most cases, the mobile user will want to be connected with the PSTN, which is provided through a GES that employs FSS spectrum at C- or Ku-band. We must consider the fading environment on this side of the link in the same manner as with any FSS applications. The resulting mobile-to-GES service can be engineered to meet almost any availability requirement, although the cost of doing so could be prohibitive.

11.1.1 Radio Frequency Spectrum Availability

Radio spectrum has been allocated by the ITU to MSS just as it has to the other satellite and terrestrial radio services. An illustration of the allocation of spectrum between 1.5 and 2.5 GHz (L- and S-band) is provided Figure 11.3. Solid bars indicate GEO allocations at L-band while non-GEO allocations are indicated with white bars. The hatched bars display allocations for terrestrial wireless services: PCS, 3G cellular radio, and unlicensed wireless data, notably 802.11. Also shown is the allocation at S-band for the downlink broadcast of DARS. The impression that

Figure 11.3 Spectrum allocations for the MSS as compared to other services. Uplink and downlink directions are indicated. Note that there are differences between regions and specific countries, which are not reflected in this figure. Allocations as of WRC 2003.

we take away is that the amount of spectrum available for MSS is substantially less than what is allocated to the FSS and BSS. This reflects the popularity of and competition for these frequencies. Much of this spectrum had to be taken away from fixed terrestrial services like short-haul microwave links and radar. Some countries, in fact, do not acknowledge that MSS has any particular right to some of these frequencies. This leads to coordination difficulties that any prospective MSS operator must address.

The coordination of MSS satellite systems at L-band is formidable as compared to both the FSS and BSS. Figure 11.4 indicates how small ground antennas used for MSS may not be able to separate the signals coming from adjacent satellites that operate on the same frequency. In the typical FSS satellite system, users employ highly directive ground antennas that can discriminate among the satellites, supporting 2° satellite spacing that promotes much higher orbit capacity. As shown in Figure 11.4, the closest orbit position that can reuse the spectrum is all the way over at C, which is effectively over the horizon from the user. The only reuse that we can count on would come from nonoverlapping satellite coverages, which is feasible if the footprints are sufficiently separate on the face of the Earth. As a result of broadbeam UTs, coordination among MSS satellites in the same spectral allocation amounts to segmenting the band. This can only be determined through the ITU coordination process, where the parties look in detail at the satellite coverage designs, Earth station characteristics, and the types of carriers that are transmitted over the network.

11.1.2 MSS Link Design

The satellite links involved with MSS services have many similarities to their counterparts in FSS and BSS. All employ microwave frequencies, experience the standard spreading loss in propagating at the speed of light through free space, and employ conventional low-noise and high-power amplifiers. The principal differences, addressed previously, are due to limited bandwidth available at L- and S-bands and the dynamic behavior of mobile link fading. Also, user terminals

Figure 11.4 For a mobile broad-beam Earth station antenna, signals from satellite B cause unacceptable interference even though it is many degrees away in orbit space. Satellite C is nearly below the horizon and cannot interfere.

employ low-gain antennas that offer little discrimination from adjacent satellites. To provide additional clarity, we include an example of a link budget in Table 11.1 for the forward link (from the GES to the UT) and return link (from the UT to the GES) of a hypothetical bent-pipe MSS satellite. Many MSS systems, such as the one illustrated here, assign Ku-to-L-band to the forward link and L-to-Ku-band to the return link. To maximize capacity and minimize uplink power from the UT, we employ speech compression, which reduces the basic transmission rate to 4.8 Kbps. The actual data rate on the link is 5.4 Kbps, which includes the extra bits to support supervisory signaling and forward error correction. The access mode is TDMA with up to 10 channels per 64-Kbps carrier, and telephone channels are assigned to individual bursts that are voice activated. The detailed review of Table 11.1 provided in the next paragraphs can be skipped by nontechnical readers.

In the forward direction, the GES uplinks 100 FDM channels with an EIRP per channel of 52.6 dBW at 14.25 GHz from an 11-m antenna. The uplink path loss to the satellite consists of 207.0 dB of free-space loss and 0.25 dB of atmospheric and other minor losses. The L-band downlink in the second column has an EIRP per carrier of 47.1 dBw to reach a handheld UT. The "Other losses" of 0.1 dB is from the polarization mismatch between the UT and the satellite. At L-band, it is common

Table 11.1   Simplified Link Budget for a GEO MSS Service to a Handheld User Terminal

Parameter	Forward Link		Return Link		Footnote
	Up	Down	Up	Down	
TDMA carrier bandwidth, kHz	48	48	48	48	
Frequency, GHz	14.25	1.54	1.64	12.45	
Transmit antenna	11m	12m	Helix	2m	(1)
EIRP, total, dBW	72.60	63.00	2.70	42.00	
EIRP, per carrier, dBW	52.60	47.10	2.70	11.40	(2)
Path loss, dB	−207.00	−187.6	−188.1	−205.90	
Other losses	−0.25	−0.10	−0.13	−0.23	(3)
Receive G/T, dB/K	4.00	−23.90	20.00	40.00	
Receive C/T, dBW/K	−150.70	−164.50	−165.50	−154.70	
Noise bandwidth, dB(Hz)	−46.81	−46.81	−46.81	−46.81	
Boltzmann's constant	−228.60	−228.60	−228.60	−228.60	
C/N thermal, dB	31.14	17.29	16.26	27.06	
C/IM, dB	30.00	20.00	30.00	17.00	(4)
C/I, dB	20.00	20.00	23.00	20.00	
C/(N + I + IM) total, dB	19.29	14.13	15.28	14.96	
Combined C/(N + I + IM) (up and down), dB		12.97		12.10	
Required C/(N + I + IM), dB		5.00		5.00	
Margin, dB		7.97		7.10	

Note: The IF routing repeater on the satellite transfers a 64 Kbps TDMA carrier in a bandwidth of 48 kHz using R = 0.8 turbo coding and QPSK modulation. Since the repeater provides a bent-pipe transmission path, uplink noise must be added to the downlink noise to determine overall C/N for the link.

(1) Up is from Earth, down is from space.

(2) 100 carriers in forward link; 4.1-dB compression due to amplifier saturation effects on forward downlink; ~1,000 carriers on return downlink.

(3) Clear sky: variable propagation margins will be taken at the end of the link.

(4) These values come from amplification chain characteristics.

practice to use circular polarization to make the link insensitive to Faraday rotation, which is more prominent at lower frequencies like L-band. The satellite provides a Ku-band footprint over a coverage area that contains all of the GESs in the network. In this example, this beam is regional in size and has a value of G/T equal to 4.0 dB/K. The uplink C/N (thermal) of 31.1 dB results from the uplink EIRP, path loss, G/T, and noise bandwidth.

The Ku-band uplink experiences interference from a variety of sources, including a cross-polarized uplink if it exists and other beams that operate on the same satellite. Interference from other satellites is very difficult to assess and so we need to treat it as a separate item during coordination (discussed in Chapter 12). The satellite repeater employs microwave power amplifiers that at L-band are typically SSPAs. With multicarrier operation, there would be a contribution to the total noise from intermodulation distortion. Because the SSPA is loaded with a large quantity of small TDMA carriers, the intermodulation can be treated as noise through the use of a C/IM entry. Uplink C/IM is produced within the Ku-band GES because it also transmits multiple carriers. The value of 30 dB in the uplink is relatively high and so it has less of an impact than the downlink C/IM of 20 dB, produced within the SSPA. By combining each of these (noise-like) contributions, we obtain a total C/(N + IM + I) for the forward uplink of 19.3 dB.

Examining the GES-to-UT downlink, the operating frequency at L-band is 1.54 GHz. The spacecraft antenna in this example is sufficiently large to produce adequate gain for the link to a handheld UT. The L-band EIRP per carrier of 47.1 dBW results from a transmit gain of the assumed 12-m antenna of 44 dBi combined with a transmit power of 3.1 dBW. The mobile phone in this example has a G/T of –23.9 dB/K, which is very low in comparison to even the smallest of dish antennas used in the BSS. This is achieved with a noise figure of 1.3 dB and gain of 2 dBi. The main reason for this low gain is that the handheld UT uses a nondirectional antenna to allow the subscriber to be able to move around freely. The UT antenna is assumed to be a compact design that produces a pattern that is tolerant of the user's orientation. The gain of 2 dBi is an average transmit/receive value that would be achieved in the direction of the satellite with the UT properly held in the user's hand.

The Ku-band downlink of the return link is transmitted by a 2-m diameter area coverage spacecraft antenna, producing a C/N of 27.1 dB. Combining this with the indicated values of C/I and C/IM produces a total downlink C/(N + IM + I) of 15 dB. The last step in the analysis is to combine the total uplink and downlink C/(N + IM + I) values, which gives 14.2 dB if all contributions can be treated as noise. The threshold C/(N + IM + I) is determined from the vendor specification or, preferably, from laboratory measurements performed by the network implementer. In this example, the value of 5 dB considers the particular compression and coding system employed with the link operating an error rate of 0.001. The bottom line of Table 11.1 indicates an overall link margin of 7 to 8 dB, which is available to protect the link from multipath fading and some shadowing losses as the UT moves from place to place. The Ku-band links will experience rain fade as well.

The following comments highlight the factors in Table 11.1. In an overall sense, the forward and return links are balanced, meaning the resulting combined C/N values are nearly the same. This is because of the dominant effect of the simple antenna on the handheld UT, which directly impacts the receive gain (forward link) and the

transmit gain (return link). The uplink from the handheld UT is perhaps the most critical link in the system. The EIRP is limited by the low antenna gain of the omni-directional type of antenna used with the UT. Transmit power to the antenna is constrained by two important factors. The first is due to battery power, which must be held to a minimum to extend the duration of use between recharging. The UT is no good, after all, if the battery is dead after only one phone call. A general guideline is that the phone should have enough battery power for 2 full hours of talking (which includes dialing and holding time) and, alternatively, 24 hours of standby time when the phone can potentially receive calls.

The second factor is the concern for a potential hazard to the human body from RF radiation. While this risk has not been quantified at the time of this writing, there are some specifications for power levels that are considered to be safe. The average power that is applied to the antenna should be under one-third watt, with the peak power allowed to be some factor times the average. This is not a recommendation by the author, and readers wishing to apply the standard should review the relevant documentation provided by the IEEE as well as the FCC of the United States. The resulting EIRP from the handheld UT is 2.7-dBW peak-power in the direction of the satellite. Since this is TDMA with 10 users sharing the same frequency, the average EIRP is 10 dB less, or −7.3 dBW.

Because of its large antenna, the GES produces very strong values of uplink C/N in the forward direction and downlink C/N in the return. Thus, the impact of the GES on the overall link is kept small. The Ku-band downlink to the GES, indicated in the right-hand column of Table 11.1, requires a large Earth station antenna communicating through an area coverage beam on the satellite. We see that intermodulation distortion and interference are much larger contributors to the overall GES links than the downlink thermal noise.

The significance of a link margin of 7 dB is that this link will work in a satisfactory manner for line-of-sight conditions. It can tolerate some additional loss from antenna misalignment, absorption by the body, multipath fading due to a reflected signal interfering with the direct path, and a limited amount of shadowing by trees and low buildings. The actual link will be exposed to a few additional losses and sources of degradation. For example, carrier levels from the UTs will vary over a range of measurement and adjustment. This means that some carriers will arrive at the satellite lower than the average, producing a loss of link margin. An allocation of 2 dB might be sufficient for this factor. Another such item is fading on the Ku-band links due to rain attenuation. In reality, the Ku-band GES can increase power during heavy rain through uplink power control.

Microwave link engineering is both a science and an art. In the case of MSS, there are several critical factors to consider, some of which vary statistically. What we are left with is a service that has great potential utility but which requires a certain degree of user cooperation.

11.1.3   Orbit Selection

MSS systems involve mobile Earth stations, many of which do not employ directional antennas. The satellite need not be stationary relative to the user, and thus inclined geosynchronous and non-GEO orbits are candidates. We mentioned

Motorola's Iridium system satellite that introduced the handheld UT. Motorola was joined by Loral and its Globalstar system and perhaps ICO Communications, a spin-off of Inmarsat, as implementers of non-GEO orbit constellations. While none of these original operations survived the tech wreck of 2001, the non-GEO systems are interesting in their approach to reaching mobile users around the globe.

The general range of candidate orbits for use in a global MSS context is shown in Figure 11.5. The LEO constellation is exemplified by the Iridium system, which currently contains 66 satellites (the original FCC filing by Motorola showed 77 satellites, but they later reduced the number). The polar orbits of Iridium cause the satellites to provide the best coverage of the North and South Poles, with the least favorable operation occurring along the equator where only a single satellite is visible on the ground at a time. The other leading LEO system, Globalstar, contains a reduced number, 48, made possible by tilting the orbits to focus the coverage away from the poles and toward the regions of interest to more users. At a higher altitude, the MEO or intermediate circular orbit (ICO) provides coverage comparable to Globalstar but with fewer satellites. The satellites of ICO Communications, taking its name from the intermediate circular orbit they employ, also see a larger amount of the Earth. A summary of the properties of these orbits is provided in Table 11.2.

The issue that leaves out extensive ranges in Table 11.2 is the Van Allen belts, which are regions where charged particle radiation levels are deemed to be unhealthy for the typical commercial satellite. Figure 11.6 presents the basic topography of the charged regions around the Earth. The magnetic field of the Earth traps much of the incoming charged particle radiation (which is good in terms of protecting us from an excessive dose) and presents it to satellites that would occupy the ranges of altitude known as the Van Allen belts. The LEO, MEO, and GEO altitudes are such that the radiation field is within design constraints of modern satellites, consistent with achieving operating lifetimes in the range of 10 to 20 years.

Figure 11.5 The three orbits that are applied to MSS: GEO, MEO (also referred to as intermediate circular orbit), and LEO. As the altitude of the orbit is decreased from GEO, the number of satellites required for continuous coverage increases. The inclined geosnychronous orbit (not shown) is applied as well (e.g., Sirus Satellite Radio), since mobile antennas either are broad in beamwidth or have tracking mounts.

Table 11.2   Options for Earth Orbits for Use in Mobile and Other
Satellite Communications Services

Orbit Definition	Altitude Range (km)	Period (Hours)
LEO	150 to 1,000	1.5 to 1.8
MEO	5,000 to 10,000	3.5 to 6
Geosynchronous orbit (e.g., synchronized to 24-hour rotation of the Earth but generally elliptical in shape; may or may not be inclined with respect to the equator)	36,000 mean altitude	24
GEO	36,000 precisely, in plane of the equator	24
Highly elliptical Earth orbit (HEO)	1,000 to 40,000	12 to 24

Why the orbit altitudes produce particular orbit periods is the result of basic physics, in particular, orbital mechanics. At the most basic level, we have the laws of planetary motion, formulated by Johannes Kepler (1571–1630). These are summarized in Figure 11.7, which shows how every orbit is basically an ellipse with the principal body (the Earth in the case of communications satellites) at one of the two foci. These laws dictate that the higher the orbit, the longer its period; this is independent of the size and mass of the satellite itself. What we have is a small object, the satellite, whose path is controlled by the gravitational force of the large body. In reality, there are other gravitational forces on the satellite, principally from the Moon and Sun, which alter the orbit over time. Thus, if we start out with a perfectly circular GEO, which is in the plane of the equator (the optimum conditions for

Figure 11.6   The radiation environment in Earth orbit is produced by electrons and other charged particles from the solar wind that are trapped by the Earth's magnetic field. This gives rise to the Van Allen radiation belts, which can produce substantial radiation damage to unprotected electronics.

- First Law: The orbit of each planet is an ellipse, with the Sun at one focus.
- Second Law: The line joining the planet and the Sun sweeps out equal areas in equal times.
- Third Law: The square of the period of a planet is proportional to the cube of its mean distance from the Sun.

$$P = 1.659 \times 10^{-4} \times (6378\,km + h)^{3/2}$$
$$\text{minutes}$$

Figure 11.7   Kepler's laws of planetary motion.

remaining stationary with respect to the ground), these extraneous gravitational forces cause the orbit to elongate (become elliptical) and tilt (become inclined). A GEO satellite will drift away from the assigned orbit slot, which is an effect of the Earth not being a perfect sphere. All together, these forces require that satellite operators perform touch-up maneuvers from time to time to maintain a GEO satellite in its "box" (typically 0.2° on each side) and non-GEO satellites on their proper orbit track.

According to the formula in Figure 11.7, the orbit period is proportional to the center-to-center distance of the satellite from the Earth (e.g., the altitude plus the Earth radius of 6,378 km) to the 3/2 power. Figure 11.8 presents the period of the orbit as a function of satellite altitude (not distance from the center of the Earth); the curve on the left is specifically for LEO altitudes and shows an abbreviated range of period between 1.4 and 1.8 hours. The full range of orbit altitudes is indicated in the curve on the right, which also presents the one-way propagation delay (uplink plus downlink, but not round trip). The latter states one principal benefit of LEO and

Figure 11.8   Orbit period and one-way (single-hop) time delay versus altitude.

MEO orbits: the one-way delay is less than 75 ms. This is significantly shorter than for GEO (260 ms, or about a quarter second) and presents a potential benefit for interactive voice and other services. However, the real performance of LEO systems like Iridium and Globalstar shows end-to-end delay in excess of 100 ms due to a variety sources, such as voice compression, multiple-hop transfers, and handovers.

At 36,000 km, the GEO approach to a global MSS system allows the service to be implemented with only three satellites (four are shown in the figure to allow for some backup capacity). Unlike LEO and MEO systems that require a constellation to provide continuous service to any point on the Earth, a single GEO satellite may be used to initiate MSS in a particular country or hemisphere. The type of satellite to be employed is demonstrated in the previous discussion of the MSS link budget. By employing such a large reflector and high-power spacecraft, the GEO system still closes the link to a handheld UT comparable to the lower altitude. An overall summary of the properties of the three orbit regimes is presented in Table 11.3. At the end of the chapter, we compare the systems on an economic basis using design and cost models developed by the author.

11.2   GEO MSS Systems

Historically, GEO satellites have provided most of the MSS capabilities, in terms of land, sea, and air. The economy and simplicity of a single satellite along with the ability to use fixed antennas on the ground have allowed GEO to reach critical mass for the applications described in previous chapters. In addition to the global capability of Inmarst, a number of GEO MSS networks capable of serving handheld satellite telephones are in service. The major benefit of the lower orbits is reduced time delay for voice services. This factor is very important in terrestrial telephone networks, particularly with high-quality transmission as provided through fiber optic technology. As will be discussed later in this chapter, time delay is less of a factor in mobile communications.

We begin with a review of the established GEO systems for global, regional, and domestic service. The latter two types of systems demonstrate that GEO has the added benefit of allowing one country or a group of countries to implement and operate a system without having to enter the global market.

Table 11.3   A Summary of the Key Attributes of LEO, MEO, and GEO Orbits

LEO	MEO	GEO
20-dB net advantage over GEO; reduced latency favored for voice	Medium altitude is compromise between LEO and GEO; reduced latency relative to GEO	Simplest and lowest in cost to implement and operate; latency an issue in some applications
Large constellation needed	Small constellation or pairing	Single satellite
Limited coverage; favors cross-links	Each satellite covers large landmass or ocean; cross-links of limited value	Each satellite covers a hemisphere; little or no use for cross-links
Nearly three-quarters of satellites over oceans at a given time	Satellite coverage extends across oceans	Satellite coverage extends across oceans and continents

11.2.1 Inmarsat (Generations 3 and 4)

The Inmarsat system evolved from an entrepreneurial start by COMSAT in the late 1970s. COMSAT used its win of a U.S. Navy contract for UHF capacity to launch L-band MSS service on a hybrid UHF/L satellite for commercial vessels. The first satellites, called Marisat, were built by Hughes and at least one was still in operation at the time of this writing. Having established the INTELSAT consortium for international FSS, COMSAT organized a private company, Inmarsat, as an international MSS joint venture of governments and telecommunication operators. Inmarsat is headquartered in London where it manages and operates a global GEO MSS system that serves L-band UTs and C-band GESs. Inmarsat owns and operates the space segment, while partners provide end-to-end services through land Earth stations (LESs).

A selection of Inmarsat terminal types is provided in Table 11.4, beginning in 1982 with the classic Standard A analog terminal for shipboard use. The antenna in this case is a parabolic dish about 1m in diameter, with a gimbal system to maintain pointing during ocean travel. Directional UT antennas were necessary to close the link to the low-power Marisat satellites, which were constrained by the small size of the spacecraft and the need to provide hemispheric coverage of ocean areas (recall that the oceans cover three-fourths of the Earth's surface). Digital communications were introduced first to improve service quality and then later, through the innovative Inmarsat M standard, as a means to reduce UT size, power, and cost. The Inmarsat M terminal, costing around $20,000 in 1996, introduced truly portable satellite voice service. Many who purchased such devices have found the quality to be very satisfactory and the cost per minute to be tolerable in light of the fact that worldwide telephone service is provided. With the launch of Inmarsat 3 and an agreement with Thuraya to lease capacity, higher speed service to very compact terminals is a reality. Examples of some of these terminals are shown in Figure 11.9.

The performance of Inmarsat GEO satellites has grown in power terms, culminating with the Inmarsat 4 series being constructed at the time of this writing and planned for introduction in 2005.

Table 11.4 Inmarsat Standards for Services and User Terminals

Year of Adoption	Satellite Series	Service Capabilities
1982	Inmarsat A	Analog telephone (FM-SCPC) and telex
1990	Inmarsat Aero	Aeronautical digital voice and low-speed data
1991	Inmarsat C	Low-speed data, briefcase keyboard terminals
1993	Inmarsat B	ISDN digital voice and data (64 Kbps), suitcase-sized land mobile terminals
1993	Inmarsat M	Compressed digital voice, briefcase terminals
1994	Global paging	Pocket size pagers
1995	Navigational services	Variety of specialized devices
1997	Mini-M	Laptop voice terminals
2003	Regional Broadband Global Area Network (B-GAN)	Palmtop medium data rate modem, ~144 Kbps, Europe, North Africa, the Middle East, and southern Asia
2005	B-GAN	Palmtop medium data rate modem, 144 Kbps, global footprint using Inmarsat 4

Regional BGAN
IP Modem (HNS)

1. Integral antenna
2. Compass
3. SIM card
4. Battery
5. External power
6. USB
7. Indicators
8. Ethernet

Figure 11.9 A selection of user terminals for use on the Inmarsat system.

Services over Inmarsat satellites are actually offered by specialist providers like Stratos Global, Telenor, and Singapore Telecom. These companies own and operate the land stations through which a majority of Inmarsat phone calls and data connections are delivered. Users arrange directly with these companies to use their UTs and are billed directly based on minutes or data transferred.

11.2.2 North American and Australian MSS Systems

American Mobile Satellite Corporation (AMSC) was first to launch a GEO MSS system in North America, using a medium power satellite called MSAT supplied by Hughes [1]. The company went public, reorganized several times, and is now operating as Mobile Satellite Ventures. The same MSAT platform was adopted for Canada by Telesat. In the early 1990s, the FCC recognized that L-band spectrum cannot be reused with orbit spacing, so they insisted instead that the applicants form a joint venture company to which the single license was awarded. More recently, the FCC used auctions to make such decisions, which had the added benefit of producing substantial revenues for the U.S. government. Telesat Mobile, on the other hand, began as a subsidiary of Telesat Canada, the monopoly satellite operator in Canada. As such, they were awarded the only Canadian license for L-band. The two companies entered into a joint purchase of a pair of satellites built in turn through cooperation between Hughes Space and Communications of the United States and Spar Canada. Figure 11.10 presents the 7-beam coverage of MSAT. The systems are identical and offer SCPC voice and vehicle position determination and low-speed

Figure 11.10   Typical MSAT antenna gain showing six area beams that cover the 50 states, Canada, and Mexico. This arrangement offers limited frequency reuse (e.g., between the easternmost and westernmost beams only).

data services. Typical user terminals are shown in Figure 11.11. The limited satellite EIRP and *G/T* precludes the use of handheld units.

AMSC entered service at the end of 1995 after the launch of the first MSAT. Telesat Mobile used this MSAT until theirs was launched in 1996. The architecture of the network is such that calls must connect through a GES to the PSTN, allowing mobile users to make and accept calls from subscribers on the fixed terrestrial network. The main market for service, determined by the technical capability of the network, is vehicular telephone service in areas not covered by cellular.

Australia was the first nation to introduce domestic MSS services through the MobileSat system. Implemented and operated by Optus Communications, the MobileSat service employs the L-band repeater onboard the Optus B satellites [2]. With a relatively small antenna system and low amplifier power, Optus B can service up to 500 simultaneous subscribers per satellite. Figure 11.12 illustrates the single-beam downlink coverage of Optus B. Users with vehicular or portable terminals can

Figure 11.11   Typical MSAT mobile UTs.

Figure 11.12   Optus B L-band downlink EIRP.

obtain mobile telephone service from anywhere in Australia. Currently, Australian terrestrial cellular is restricted to the larger metropolitan areas and corridors, leaving MobileSat a captive market in the Australian outback.

The supporting ground segment for MobileSat is very similar to that of AMSC, with the exception that a pair of GESs provide a fully redundant network. In addition to reliability, the scheme also saves on long distance costs since calls can be directed to the GES closest to the PSTN end of the connection. The hardware and software for the MobileSat ground segment were provided by NEC Corp. and CSC of Australia, respectively. User terminals were supplied by NEC and Westinghouse, the latter also being a supplier for the AMSC system.

11.3   GEO MSS Systems Serving Handheld Terminals

In June 1993, Hughes Communications, Inc., introduced the concept of handheld service from a GEO MSS system, providing the encouragement for a number of systems that began construction during 1995–1996. A brief summary of the characteristics common to the GEO MSS systems in production at the time of this writing is provided in Table 11.5; a similar link budget was presented in Table 11.1. The capacities listed in Table 11.5 represent the maximum number of simultaneous telephone calls that can be conducted over a single satellite. Each subscriber might use the service about 60 minutes, on average, per month. Assuming 8 hours for traffic purposes, a satellite can serve a subscriber population in the neighborhood of 1 to 2 million.

Table 11.5   Technical Summary for a Typical GEO MSS System Capable of Supporting Handheld User Terminals and Satellite Modems Capable of Medium- to High-Speed Two-Way Data Services

Characteristic	Value	Units or Comments
Frequency band		
Forward downlink	1,525–1,559	MHz (GEO MSS allocation)
Forward uplink	13,750–14,400	MHz (GEO MSS feeder link)
Return link downlink	10,250–10,900	MHz (GEO MSS feeder link)
Return link uplink	1,610–1,644	MHz (GEO MSS allocation)
Connectivity	Mobile to gateway	Connection to PSTN
	Mobile to mobile	Single hop, on demand
Services provided	Telephone	
	Fax	
	Circuit-switched data	Virtual private network
	Packet-switched data	Internet access
Satellite EIRP	72	dBW (aggregate, per beam)
Satellite G/T	15	dB/K
Polarization	Circular	
Station-keeping	0.1	Degrees, north/south and east/west; alternatively inclined orbit up to 6°
Channel capacity	13,750	Channels, mobile to PSTN
Call setup time	6	Seconds, to domestic PSTN

A critical part of the design of this generation of GEO MSS satellite is the technique for routing channels and calls between the beams (which number in the hundreds). Improving on the analog processing concept of MSAT, these advanced satellites take the next step of performing the operation in digital form using an onboard digital processor such as that suggested in Figure 3.8. The L-band spectrum in each beam is translated to IF as in the analog approach and converted to a digital representation in an analog-to-digital converter. From this point, the digitized information can be filtered, routed, and applied to the downlink with dynamic beam forming. After the necessary operations are performed, the selected information is routed to the appropriate transmission channel where it can be converted back to analog form. The resulting band of carriers is translated to either C-band (for mobile-to-gateway service) or again to L-band (mobile-to-mobile service) and transmitted through the appropriate amplifier and antenna feed.

The concepts previously described were employed to bring the first regional MSS system to the Middle East, Africa, Europe, and Southern Asia, that is, by Thuraya of the UAE. The name means "constellation" in Arabic and defines how the satellite services the region with a single satellite, shown in Figure 11.13, that covers literally every corner of the land with more than 200 spot beams. Two satellites, Thuraya 1 and 2, operate at 44 EL and 28.5 EL, respectively. The overall contract by Boeing (formerly Hughes Space and Communications Company) included two spacecraft, launch services (Thuraya 1 was successfully launched on October 20, 2000, and Thuraya 2 was launched in 2003), a ground segment supplied by Hughes

Figure 11.13   The Thuraya GEO MSS satellite, with its 12.25-m deployable antenna and phased array feed system.

Network Systems, and 235,000 handsets supplied by Hughes and Ascom of Switzerland. The 200 spot beams are generated using onboard digital processing and a sophisticated 12.25-m deployed reflector and phased array feed system, which are illustrated in Figure 11.14. An example of how the spots cover a region as large as Africa is shown in Figure 11.15 (not the specific coverage chosen by Thuraya but nevertheless feasible). Beams toward the western coast appear to be elongated because the map is a Mercator projection.

One of the more advanced features of Thuraya is its digital onboard processing repeater, shown in basic form in Figure 11.16. Transmissions between UTs of the handheld or mobile variety are made at L-band at 1,626.5 to 1,660.5 MHz through the 12.25-m antenna and the uplink from the gateway Earth station is at C-band at 6,425 to 6,725 MHz. The L-band downlink at 1,525 to 1,559 MHz employs 128 active SSPAs at 17W each, while the C-band return downlink at 3,400 to 3,625 MHz employs two 125-W TWTAs. The spot beams are created from only 128 individual dipole elements in the feed assembly; these are energized by the processor with appropriate amplitude and phase to produce the desired spots. In addition, digital beam forming allows Thuraya to enlarge beams and add new beams to respond to hot spots of demand. There is also the flexibility to allocate 20% of the total power to any spot beam. The system reuses spectrum by up to 30 times, based on division of the 34 MHz of L-band downlink bandwidth over a 200 spot beam pattern with a division by 7 to account for adjacent beam isolation. The latter provides adequate C/I between beams operating on the same piece of spectrum.

The specific signal characteristics are summarized as follows:

- Channel bandwidth of 27.7 kHz, capable of supporting a bit rate of 46.8 Kbps;

Stowed   Deployed

Figure 11.14   The 12.25-m deployable antenna for Thuraya.

- Modulation with $\pi/4$ QPSK;
- TDMA within the individual FDMA channels for up to eight multiplexed voice channels;
- Provision of data transmission in increments of 4.3 Kbps up to the carrier maximum of 46.8 Kbps.

Figure 11.17 presents how the processor receives, selects, and routes individual frequency channels. Any beam can be connected to any other beam directly for mobile-to-mobile services. Mobile-to-gateway service is provided through the C-band portion of the repeater shown in Figure 11.16.

Figure 11.15   Example of GEO MSS coverage beams.

Figure 11.16   GEO mobile payload with low-level digital beam forming and mobile-to-mobile channel routing.

The Thuraya system relies on a GPS receiver in the handsets to allow determination of which beam the handset should be considered to be in. During the recent Iraq war there were some concerns about confidentiality of the user locations. Thuraya system management made special efforts to assure their customers that both their identities and locations would remain confidential.

11.4   Non-GEO MSS Systems

At the time of writing our first edition, non-GEO systems were in development but not in operation. The situation at the time of this writing is that two systems, Iridium and Globalstar, are, in fact, operating and providing services to users in different parts of the globe. That neither succeeded in business terms is a well-known fact and both have been restructured. We continue to discuss non-GEO systems because they exist and represent an interesting challenge for those who would exploit them for mobile and fixed services.

Non-GEO orbits, as shown in Figure 11.18 for the Iridium system, give up the synchronization of satellite revolution to Earth rotation in exchange for the advantages of being closer to the Earth. The advocates of the non-GEO MSS systems emphasized the benefits of reduced propagation delay, which certainly is advantageous for interactive voice service. The path loss might be lessened by 30 dB, which is useful in reducing demands on the UTs; however, there are several inefficiencies in system application for LEO and MEO as opposed GEO systems. This 30-dB advantage is not fully realized due to lessened spacecraft antenna gain and greater near-far effect for users at the edge of coverage. Real-time services like telephone and circuit-switched data require that at least one satellite be in view between the link at any given time. Developers of non-GEO systems must deploy significantly more

Figure 11.17   The GEO mobile processor accepts a group of narrowband channels within an uplink beam of 30 MHz and rearranges them in the process of cross-connecting to the downlink beams. The rearrangement is done by selecting, frequency translating, cross-connecting, and then translating and summing them into another band of 30 MHz. This allows any user in a particular uplink beam to be connected to a different user in any other downlink beam.

satellites in order to maintain continuous coverage of the user. The consequence is that all of the non-GEO MSSs were expensive to implement and had to target a global market.

The non-GEO MSS systems that are intended for real-time applications in the global mobile telephone market are collectively referred to as the "Big LEOs." This is a bit of a misnomer because two of the contenders, ICO and Odyssey (which

- 66 satellites
- Six polar orbits
- Intersatellite links (shown
- in dashed lines)
- User links (shown in
- dotted lines)

Figure 11.18   The polar LEO constellation for Iridium, designed to provide true global coverage with provision of intersatellite links. Single-frequency L-band in used for user links; intersatellite and gateway links are at Ka-band.

merged and had not entered service by the time of this writing), are MEO systems. By way of contrast, the little-LEO systems are smaller in terms of the size and number of satellites, concentrating on market niches in applications such as low-speed two-way burst data messaging, radio broadcasting, and positioning. The economic viability of the latter has never been demonstrated. A company called Orbcomm began launching little LEO satellites in 1995 and initiated commercial service in February 1996 by offering two-way data messaging to inexpensive handheld radio transceivers.

In the following sections, we review the basic strategies available and the non-GEO systems that were constructed.

11.4.1   Iridium

The LEO approach, adopted by Motorola, locates the satellites in closest proximity to Earth, with an orbit period of about 90 minutes. This provides a time delay of less than 10 ms due to propagation but increases the number of satellites needed for continuous service. Importantly, Iridium represents the first system to offer ubiquitous service from space to handheld phones. The Iridium system provides location flexibility for the user by relaying calls from satellite to satellite through a network of intersatellite links. Except for control and tracking stations, the network is independent of the terrestrial infrastructure, allowing a user in, say, the Gobi Desert to speak directly to another user in the Sahara Desert. The system, however, did not go into service without its technical and financial challenges. The complex call handling, packet switching, and intersatellite link routing took a year of debugging. This factor, to be reviewed later in the chapter, requires that the user see the satellite as it rises above the horizon on the upward direction of its pass and follow it until it "sets" below the horizon on the downward direction.

The basic orbital arrangement of Iridium is shown in Figure 11.19, and some of the basic characteristics of the system are presented in Table 11.6 [3]. The constellation consists of six polar orbital planes (60° apart), with 11 satellites in each. The actual inclination is 86.4° rather than 90°. The benefit of this architecture is that is assures 100% coverage of the globe, including the poles. In fact, the polar coverage is better than that over any other region. Satellites are constantly in motion relative to the Earth and are interconnected through a system of Ka-band intersatellite links. Each satellite, illustrated in Figure 11.19, has 48 spot beams generated by a phased array antenna that links to the mobiles at L-band (see white bars in Figure 11.3). Unique for Iridium is that both the user uplinks and downlinks are in the same band.

The Iridium satellites permit users to communicate anywhere in the world, without passing through ground facilities. Alternatively, the network includes GESs that support calls to the PSTN and thereby to fixed wireline and cellular subscribers. Each beam covers a small area that, of necessity, moves across the surface of the Earth. Telephone calls must be handed off from beam to beam and satellite to satellite, whether the user is in motion or stationary. The GESs all use Ka-band tracking antennas to follow the satellites as they orbit the Earth. Although the L-band links with the users are in a narrow band, there is 20 times this bandwidth to and from the gateways and on the intersatellite links.

Figure 11.19 The Iridium spacecraft in orbit, indicating the L-band array antennas and the associated cellular-style beams, and Ka-band reflector antennas used for intersatellite links and space-ground links to the gateway Earth stations.

Motorola, prime contractor for the entire system, developed its own packetized voice transmission system using the same L-band frequency for transmit and receive. This TDD approach has a 90-ms time frame and allows both transmit and receive to be on the same frequency to reuse spectrum and allow the same propagation characteristics for both links. Motorola supplied the communications payload, utilizing the PowerPC microprocessor among other innovations, and entered into a partnership with Lockheed Martin for the construction of the satellites. Raytheon and ComDev supplied other RF and antenna components. Many of the original national service providers were existing cellular joint ventures between Motorola and local companies. At the time of this writing, the reorganized Iridium is in full business serving the U.S. government and a variety of industrial users. The number of GES has been greatly reduced to focus on the U.S. market. Replacement of the system at end of life has been under discussion.

11.4.2 Globalstar System

Motorola picked up some early imitators, the most ambitious of which was Loral's Globalstar. As reviewed earlier in this chapter, the FCC issued operating licenses to Iridium and Globalstar, and the ITU allocated the necessary feeder link frequencies at C-band. Figure 11.20 provides the basic outline of the satellite and the arrangement of the mobile user transmission links [4].

A summary of Globalstar technical characteristics is provided in Table 11.7. Coverage of the poles is not provided, while coverage of populated areas in northern latitudes is enhanced. The higher altitude reduces the required number of satellites,

Table 11.6   Summary of Key Characteristics of the Iridium System and Satellites

Characteristic	Value or Comments
Orbit altitude	780 km
Geometry	Polar orbits at 86.4° inclination
Number of orbits	6
Satellites per orbit	11
Total number of satellites	66 plus spares
Number of beams per satellite	48 at L-band
User links	1,616–1,626.5 MHz both up and down
Gateway downlinks	19.4–19.6 GHz
Gateway uplinks	29.1–29.3 GHz
Intersatellite links	Ka-band at 23.0–23.4 GHz to adjacent satellites in same plane and adjacent planes (total four ISLs per spacecraft)
Repeater design	Onboard digital processing of packets
Multiple access	TDMA (Time division duplex)
Satellite lifetime	6 to 8 years, subject to available fuel and battery performance
System capacity	72,600 circuits worldwide (effective capacity of 16,700 circuits)
Channel bandwidth	31.5 kHz
Channel data rate	50 Kbps
Modulation	QPSK
Channel coding	$K = 7$, $R = 3/4$
BER	10^{-5} after decoding
User link margin	16 dB

and the inclined orbit raises the operating elevation angles. Globalstar, like Iridium, provides too much coverage of the ocean regions, resulting in less effective capacity over land areas than can be obtained with either a MEO or GEO strategy. Without intersatellite links, Globalstar cannot serve users out on the oceans since an operating GES must simultaneously be in view. This is in contrast to Iridium's ability to serve users no matter where on or above the planet they may be. Through a relationship with Qualcomm, Inc., Globalstar provides service using of the CDMA cellular standard.

The satellite uses a simple bent-pipe repeater rather than a digital processor such as that being flown on Iridium. A phased array antenna with 16 spot beams provides gain and a degree of frequency reuse, as illustrated in Figure 11.20. All transmissions are relayed at C-band (7/5 GHz) through GESs that track the satellites as they pass by. This places a constraint that a given LEO satellite must simultaneously see both the mobile user at L/S-bands and the GES at C-band.

11.4.3   ICO Communications

The last non-GEO system we review is the ICO system that was originally a spin-off from Inmarsat. The Inmarsat Council, the governing body of Inmarsat, established

Figure 11.20   Globalstar mobile user links using L-band for the uplink and S-band for the down-link. The satellite employs a hardwired phased array system with 16 beams.

ICO Communications Ltd. in the United Kingdom to implement and operate the project that previously had the working title of Inmarsat P (personal). With 37 supporting founders representing as many countries, ICO was potentially in a strong business and financial position to reach the operation stage and, ultimately, success in the market. However, the company ran into insurmountable difficulties in finding the funding needed to complete the system. The company underwent several reorganizations involving Teledesic and Craig McCaw; further, it was announced in August 2003 that Boeing would complete satellite construction.

Inmarsat began its evaluation of how to provide handheld service by making a thorough comparison of LEO, MEO, and GEO constellations. Their first decision was to drop the LEO approach in favor of either MEO or GEO. They based their decision on the cost and risk associated with each approach. Then, after working closely with Hughes Space and Communications and TRW, as well as other suppliers of satellite and ground equipment, they made their final selection of the MEO. The use of the MEO brings with it a greatly reduced quantity of satellites that operate at higher altitudes. The lengthened propagation delay and increased path loss are accounted for in the network, but the satellite moves much slower and would normally be visible for the duration of a very long telephone call. Elevation angles also are favorable for the majority of the time. There is global coverage of land and ocean, using GESs connected by terrestrial leased fiber.

The 12 satellites that are needed to meet the service objectives were constructed by Boeing Satellite Systems; the program had been started and stopped and appears to be back on track. The HS-601 spacecraft was modified for the MEO mission, and

Table 11.7   Summary of Characteristics of the Globalstar System and Satellites

Characteristics	Value or Comments
Orbit altitude	1,401 km
Geometry	Inclined 52°
Number of orbits	8
Satellites per orbit	6
Coverage	Latitudes up to about 70°
Total number of satellites	48 including spares
Number of beams per satellite	48 for Earth to space at L-band
User links	2,483.5–2,500 MHz downlink 1,610–1,626.5 MHz uplink
Gateway downlinks	5 GHz
Gateway uplinks	7 GHz
Intersatellite links	None
Repeater design	Analog bent pipe
Multiple access	CDMA
Satellite lifetime	7.5 years
System capacity	Sensitive to assumptions about CDMA loading
Channel bandwidth	Approximately 1.2 MHz (Qualcomm CDMA standard)
Channel data rate	1.2 to 9.6 Kbps (compression: 9.6, 4.8, or 2.4 Kbps to accommodate variable power margin)
Modulation	QPSK
Channel coding	$K = 9$, $R = 1/2$

the repeater is based on digital onboard processing. The frequencies to be used for the links to UTs are at the upper end of S-band, being taken from an ITU allocation for land mobile services (see the wide white bars in Figure 11.3). Feeder links operate at C-band, which is the Inmarsat standard. The system employs TDMA, and all calls are between the mobile user and the GES for access to the global PSTN. Direct mobile-to-mobile calling requires a double hop, which should still be acceptable due to the shortened propagation delay relative to a GEO MSS system. Calls may also be routed between regions over the interconnecting landlines.

ICO Communications anticipated that most users would want to employ hand-held UTs. For this reason, they opened up the specifications to industry to draw as many sources of UTs as possible. Like Thuraya, discussed in the section on GEO MSS, ICO uses GSM as its foundation.

11.4.4   Comparison of the Performance of Non-GEO Systems

The most technical success factor in the operation of an MSS system designed for real-time voice communication is the continuity of service. Nothing is more disturbing than for a call, once connected, to be terminated due to a lack of a satellite to access. Assuming that the user is standing still during the call, the only way for the

connection to be dropped is if the current satellite goes out of view before the next one rises sufficiently above the horizon for a line-of-sight path to exist. The situation with a GEO satellite is that once you have a connection, it is not likely to be lost unless the user relocates where a line of sight does not exist or multiparty cancellation occurs.

The best way to illustrate this effect is to examine the variation in local elevation angle between the user and satellite for some of the systems just described. Figures 11.21, 11.22, and 11.23 provide the elevation angle from the ground to the satellite for Iridium, Globalstar, and Odyssey (a system designed by TRW that is very similar to ICO) [4]. The diamonds indicate the minimum elevation angle that would occur during the pass of each satellite. For consistency, the charts are plotted for a common local latitude of 55 N, corresponding to locations in the United Kingdom, Germany, Russia, Canada, and Alaska. Iridium, which this scenario tends to favor, still experiences a rather low minimum elevation angle of 15°. Globalstar and Odyssey do better, where the minimums are 20° and 25°, respectively. A compensating factor for Iridium is its 16-dB link margin, 10 dB more than the other systems. This helps overcome terrain blockage near the horizons. In summary, non-GEO systems have not achieved their goal of a true global mobile personal communications service; however, that they represent a milestone in new system development cannot be denied.

11.5 Intelligent MSS Services

The typical MSS system is capable of delivering a comprehensive package of advanced telephone services, similar to the offerings of GSM and other digital

Figure 11.21 Iridium coverage at 55 NL (66 satellites, 86.4° inclination, 785-km altitude). (Courtesy of David Bell.)

Figure 11.22   Globalstar coverage at 55 NL (48 satellites, 52° inclination, 1,406-km altitude).
(Courtesy of David Bell.)

telephone systems. This means that a subscriber with an appropriate UT can engage in a variety of activities from anywhere in the coverage area. These services mirror the intelligent network, affording users the freedom to conduct business from remote locations just as if they were in their home offices.

Figure 11.23   Odyssey coverage at 55 NL (12 satellites, 55° inclination, 10,354-km altitude).
(Courtesy of David Bell.)

11.5.1   Mobile Telephone and Data Services

The foundation of any MSS system is the provision of mobile telephone service that is capable of connecting calls to the PSTN. In Chapter 10, we reviewed how a satellite network can provide telephone services among fixed locations on the ground. The concept behind mobile telephony is to carry out the same function but to allow the added dimension of mobility. It should be no surprise that much of the same architecture and technology that was proven in FTS networks has been taken over to the MSS field. Furthermore, terrestrial cellular telephone standards like AMPS, TACS, GSM, and CDMA2000 have achieved total acceptance by the user community. This is because the subscriber can conduct business over the telephone without being tied to the wireline network. Satellite mobile telephone provides similar mobility but over a much broader area.

In circuit-switched services, voice calls and fax calls are the dominant applications. A voice call provides a point-to-point connection between mobile user and either a PSTN subscriber or, if provided, another mobile user. Subscribers must be able to originate calls in either direction. The network must be able to locate a mobile subscriber, ring the UT, and establish the communication path. To make this happen, the satellite broadcasts an alerting channel throughout the coverage region. The UT, when first activated, detects this channel and responds by transmitting a message that indicates it is prepared to receive calls. The MSS network then goes through an authenticating process before allowing service. From that point, a GES can direct calls to this particular UT when entering the network from any other direction or location. The origination of calls from the UT is very straightforward, since it is already in contact with a GES.

The quality of voice calls is determined by the type of digital compression that is employed as well as the local fading environment. All MSS systems that support handheld UTs use highly compressed speech to get the data rate below 5 Kbps. Speech quality is rated as fair to good (numerically, 3.5 out of 5), meaning that the subscribers consider it to be significantly poorer than modern PSTN services using fiber optics and uncompressed speech and only slightly worse than terrestrial digital cellular. This rating probably means that subscribers will gladly make telephone calls over the satellite network when nothing else is available. Old technology like analog cellular and high-frequency radio would receive a lower rating still. Additional information on compression standards and how they are rated is presented later in this chapter.

The intelligent nature of the MSS network will allow some attractive voice services to be introduced quickly and efficiently without added cost. Voice-mail (also called call answering) is very popular on cellular and wireline networks. It is provided through the mobile switch that is associated with the GES. Each subscriber is given a voice mailbox to retain all messages. A call will be directed to the voice mailbox if the subscriber's phone is off, out of range, or engaged in a call. The function is exactly the same as voice mail over a PBX or local telephone network. Voice-mail is potentially very valuable to a mobile subscriber, allowing access to the mailbox from anywhere within the coverage area of the satellite.

Other services through the GES include call forwarding, call blocking or barring (to filter or block calls based on the caller's number), caller identification (subject to local regulatory rulings), call holding (e.g., putting a caller on hold to allow another

call to be taken or placed), and call waiting (e.g., you hear a tone when a second caller is ringing your line). While any and all of these are relatively easy to implement, there may be circumstances where the network operator will not make them available to subscribers. Satellite phones usually employ the same type of subscriber identification module (SIM) card as GSM and GPRS land mobile services.

Data and fax calls represent an important and growing application of mobile telephone networks. A satellite MSS service must be capable of supporting these needs in order to compete effectively. The speech processing and compression must be bypassed so that the data is applied directly to the UT transceiver subsystem, a technique already applied in FTS networks (Figure 10.10). At this point of interface, the channel unit can apply forward error correction to the unmodulated data. The speed of transmission will depend on the operating mode of the UT. In one system, a fixed data rate of 2,400 bps is available to data and fax. However, this speed is much slower than that which is available over the PSTN, so newer MSS systems are pushing to provide 9,600 bps to meet user needs for reasonably fast fax transmission and computer network access. GEO MSS systems are already offering 100-Kbps high-speed data access using specialized satellite modems.

SMS is a direct competitor to the offerings of little LEO systems. As a one-way receive only service, it effectively implements satellite-based paging. A connection between sender and recipient is not required because the network will forward the message. The time to accomplish this will depend on the quantity of messages being processed at the same time and the data transfer rates along the information path. While not as effective as terrestrial paging in terms of its ability to penetrate buildings, satellite message broadcasting would be available over an extremely wide area where paging is not available commercially. A two-way messaging capability is provided by the UT using a burst mode of transmission.

11.5.2   Handheld User Terminals

There are a few issues regarding the use of handheld UTs in a GEO MSS system. Because the link requires a near-line-of-sight path to the satellite, the user must employ the phone either outside in the open or, if within a building, at a window that faces toward the satellite position. Users who are accustomed to cellular service where it is possible to make and receive calls from within a building and on the outside without a direct path to a base station may find MSS to be more constraining. The end result is that any GEO MSS system providing service to handheld phones requires a certain degree of user cooperation. By this we mean that the user must be aware of the unique constraints of MSS and act accordingly. For best performance, the user should remain stationary after acquiring the satellite. Since the path to the satellite is stable under this condition, the user can continue to talk for an indefinite period, limited only by the available battery power. The situation for non-GEO MSS links is less favorable because the path changes even if the user is stationary. Operation from within a moving vehicle, a popular mode in cellular systems, is not recommended because of the difficulty of obtaining and maintaining the line-of-sight path. For vehicles in motion, the preferred approach is to use an external antenna and a docking set that transfers signals between the antenna and the handheld phone.

11.5.3   Vehicular Terminals

The vehicular UT will normally consist of a standard handheld phone and the dock-ing set mentioned in the last section. This allows the unit to be powered by the vehi-cle battery and the antenna to be attached to the exterior of the vehicle where a view of the entire sky is possible. RF output power will be limited to what the handheld UT can produce. The net EIRP and G/T will not be appreciably different from the handheld by itself because the loss in the cable will be compensated by the improved visibility of the exterior-mounted antenna.

There are applications in commercial transportation involving trucks and buses where the docking set is inappropriate. Instead, an installed transceiver and improved antenna will be the likely approach. Within the vehicle cab, the user will employ a mobile phone handset of the type currently in use in cellular vehicular installations. A hands-free capability can be included, as it is unlawful in many coun-tries to hold a handset while driving. The transceiver, with its power amplifier, will be installed in the trunk or boot of the vehicle, where it is out of the way and in close proximity to the antenna to minimize RF losses. A control and signal cable connects from the handset to the transceiver.

Both AMSC and MobileSat began their services with this type of installation. The service quality from such an installation is enhanced by a more stable link and features like hands-free operation. Another benefit of the permanent approach is that other facilities can be added, such as a fax interface, data communication adapter, position location, and imaging. The vehicular terminal could cost signifi-cantly more than the consumer-oriented handheld UT; however, the user has access to potentially lower airtime charges from the GEO MSS service provider.

11.5.4   Fixed Telephony User Terminals

We introduced in Chapter 10 the concept of using an FSS satellite as a networking resource for thin-route telephony services. The types of FTS terminals discussed pre-viously are probably too expensive to be used for home services; they likewise may be out of the reach of small businesses wishing to use them as part of a private net-work. The associated FSS space segment costs also tend to favor bulk transmission of telephone channels as opposed to selling minutes of transponder time to individ-ual users. As a consequence, FTS has not been extended to the general public in the same manner as DTH or broadband Internet access. This may change one day, but as of the time of this writing, only MSS systems offer the potential of direct-to-user telephony service, VoIP notwithstanding.

To address the individual user in need of basic telephone service, a GEO MSS system has a number of attractive features. First, and most important, the basic handheld UT design must be produced to sell at a price that is affordable to middle-income individuals. Airtime is offered on a per-minute basis, and access to the PSTN is seamlessly provided through the facilities of a GES. A fixed telephony installation only requires an antenna aligned with the satellite and the necessary elements of the UT. The subscriber could employ a four-wire telephone that interfaces with the UT. The entire installation would cost only slightly more than the handheld unit. The antenna, about half the size of a DBS TV antenna, would be attached to the exterior of the building and pointed in the direction of the satellite. A low-noise L-band

preamplifier on the antenna provides gain to overcome the loss of the cable that connects to the actual UT/telephone. Transmit RF output from the UT would, when combined with the enhanced gain of the antenna, be sufficient to deliver adequate EIRP.

The quality of service for this type of FTS installation would be superior to either the handheld or vehicular UTs. The reason for this is that, with constant alignment of the antenna to the satellite, the link will rarely fade below threshold. Voice quality will be limited only by the selected digital compression system. The UT can support fax and circuit-switched data through the appropriate access port.

The attractiveness of this approach to FTS results from the availability of a low-cost UT. Satellite air time would be charged based on the quantity of users and the amount of power and bandwidth per call. With the improved gain of the FTS antenna installation, it is likely that airtime can be sold to such subscribers for about half of what a handheld user would experience. This might still be expensive in comparison to basic PSTN telephone service within a city but might be competitive with what is charged for international long distance service.

A special application of FTS via MSS would be to install a shared terminal in a village or industrial site. Individuals would gain access to the UT through a manual switchboard or a small digital exchange. The cost of installing this type of UT is substantially less than that of extending the PSTN. Therefore, it might be effective to subsidize the price of making calls in order to offset capital commitments. PSTN service can be extended terrestrially once the demand exceeds a particular level. Commercial installations would use a PBX to permit users to speak to each other within a facility (e.g., factory, construction site, and petroleum production facility) and to access the FTS terminal to make long distance calls.

Aeronautical and maritime UTs offer special opportunities to improve communications to passengers and crew. In this case, the cost of the UT is substantially more than that of the handheld or vehicular models. This is because of the complexity of installation of the antenna and cabling system within the structure and electrical system of the aircraft or vessel. The operation of an aeronautical or maritime service might generate some nice revenues that justify the incremental investment and could be a differentiating factor for premium class travelers.

11.5.5   Broadband Data Terminals

Needless to say, one of the most attractive data applications is for broadband access to the Internet and private intranets. MSS systems of the past were designed primarily for voice services, and the majority of connections are for that purpose. Since the first edition of this work was written, the Internet has become as much a facet of telecommunications as voice, and so MSS systems now take it very seriously. With the writing of this second edition, only a few such terminals and services have appeared on the commercial scene. As a result, the number of credible examples is somewhat limited.

Perhaps the first true MSS broadband data terminal is that introduced by Inmarsat for the Regional Broadband Global Area Network (R-BGAN). This is a precursor to Inmarsat 4, which will focus on providing better broadband access throughout the world. For the time being, Inmarsat has leased L-band spectrum on

Thuraya 1 and is offering the service directly to users within this satellite and coverage area. The R-BGAN terminal illustrated at the right of Figure 11.9 provides up to 144 Kbps of bidirectional IP data service. Developed by HNS, the device is described as an IP modem and would be connected to an IBM-compatible computer using the Universal Serial Bus (USB) interface. The lightweight unit (1.6 kg) sells for around $1,500 and includes these features:

- "Always-on" service availability at up to 144 Kbps;
- Integral antenna, which is aligned with the GEO satellite;
- Compass and GPS navigation;
- SIM card (GSM/GPRS);
- Battery;
- External power connection;
- USB connection;
- Ethernet (RJ-45 jack);
- Integrated Bluetooth wireless access.

In a recent review by *PC Pro* magazine, the unit performed its function well and provided a solid 64 Kbps of data throughput (e.g., 7.5 to 8.5 KBps) [5]. Owing to the location in the United Kingdom, the local elevation was only 17°. The ping latency was about 800 ms, yet the unit performed well in terms of interactivity for basic office applications. The connection was "a lot more reliable than a GSM/GPRS connection." The cost of the service is measured in data quantity transferred, currently being offered at $2.50 per megabyte. Using the Thuraya satellite, Inmarsat advertises that the service is available in 99 countries across Europe, the Middle East, the Indian subcontinent, and the northern half of Africa. As is always the case, the ability to import the modem and operate it may be subject to local approval.

11.6 Multiple Access in MSS

The selection of the multiple access method is always an important technical and operational decision because it can influence critical parameters like service availability, system capacity, and network connectivity. From a financial standpoint, capacity is very critical since it directly determines the quantity of users that a fixed investment in space can support. MSS systems generally have a limited amount of available spectrum and are subject to the complicating factor of mobile link fading.

The standard choices in increasing order of bandwidth per carrier are FDMA, TDMA, and CDMA. FDMA signals are assigned to individual SCPC frequencies and are active when users are talking. In TDMA, there are fewer transmissions (carriers) visible, but each is the result of several users time-sharing the same channel. Finally, in CDMA, the bandwidth of each single channel of communication is expanded by a digital code. Because the code is noise-like, users can transmit one on top of the other, as will be explained next. All three access methods are at least 30 years old in terms of actual use in radio communication systems. There is really nothing new here; rather, it is a case of how effective the scheme performs under

specific conditions (and these vary greatly) and how the associated hardware can be implemented within, say, portable and handheld phones.

11.6.1   Applying FDMA to MSS Service

The first multiple access scheme to be applied to satellite communication and the one with the most established track record (including in terrestrial cellular) is FDMA. An example of spectrum utilization, along with the three key benefits, are presented in Figure 11.24. The simplicity of FDMA lies in the fact that each user transmission is separated from every other by using a different frequency. A user terminal can receive a selected signal by simply tuning to the associated frequency in the same manner that you tune a television or FM stereo radio. The frequencies are assigned to users from a pool that is under the control of a central management resource that is also in direct contact with all users through a common signaling channel.

An MSS FDMA network and its FDMA terrestrial cellular counterpart can multiply the available channel space by reusing frequency channels across geography. The available frequency bandwidth (typically in the range of 10 to 30 MHz) is divided equally among the seven cells in the pattern so that adjacent cells do not use the same frequencies. Terrestrial cells are defined by the radio transmission range of cell site towers, while MSS cells are created by contiguous spot beams from the spacecraft antenna. This is illustrated in Figure 11.25 for an array of 14 spot beams and a seven-segment division of the usable spectrum. Dividing the total number of beams by the number of segments, we find that the effective frequency reuse is 100% in this example (e.g., 14/7 = 2). The seven-cell reuse pattern requires that there be a gap of two cells between any cells that use the same band segment. This is evident in Figure 11.25 in that segment 1, for example, is used in the center of the pattern and not again until the lower right-hand corner. A more intensely replicated pattern, like that shown in Figure 11.15, employs hundreds of much smaller spots. These greatly increase the frequency reuse at the same time that the higher gain enhances link capability.

FDMA can be practiced with analog or digital modulation because the individual carriers are completely separate from one another. The flexibility of channel assignment also means that different modulation methods can coexist within the same piece of spectrum. This provides an easy way to integrate nonhomogeneous networks and to provide for a smooth transition from analog to digital.

Digital SCPC with demand assignment was introduced in the early 1970s in the INTELSAT system to provide full connectivity with a common satellite. Voice and data service at high quality was afforded by a channel rate of 64 Kbps. This eased the problem of providing international telephone service among all of the countries

6 to 30 kHz

Flexible
Widely used in satellite and terrestrial networks
Low complexity

Figure 11.24   Application of FDMA for MSS networks. FDMA is applied in the Inmarsat M, Optus, and AMSC systems, using digital speech compression.

Frequency divided among beams (100% reuse)

Figure 11.25   Satellite antenna multibeam coverage with seven frequency segments.

of a region, which would have required literally thousands of permanently assigned links. The early networks did not pursue high degrees of digital compression simply because of the limitations of microchip technology available at the time. Today, digital compression allows the voice data rate to be reduced to about 5 Kbps, which greatly benefits a link compared to analog FM. The channel data rate, including overhead and FEC, would be approximately 6 Kbps. Using QPSK, the bandwidth of a typical SCPC carrier using speech compression and FEC is under 4 kHz.

The ultranarrow bandwidth of compressed digital SCPC is a potential source of concern when it comes to assuring adequate frequency stability. The guard band between carriers must be held to a small number, perhaps only 500 Hz, to achieve good bandwidth utilization. The necessary stability can only be obtained with a system of frequency tracking within the Earth stations and the use of a very stable local oscillator onboard the satellite. Another source of error is Doppler shift due to relative motion between the satellite and the various terminals on the ground. A GEO system experiences low Doppler rates and therefore can employ FDMA without much loss of bandwidth. For LEO systems, this is a serious problem, and for that reason FDMA has not been adopted. A MEO system could be a possible candidate for FDMA, provided that the relative velocities are within certain bounds. Implementers of LEO and MEO systems have chosen TDMA or CDMA, which are of sufficient bandwidth not to suffer substantial loss of bandwidth due to Doppler or oscillator frequency error.

In terms of hub utilization, FDMA requires a separate channel transceiver per active mobile user. Therefore, the hub must be equipped with a number of such

transceivers to support the peak expected voice calling demand. Channel transceivers are tuned to appropriate frequencies under control of the demand assignment system. Other parts of the hub, notably the antenna, transmitter, switching equipment, and support systems, can be shared.

11.6.2   TDMA in MSS

TDMA is inherently digital because information must be stored and subsequently transmitted in the form of a burst at high speed. It was developed for the INTELSAT service in the late 1960s to efficiently use a 36-MHz transponder at a transmission rate of 60 Mbps. This is 1,000 times faster than we need to apply to MSS. Figure 11.26 indicates how spectrum is occupied by TDMA carriers, each of which supports several users. For example, a channel that is shared by eight users would operate at eight times the data rate and bandwidth of a single FDMA user. This amounts to about 48 Kbps, assuming 6 Kbps per user; the channel bandwidth would be approximately 30 kHz. TDMA is currently used in VSAT satellite networks and is the core of the GSM digital cellular standard. As noted in Section 11.4.1, the Iridium uses a form of TDMA called time division duplex wherein both directions of transmission—from multiple users—share a common channel.

   TDMA systems demand precise timing among users that share the same channel. The necessary synchronization is accomplished with circuitry at the hub that locks to the individual transmission from the UTs. In reality, there are several TDMA channels in operation at the same time, each on a different frequency. This reduces the required power per carrier and simplifies the design of the receiver. Multiple carriers, therefore, interpose a form of FDMA on top of which is TDMA. Due to its wider bandwidth, TDMA is more tolerant of frequency errors and Doppler shift. However, burst timing must allow for time errors introduced by relative motion. The multiple carrier approach provides the opportunity to improve channel utilization through frequency hopping (e.g., F-TDMA). A UT can jump between carrier frequencies to grab unused time slots. TDMA also facilitates handoff in non-GEO networks.

11.6.3   CDMA

CDMA approaches multiple access in a unique way by transmitting signals literally one on top of another. A basic introduction to CDMA is provided in Chapter 2.

Figure 11.26   The TDMA format may be applied to MSS systems in a similar manner as digital cellular (e.g., GSM). The specific frame time and bit rate may be altered to suit the unique conditions of the satellite link.

Such direct jamming would render FDMA and TDMA useless, but for CDMA this kind of interference is managable because signals are uncorrelated. This is achieved either by using a different code for each transmission or by offsetting the transmissions in time (the latter being termed rode reuse multiple access, discussed in Section 9.1.3.3). The code is in the form of a random bit sequence that is multiplied by the digitized data. The speed of this random bit stream, called the chip rate, is tens or hundreds of times faster than the original data, thus expanding the signal bandwidth by the same ratio. The bandwidth-expanding property of CDMA has caused it to be called discrete sequence spread spectrum modulation. (Another CDMA approach called frequency hopping spread spectrum takes an FDMA carrier and moves it according to a random pattern across a much wider frequency band.) The information can only be recovered by multiplying the incoming CDMA signal by the original high-speed chip stream (a process called correlation detection). If the two streams are precisely synchronized, the user data literally pops out. Otherwise, the output is indistinguishable from noise.

CDMA is actually older than TDMA in terms of the art and practice of radio communication. Much of the research and application engineering work on CDMA has been shrouded in secrecy since World War II. This is because CDMA has some particular benefits in military communications and air defense. For example, a signal can be spread so much that it disappears in the ambient noise present in radio receivers. Another military benefit is that a narrowband jammer that is present within the band of the CDMA signal is spread out when mixed with the chip rate in the receiver. In contrast, the spread CDMA signal shrinks back into its original form in the same action so that the jammer has practically no effect. The ratio of chip rate to information rate, when converted to decibels, is called the jamming margin because it indicates how well a CDMA receiver can suppress narrowband interference. The other side of the coin is that if, say, 100 jammers were within the bandwidth of the CDMA signal, the interference in the above example could potentially be harmful.

The following is a typical example. The user data is at 6 Kbps, while the random bit stream (and hence the CDMA signal) is 1 mega chip per second. The ratio of chip rate to data rate is approximately 167, which is also the ratio by which the bandwidth is expanded. This factor determines how well the correlation receiver will reject the jamming and interference (e.g., the jamming margin previously defined). The spreading of the carrier reduces both its tendency to cause interference (because the power spectral density is reduced) and its sensitivity to interference from any signal, which is narrower than the spread spectrum signal. A narrowband interference within the bandwidth of the CDMA signal would have to be approximately 167 times as powerful to produce the same effect as two equal FDMA or TDMA signals interfering with each other. Another impact of spreading is that the ratio 167 is also approximately equal to the channel capacity.

Bandwidth utilization, as well as some of the applicable characteristics of CDMA, are presented in Figure 11.27. The narrowband form of CDMA occupies a channel bandwidth of 1.25 MHz, similar to the previous example. This is the approach taken by Qualcomm in the Globalstar system. Multiple CDMA channels can be assigned across the allocated spectrum, which allows the network to use frequency reuse through beam isolation techniques. Alternatively, the network might

Figure 11.27 Application of CDMA for MSS networks. Both the narrowband and the wideband cases are illustrated. The Globalstar non-GEO MSS network uses narrowband CDMA.

assign the same channels to adjacent beams and use the inherent isolation afforded by CDMA itself. Wideband CDMA, shown on the second line of Figure 11.27, is based on allowing the carrier to occupy the entire allocated spectrum. This might amount to 10 MHz, for example. All signals would have to be transmitted on top of each other.

There are some important issues with regard to the performance of CDMA in a fully loaded system. If the CDMA carriers within the satellite repeater are all equal in power, then it is very easy to calculate the capacity. Because it is very difficult to maintain carriers within tight bounds, power levels are not equal. As a user moves from position to position, the power reaching the satellite can vary over a wide range due to multipath, shadowing, and antenna misalignment. The system must use automatic power control to compensate for variations in path and shadowing loss, but, due to propagation delay, this cannot be perfect. This is the defining factor for capacity, which can be degraded by a factor of two or more. Another consideration is the degree to which CDMA networks can share the same bandwidth. While cohabitation with TDMA was advanced by CDMA proponents, the FCC ultimately decided that the three successful Big LEO applicants were to be given their own non-shared band segments.

11.6.4 Comparison of FDMA, TDMA, and CDMA

A summary of the comparative properties of FDMA, TDMA, and CDMA is provided in Table 11.8. The properties in which we are interested concern the use and reuse of frequency spectrum and the ultimate capacity that the multiple access technique can deliver. FDMA is straightforward in terms of how it is applied to MSS and the performance that results. TDMA has some benefits from its ability to hop among carrier frequencies (e.g., F-TDMA, discussed in Chapter 9) and since it can be made to only transmit when there is data to be sent. Furthermore, its wider bandwidth is more tolerant of frequency error and doppler shift than FDMA. Finally, CDMA offers superior interference rejection properties. In exchange, the performance of CDMA in a fully loaded MSS system has not been determined on a practical basis at the time of this writing. Each technique can provide a viable access methodology for terrestrial cellular and satellite mobile systems. There are many trade-offs

Table 11.8 A Comparison of the System Characteristics of Multiple Access Techniques as Related to Mobile Satellite Service

Characteristic	FDMA	TDMA	CDMA
Bandwidth utilization	SCPC	MCPC, partial allocation	SCPC partial or full allocation
Interference rejection	Limited	Improved using frequency hopping	Can suppress interference, up to noise or capacity limit
Frequency reuse	Requires seven-cell pattern	Requires seven-cell pattern	Zero to four cell reuse possible, up to noise limit
Intermodulation effects	Most sensitive (backoff required)	Less sensitive (less backoff required)	Least sensitive (least backoff required)
Doppler frequency shift	Establishes lower limit to bandwidth	Burst-time limiting	Removed by receiver
Spectrum flexibility	Most flexible since uses least bandwidth per carrier	Moderate bandwidth flexibility	Least flexible as requires large contiguous bandwidth
Capacity	Easy to access	Can provide capacity improvement through F-TDMA	Capacity may be indeterminate due to loading unknowns

in the selection, but there does not appear to be an overriding benefit to one with respect to the others.

11.7   Digital Speech Compression

Voice coding and compression is a process whereby speech is converted into a digital bit stream and compressed to reduce the amount of bandwidth needed for transmission. It is a technique that has many alternative implementations, some of which we will review later. The goal is that the number of bits per second is cut substantially; yet the listener may not detect that there has been an alteration in the signal. The reality for highly compressed speech is that what is actually recovered at the receiver has been changed to some extent. Compression of this type can increase the effective capacity of an MSS or cellular network by a factor of 4 to 20. Coding algorithms fall into three categories (and these are reviewed next): waveform coders, vocoders, and hybrid systems.

There is a wide range of coding and compression systems that have evolved over the decades. As digital technology has improved on the microchip side, so have the algorithms that provide the theoretical basis for efficient and high-performing compression systems. Table 11.9 provides a summary of the progression of coding standards, beginning in 1972 when PCM became firmly established as the 64-Kbps standard for the PSTN worldwide [6]. Low-delay coder/decoders (codecs) were introduced over the years, bringing the rate down to 16 Kbps without seriously affecting the perceived speech quality. Breaking the 12-Kbps barrier occurred with the Skyphone and North American digital cellular systems, which became available in the early 1990s. While the commercial systems pushed the envelope further to the 4.8-Kbps level by 1995, the U.S. government had long ago been using high degrees of compression down to 2.4 Kbps as far back as 1977. The quality of these vocoder

Table 11.9   Digital Speech Coding Systems and Standards in Use Since 1973

Data Rate (Kbps)	System or Standard Applied	Compression System or Standard	Class	Year First Applied
64	PSTN	PCM – G.711	Waveform	1972 (and before)
32	PSTN	ADPCM – G.796	Waveform	1984
16	Inmarsat Standard B	APC	Hybrid	1985
16	VSAT networks (HNS)	RELP	Hybrid	1987
13	GSM – full rate	Regular-Pulse Excitation Long-Term Prediction (RPE-LTP) – GSM 05.03	Hybrid	1988
9.6	Skyphone (aeronautical)	Multiphase Linear Predictive Coding (MPLPC)	Hybrid	1990
8	North American Digital AMPS (IS-54)	CELP	Hybrid	1992
7	GSM – half rate	Vector-Sum Excited Linear Predictive (VSELP)	Hybrid	1994
4.8	Inmarsat M, AMSC	Improved Multi-Band Excitation (IMBE)	Hybrid	1993
4.2	Inmarsat Mini-M, Iridium	Advanced Multi-Band Excitation (AMBE)	Hybrid	1996
3 to 4	—	AMBE+	Hybrid	2002
2.4	U.S. government federal standard	LPC-10	Vocoder	1977

systems was far from satisfactory for commercial service because of the robotic nature of the speech sound. No network or cellular quality codec was on the market at 2.4 Kbps at the time of this writing, although one supplier, DVSI, has published characteristics.

Human speech is a remarkable communications facility. Often, our communication involves more than just words and phrases—that is, content that can be transferred more quickly in text form. We also communicate our respective personalities and feelings in the tone, velocity, and articulation of what we say. A standard telephone channel, which occupies 3,000 Hz of bandwidth and requires 32 Kbps of data rate for transmission with good fidelity, will convey essentially 100% of this information without degradation. Specialists in this field refer to this level of service as network quality or toll quality. Essentially 100% of users would rate this quality as good to excellent.

Voice quality is measured using a scale of 1 to 5, as follows: 1 = bad, 2 = poor, 3 = fair, 4 = good, and 5 = excellent. This scale, called the Opinion Score, is published by ITU-T in their P-series of recommendations [7]. The opinion score technique is widely accepted as the laboratory technique to evaluate subjective quality of wireline and radiotelephone systems. In a typical test, a group of human subjects is allowed to listen to the quality of speech that has passed through the trial link. Depending on the nature of the test, it could take anywhere from 20 to 100 subjects to get a meaningful result. The MOS is obtained by averaging the numerical results for the subjects who have been exposed to the same test conditions. This is a rather complex test because there are many variables. In the simplest procedure, a

recording is made of the output of the channel from the speech of a talker. The subjects are then allowed to listen and rate the speech quality according to the five-point scale. In a more accurate simulation, the channel is also subjected to impairments such as bit errors and interruptions. A very realistic simulation results when there are two subjects who talk to each other over a full duplex link. Various impairments such as fading and time delay can be introduced to test their relative impact on the MOS score.

A performance comparison of various voice coding and compression systems is presented in Figure 11.28. Speech quality, as measured by the MOS score for listening only, is plotted against the output speech bit rate. The curves represent three different classes of speech coders. These categories are waveform codecs, vocoders, and hybrid systems.

The least effective in terms of compression but best in terms of voice quality is the waveform coder. Due to its relative simplicity, the waveform coder introduces the least delay as well. Included is the familiar PCM coding scheme, which is the standard for all telephone networks. The standard 64-Kbps rate of PCM can be halved by employing ADPCM, which has also been adopted as a world standard. ADPCM is indistinguishable from PCM for voice communication, as indicated by the flat portion of the upper curve. Note how quality degrades rapidly as data rate drops below 16 Kbps.

The compression techniques indicated by the heavy solid and dotted curves require a considerable degree of complexity within the coder and decoder. As indicated in Table 11.9, there are a variety of techniques with their associated names and acronyms. A more comprehensive discussion and comparison of these techniques can be found in [6, 8]. To achieve a data rate of 12 Kbps or less, one needs to filter out nonspeech information and compress the rest, which is the approach taken in the vocoder and hybrid systems. A vocoder actually substitutes synthesized speech

Figure 11.28 A comparison of the performance and quality of various digital speech compression systems.

elements for those of the sender. Hybrid systems do a little of both by breaking speech into two parts, one of which is transferred in terms of its time waveform and the other in form of a coded pattern. The intent of hybrid coding is to produce life-like speech (which tends to increase the data rate) while compressing much of the data through some of the techniques found in vocoders.

A popular class of speech coders uses linear prediction, which requires that a segment of digitized speech be read into a memory so that analysis can be performed on it. What results is a compressed datagram that can be forwarded to the distant end. These work very well in practice but increase the time delay due to the requirement to store a segment of speech at the sending and receiving ends. In comparison, the feed-forward technique found in RELP tends to have much less time delay, but at the expense of delivering less compression. Currently, coders that operate around 4.8 Kbps impose the memory requirement, while those that operate at 16 Kbps or more do not. Between these limits, we incur some delay. Another consideration is that the greater the complexity of the coder, the more physical circuitry and electrical power that must be accommodated. This tends to work against allowing a very compact handheld instrument but is not a problem for portable and vehicular units.

Hybrid coders are clearly the best choice for MSS applications at the time of this writing. At data rates around 4 Kbps, they deliver communication quality at an MOS score in the range of 3.2 to 3.6. However, anyone who is considering the development of a new MSS system for telephone service must select the speech codec very carefully. Some technologies listed toward the bottom of Table 11.9 could be obsolete within a few years of publication of this book because of the rapid pace of advancement in digital electronics and compression.

11.8   Ground Segment Architecture in MSS

National, regional, or global MSS offers many options for subscriber services, beyond making traditional telephone calls. A majority of calls will no doubt be placed within the borders of a country, between a mobile subscriber and a fixed subscriber connected through the PSTN. Because the network provides PSTN access, MSS is an extension of the national or regional infrastructure. Whether the mobile subscriber is truly mobile or not is immaterial to the network. Figure 11.2 suggests how the ground segment relates to the other elements in the overall system architecture. This section is meant as an introduction; greater detail can be found in our other work [9].

11.8.1   Network Control

Network control in MSS combines many of the functions discussed in Chapter 10 for the fixed telephony service. Added to this is the complexity of dealing with mobile and transportable users who roam from area to area and potentially country to country. Some users may, in fact, be airborne and on ships outside of the legal territory of any country. The key to achieving operational control is to use the same satellite system that provides service. Thus, as long as the user can "see" the

satellite, control is feasible. The number and sophistication of network control functions will depend on the nature of the service and user.

Table 11.10 provides a detailed listing of control functions, and their application and impact on the user. The objective is to provide a well-managed service that efficiently assigns bandwidth, completes connections, accounts for and collects payments, manages subscribers, and ultimately makes money for the operator. The range of tasks indicates how complex and critical these functions are. The computing systems to do this are at the leading edge of business information systems provided by major vendors like IBM and HP as well as specialists who supply the software. Some of the software is supplied as part of standard switching systems by Ericsson and Nortel; other software is part of systems for customer relationship management (CRM) from Siebel, Nortel, Oracle, and others.

11.8.2 Subscriber Access and Connectivity

The various categories of calls that subscribers can make over a regional or global MSS system are indicated in Figure 11.29. The most basic and common connection is that of a UT to a GES within the same country, indicated as the top double-headed arrow. UT1 can place and receive calls from any PSTN subscriber on the domestic PSTN. It is also possible for the terrestrial side of the call to connect over the international PSTN infrastructure to a subscriber anywhere in the world. The second most

Table 11.10 Network Control Functions and Issues as Applied to MSS

Overall Function	Capability	Where Provided	User Impact
Bandwidth management	Assign channel capacity to specific requests for service	Satellite bandwidth and gateway service	Command UT to particular frequency and time slot
Call management	Accept dialing instructions and complete call to destination	UT, gateway, and PSTN	Provide end-to-end connection within acceptable time and with acceptable grade of service
Mobility management	Allow user to activate UT anywhere in coverage area	Within the accepted coverage area (subject to regulatory restrictions)	Simplified access for service, dependent only on line of sight to the satellite
Service provisioning	Allow the network operator to add users to the network; allow users to modify their service to add and remove features	In countries where the service may lawfully be sold	Depends on local rules and customers—could use credit card, bank reference, or prepaid account
Billing	Collect call detail information and transform into bills	Part of business support system	Get bills by mail or on-line; alternatively, prepaid service
Rating	Introduce new calling plans and other services without extensive reprogramming	Part of the business support system	Be able to change the service plan to attract and retain customers
Settlement	Allow transfers of charges to other operators, such as for PSTN services	Part of the business support system	Allow the user to make domestic and international calls using the existing account

UT1 in country A	UT2 in country A	PSTN subscr in country A	PSTN subscr in country B	UT3 in country B

Figure 11.29   Potential connectivity in a regional or global MSS system; solid lines represent domestic connections, while dotted lines represent international connections subject to regulatory hurdles.

important connectivity is between mobile subscribers within the same country. As the second solid arrow indicates, UT1 and UT2 are located in the same country and may call each other as if they were fixed line subscribers. The quality of the call will depend on whether the network provides single-hop or double-hop calling (e.g., the call setup time and total time delay). The remaining two broken arrows indicate categories of calls that may be technically feasible but will depend on the regulatory environment. This would allow a mobile subscriber in country A to place a call to a PSTN subscriber in country B, permitting the call to bypass the national PSTN gateway. The call from UT1 would be landed at the GES in country B. A similar type of scenario might allow two UTs in different countries to connect to each other without passing through GESs or international gateways. This represents true borderless communication, which is on the horizon.

The MSS network can be managed as a single, unified entity, much like a cellular system within a city or country. This is the most straightforward approach and provides for the greatest degree of control and ease of management. There might be one or two gateways, such as provided in the AMSC and MobileSat systems, respectively. The other approach is to distribute the control and management of the network among multiple (national) operators. They, in turn, have their own respective GESs that control and manage the domestic subscriber base. Calls that traverse borders must be coordinated among the respective operators, who may or may not wish this to occur in practice. There might still be a need for a higher level of network control and coordination. For example, the operator of the space segment portion of the network will need to allocate capacity among the members of the system as well as arrange for charging for their relative usage. Also, calls that cross borders may also need coordination at the system level.

11.8.3   Network Security

Another important requirement is security of communications and network operation. Since this is a radio communications system, it is a relatively simple matter to

receive the power of a given transmission in the downlink from the satellite. Subscribers on analog cellular networks should be aware that their conversations can easily be monitored. The only protection in this case is that the origin of the speaker is not directly known to the eavesdropper. With the advent of digital speech processing, it becomes much more difficult to overhear conversations with simple receiving equipment. The security is further enhanced with the introduction of encryption either on an end-to-end basis or simply over the radio path. The latter approach is taken in most digital MSS networks.

Advances in network security are available to the developers of MSS systems. In particular, medium-strong security plays a significant role in the architecture of the GSM standard. This is a good starting point for its inclusion in a compatible MSS system such as the regional and global networks under development at the time of this writing. The two important provisions of security are authentication and encryption, both of which are reviewed in the following sections.

11.8.3.1   Authentication of UTs

Authentication is the process where the network verifies that a UT is legitimate and therefore authorized to access its resources. If the authentication codes and processes can be duplicated, then an unauthorized user with the right terminal equipment can access network services without permission and therefore without paying. This can be countered with some form of password and a system of encryption keys to block monitoring of password transmission. If encryption is not used, then it would be possible to monitor the password when it is sent over the network and then simply replay it by the unauthorized user to gain access.

A very effective type of password and key system is called public key encryption. Now popularized as the RSA algorithm (from RSA Data Security of Redwood City, California), public key encryption is a readily available but very strong technique for protecting passwords from theft and abuse. While it is called public key, there are actually two different keys used on each end of the data link: one is the public key while the other is a private key. The roles can be reversed, depending on the application. For example, if we wish to insure that no one except the intended recipient gets the data, then we encode with the recipient's public key (obtained from an open directory) and the recipient decodes with his or her private key. The other possibility is if we want to authenticate the sender, in which case the sender encodes using his or her private key and we use their public key to decode.

In applying public key encryption in GSM, each subscriber device has a private key that is contained in a chip called the SIM. These SIM chips or cards are common to users of GSM phones. The private key is used to encrypt an authorization request from the UT, resulting in access to network services. Decryption of the authentication is accomplished in the appropriate GES using a complementary public key. Since the private key portion of the encryption scheme is kept secret, the authorization center will know that only the authorized SIM card could have been used to encrypt the message. The most obvious way around this system is to duplicate or clone the SIM card.

In the case of a regional or global system, the subscriber's authorization is granted by his home service provider, who could be located in a different country. When the subscriber roams to another country or service area, the operator in that

area retransmits the authorization request over the terrestrial network to the home system. There, the authorization is verified and access granted for roaming services.

11.8.3.2   Encryption of MSS Transmissions

Privacy of communication can be provided through the familiar technique of encryption. It would be possible to use the public key approach previously mentioned, but it is generally accepted that the symmetrical private key is more computationally efficient. The important point is that it is very difficult to listen in on the communication once it has been encrypted using a private key system. Theory states that with enough computation resources it would be possible to obtain the information. From a practical standpoint, the information involved may not be worth the effort. Steps must be taken to strengthen even private key encryption due to the ever-increasing ability of hackers and pirates to find a way around the security. It is common knowledge, for example, that the lowest level of encryption in 802.11b Wi-Fi, called Wired Equivalent Privacy (WEP), may be broken using a software tool that can be downloaded from the Internet. Thus, users of Wi-Fi must employ encryption of greater strength, such as longer keys and changing of keys at frequent intervals.

A popular private key system is the DES that was developed for commercial encryption applications by IBM and the U.S. National Institute of Standards and Technology (NIST). Other private key techniques are available from commercial suppliers and over the Internet for free. In private key encryption, the same key is used to encrypt and decrypt the information. The basic approach is for the UT to select a random number as the encryption key. This itself is encrypted using the public key technique and sent over the link to the GES. From there, the randomly selected private key is decrypted and used for subsequent decryption of the communication.

Breaking the authentication or encryption system is generally beyond the means of the majority of potential abusers. However, it is worthwhile to find a way to duplicate or clone the SIM card and thereby be able to obtain service for free. SIM card technology is somewhat difficult to duplicate but not impossible. There were reports of the reverse engineering of the SIM card used for DBS service. Such compromise of the GSM security system has not been made public but would be a source of concern if it occurred. Fortunately for MSS, the system is under direct control of the service provider and so it would be possible to shut down a user once an abuse is detected.

References

[1]   Whalen, D. J., and G. Churan, "The American Mobile Satellite Corporation Space Segment," *14th Int. Communications Satellite Systems Conf.*, Washington, D.C., AIAA, March 22–24, 1992.

[2]   Wagg, M., "Regional Satellite Developments—MOBILESAT," *Mobile Satellite Communications in Asia Conference*, AIC Conferences, Hong Kong, December 6–7, 1993.

[3]   Dietrich, F. J., "The Globalstar Satellite Cellular Communication System: Design and Status," *17th AIAA International Communications Satellite Systems Conference*, Yokoyama, Japan, AIAA, February 23–27, 1998, p. 47.

[4]   Bell, D., "Mobile Satellite Communications," presented at UCLA Extension, January 2003.

[5]   "Inmarsat Regional BGAN Satellite IP Modem," *PC Pro*, February 2003, p. 195.

[6]   Kondoz, A. M., *Digital Speech—Coding for Low Bit Rate Communications Systems*, New York: John Wiley & Sons, 1994.

[7]   "Recommendations for Testing of Service Quality," *Consultative Committee on Telegraph and Telephone (P-Series),* Geneva, Switzerland: International Telecommunication Union, 1992.

[8]   Macario, R. C. V., (ed.), *Modern Personal Radio Systems*, London, England: The Institution of Electrical Engineers, 1996.

[9]   Elbert, B. R., *The Satellite Communication Ground Segment and Earth Station Handbook*, Norwood, MA: Artech House, 2001.

PART IV

Service and Business Development

Frequency Coordination and Regulation of Services

Operators of communications satellites and Earth stations are dependent on national governments and the rules of international frequency coordination to obtain the authority needed to deliver services. Consistent with international laws and regulations, each country controls the use of the radio spectrum within its borders. The government is the entity that obtains international recognition for its satellites and Earth stations. The international rules of the road for obtaining and using frequency spectrum and either GEO or non-GEO orbits provides assurance that the new system will neither cause nor receive harmful or unacceptable interference when it goes into operation. The ITU lexicon differentiates among various classifications of interference [1]:

- *Interference:* The effect of unwanted energy due to one or a combination of emissions, radiations, or inductions upon reception in a radiocommunication system, manifested by any performance degradation, misinterpretation, or loss of information which could be extracted in the absence of such unwanted energy.
- *Harmful interference:* Interference that endangers the functioning of a radionavigation service or of other safety services or seriously degrades, obstructs, or repeatedly interrupts a radiocommunication service operating in accordance with the Radio Regulations (discussed at length in this chapter).
- *Permissible interference:* Observed or predicted interference that complies with quantitative interference and sharing criteria contained in the Radio Regulations or in ITU-R Recommendations or in special agreements as provided for in the Radio Regulations.
- *Accepted interference:* Interference at a higher level than that defined as permissible interference and which has been agreed upon between two or more administrations without prejudice to other administrations.

Interference in satellite communications is one of the key determinants of system capacity and performance. Therefore, the precision with which the ITU considers it is beneficial for proper sharing of the spectrum. Even after the orbit positions have international recognition (a process we will review in detail), any operator or major user must obtain the approval of the local or national government before services can be provided. Our objective is to understand the regulatory hurdles that a new system or Earth station faces and to implement an appropriate strategy to meet these challenges.

The international regulatory environment established under the ITU, a specialized agency of the United Nations, provides a broad framework for the approval process for obtaining orbit positions and authorization to transmit to and from satellites. This process is covered in detail in Sections 12.3 through 12.5. Individual governments, however, directly control these provisions within the domestic borders of the respective countries. The regulation of satellites and the services they provide has matured over the past decades and is driven by the ability of satellites to simultaneously cover multiple countries. Established spectrum users like radio and TV broadcasting, terrestrial microwave links, and special services like radio location (radar), radio astronomy, and amateur radio, have procedures applicable to them as well.

Four decades ago, the leading governments of the world recognized the potential value of using the radio spectrum with satellites, and so the frequencies were allocated and the procedures developed even before HBO and BSkyB existed. On the domestic front, governments in the United States, Canada, and Western Europe were the first to allow privately owned satellites and Earth stations to be brought into service. For this to happen, an array of technical, business, and political interests had to be balanced. A trend in the late 1990s and past 2000 is the auctioning off of spectrum and orbit slots by governments to the highest bidder, as well as the hoarding of slots for resale. To encourage satellite network development, progressive regulators institute blanket licensing wherein specific Earth station locations are not licensed. Once the characteristics of typical terminal equipment are approved, individual Earth stations may be installed and activated at will.

12.1   Sharing Radio Frequencies

Satellite systems, cellular telephone, and wireless digital networks all employ the microwave bands between about 0.1 and 30 GHz, and must compete with each other for this valuable spectrum. Communications satellites are unique in their ability both to cause interference over a wide area, and to receive interference as well. Operation of Earth stations on land, sea, and in the air likewise poses a frequency coordination challenge. Much of our attention is focused on the techniques for resolving the interference between satellite networks that operate on the same frequencies. This process is what has allowed satellites to proliferate in North America, where an orbit spacing of 2° is commonplace. In other regions, such close spacing is now a necessity because many countries wish to exploit satellite technology. What we find is that coordinating the use of radio frequencies among satellites is growing in complexity. On the other hand, the regulations and regulators have responded to this complexity by simplifying some of the rules and processes.

The mechanisms for how satellites and Earth stations can interfere with one another are covered elsewhere in Chapter 2 and in [2] as well. Satellite systems are also able to cause interference to terrestrial radio stations that employ the same frequencies. The severity of the problem is reduced for GEO because the satellites are usually far above the local horizon at a given location—that is, above the direction that terrestrial microwave antennas normally point. The tendency has been to segment the particular band so that each can pursue their respective business

unimpaired by the other. On the other hand, there are frequency bands that began their existence as terrestrial radio media, with satellite usage representing the intruder. This is the case at L-, S-, C-, and X-bands and to some extent at Ku- and Ka-bands. Thus, coordination procedures to share bands are required prior to beginning service.

Figure 12.1 illustrates the possible interference paths that can leak between the operation of a satellite link and a terrestrial microwave link. Normal operation of the respective links is indicated by the solid arrows, which provide full-duplex transmission. The terrestrial link in this illustration could operate in the satellite's uplink band (assumed to be 6 GHz) or the downlink band (assumed to be 4 GHz). Other bands, such as X (8 GHz), and portions of Ku (11 GHz) and Ka (18 GHz) are shared between terrestrial and satellite and would have similar interference paths. Specific segments are defined in the Table of Frequency Allocations (Article 5, Section IV of the ITU Radio Regulations). The interference paths, indicated by the dashed arrows, can be described as follows.

- Path (1): 6-GHz terrestrial radio interference into the uplink receiver of the satellite. The satellite is protected by a maximum radiated power limit from the terrestrial microwave antenna and by a stipulation that these antennas should not be directed at the GEO. As a result of the low radiated power from terrestrial stations, interference of this type is not experienced in practice. The one possible exception is high-power tropospheric scatter links that employ billboard-sized antennas. These must not be pointed at the GEO in any instance.
- Path (2): 6-GHz Earth station interference into the 6-GHz terrestrial radio receiver. This is the most important interference case for determining if a transmitting Earth station can be operated in a particular location. The radiation along path (2) is in the sidelobes of the Earth station antenna and propagates along a variety of paths to the terrestrial receiver. Later in this chapter, we will discuss how this radiation is assessed and the process for coordination that results.

Figure 12.1   Illustration of the possible sources of interference between a satellite network and a terrestrial microwave link operating at the same frequencies.

- Path (3): 4-GHz terrestrial radio interference into the downlink receiver of the Earth station. A common term for this is terrestrial interference (TI), and because of this, DTH services are more popular in TI-free portions of Ku band. It might be possible to operate the downlink at a different frequency from the local source of TI. Alternatively, the Earth station antenna might be shielded from the terrestrial transmitter—a technique that has been applied to path (2) as well. Shielding may be natural from terrain features or manmade from buildings, metal or concrete walls, or earth in the form of an embankment or trench.
- Path (4): 4-GHz satellite interference into the 4-GHz terrestrial radio receiver. The radiation level from a satellite is typically too weak to be of much concern to terrestrial receivers. Still, the ITU has placed *power flux density* (PFD) limits from the satellite on the surface of the Earth to provide a high degree of protection margin. We consider this aspect in Section 12.3.5.

The ITU-R administers the worldwide radio frequency spectrum according to the Radio Regulations, providing the primary mechanism for the subdivision of the spectrum for the various radio communication services that are in use today. Their particular challenge is to oversee the allocation of spectrum in such a way that the needs of all members are met and that new technology can be accommodated. Historically, the area of greatest interest was in medium- and high-frequency channels (below 30 MHz) used for AM radio broadcasting, particularly because these short-wave signals can propagate well beyond the borders of a country and can interfere with one another. Radio frequency interference, whether harmful or unacceptable, is something that the ITU seeks to avoid. This was extended in recent years to microwave frequencies, which are governed by line-of-sight propagation, when transmitters are physically close to borders or on satellites with extensive footprints.

High-power technology like DTH and mobile satellite services have challenged the ITU to adapt its ways and produce workable regulations. It has been able to do this for the most part. A political force is the interest of developing countries, particularly those in Africa, to obtain special privileges. In the 1990s, there were thrusts from the United States and Europe to open up new spectrum for LEO satellite systems and broadband wireless services on the ground. Some viewed these as efforts to grab spectrum for special interests.

12.2   Structure of the ITU

Founded in 1865 as a postal union, the ITU is the oldest and most prominent of international organizations in the field of telecommunication regulation. As a specialized agency of the United Nations and with origins dating back before the age of radio, the ITU has proven itself in a changing and turbulent world. Today, it has more than 150 member nations and has its headquarters in Geneva (Switzerland is a member of the ITU but not the United Nations). Membership is restricted to government telecommunication agencies, called administrations; each is required to contribute to the Union's financial support more or less in proportion to its economic strength. Telecommunications carriers, equipment operators, and users are

represented through their respective governments and often attend meetings along with their government counterparts.

While the ITU is composed of countries and their respective administrations, the world as a whole is divided up into three regions. As illustrated in Figure 12.2, these correspond to what might be called the old world (Europe, including Russia, Africa, and Mongolia), the new world (North America and South America), and the Asia-Pacific region. Countries in a particular region may share cultural or political origins or relationships. In many respects, the boundaries have become blurred and these distinctions are no longer particularly strong. The designation of the three regions persists to the most current version of the ITU Radio Regulations, where they are specified and used to arrange allocation of frequency bands and other aspects.

12.2.1   Objectives of ITU Regulations

The objectives of the ITU, reflecting the shared purposes of administrations, are:

1. To maintain and extend international cooperation for the improvement and rational use of telecommunications of all kinds;
2. To promote development of technical facilities and their efficient operation so as to improve telecommunications services and increase their usefulness and availability;
3. To harmonize the actions of nations in attainment of these goals.

Generally speaking, the ITU does not have direct power to regulate telecommunication and radio transmission since these are considered to be the right of sovereign states. This means that the government of a country has total regulatory control of how the radio spectrum is used within its borders. Radio transmissions, however, do not honor borders and therefore can easily spill over to interfere with the legitimate radio communication activities of other nations. It is this interface with which the ITU is concerned. Because the ITU is a common body among

Figure 12.2   The three ITU regions.

governments, its regulations and recommendations are followed by the member nations out of the practical need for a realistic framework.

Developing countries have a special place in the ITU because of the shared objective of improving worldwide telecommunications services. A country with a poor telecommunication infrastructure cannot provide international telecommunications services of good quality.

12.2.2 Regulatory Philosophy

There are two basic areas in which the ITU regulates international telecommunications: spectrum management (ITU-R) and telecommunication standards (principally interface standards) (ITU-T). The area of interest to satellite communications is the ITU-R management of the radio frequency spectrum, which is a critical role because individual countries and operators of radio transmitters need viable rules of the road. Otherwise, radio *stations* (i.e., transmitters with antennas for any purpose) would operate on any convenient frequency they choose and interference would be very common; it would be like not having directions specified for traffic on a highway. Therefore, no one would be able to count on reliable broadcasting and point-to-point radio communication. Through international conferences, technical recommendations, and its full-time spectrum management activities, the ITU oversees the use of radio frequencies and provides a forum for the resolution of interference difficulties. They do not actually police the airwaves, which is a function left to member nations. Rather, they confer "recognition" of a given country's use of the frequency, which in turn offers "protection" in the form of "registration" of the particular frequency assignment in the "master register" of the ITU.

To summarize, an administration assigns a frequency (or range of frequencies) from the bands that the ITU allocates to a particular service. To gain recognition for the assignment, the administration must notify the ITU-R of the assignment so that it may be recorded in the Master Register. However, before it may be notified, the administration must have successfully completed the process of international frequency coordination with other administrations that might have objections.

The second principal area is in the interfacing of telecommunications networks with one another, promoted by ITU-T (e.g., the Telecommunication Sector). In previous chapters, we discussed the interfaces between telephone, data, and other networks. Without these definitions, telephone calls could not be completed automatically across borders. With the digitizing of public networks and expansion of new data communication services, the importance of ITU activity has increased significantly in the last few years. Popular standards like V.90 and V.92 for dial-up data modems and H.323 for IP-based videoconferencing are among the outputs of ITU-T.

12.2.3 ITU Sectors and Bodies

The three areas of focus of the ITU are evident in its organizational structure, shown in Figure 12.3. The Radiocommunication Sector (ITU-R) ("radiocommunication" is written as one word) carries out the mission with respect to the use of the radio spectrum by ITU members and is the area of greatest interest in satellite

Figure 12.3   The overall structure of the ITU, indicating the main branches of the Radiocommunication Sector.

communication. Next in importance is the Telecommunication Standardization Sector (ITU-T), where the effort concerning the interfacing of networks is centered. The third arm, Telecommunication Development (ITU-D), is that part of the ITU that carries out the United Nation's mission for improvement of the telecommunications infrastructure in developing countries. The programs and processes that the three sectors carry out are set at their respective world conferences. We will speak more about the World Radiocommunication Conference (WRC) and the Regional Radiocommunication Conference (RRC), which are very important to the use and regulation of the spectrum.

Figure 12.3 does not indicate two other administrative areas that are nevertheless important to the functioning of the ITU. These are the Plenipotentiary Conference (the Plenipot) and the General Secretariat, which is the full-time headquarters. The charter and administrative operations of the ITU are established by international treaty. Periodically, these are modified by a Plenipot, where new officers are elected and the budget and agenda for the Union are set for the future. While the international conferences have become more and more important over time, the cost of the operation of the ITU has continued to escalate. Because administrations ultimately pay for the operation of the Union through allocations, there is continued

pressure to reduce the number and duration of the conferences and to reduce the full-time staff. This is counterbalanced by the rapid changes in telecommunications technology that drive the need for flexible regulations. The upshot is that the ITU's budget keeps increasing as its activities are expanded.

The General Secretariat in Geneva runs the ITU on a day-to-day basis, maintaining a full-time staff and arranging future conferences. The Secretary General who heads the organization is elected at the Plenipot. All of the regulations, reports, and supporting documents are published and distributed to members. Copies of documents, which are quoted in Swiss Francs and are expensive, can be obtained from the ITU in Geneva. However, sections of documents and specific recommendations can be downloaded from the ITU Electronic Bookshop at http://ecs.itu.ch/cgi-bin/ebookshop/. These may be purchased directly from this Web site using a credit card. Another noteworthy activity of the General Secretariat is the biannual Telecom Expo held in Geneva. This is the biggest telecommunications show in the world and is attended by dignitaries from the majority of countries and telecommunications organizations, including operators and manufacturers. The total amount of money spent on just one of these shows would probably pay for a satellite and launch.

Use of the radio spectrum comes under the auspices of the Radiocommunication Sector and, in particular, the Radiocommunication Bureau (BR, using the French abbreviation) and the Radio Regulations Board (RRB). There can be confusion between the purposes of these two very important elements of the ITU. As the new name and face for the CCIR, the BR has the task of setting down the technical specifications and criteria that apply to the ITU's management of the radio spectrum. These recommendations are followed by all administrations and the full-time elements of the ITU. The way that these recommendations come into being is through the Radiocommunication Assemblies that the BR conducts from time for time. There will be an assembly just prior to any WRC in order to establish the technical ground rules for any proposed changes to the Radio Regulations, discussed next.

12.3 The ITU Radio Regulations

The Radio Regulations of the ITU contain the rules and procedures for the planning and use of all radio frequencies by administrations. Under international law and treaty, member nations are obliged to follow the rules, procedures, and allocations contained in these regulations. Not following the rules risks the loss of protection of frequencies that an administration itself might require. The administrations have the right and responsibility, in turn, to assign individual frequencies to their respective government and private users. Experience has shown that the regulations provide a reasonably effective and flexible framework for this purpose.

12.3.1 Objectives of the Radio Regulations

The following is the preamble from the Radio Regulations, Version 2001, which will provide readers with a clear understanding of their scope and intent.

Administrations shall endeavor to limit the number of frequencies and the spectrum used to the minimum essential to provide in a satisfactory manner the necessary services. To that end, they shall endeavor to apply the latest technical advances as soon as possible.

In using frequency bands for radio services, administrations shall bear in mind that radio frequencies and the geostationary-satellite orbit [the ITU term for GEO] are limited natural resources and that they must be used rationally, efficiently, and economically, in conformity with the provisions of these Regulations, so that countries or groups of countries may have equitable access to both, taking into account the special needs of the developing countries and the geographical situation of particular countries.

All stations, whatever their purpose, must be established and operated in such a manner as to not cause harmful interference to the radio services or communications of other administrations, of recognized operating agencies, or of other duly authorized operating agencies that carry on a radio service or that operate in accordance with the provisions of the Radio Regulations.

With a view to fulfilling the purposes of the ITU set out in Article 1 of the Constitution, the Radio Regulations have the following objectives:

- To facilitate equitable access to and rational use of the natural resource of the radio-frequency spectrum and the geostationary-satellite orbit;
- To ensure the availability and protection from harmful interference of the frequencies provided for distress and safety purposes;
- To assist in the prevention or resolution of cases of harmful interference between the radio services of different administrations;
- To facilitate the efficient and effective operation of all radiocommunication services;
- To provide and, where necessary, regulate new applications of radiocommunication technology.

The application of the provisions of the Radio Regulations by the ITU does not imply the expression of any opinion whatsoever on the part of the Union concerning the sovereignty or the legal status of any country, territory or geographical area.

12.3.2   Pertinent Content of the Radio Regulations

The Radio Regulations are contained in a set of three loose-leaf volumes, informally called the Red Books, organized more or less by major topics. The articles that have a direct relevance to satellite communication are listed in Table 12.1. Unlike laws, the Radio Regulations are not drafted by attorneys but rather by experts in frequency management and telecommunications. On first examination, the books are very imposing and it might seem that using them is an impossible chore. The language is not entirely clear and the logical flow at times is difficult to follow. Many of the paragraphs seem to read exactly the same way, differing only by a few words or numbers.

The best way to approach these volumes is with a question or issue that needs resolution. For example, an issue concerning the time limit of the advance publication can be researched by looking for the applicable article and regulation. This is

Table 12.1   Key Provisions of the ITU Radio Regulations (Taken from the Table of Contents)

Article	Content
Chapter I	Terminology and technical characteristics
Article 1	Terms and definitions
Article 3	Technical characteristics of stations
Chapter II	Frequencies
Article 4	Assignment and use of frequencies
Article 5	Frequency allocations
Chapter III	Coordination, notification and recording of frequency assignments and Plan modifications
Article 7	Application of procedures
Article 8	Status of frequency assignments recorded in the Master International Frequency Register
Article 9	Procedure for effecting coordination with or obtaining agreement of other administrations
Article 11	Notification and recording of frequency assignments
Article 14	Procedure for the review of a finding or other decision of the Bureau
Chapter IV	Interferences
Article 15	Interferences
Article 16	Internal monitoring
Chapter V	Administrative provisions
Article 18	Licenses
Article 19	Identification of stations
Chapter VI	Provisions for services and stations
Article 21	Terrestrial and space services sharing frequency bands above 1 GHz
Article 22	Space services
Article 29	Radio astronomy services

shown in the table of contents as being part of Article 9, Section I, paragraph 9.1. You would then proceed to these pages and read the sentences one by one until you discover the particular time period. (An advantage to downloading the document is that you can perform a text search using the Find command of either MS Word or Adobe Acrobat.) As you take the time to follow the development of the particular topic, the logic and application always come through. A complicating factor, however, is that one needs not just the latest version of the Red Books but the Final Acts of the most recent WRC as well. This is because there is typically a lag of 1 to 2 years in the publication of updated Red Books.

The main areas of focus of the Radio Regulations are: (1) definitions that are used throughout; (2) general rules for using frequencies; (3) the Table of Frequency Allocations (Article 5), which is central to the regulatory process; (4) special regulations for particular types of services, particularly for GEO and non-GEO satellites; (5) procedures for frequency coordination (Chapter III); and (6) techniques and rules for sharing bands between space and terrestrial services (Article 21). There are literally thousands of specific regulations.

There are a total of 59 articles in the Radio Regulations (the quantity will vary as a consequence of the WRCs). In addition, there are a number of appendices that lay out particular plans and analysis techniques referred to in some of the articles. For example, Article 30 contains the plan for the assignment of orbit positions and frequency channels for the Broadcasting Satellite Service in Regions 1 and 3. Other portions of the Red Books cover resolutions of past WRCs, dealing with a variety of policy issues. The exact interpretation and application of the resolutions are not as clear as the rules and procedures that precede them.

12.3.3   Table of Frequency Allocations

Definitions of how the frequency bands may be used by particular services are summarized in Table 12.2. To determine precise frequency ranges for a particular service, always refer to the current edition of the Radio Regulations. There are also literally hundreds of details contained in footnotes that provide additional uses and restrictions. An allocation is made to a particular service for a particular region in a particular band.

Table 12.2   Definitions of Radiocommunication Services by the ITU Radio Regulations (General Frequency Ranges Not Specific to Regions; * Indicates Non-GEO Allocations)

Service	Definition	Band	Typical Frequencies (GHz)
FSS	Used between a specified fixed point or points within specified areas when one or more satellites are used; in some cases this service includes satellite-to-satellite links, which may also be operated in the intersatellite service; the fixed-satellite service may also include feeder links for other space radiocommunication services.	C	Uplink: 5.85–7.075 Downlink: 3.4–4.2
		X	Uplink: 7.90–8.40 Downlink: 7.25–7.75
		Ku	Uplink: 13.75–14.8 Downlink: 10.7–11.7
		Ka	Uplink: 28.0–30.0 Downlink: 17.7–19.7
BSS	Used to transmit signals by space stations that are intended for direct reception by the general public. In the BSS, the term "direct reception" shall encompass both individual reception and community reception.	S	Uplink: 2.65–2.69 Downlink: 2.5–2.54
		Ku	Uplink: 17.7–18.2 Downlink: 11.2–12.2
		Ka	Uplink: 24.75–25.25 Downlink: 21.4–22.0
MSS	Used between mobile Earth stations and one or more space stations, or between space stations used by this service. This service may also include feeder links necessary for its operation.	L	Uplink: 1.626–1.66 Downlink: 1.525–1.56
		L/S	Uplink*: 1.61–1.626 Downlink*: 2.483–2.5
		S	Uplink*: 2.67–2.69 Downlink*: 2.5–2.52

The Table of Frequency Allocations recognizes two levels of priority: primary and secondary (in the Table of Allocations, primary services are indicated all in capital letters). A primary service has the highest level of rights to operate in the particular band, and according to the Radio Regulations:

> A secondary service shall not cause harmful interference to stations of primary services to which frequencies are already assigned or to which frequencies may be assigned at a later date; furthermore, it cannot claim protection from harmful interference from stations of a primary service to which frequencies are already assigned or may be assigned at a later date; and finally, it can claim protection, however, from harmful interference from stations of the same or other secondary service(s) to which frequencies may be assigned at a later date.

There can be more than one primary service in the same band, in which case the two must coexist through the prescribed sharing criteria and frequency coordination procedures. In previous years, users stayed away from secondary service assignments, preferring instead to use bands in which the particular service is allocated as primary. If you are in the position of being a secondary service, then you must be prepared to cease transmission on first complaint from a primary service user. However, if no complaints are received, then the secondary user can continue to operate. More recently, unlicensed applications [like cordless phones, wireless data networks (802.11b), and family radios] are operating as secondary services and thus are under these rules.

The Table of Allocations contains hundreds of footnotes that allow administrations to take exceptions to the rules when assigning frequencies to particular services. These are of two classifications: additional allocations, which add spectrum in particular countries or a subregion (but not an entire region); and alternative allocations, which replace an allocation in the table for particular countries or a subregion. For example, a footnote may permit a country in Region 1 to assign a terrestrial microwave radio user to a frequency in a microwave band allocated exclusively to satellite communications. These exceptions are taken at every WRC, where countries agree to the overall principle but must protect an existing operator in their country that is doing something that is in conflict with the regional allocation. As a general statement, the footnotes take away what the Table of Allocations gives (sometimes it is the other way around).

12.3.4   Coordination Procedures

Through the procedures in Chapter III, the Radio Regulations define how an administration can employ a particular allocation of the spectrum and assign a frequency to one of its users. The procedures are often quite complex and may even lead to confusion and a lack of resolution (this has been resolved to some degree through a simplification of the procedures undertaken during the 1990s). Theoretically, an administration checks the allocation to see if the frequency fits a predefined category. If so, it follows the procedure for informing the BR and other administrations whose radio communication services could be interfered with by the operation of the proposed radio station or stations. The affected administration may agree or disagree to the use of that frequency or band. Any disagreement must

be resolved through the process of frequency coordination, which involves bilateral negotiations between the applicant and incumbent and multilateral procedures that might include a special conference. The ITU-BR will provide assistance to either administration to help resolve a disagreement as to the rules and criteria for interference.

Space-based radio transmitters are more of a concern because of their greater potential to radiate into the territory of more countries. If the transmission could cause interference, then the administration must follow the convoluted process of frequency coordination, which is defined in the Radio Regulations. This gives the other administrations a chance to decide if they want to allow this particular station to go on the air. A successfully coordinated frequency assignment can be recorded in the Master International Frequency Register (the Master Register) of the ITU and thereby gain international status.

The basis by which frequencies were assigned and coordinated for the GEO was originally the principle of first come, first served. This was adequate when the number of systems and administrations was small. During the 1970s and 1980s, it became apparent to some administrations that this process might eventually cause new entrants to be frozen out of the orbit and spectrum. It was at WARC-77 that the principle of a priori planning was firmly established as a way to guarantee entry for countries that were late to construct their own satellite systems. This conference produced two worldwide Ku-band BSS plans that assigned at least one slot to every member of the ITU for the purpose of satellite broadcasting (not two-way interactive services, which belong in FSS and MSS bands). Over the years, very few systems that exploited these plans were actually installed and put into operation—most are in North America, Europe, and Japan.

Other bands were opened up for a priori planning, but the nature of the plans was less demanding. In particular, new segments of C band and Ku band were allocated to the FSS and preassigned to administrations through the Allotment Plan. An allotment is made of a particular frequency or band segment to a particular country or area. This gives each administration an orbit position and access to the spectrum for a variety of purposes. The need for coordination is reduced but not eliminated. The use of allotment band frequencies is increasing slowly, as the more popular segments of the FSS become very crowded.

A frequency assignment is usually made by an administration to one of its domestic users. These in turn can be registered with the RRB in order to gain international recognition and protection. Frequency assignments consider the particular power, frequency, bandwidth, and modulation type. From a domestic standpoint, a user obtains permission to use a particular frequency through a license granted by the administration. Licenses are required—but not granted—by the ITU. The way the frequencies get registered is that the appropriate administration follows the procedure for coordination, which is summarized in Section 12.4.

12.3.5   Rules for Satellite Operations

Article 22 identifies special rules for the use of the spectrum, some of which are key to the operation of GEO and non-GEO satellites, as appropriate. The following provides a selection of these rules.

All space stations (GEO and non-GEO) must be fitted with devices to cease emissions by telecommand, whenever such cessation is required under the provisions of the Radio Regulations.

- Transmissions to and from non-GEO satellites shall cease or be reduced to eliminate interference to FSS networks operating in the GEO (this provision was modified at WRC-95 to allow non-GEO Ka-band systems to operate on a coprimary basis with GEO Ku- and Ka-band systems under certain constraints, described in detail in Article 22, Section II).
- Nongeostationary-satellite systems shall not cause unacceptable interference to geostationary-satellite systems in the fixed-satellite service and the broadcasting-satellite service operating in accordance with the Radio Regulations.
- Space stations on GEO satellites shall have the capability to maintain their nominal longitudinal position to the tolerances of 0.1° in bands allocated to the FSS and BSS and 0.5° in other bands (e.g., MSS). A satellite may exceed these limits only if this does not cause unacceptable interference to satellite networks that adhere to this rule.
- Space station antenna pointing shall be maintained to be the greater of 10% of the half power beamwidth or 0.3°. A satellite may exceed these limits only if this does not cause unacceptable interference to satellite networks that adhere to this rule.
- In the event that the satellite coverage beam is not rotationally symmetrical about the axis of maximum radiation (e.g., a circle in the direction of the Earth), the tolerance in any plane containing this axis shall be related to the half power beamwidth in that plane. A satellite may exceed these limits only if this does not cause unacceptable interference to satellite networks that adhere to this rule.

The final sentence for the last three rules seems innocuous; however, it can be troublesome for a new operator to coordinate with an older satellite that is not well managed.

Sidelobe emissions (called off-axis emissions) on Earth stations in the FSS are restricted to manage the overall uplink interference in the direction of GEO. The following is one of a series of rules under Article 22, Section IV:

The level of EIRP emitted by an Earth station of a geostationary-satellite network shall not exceed the following values for any off-axis angle (φ), which is 3° or more off the main-lobe axis of an Earth station antenna and in the bands listed below:
Off-axis angle Maximum e.i.r.p.

$3° \leq \varphi \leq 7°$	$42 - 25 \log \varphi$	dBW/40 kHz
$7° < \varphi \leq 9.2°$	21	dBW/40 kHz
$9.2° < \varphi \ 48°$	$45 - 25 \log$	dBW/40 kHz
$48° < \varphi \leq 180°$	3	dBW/40 kHz
FSS bands where this is applicable:	12.75 – 13.25 GHz	
	13.75 – 14.5 GHz	

Other limits apply for analog FM TV, with and without energy dispersal.

12.3.6   Power Flux Density Limits

There are specific provisions of Article 21, Section V, that limit satellite power in bands where FSS is coprimary with the fixed service. This is based on the power per unit of bandwidth on the surface of the Earth, called the power flux density. Referencing path (4) in Figure 12.2, the PFD defines how the satellite might interfere with a terrestrial microwave station through the latter's receive antenna on the ground. This is somewhat different from the saturation flux density, which is the total uplink power per unit area at the satellite. The general way to calculate the PFD is provided in the following formula:

$$PFD = EIRP - 10\log_{10}\left(4\pi D^2\right) - 10\log_{10}\left(bw\right) \tag{12.1}$$

where D is the path length in meters and bw is a reference bandwidth in hertz used in the regulation. D depends on the location of the point on the ground where the limitation is checked; while the bandwidth, bw, takes account of the modulation format of a terrestrial microwave system that could receive interference. In most cases, bw is equal to 4 kHz (4,000 Hz in the formula) to correspond to the bandwidth of a telephone channel in an analog FDM/FM microwave system. Another value for digital modulation of the interfering carrier is 1 MHz of bandwidth.

Table 12.3 provides the PFD limits for the satellite bands in common use. These limits apply for all conditions and all methods of modulation and relate to the PFD that would be measured on the surface of the Earth under assumed free-space propagation. The first column in Table 12.3 identifies a particular frequency range corresponding to an allocation to FSS, BSS, or another satellite-related service. In the second column we find the PFD limit that applies for ground receiving elevation angles between 0° and 5° (e.g., at or near the local horizon). It is the lowest value permitted in the particular frequency band because terrestrial microwave antennas could be pointed along low elevation angle paths that align with a satellite (a condition discouraged by the Radio Regulations). Between 5° and 25°, the PFD is allowed to increase in direct proportion to the elevation angle according to the linear escalation rate shown in the third column. The PFD limit reaches a maximum at 25° and remains constant all the way to the subsatellite point where the elevation angle equals 90°. The reason why the PFD limit is a constant maximum over the higher elevation angles is that terrestrial microwave antennas seldom, if ever, point in such directions. The table is rather formidable in its detail; however, it is a relatively simple matter to identify the particular segment of the band and determine the PFD limit corresponding to the range of ground elevation angles. The value calculated using (12.1) must be less than the prescribed limit (e.g., it must be a more negative number in terms of decibels).

12.4   International Frequency Coordination

The process of international frequency coordination is required because the spectrum and orbit space are limited resources and must be shared by all nations and users. Because this process is technically complex and often political, cooperation with the ITU-R could potentially be very helpful. However, this cooperation can

Table 12.3 The ITU PFD Limits, in dBW/m² Measured in Either 4 kHz or 1 MHz Reference Bandwidth (bw), Produced by a Satellite, as Measured on the Ground at a Given Elevation Angle (δ) for All Conditions and for All Methods of Modulation

Frequency Range (GHz)	PFD Limit, Below 5°	Rate of Increase Between 5° and 25° (dB/m²)	PFD Limit above 25° (dBW/m²)	Ref bw
2.50–2.69	–152	–152 + 0.75(δ – 5)	–137	4 kHz
3.40–4.20 4.50–4.80 5.67–5.725 7.25–7.85	–152	–152 + 0.5(δ – 5)	–142	4 kHz
5.150–5.216	–164	–164	–164	4 kHz
6.700–6825	–137	–137 + 0.5(δ – 5)	–127	1 MHz
6.825–7.075	–154 and –134	–154 + 0.5(δ – 5) and –134 + 0.5(δ – 5)	–144 and –124	4 kHz 1 MHz
10.7–11.7 (GEO)	–150	–150 + 0.5(δ – 5)	–140	4 kHz
10.7–11.7 (non-GEO)	–126	–126 + 0.5(δ – 5)	–116	1 MHz
11.7–12.5 (Region 1) 12.5–12.75 (Region 1 listed countries) 11.7–12.7 (Region 2) 11.7–12.75 (Region 3) (Above non-GEO)	–124	–124 + 0.5(δ – 5)	–114	1 MHz
12.2–12.75 (Region 3) 12.5–12.75 (Region 1 listed countries) (Above GEO)	–148	–148 + 0.5(δ – 5)	–138	4 kHz
15.43–15.63	–127	5° – 20°: –127 20°–25°: 127 + 0.56(δ – 20)	25°–29°: –113 29°–31°: –136.9 + 25 log (δ – 20) 31°–90°: –111	1 MHz
17.7–19.3	–115 or –115 – X (X determined by number of non-GEO satellites)	–115 + 0.5(δ – 5) or –115 – X + ((10 + X)/20)(δ – 5)	–105	1 MHz
19.3–19.7 22.55–23.55 24.45–24.75 25.25–27.5	–115	–115 + 0.5(δ – 5)	–105	1 MHz
32–33 Intersatellite link	–135	–135 + (δ – 5)	–115	1 MHz

Note: The limit relates to the PFD, which would be obtained under assumed free-space propagation conditions, and applies to emissions by a space station of the service indicated where the frequency bands are shared with equal rights with the fixed or mobile service (WRC 2000). All limits apply to GEO satellites unless otherwise stated.

break down when satellite operators enjoy or behave as if they have a monopoly position, acting as if they own the orbit positions and spectrum that they employ. Likewise, users of satellite capacity sometimes violate various provisions of the Radio Regulations and cause harmful interference to their neighbors or otherwise abuse their access privileges. Such obstructions are often overcome through the

process of diplomacy, which can transcend some of the self-interest that at times seems to dominate the process. For this reason, a prospective satellite operator must make a good-faith effort to follow the ITU rules for frequency coordination and, at the same time, conduct its own campaign for the support of its domestic administration and possibly other administrations in the countries that could impact the outcome.

Stories abound of how the rules and procedures are either ignored or directly violated. Here are some examples:

1. A new satellite operator "rents" an orbit position from the government of another country and then intentionally interferes with an adjacent satellite of a third country who had not concluded the proper coordination procedure.
2. A user of FSS capacity operates from a moving vehicle to transmit live video, causing cross-polarization and adjacent satellite interference to legitimate fixed Earth station users.
3. A disgruntled Earth station operator decides to jam a popular satellite TV channel because the channel started to scramble its signal to prevent piracy (the operator had a side business to sell unauthorized dishes).
4. The owner of a multipurpose satellite, having ignored the practice of coordinating with satellite owners, left an FSS payload turned on as the satellite drifted past the orbit locations of very popular TV distribution satellites, causing harmful interference.

In all of these cases, authorized users had to suffer difficulties until the wrong-doer was identified and held accountable.

12.4.1   The First Step in the Process

The overall regulatory process is shown in Figure 12.4. Beginning at the top, the ITU develops the frequency allocations and rules for coordination; next, the administration participates in the ITU activities at the WRC and oversees the assignment and user of frequencies within its domestic borders; and finally, the user applies for and obtains frequency assignments and the authority to operate from the administration. A planned user of the orbit and spectrum, such as a company proposing to launch a GEO satellite, must therefore apply to the domestic regulatory body, which acts as the administration for the particular country. The user will usually be required to prepare the actual applications that are forwarded to the RRB by the administration. As reviewed previously, the ITU has created the international frequency coordination framework embodied in the WRC, BR, and RRB components of its structure.

There are basically two types of frequency coordination: terrestrial coordination, for land-based microwave transmitters; and space coordination, for radio transmitters and receivers on satellites. Terrestrial coordination involves any terrestrial radio transmitter that can potentially radiate signal power across a border into a neighboring country [e.g., paths (2) and (3) in Figure 12.1]. In particular, the Radio Regulations and Recommendations provide technical analysis procedures to

Figure 12.4   The relationship between the application for a satellite system, the associated domestic administration who submits the application, and the ITU, where the regulatory process is developed and managed.

compute the coordination contour, which is a graphical depiction of the expected and worst case power levels from a transmitting Earth station after propagating through the atmosphere. This is covered in more detail in Section 12.4.3. A neighboring administration analyzes the coordination contour to determine whether or not this level of power could interfere with the operation of domestic radio receivers that employ the same frequency band. If so, then the two administrations would, on a bilateral basis, make an agreement as to which frequencies would be used or how the transmit radiation pattern of the offending Earth station should be altered. After coordination is complete, the administration can register the new frequency assignment with the RRB in Geneva.

12.4.2   Frequency and Orbit Coordination

The process for coordinating a new communication satellite is described in Radio Regulations, Chapter III, "Coordination, notification and recording of frequency assignments and Plan modifications." Figure 12.5 summarizes the regulatory process for a typical application. This chart is intended to provide a conceptual flow and does not include all of the details of the Radio Regulations. In fact, it is highly recommended that any reader who needs to understand the specifics of the process study the text of Chapter III in its entirety. The documentation is formal and difficult to understand in a first reading. After consideration and discussion with regulators, the process begins to make sense and can be followed in a rather straightforward manner.

Before an administration can launch a satellite and begin operation, it must go through the three-step process: advance publication, coordination, and notification

Figure 12.5   A summary of the international satellite regulatory process. Specific important details are ignored to simplify the schematic.

and registration. The Advance Publication of Information (API), provided to the ITU in the first step no earlier than 5 years nor later than 2 years before the start of operation, is a general description of the proposed satellite network to allow other administrations to assess the potential impact on their existing or planned satellite networks. After review by the RRB for compliance with the Radio Regulations, the information is published in the Weekly Circular, which is a special publication containing these applications as well as requests for coordination and notices of frequency assignment (discussed later). The required information is specified in Appendix 4.

The API step in Article 9 is a prerequisite for coordination, allowing any administration to submit comments on how the new satellite network could

interfere with the operation of their own existing or planned satellite networks. If this is the case, they are to respond with their comments within 4 months of publication to the filing administration. The filing administration is supposed to try to resolve any difficulties raised in these comments by altering the orbit position and/or transmission characteristics of the proposed satellite. This could involve a change in the coverage pattern, power levels, or specific frequency bands. The Radio Regulations require that the filing administration make an effort to bring this to resolution; however, if the administration making the comments does not remove its complaint, the rules do not explicitly obligate the filing administration to do much else. At the end of the 4-month period of the comments and attempts to resolve the problems, the filing administration must submit a progress report to the RRB. Additional reports may be submitted on 6-month intervals. In any case, the filing administration may move on to the coordination phase when 6 months have passed after the publication of the API in the Weekly Circular.

After the prescribed period, the administration can initiate the coordination process for the satellite and Earth stations in the network. To get a head start, the filing administration can submit coordination information along with its API; however, the BR will indicate a filing date 6 months later. The API filing will be canceled after 24 months if the filing administration fails to submit coordination information. Coordination is the most critical phase as it determines the priority that the new network will have over new applicants who come later. Coordination is only required with other registered satellite networks that fall into one of the following two categories:

- For a station in a satellite network using the geostationary-satellite orbit, in any space radiocommunication service, in a frequency band and in a region where this service is not subject to a plan, in respect of any other satellite network using that orbit, in any space radiocommunication service in a frequency band and in a region where this service is not subject to a plan, with the exception of coordination between Earth stations operating in the opposite direction of transmission;
- For a nongeostationary-satellite system in the fixed-satellite service in certain frequency bands, in respect of a specific Earth station in a geostationary-satellite network in the fixed-satellite service.

The fact that any satellite network that has at least entered coordination gains priority over newcomers is very significant to the entire process of launching a new satellite. Because of the importance of being early, the manager of a new satellite and Earth station project should push the respective administration to initiate the API and coordination process as soon as possible.

The information to be supplied for coordination is contained in Appendix 4. It is very much like the API, except that specific frequency assignments are requested. This can be accommodated by filling in the characteristics of the transponders (center frequencies and bandwidths) along with the basic specifications of typical Earth stations and services. In previous editions of the Radio Regulations, the filing administration was required to send a copy of these forms to each administration with which it needs to coordinate. The 2001 edition places the burden of

determining which administrations could be affected by the new assignments upon the BR itself. The determination of the affected administration uses the calculation of the percentage increase in equivalent link noise temperature ($\Delta T/T$), defined in Appendix 29. The threshold for coordination was initially set at 2% but was subsequently increased to 6% to reduce the quantity of coordinations that a new applicant would have to undertake. The analysis technique is very conservative in that it will indicate that unacceptable interference could occur even when it would not be the case in practice. The idea is to assure the priority given by the Radio Regulations to any other administration that has a system that meets one of the two conditions cited above.

The requested administration must evaluate the data in the coordination filing to see if the calculated interference levels would be acceptable. Specifically, it must:

1. Acknowledge receipt of the data;
2. Examine the data to determine if interference would be caused to its lawful frequency assignments;
3. Within 4 months of receipt, inform the requesting administration of its agreement that the interference is acceptable or its disagreement, giving also the technical details upon which its disagreement is based, including relevant characteristics about its system not previously notified and its suggestions with a view to a satisfactory solution to the problem.

An affected administration will evaluate the technical characteristics of the proposed satellite network and inform the filing administration if it agrees to its operation. This is not a unilateral process, and the techniques for making the assessment have become fairly standard over the years. The basic criteria for acceptable interference levels are set by the BR at relevant WRCs or through recommendations. Also, if the two administrations have come to agreement on criteria in the past, then the same criteria should be used unless the two sides agree to something else. Any special agreements between the two administrations will not prejudice the negotiations with or among other administrations.

Under the new rules, an administration with which coordination is sought must issue its opinion within 30 days of the final request from the BR; otherwise, it is presumed to agree that no harmful interference would occur. This is a significant improvement because it eliminates an endless loop that existed in previous versions of the rules. On the other hand, an administration believing it should have been included in the coordination process can demand of the BR that it be allowed to provide its review as well.

Coordination of satellite networks is a bilateral activity where the newcomer must approach the incumbent and obtain their agreement regarding the potential for interference between systems. Such discussions and negotiations will take a year or more, particularly in difficult or acrimonious situations. This difficult process could be helped by an arrangement under the WARC 88 Final Acts, wherein a multilateral planning meeting (MPM) can be called by an administration. The concept is that the parties involved would get together more or less outside of the Radio Regulations and work to achieve a settlement. Meetings of this sort were called to segment L-band spectrum by MSS satellite operators in the same region.

Following a successful coordination, a filing administration submits its frequency assignments to the RRB for recording purposes. The RRB examines the assignment to make sure that it fits the Table of Frequency Allocations and that the coordination procedures have been followed. In particular, the new assignment should not cause unacceptable interference to an existing assignment that has already entered or completed the review and coordination. Although it has not been given policing authority, the RRB does have power over administrations because of the status given to frequencies that have been recorded in the Master Register. When a frequency assignment is in coordination between two administrations, the RRB can assist by performing calculations of the expected level of interference and can make recommendations to the parties on how the interference could be prevented.

In the event of a deadlock, and provided the BR has already been asked to assist in resolving the conflict, the filing administration can proceed directly to the notification phase (discussed later in this section). This particular provision is covered in its own subsection (IID) in order to provide a means to overcome the kind of obstacle that has been used in the past to attempt to block a competitor. Under IID, the BR would examine these notices with respect to standard technical criteria that assure that appropriate frequencies were chosen and that harmful interference will not occur. A positive conclusion will result in recording of the assignments in the Master Register, indicating the names of administrations with which coordination was successfully completed and those that were not but where the technical finding by the BR was favorable. This provision happily provides the means to obtain protection where an uncooperative administration would attempt to block the new entrant.

Most of the time, coordination and recording are accomplished in a straightforward manner, taking anywhere from 6 months to 3 years, depending on the number of administrations involved and the complexity of the technical analysis of potential interference. Occasionally, an administration refuses to allow coordination of a new frequency assignment by another administration. The question arises as to what motivates agreement on coordination in an environment of newcomers who have to obtain agreement from entrenched providers. Experience has shown that the following principles, while not necessarily written down, apply in practice:

1. The force of the rules and their interpretation by the RRB is such that all administrations shall make all possible mutual efforts to overcome the difficulties, in a manner acceptable to the parties concerned.
2. Various types of interactions and meetings can be employed, particularly:
   - Bilateral meetings between the administration requesting coordination and the Administration that has raised the objection;
   - Multilateral meetings as needed and where appropriate (not necessarily an MPM introduced by WRC-88).
3. All parties recognize that:
   - Every administration has a legitimate right to put up satellites;
   - Any administration must be willing to make adjustments in its planned or existing operation to allow a new applicant to enter into service;
   - A new entrant must approach the existing operators with the proper attitude, recognizing that the principle of first come, first served is firmly embodied in the Radio Regulations;

- An established operator should do what it can to allow a new entrant to move forward with a plan for satellites;
- That today's new entrant becomes tomorrow's established operator—also, any established operator will eventually want to introduce a new satellite network and hence will find itself in the position of being a new entrant.

The completion of the coordination process will take time and effort. There are often difficult problems with one or more other administrations, and some of these problems may involve a serious incompatibility between the new and the existing systems. The important point is that the new applicant and the associated administration must continue to work through problems and to creatively find solutions with which everyone can live. The following concepts and techniques are offered as loose guidelines for reaching agreement.

- Hold the meetings with the other administrations and be sure to come prepared to discuss all of the technical issues and options.
- Request that the other party do their part.
- Involve the BR and the RRB, and seek their counsel and help by telephone and in person.
- Be assertive, but not overly demanding.
- Prove to the other administrations that your side is actually going ahead and therefore must be taken into account.
- Make proposals on how to allow safe operations without unacceptable interference.
- Be willing to make reasonable adjustments, but do not sacrifice the economic viability of your system.
- Be prepared to use as much influence as possible to aid the process.
- Be patient—coordination can be a lengthy process.

Notification of frequency assignments, the last step in the regulatory process, should be accomplished before the service is initiated. This is done by submitting the same data items with the characteristics of the satellite network after the completion of coordination. The RRB will verify that the entire coordination process is complete with all of the appropriate administrations. If so, they will record the assignments in the Master Register and the new operator is free to begin using them to provide service.

12.4.3   Terrestrial Coordination of Earth Stations

Terrestrial coordination is the process intended to protect terrestrial microwave stations in other countries from transmissions by new Earth stations that operate in a shared frequency band. The bands where this is required are those where the FSS and the fixed service are coprimary. Referring to Figure 12.2, we are trying to control interference along path (2) where the Earth station interferes with the terrestrial station, and path (3) where the inverse situation can exist. An applicant for a new Earth station starts with data that describes the radiation pattern around the full

360° of azimuth along the horizon. This is used to compute the amount of RF energy that could propagate from the Earth station location to locations in a neighboring country where terrestrial receivers could be located. Additional shielding is provided by the existing terrain and buildings around the Earth station site.

The mechanism used to perform this assessment is the coordination distance, which is the calculated distance over which interference could potentially result. This distance is calculated with formulas in Appendix 28 to the Regulations and considers attenuation produced by propagation over the surface of the Earth along the great circle path, which introduces greater path loss than simple line of sight. The techniques are complicated and are best carried out with computer software available for download from the ITU Web site.

The technique computes the two different coordination distances; that is, d1 corresponds to the great circle, and d2 is that produced from rain scatter propagation. Formulas and graphs for computing d1 and d2 are found in Appendix 28. These consider the radiation pattern of the Earth station antenna, where the longest distance occurs along the axis of the main beam. The best way to view the coordination distance is in the form of a coordination contour, which is a polar plot of d1 and d2 as a function of azimuth around the Earth station. A simplified example is shown in Figure 12.6. The minimum distance provided for in the rules is 100 km, while 200 km is more typical (as shown in the figure). The elongation of the contour is produced by the main beam of the antenna. Unlike satellite coordination, it is up to the requesting administration (and not the BR) to determine which administrations need to be included—a relatively simple process once the coordination diagram for the new Earth station is complete.

An administration that receives the request for coordination, including the coordination contour along with the other information in Appendix 4, must determine if interference could result from the operation of the new Earth station. Receipt of this data must be acknowledged within 30 days of receipt. The examination would consider:

- Interference to terrestrial stations existing or to be operating before the Earth station enters service or within 3 years, whichever is longer;
- Interference to the Earth station by such terrestrial stations.

Each affected administration notifies the requesting administration within 4 months of one of the following:

- Its agreement to the proposed Earth station with a copy to the RRB;
- Its desire to include specified terrestrial stations in the coordination;
- Its disagreement.

In the last two cases, the notification should include a diagram showing the location of existing or planned terrestrial stations within the coordination area (which is the area inside the coordination contour) and suggestions for solving the interference problem. A copy should also be provided to the RRB.

Terrestrial coordination is considered to be very routine by all administrations due to the popularity of both satellite and terrestrial microwave communication.

Figure 12.6   Example of Earth station coordination area. (*From:* [3]. © 2002 ITU. Reprinted with permission.)

The requests for terrestrial coordination are generally handled on a routine basis and should produce agreement in a reasonably short period of time. This also considers the fact that all countries have some previously recorded terrestrial links and Earth stations and so no one is in a position of control of the spectrum based on first come, first served. In cases where it is difficult to get agreement, the RRB can be counted on to provide assistance. They have even been in the position of acting as a disinterested party and honest broker in finding the needed solution to the problem.

12.5   World Radiocommunication Conference

A WRC is held every 2 years or as required to deal with the changing needs of administrations for radio communications services. These conferences are where the Radio Regulations are established and rewritten periodically. Less frequently, a

RRC is held by one of the three ITU regions, indicated in Figure 12.2. The last RRC met in 1985 to develop a plan in Region 2 for the BSS. Regions 1 and 3 already had developed their plans at the WARC-77, which was convened to create the planning process for this new service.

A typical WRC lasts 1 month and is usually held at the conference center immediately adjacent to ITU headquarters in Geneva. Preparation for the next conference begins almost immediately after the last one is over. Preparatory committees meet at the national and global levels. Since many of the issues have to do with sharing frequency bands and coordinating systems, there is a need to provide a solid technical foundation for the conference. This includes the development of technical criteria and possibly special computer software to analyze the impact of a particular approach or plan.

After the conference begins, it takes many weeks to refine the two complex aspects until a general consensus can be worked out. Each country can express its position at general plenary meetings and can participate in smaller working groups that delve into critical issues. The Final Acts of the conference must be written, translated into the various languages, and approved by the entire body. Recommendations and studies of the BR are used as a basis for technical positions that the WRC takes and incorporates into future provisions of the Radio Regulations. Some sections of the Final Acts become provisions of the Radio Regulations, and others become instructions to the RRB and the BR. Following the WRC, the respective governments have the opportunity to confirm the Final Acts, incorporating them into domestic law. There are instances where a particular government was not in agreement with the overall conference and so took exception to some provision or perhaps the Final Acts in their entirety.

12.6   Additional Regulatory Approvals

The discussion up to this point has concerned the international regulatory process as it relates to the radio spectrum and orbit resource. There will be other regulatory approvals that will need to be obtained in each and every country where satellite communication services will be provided. Unlike the ITU with its forced consistency, the domestic approval process varies greatly from country to country. As discussed in [4], the degree of complexity and potential for success depend on the nature of telecommunication policy and degree of openness to competition. These are summarized in Table 12.4.

The information provided in this section is for general guidance and should not be considered to be recommendations on how to enter a market in a country. Obtaining the necessary approvals is a very complex and involved process, requiring a commitment of the necessary resources. Readers who wish to approach this topic for a project should seek the aid and assistance of people and organizations who have experience in the country and type of service.

It is vital that new entrants identify the particular environment in question and gain a thorough understanding of the rules that apply to the type of services to be provided. In the following sections, we briefly review some of the generic types of approvals that should be experienced in a majority of cases. Just how the particular

Table 12.4   Six Telecommunication Market and Regulatory Structures That Can Be Encountered During the Process of Obtaining Approvals in Different Countries

Level	Classification	Characteristics
1	Government monopoly	The classic case of a monopoly vested in a government ministry or department; this monopoly maybe based in law or by administrative fiat or convention
2	Public corporation government monopoly	A government monopoly that is instituted through a government-controlled corporation rather than an administrative department
3	Government competition	A structure wherein both the government and private sector entities compete with each other in the market
4	Regulated monopoly	The classic case of private ownership of facilities (one entity) with regulation by a government department
5	Regulated competition	The extent of competition, the number of competitors, the products or services that can be offered by each supplier are subject to authorization, licensing, type approval, and the like, by a government department or agent of the government
6	Liberalized entry	A market situation characterized by the absence of all official government rules and regulations that serve to regulate market entry, structure, or conduct coupled with complete reliance on private sector entities for the provision of goods and services

government chooses to address these areas will depend on the type of environment, as listed in Table 12.4. In protected environments where the government exercises a high degree of control, it will be much more difficult to find a workable solution. This could involve the use of forms of influence, which are outside of the regulatory process. Some strategies of this are suggested in our previous work [4].

12.6.1   Operation of Uplink Earth Stations

The terrestrial coordination procedures discussed previously are carried out by the administration for the benefit of an Earth station operator. Many of the satellite applications covered in this book require some form of large transmitting Earth station. In all cases, the applicant will have to satisfy the domestic regulator that the operation of the Earth station will be safe and in accordance with local standards. Part of this safety relates to RF radiation, which may not be a problem in reality but still represents a significant community relations issue. To quell this, the Earth station developer may have to locate transmitting antennas far away from populations or introduce shielding not needed for TI purposes.

Many countries require that the government or government monopoly, as appropriate, actually own and operate the station. This can be arranged through the build, operate, and transfer (BOT) or build, transfer, and operate (BTO) schemes that are common around the world. With BOT, the service provider or user constructs the Earth station and completes its acceptance test. It may then operate the facility for a specific period of time (say, 5 years), after which ownership is transferred to the government entity. Service may or may not be provided thereafter depending on the terms of the original deal. The BTO scheme reverses the sequence of the last two steps. Another situation is where the government or monopoly operator must actually provide the uplink or hub capability as a service to the user.

This would be paid for by a combination of initial construction charges followed by monthly or annual operating fees. The equipment would always be owned and operated by the organization providing the capability. Combinations of direct ownership, BOT, BTO, and service provision are possible.

12.6.2   Type Acceptance of Terminals

The terminals required in applications like VSAT networks and MSS services are too numerous to expect to obtain individual operating permits as one would have to do for a major Earth station. Type acceptance is an expedient approach for clearing user-oriented Earth stations through the regulatory approval process. By type acceptance we mean that the regulatory body is given a sample (or samples) of the type of terminal so that they may conduct whatever tests they think are appropriate. This would consider electromagnetic compatibility with other devices and services, safety, and orbital interference potential.

Once type acceptance is obtained, the network provider can begin to introduce the terminals into the market. Follow-up inspections and tests may be required to ensure that terminals continue to satisfy the requirements of the original type acceptance. The tests may be conducted in a government laboratory, by a third party who engages in this activity for a business, by the manufacturer, or by the network operator. For personal terminals, whether handheld or modem-style, part of the certification will relate to user safety. This gives consideration to hazards from high voltages and currents, as well as RF radiation.

12.6.3   Importation of Equipment

User terminal or other equipment that is produced within the country where service is to be provided can usually be introduced into the network once type acceptance is obtained. It could be a completely different story if the equipment must be imported. One reason for this concern is that the inflow of possibly advanced equipment may harm domestic manufacturers. Also, the government may not want the particular type of device or service to proliferate without adequate control on distribution and use. These concerns are heightened with satellite communication because of the ease with which the terminals can transmit data out of the country and receive information of a type that the government may need to restrict. Reasons for this involve business, economic, social, and economic issues, which are beyond the scope of this book.

To clear this question, the service provider or distributor must obtain an import license. This could be nothing more that an addendum to a type acceptance certificate, or it may involve a totally different process and part of the government. The import license would identify the rules and restrictions as well as the particular import duties that would be levied on the overall operation and each user device as well.

There is generally one area where terminal equipment may be brought in under fewer restrictions. This is where the equipment would be for temporary use, after which it would be removed from the country. Activities where this is applied would be in emergency situations where communications need to be restored in a hurry as

well as for demonstrations of new technologies or services. The latter is important when the operator is in the process of obtaining the approvals and needs to show the government what is involved in providing service in the first place. Such temporary importation may or may not be subject to duties and taxes.

12.6.4   Approval for Construction and Installation

A last step in the regulatory process occurs with the approval to construct, install, and ultimately operate a satellite communications Earth station or network. This would seem to be a formality, but in reality could end up being the main stumbling block. A company may obtain a license for a particular frequency and location, based on satisfying general guidelines and payment of fees. Later, the actual construction of a facility could be under a totally different regulatory regime. It is not this way in all countries and cases, but it could arise to the surprise of the developer.

Once constructed, the developer may also need to obtain approval to actually go on the air. Traditionally, this is the domain of the satellite operator who places technical requirements and other specifications on the Earth station. This could come from the same country or another country, and usually is stated in the service agreement. It is possible that a particular characteristic of the ground antenna or the required uplink power might not pass this type of certification, resulting in delays until problems are corrected.

12.6.5   Usage and Content Restrictions

Having obtained all of the approvals needed to introduce the equipment and network operation, there remains the possibility of usage restrictions on the service. An example would be for a DTH service in a country that does not allow selling movies over pay TV. In this case, the business must be based on advertiser-supported programming and possibly flat subscription movie channels, but a PPV service would clearly be out of the question until new procedures are introduced.

Another example taken from MSS would be a restriction of international calls over the satellite network. In this instance, all calls would have to be connected to the domestic GES or, if technically possible, to another UT within the borders of the same country. A restriction of this type might be required by law in order to protect a monopoly international telephone service provider. The network might be technically capable of making international connections, but the government might have to be satisfied that such calls are inhibited through hardware or software blocks that can be audited from the outside.

12.6.6   Competitive Entry

Over and above the operational issues cited in the previous sections, there is still the overriding question about whether a particular government will allow a new and potentially competing service to be introduced. This goes back to the levels of competitive entry cited in Table 12.4. If the service is particularly advanced and of value to the new market, then there is more willingness to take on the entrenched

monopoly. This will involve the political processes of the particular country, possibly including the introduction of new laws and regulatory procedures. This was the case in Japan prior to the creation of their domestic satellite communications service industry in the late 1980s.

12.6.7 Licensing

The licensing of individual terminals may be required since, as a transmitting device, it is a form of radio station. Type acceptance could be used in lieu of licensing, although the ITU Radio Regulations require that transmitting stations be licensed. However, it is up to the country in question as to how they would handle the matter. A license from the government might be obtained by the user when registering the unit as an operating terminal. This would apply whether the unit was imported or domestically produced and provides a vehicle for the government to obtain additional revenue from the license fee.

12.6.8 Other Roadblocks

The final category of additional regulatory requirements is referred to as roadblocks. Governments and monopoly service providers can be very innovative and persistent in the way that they interact and control the new operator. The potential list could be very long, being different from country to country. Actually, some of the advanced and most industrialized countries of the world have the most difficult and convoluted roadblocks. An area that continues to get much attention is the control of cross-border data flows to protect the privacy of domestic citizens. As mentioned previously, satellite transmissions do not heed national borders.

Just how one identifies and resolves these other roadblocks will depend on the particular situation. There is no substitute for a thorough examination on the ground in the country in question. This might be accomplished with a series of field trips. In many other situations, one will actually have to establish a permanent activity to learn about and subsequently address the concerns of the domestic authorities.

12.7   Regulatory Environments in Different Countries and Regions

We review the regulatory environment with regard to satellites in a number of the leading countries. This provides the basis for understanding how the regulations can be adopted and applied in other countries where the process is only beginning. Later, we move from this specific experience base into the broader aspects of how the rights to the use of spectrum and the orbit arc are shared on a global basis.

12.7.1   The U.S. Regulatory Environment

The United States went from being a tightly regulated environment regarding satellite communications under President John F. Kennedy to a policy of nearly free access to the orbit under President Richard M. Nixon, under his "Open Skies"

policy. The FCC is the civilian agency that has authority over the U.S. satellite industry, assigning orbit positions, granting construction permits, and licensing Earth stations (considerable valuable and current information may be viewed on the FCC's Web site, http://www.fcc.gov). Government usage of the spectrum is handled by the National Telecommunication and Information Agency, a completely different agency that cooperates with the FCC in the context of the ITU. In the past, the FCC has pre-assigned slots to companies for commercial development. The reason why the FCC is the most expert in satellite regulation is that the first satellite operator, COMSAT, sought and obtained its satellite and Earth station licenses in the United States. FCC rules and procedures, while often lengthy, allow satellite operators to plan, construct, and market satellite capacity with reasonable confidence. Provided that the proper regulations and laws are followed, the availability of slots has rarely hampered a legitimate, adequately funded business from proceeding. On the other hand, the most useful slots to serve North America and Europe have long been taken and newcomers must be satisfied with providing services to these areas at relatively low elevation angles.

Obtaining an orbit slot in the United States involves a lengthy and sometimes costly process. Satellites in the FSS are granted slots after a demonstration of business need, technical capability, and more recently, financial capability. Some slots that were authorized actually never got used; in these circumstances, the FCC withdraws the license after the prescribed period (usually 5 years). Historically, the FCC has allowed operators to replace satellites at end of life to preserve their services and businesses. Deals to exchange orbit slots between operators have been allowed, subject to public notice and debate in an open forum of written comments. The public notice process is surprisingly effective in identifying and resolving differences.

More recently, the FCC has chosen to use auctions to decide who would be granted the license to develop a business around the use of a particular piece of the microwave spectrum. The first such auction occurred in 1996, when MCI Corp. bid and subsequently paid $682 million for 28 of the 32 Ku-band channels at the last remaining U.S. BSS orbit position (110 WL). This price established a high value for the spectrum that provides access to an attractive market like the United States. In comparison, Hughes and Hubbard paid little more than nominal filing fees for the 32 channels that they employ at 101 WL.

Auctions appear to improve upon the use of the lottery, which in the past resulted in some recipients not having the financial resources to follow through. In those instances, the "winners" simply sold their licenses through a secondary market where values were continuously bid up. This market mechanism works to a fashion but has the disadvantage that the government and taxpayers receive none of the real money. In comparison, the auction prices received for licenses in PCS and BSS have been very high, demonstrating in clear terms the value of spectrum. Auctions, however, are only used within the country's borders and have no bearing on the ITU coordination process.

The North American market is the most developed in the world, with over 50 satellites serving the United States, Mexico, and Canada. The FSS systems facilitate the cable TV and broadcast TV industries, and true BSS serves the United States and Canada. Other services delivered through VSATs and hub Earth stations have given a strategic advantage to some of the more successful companies in a variety of fields.

This environment provided the motivation for other regions to follow suit, including the adoption of liberalized regulations for satellite services.

One of the most important FCC innovations is blanket licensing of Ku-band VSATs. Rather than suffer individual station clearance and licensing as at C-band, an operator need only obtain one license that relates to the overall installation of facilities. The properties of a typical VSAT are included with the license to define the nature of the transmission. From then on, VSATs can be added at will. This is one of the biggest impetuses to VSAT applications within the U.S. market.

The way the FCC assigns FSS orbit slots is in response to applications from incumbent and prospective satellite operators. In principle, any U.S. corporation can file an application with the FCC, who then evaluates it for technical, business, and policy considerations. In the early days applications could be considered individually, as few companies seemed to consider this to be a business opportunity. In time, license applications became a commodity as individuals threw applications over the FCC's transom; some of these were actually photocopies of others with the new name substituted. The FCC must, by law, treat all applications equally, by publishing each for comment by the U.S. community. To deal with the growth in applications, the FCC instituted filing in "processing rounds" where a particular frequency band was put up for consideration. They would subsequently make selections that would be published for comment. This process could take a year or more, and the large volume of applications in a given round would slow it down further.

On April 23, 2003, the FCC modified its rule, abandoning the processing round in favor of the ITU style of first come, first served [5]. Under the new rule, a queue for satellite applications will be established and each application will be considered under the commission's public interest standard in the order in which it is filed. The new FCC framework also includes:

- Different procedures for licensing nongeostationary-like satellite systems and geostationary-like systems, in recognition of the technical distinctions between these types of systems;
- Streamlined license processing round approach for non-GEO-like systems;
- "First-come, first-served" licensing for GEO-like satellites.

The FCC also adopted safeguards to discourage speculation (also called trafficking in orbit slots), including a requirement that licensees post a $5 to $7.5 million bond within 30 days after receiving a license, payable upon revocation of the license for missing a milestone; a limit of five pending GSO applications and unbuilt GEO satellites and one pending non-GEO application or unbuilt system in a particular frequency band; and an attribution rule so that licensees cannot evade these limits. Also, additional implementation milestones and stronger enforcement of milestones:

- Replaced the current financial qualification requirement with the bond requirement;
- Streamlined the replacement satellite application procedure;
- Eliminated the satellite antitrafficking rule;

- Confirmed that the FCC retains the discretion in reviewing assignments and transfers of control to determine whether the initial license was obtained in good faith with the intent to construct a satellite system.

To implement the new framework, the FCC adopted mandatory electronic filing for satellite applications. It also revised the licensing process for non-U.S.-licensed satellites to make them consistent with the new procedures.

12.7.2   The European Experience in Orbit Assignments

Europe, a collection of many individual countries, languages, and governments, has pursued a dual approach to satellite regulation. The first operator, Eutelsat, was patterned after Intelsat as a cooperative of European telecommunications operators. Conversely, several European countries proceeded on their own for domestic and regional satellites and followed the standard coordination procedures of the ITU. While this lead to less efficient orbit utilization and more political horse trading between governments, it created a degree of diversity inside what was otherwise a coalesced continent in regulatory terms. The most successful national operator is SES Astra (the cornerstone of SES Global), which relied on the friendly Luxembourg government to back the system both financially and through its regulatory framework. Successful DTH and cable clients in the United Kingdom, France, and Germany provided a very substantial foundation for what today is the largest satellite operator in the world.

The framework of the European Union helps resolve pan-regional difficulties by providing a regulatory forum for its members. Another area of difficulty is the various rules and laws that are designed to protect domestic industry from foreign (and even domestic) competition. Cable systems in one country want to be protected from satellite broadcasting; satellite broadcasting in another way wants to be protected from privately owned broadcasting stations. As the EU grows, this process may improve the general situation, particularly with regard to new satellite networks in central Europe where telecommunications infrastructure is generally weak. An issue that continues to daunt Europe is blanket licensing of VSATs, which was mentioned in the previous section for the United States. From time to time, the issue is brought up within the EU, studied by various committees, and then tabled. A contributory factor may be that in Europe, unlike in the United States, some countries also use Ku-band for terrestrial point-to-point transmissions.

12.7.3   Satellite Regulation in Japan

Japanese regulatory policy has changed rapidly from a government monopoly to near-open competitive entry. The Ministry of Post and Telecommunications (MOPT) is the regulator, having a very high degree of control of spectrum use and the provision of telecommunication services within Japan. NTT, the internal public network operator, was privatized and now has several domestic competitors. Called Type I and Type II carriers, these companies use terrestrial or satellite systems to compete head-to-head with NTT. These competitors are backed by the major

trading, electronics, and utility companies who have been allowed to enter the tele-communications marketplace.

There are three FSS satellite operators in Japan: Japan Satellite Corp. (JSAT), Space Communications Corp. (SCC), and BSat. All three are currently operating Ku-band satellites and provide a variety of services in the video, data, and telephony fields. The market is still under the control of MOPT, who wishes to maintain some order and discipline. Likewise, there are international communications firms that compete with KDD, the previous monopoly international service provider for Japan. All use satellites and fiber optic links to provide connectivity from Japan to other countries of the world. Through these initiatives, Japan has demonstrated a special ability to make a rapid conversion to introduce a competitive market in satellite and other services.

The opening up of the Japanese economy has aided making satellite communications systems operational. Satellite dishes now appear on many rooftops in major cities and in the countryside. During the first generation of Japanese FSS systems, many new and innovative applications were introduced. Several have continued into the second and third generation of satellites. These have relied on Ku-band frequencies to deliver one-way broadcast services, which require no special regulatory approvals on the ground. This frequency band was selected in the first place because of the extensive use of C-band terrestrial microwave links by NTT. It was a given in Japan that sharing of terrestrial and satellite services at C-band would not be workable. With the introduction of foreign satellites covering Japan with good C-band performance, there have been inroads in the delivery of point-to-multipoint applications. Experience indicates that the problems associated with locating C-band receiving terminals is no more difficult than in the United States. Unlike in the United States where receive-only dishes can be licensed to obtain protection against terrestrial microwave transmissions, the Japanese government does not protect C-band receive services. This means that users who install C-band dishes to receive foreign satellite downlinks must be prepared to provide greater isolation if new microwave transmitters appear.

12.7.4 Satellite Operations in Asia and the Pacific

The Asian environment outside of Japan has been a hotbed of activity for satellite operators and network developers, with government and commercial operators competing for orbit positions, spectrum, and ultimately the business. The ITU is being pushed to its limits because literally every Asia and Pacific region country is pursuing a satellite system of one type or another. It does not matter how large or small the country may be. Tonga, for example, may extend across a thousand kilometers of ocean but hardly needs 12 orbit slots. These, it turns out, are rented out to other countries that failed to obtain prime slots over their own territory.

The first country in Asia to implement a satellite system was Indonesia. Having gone through the ITU regulatory process, this country is a stalwart of applying satellite communications in place of terrestrial networks. As the largest country in the world in terms of population, China has a dominant position due to its demand for service and technical capabilities in the way of design, operation, and launch of satellites. Its regulatory environment is evolving from one of the most complicated to

one where FCC-like processes and procedures are being introduced. This is due to both the interest of Beijing to advance its economic position and to satisfy demands by the World Trade Organization.

On the other hand, Hong Kong and Singapore are demonstrating their unique ability to gather businesses and produce products and services that are of value worldwide. Hong Kong, in particular, has its own regulatory setup, making it relatively easy for a local company to pursue satellite communications as a business. Since the handover of Hong Kong to China in July 1997, Beijing has become more involved in Hong Kong's business and government; as a result, no new operator has appeared on the scene.

Singapore likewise has moved from a user of satellites to an operator, with at least one satellite operator on its soil. As a sovereign nation, Singapore is in a unique position to exploit its right to assign radio frequencies and engage in international regulatory affairs.

The satellite systems that have sprouted in Thailand, Malaysia, and the Philippines are all backed by commercial companies. This is more like the U.S. model, where private companies are afforded the use of the orbit for achieving both a national and a commercial purpose. Other countries are moving more slowly, in line with the ability of their economies to absorb this type of investment. Also, it continues to be attractive to rent transponders from international operators based in the United States and Europe, and from neighboring countries. Competing international satellite operators like PanAmSat, Loral Skynet, and NewSkies Satellites now provide capacity in the region, giving users even more options from which to choose.

Not technically part of Asia but certainly within reach are Australia and New Zealand. While New Zealand has not developed its own system, Australia was among the first in the region to launch domestic Ku-band satellites. Now a private company owned by Singapore Telecom, Optus has two satellites and launched a third in 2002. The system forms part of the second telecommunications operator, in competition with the former government-owned company, Telstra. Putting DTH and VSAT networks together in these countries is fairly straightforward from the national government perspective. However, local cities and town councils are notorious for placing difficult roadblocks in front of those who would install satellite dishes in clear view of citizens. An interesting exception is documented in the entertaining and accurate movie, *The Dish*, which presents how a small town adopted a radio telescope that worked as a tracking station during the Apollo Program.

The vast reaches of the Pacific Ocean with its distribution of small to medium-sized islands provides one of the greatest opportunities for satellite networks. From the beginning of our industry, the island nations, colonies, and trust territories have depended on satellites to reach populations off the cabled track. At the annual Pacific Telecommunications Council (PTC) Conference held in Honolulu, HI, the leaders of Pacific Island resources and regulation gather to discuss and resolve many issues that hamper development. The Pacific Islands Telecommunications Association (PITA) was established through the PTC to promote better acquisition of services and processes. Even with this effort, it can still be a difficult challenge to get the needed satellite services fielded in a reasonable amount of time and at acceptable cost. Regulations in the region tend to favor national operators, who themselves

suffer from the need to invest millions of dollars to support populations that number less than 100,000 inhabitants.

12.7.5   Satellite Regulation in Latin America

Another region of strong interest in satellite regulation is Latin America, which stretches from Mexico through Central America to the full extent of the continent of South America. In contrast, the area of the Caribbean is generally associated with North America, although Puerto Rico, Cuba, and the Dominican Republic are clearly Spanish speaking. Mexico and Brazil were early adopters of domestic satellites and continue their leading roles as supporters of the ITU and its regulatory processes. They were joined at the beginning of the twenty-first century by Argentina and the Andean Pact nations who now have their own satellites as well. The rapid trend toward privatization of telecommunication monopolies has influenced the development of satellite networks in this region. In particular, whereas the first domestic systems were implemented by governments, the next generations and new entrants are owned by private interests. Brazil, for example, opened up its orbit space by offering licenses for sale in subregions of the country. This has allowed Loral, PanAmSat, and others to enter the otherwise highly protected Brazilian satellite market.

12.7.6   The Middle East and Southern Asia

Long a user of Intelsat, the Middle East continues to represent a region of solid use of satellite communications. During the 1990s, satellite systems were developed for enhanced services to the Arab League nations, including Saudi Arabia, Egypt, and the United Arab Emirates (UAE). With vast reaches of sand and lush plains, the Middle East presents a fertile ground for satellite broadcasting and VSAT networks, which have seen rapid growth. The political challenges of this region, including war at times, adds to the need for communications that can be established quickly and yet can provide the full range of broadband services. However, gaining access to bandwidth from some countries has been a challenge. A focus on education, technology adoption, and social progress are factors that will allow the Middle East to move to the front of developing regions. Satellite regulation will need an impetus so that the benefits of broadband communications are not withheld from those most in need of it.

Moving to southern Asia, India was an early adopter of satellite communications and has nearly all of the necessary technologies and expertise within its borders. The Indian Space Research Organization (ISRO) continues to be the designated satellite developer and operator; however, foreign satellites are beginning to service the local market. This has taken more than a decade to develop and in some ways is still hampered by regulations that at times are difficult to understand. India has unrivaled ability in software development and technical support, strengths that again encourage greater access to domestic and international telecommunications. The tech industry in India is therefore a strong advocate of a more open policy in satellite regulation.

Pakistan, a longtime user of foreign satellites, has joined the operator club through Paksat 1. This satellite was purchased through G2 Satellite Solutions of the

United States and was originally Palapa C1. In time, it is highly likely that Pakistan will introduce a new satellite of their specification. There are several existing satellite networks in Pakistan as the technical community is quite substantial. Neighboring Afghanistan is another country that has demonstrated the value of satellites and is undergoing a transformation to an open society and regulatory environment.

12.7.7   Sub-Saharan Africa

The countries of this region, with the exception of South Africa, are all within the general classification of developing economies. As of this writing, almost all satellite capacity is offered by foreign companies not based on the continent. The important exception is the Regional African Satellite Communications Organization (RASCOM), based in Abidjan, founded in 1993 with 44 member states [6]. The idea to have a small fleet of RASCOM GEOs catering to Africa has been delayed time and again—partly because the money could not be found, partly because of political stumbling blocks.

The first RASCOM satellite, due to be launched in 2005 and located at 2.9 EL, will have 16 high-power Ku-band spot beams covering all of Africa. It is tailored to serve some 50,000 small Internet VSATs in telecenters, businesses, schools, universities, post offices, airports, and government offices. The VSATs are part of the RASCOM order, which has been given to French satellite builder Alcatel. RASCOM began operation in 2002 with a number of pilot projects, using leased satellite capacity under PASCOMSTAR-QAF, established in Mauritius. The operational phase of RASCOM 1 is 2005–2006, including RASCOM 2. As for VSATs, each vendor of equipment can tell you how to get a license or they will get it for you. In some African countries like Ghana and Uganda this is quite easy—in some others it is more difficult.

Developments in the Americas, Europe, Asia and the Pacific, the Middle East, and Africa provide the backdrop for considering the impact of regulation on the development of satellite networks. This was an introduction and review and is not a substitute for the regulations themselves or a specific study of the situation that would exist for a given system or service. What we wish to do is to provide some broad guidance to newcomers to the field. The people who engage in the regulatory process are professionals who have the requisite experience and technical expertise. They fall into three broad categories: (1) the people who write the regulations, (2) the people who administer the regulations, and (3) the people who practice systems engineering and spectrum management in order to allow a given system or service to go into operation. Many of us, particularly this author, likewise need to understand these rules so that, as practitioners, we can take the most appropriate steps in bringing the needed systems into existence.

References

[1]   International Telecommunication Union, *Radio Regulations*, Geneva, Switzerland: ITU, 2001.

[2]   Elbert, B. R., *Introduction to Satellite Communication*, Norwood, MA: Artech House, 1987.

[3]   Malaguti, N., "Coordination of an Earth Station with Stations of Terrestrial Services as well as Earth Stations Operating in Opposite Direction," *ITU-BR,* Geneva, Switzerland, 2002.

[4]   Elbert, B. R., *International Telecommunication Management*, Norwood, MA: Artech House, 1990.

[5]   "US FCC Revamps Satellite Licensing Process," *SatNews*, April 23, 2003.

[6]   Hegener, M., "Internet Via Satellite in Africa—An Overview of the Options Available," *The International Institute for Communication and Development (IICD), Advisory Note No. 6,* January 2002.

The Business of Satellite Communication

After four decades of development, we know satellite communications systems and services to be a mature industry. Literally any organization with the financial means can invest in space segment capacity as well as the associated Earth stations for communications services. As we shall see, this can be done as an integrated offering, such as in the MSS and BSS fields, or segmented to focus on a particular aspect, which is common for FSS systems. The financial rewards in this business vary widely, just like in real estate, Internet service provision and Web hosting, and oil exploration. We discuss many of the factors necessary to establish a satisfactory market for the systems and the services satellites deliver. At any given time, innovation in space and on the ground will provide new opportunities and pose threats to various incumbents in this industry.

13.1   The Satellite Marketing Challenge

Satellite communications is a capital-intensive industry, like gas pipeline operation and long-haul fiber links, where the investment cost for entering the business is high and operating cost for delivering service is relatively low. This statement is more applicable for space segment sales than DTH broadcasting, the latter paying a heavy price for content. The investment in satellites, transponder capacity, and fixed Earth stations represents a sunk cost that is very difficult to liquidate if the revenues are not satisfactory. This aspect is like constructing a large commercial building that could end up being unfilled by tenants. However, unlike a fixed building that is located in one particular part of town, a satellite system is unique in its ability to serve an entire nation or region. Its capacity may be subdivided and sold in small increments like lumber or airline seats. Another issue that impacts the business performance of a satellite is its fixed lifetime. This means that capacity that is not earning revenue due to a lack of customers represents a permanent loss for the time that it is not used.

Figure 13.1 indicates the major elements in a satellite network that is used for an integrated service such as DTH or data communication. The pair of satellites along with the TT&C stations and satellite control center are owned and managed by a satellite operator. Most of the expense is tied up in the initial capital needed to construct and launch the satellites (including launch insurance) and to install the ground control elements. The ground segment, shown below the dashed line, represents the application portion of the space-ground system. It consists of a major Earth station, called a hub or teleport, and a collection of VSAT user terminals. The

Figure 13.1 Key elements of a typical satellite system form the standpoint of marketing and business. The satellites and TT&C/satellite control facilities make up the space segment, owned and maintained by the satellite operator. The ground segment is often provided in pieces by different organizations that either use the capabilities themselves or offer their services to outside customers.

type of terminal obviously depends on the application being delivered. Whether the application elements are owned by the satellite operator or not depends on factors discussed later.

For example, the most profitable DTH service business in the world is BSkyB in the United Kingdom. The satellites for this service are owned and operated by SES in Luxembourg, who in turn has no interest in BSkyB other than to supply transponders to distribute programming. In this case, the DTH service provider concentrates on that particular business and leaves the investment and operating demands of the satellites to SES. In the United States, the DISH Network from EchoStar owns and operates both the space and ground assets; only the receiving terminals are owned by customers. EchoStar's larger competitor, DIRECTV, is similar except that it owns satellites that are operated by PanAmSat and Loral.

The ground equipment sold to space and ground segment operators represents around one-quarter of the total market. Specialists provide antennas, amplifiers, modems, and packet encapsulators. Other companies put together an integrated system to support data communication or video transmission. Design and manufacture of user terminals for the applications is a third important segment of the ground equipment market.

Perhaps the most difficult challenge to achieving satisfactory financial rewards from a satellite investment is that of creating an adequate customer base. Big projects like XM Satellite Radio and Thuraya become profitable above 1 million users; the failure of Iridium and Globalstar to reach even a quarter of this number makes this challenge clear. In transponder marketing, the users must posses their own transmit and receive Earth stations, which can be a significant investment for many organizations. In the case of cable and broadcast TV, the customer could already have the necessary antennas but may be using another satellite. The only way to acquire the business is to undercut the existing operator in price or to offer additional features at the same price. The buying opportunity comes when the existing satellite is nearing its end of life, which is a situation that presents itself to every incumbent transponder provider. This is why an operator should have an effective plan for replacing satellites without loss of continuity.

A DTH system, including sufficient satellite capacity and program content, is a large commitment by anybody's measure. An organization wishing to explore satellite communications as a value-added service but on a much smaller scale can enter the packaged video transmission or VSAT network service business. New entrants in this market still experience the hurdle of building up an adequate customer base. A new network provider must make it as easy as possible for subscribers to install the necessary user equipment and access the service. Break-even points for a VSAT shared hub operation as compared to a DTH system might be 2,000 data communication users as opposed to 1 million or more pay-TV subscribers.

13.1.1 Selling Hardware

The range of possible hardware elements covers many manufacturing sectors in aerospace, telecommunications, electronics, construction, and heavy industry. This affords manufacturers many possible opportunities to contribute to any number of equipment or service businesses. Due to the large investment involved in the launch and introduction of a major satellite system, manufacturers of spacecraft, launch vehicles, or ground systems largely deal with organizations that can make the necessary capital investment. They, in turn, must have market power in their proposed area of service. As is typical in the aerospace industry, the customer buys through the request for proposal or request for tender (RFP or RFT) process and chooses the lowest bidder that meets the requirements. There is also a certain degree of development risk that the system could fail to meet specifications, be delivered late, or cost substantially more than the original estimate. Technologies that are internal to the satellite as well as those needed to develop and manufacture the satellite make entry into this business very difficult. Competition being fierce, there are few spacecraft manufacturers that have been financially successful at it.

Manufacturing and selling Earth stations have evolved from a similar structure into something more like a mom-and-pop industry. Originally dominated by large electronics and defense firms like Northrop, Hughes Aircraft Company, and NEC Corp., most large Earth stations are now provided by small specialists like Satellite Transmission Systems (STS) of Hauppauge, New York, ND SatCom, Inc., of Dallas, Texas, and Allen Communications of Los Angeles, California. This is possibly because these stations are put together like custom houses from off-the-shelf

electronic units produced by still other specialists like Miteq, CPI, Xicom, L-3 Communications, and Tandberg. Entire networks, produced by Gilat, HNS, ViaSat, and Shiron, provide an integrated solution for an end-to-end data communication service or application. The situation with regard to user terminals is different again because we are talking about a device that is more akin to a piece of consumer electronics like a VCR. This is exemplified by the first generation of digital DTH set-top boxes by Thomson Consumer Electronics, designed for the DIRECTV and USSB networks in the United States. However, unlike a VCR, a DTH box is usually tied to a particular DTH service, rendering it useless if the service is discontinued. A standard voice-band modem used to access the Internet or World Wide Web with a PC is not so constrained, making this type of communication device almost as versatile as a fax machine or, for that matter, a telephone. Ericsson and Nera, two Scandinavian companies, provide MSS telephones that operate a lot like mobile phones.

Success in the sale of hardware is dependent on the key factors of reputation, technology base, and economy of scale. As we indicated, large Earth stations are usually produced by small companies who specialize in the type of business that is called "rack and stack." Satellite manufacturing is restricted to the most established producers, and only Orbital Sciences has succeeded in entering this business in recent history. Regarding user terminals, the field tends to be wide open to any company that either has a technical edge or can flexibly produce quality consumer electronics at affordable prices. Different players from the two ends of this particular spectrum can integrate their strengths through licensing agreements or other types of joint ventures.

13.1.2 Selling Services

Marketing services from satellites is in some ways much easier because there are a variety potential customers with which to deal. The corollary to this is that it takes many customers to make a successful business out of it. Some customers are only looking for communications and have no concept of how the satellite fits in. Their knowledge of satellite and Earth station technology may be lacking as well. The real challenge is to make the service comprehendible and very user friendly. That a satellite is used in the application is not sufficient to entice a customer onboard.

A satellite operator can sell space segment alone or offer partial or full network services. The marketing of satellite capacity in the FSS has been called "plain old transponders," a parallel to the plain old telephone service (POTS) term used to describe the conventional PSTN. The principal advantage of the plain old transponder business is that it greatly reduces the selling burden on the satellite operator and tends to maximize the range of potential needs that the satellite might satisfy. The cost of operating the system in this manner is relatively small in relation to the cost of capital. In cases where potential customers are either unfamiliar with the ground segment or have incompatible equipment, the satellite operator might need to provide or subsidize acquisition of missing pieces.

In providing value-added services, the operator adds the ground segment facilities, including Earth stations and backhaul fiber optic capacity, needed to provide end-to-end capabilities for customers. Including ground facilities in the offering increases business risk because the delivery of services is tied to a number of fixed

points on the ground and these points represent such cost. In some cases, satellite operators have found this to be a good business when anchored by a long-term contract. On the other hand, competitors that construct too many teleports within a limited area can produce a kind of glut. Price competition forces down revenues until one or more of the providers must sell out or close down the operation. Converting the installed network to enter a different service is extremely difficult; an unused Earth station may have to be closed, the equipment entirely removed, and the land or building returned to its original state before the property can be vacated.

There are some other services that would allow the satellite operator to add value for customers and potentially increase revenue and, hopefully, profits. Customers who require satellite capacity also need to install and manage Earth stations. This poses a number of problems for the newcomer to the business. A satellite operator can assist the customer by providing engineering and installation services, along with the associated maintenance after the sale. Some, in fact, offer to outsource the operation and maintenance of the facilities. Actually, this is not a new concept, just a new name for the provision of operations and maintenance (O&M) service, something that old-line contractors like Lockheed-Martin and Boeing have done in the government sector for decades. O&M service can be profitable if the customer is willing to pay the true cost along with a reasonable profit on top. The opportunity (or problem) for the satellite operator lies in the fact that many equipment vendors, in their wish to increase hardware sales, undercut this service by offering maintenance at very little additional cost to the customer.

13.2   Selling the Space Segment

Our focus in this section is on selling the space segment after the satellite is placed into service. With delivery-in-orbit satellite contracts from the leading spacecraft manufacturers and adequate launch insurance, risks to new operators are of reasonable magnitude. Once the satellite has been successfully placed into service, it is an unlikely occurrence for there to be a catastrophic failure that would bring the entire system down permanently. Rather, the operator must deal with the normal problems of building an effective marketing and operations organization along with planning for the replacement of the satellites as they near end of life.

The space segment market has so far provided the greatest net profits to operators of any particular segment of the satellite industry business. This, of course, is not guaranteed, as the key determining factor is the location of the satellite tied to the country market that it serves. This is referred to as the "neighborhood" or "hot bird" slot. A really good neighborhood is one where several services aggregate, attracting more subscriber antennas and, as a consequence, more services. If your market is based on such an excellent neighborhood and you have a leading position, then the financial rewards are likely to be outstanding. A prime U.S. example is Galaxy I, a cable TV satellite that achieved 100% fill immediately after launch with top-line services like HBO, CNN, ESPN, MTV, and the Discovery Channel. In Europe, the Astra 1 series proved that the neighborhood principle also applied in a region with several country markets. The converse is also true; that is, without a

good neighborhood, financial returns are far less satisfactory, as witnessed by several defunct satellite operators in the United States and Europe.

The primary discussion is on marketing of transponder capacity on FSS satellites. BSS satellites are typically operated by the DTH/DBS operator, and transponders are not sold individually or subdivided as discussed next. In contrast, the MSS market is very specialized and services are typically offered on a per-minute basis. Thus, there is a very tight connection between the mobile satellite operator and the provider of telephone and data applications.

13.2.1   FSS Transponder Segmentation

For a company that has selected the space segment market for its area of business, there are several ways that its capacity can be offered to prospective buyers. As indicated in Figure 13.2, capacity can be divided up in terms of time or spectrum. This then prescribes the types of applications that can employ the capacity and therefore the market segments that can be addressed.

Before considering how to sell this capacity, we first need to examine how rewarding, in financial terms, each of these options could be. This vital factor is indicated by the shading of the particular segment of the market shown in Figure 13.2. The most attractive segment shown in the upper right-hand corner consists of full transponders on a full-time basis. Since the customer is committed to all of the capacity, the problems associated with maintaining the "fill" are reduced to a minimum. Customer commitments could be for a period as short as 6 months to as long as the life of the satellite. In addition, the revenue generated from full transponder sales or rentals is typically greater on an aggregate basis than from the other segments, except possibly the one below it in the figure.

Moving along the top toward the left (maintaining the full-transponder aspect), we encounter the part-time (recurring) and then the occasional segments. When we approach the market from this perspective, we accept the possibility of lower maximum revenue in exchange for the ability to serve more customers whose needs are not full time. Part-time users commit to specific time periods on a daily or weekly basis. If you can line up enough part-time users, you may come close to filling the transponder from a financial standpoint (periods of time after, say, midnight and

Figure 13.2   Division of "plain old transponders" services, indicating their relative value and importance to the satellite operator or service provider. The most attractive services are represented by full-time commitments to customers in the TV and data communications fields.

before 6 a.m. are difficult if not impossible to sell, so the idea is to charge a premium for the more desirable hours). Occasional users make no specific commitment for recurring services, instead reserving the desired block of time as if it were a seat on a commercial airliner. Transponder resources are booked through a reservation desk, and the service is subsequently delivered at the prescribed time by the network operations staff. The aggregate revenue could approach that of the previous two full-transponder services, but only if there is a large enough pool of users and transponders.

The other way to slice up the transponder is to divide the bandwidth and power among several users. Terminology for this is partial transponder service or SCPC. One or more applications are provided in the same transponder using FDMA, where each user is assigned a separate carrier frequency. The maximum power that is available is reduced by the required output backoff (typically 4 dB to control intermodulation distortion). A small customer could take a very narrow slice of perhaps 30 kHz and only 1% of the total power (after the backoff). On the other extreme, one customer might commit to a significant fraction of the transponder, say 45%, and then subdivide it further for their own internal use. The financial reward for this is as good as part-time/full-transponder service or, in exceptional cases, as good or better than a full-time, full-transponder commitment by one customer. As will be discussed later, the difficulty comes in pulling in enough partial transponder users within a reasonable amount of time (because before adequate fill is achieved, a lot of potential revenue is lost).

The blocks in the lower center and left of Figure 13.2 indicate part-time and occasional services in the partial transponder segment. Experience has shown that it is much too complicated to make assignments and track part-time and occasional usage in this type of arrangement. There are users who may wish to employ an occasional slice of bandwidth for applications such as video teleconferencing and emergency communications. However, the business they generate is usually much less than the difficulty they cause. The better approach is to offer a dedicated FDMA channel to the user; they, in turn, employ it on a part-time or occasional basis. The frequency is therefore dedicated to them and they will not have a problem gaining access to it. As a satellite operator, you would be able to depend on a continuous and consistent flow of revenue from this customer, even though their usage is erratic. Whether it makes sense or not depends on the customer's particular situation and resources.

The typical satellite operator will market space segment services in search of the most revenues with the least effort at selling. The degree of penetration in the occasional and partial transponder service markets will depend on the amount of disposable capacity, the available market, and the technical and operational resources of the operator. Some of the fine points of space segment marketing in each of these segments are covered next. The particular types of services and the applications they address are listed in Table 13.1.

The market for transponders has evolved over the years and partly represents a commodity business because of the standardization frequency, bandwidth, and power. A key discriminator is the satellite footprint, which defines where services may be delivered. There will be instances where a customer will not be able to locate more than one operator with appropriate transponders, in which case the satellite

Table 13.1   Space Segment Services and Applications, Based on the Allocation Techniques
Indicated in Figure 13.2

Overall Market Segment	Applications	Potential Buyers	Sales Approach
Transponder marketing (full time)	Cable TV, DTH	Cable networks (HBO, Fox Sports Channel); DTH networks (Sky Italia, Multichoice)	Outright sales and leases
	Television distribution, video backhaul	TV networks (CBS, Fox TV, BBC); news channels (CNN, MSNBC, Fox News Channel)	Leases and short-term rental
Partial transponder services	Radio program distribution, data broadcasting, backbone point-to-point links	Radio networks (ABC Radio, Deutsche Wella)	SCPC marketing, based on power and bandwidth used
	Data communications, Internet access	Corporations and government agencies; ISPs	Fractional bandwidth, based on power and bandwidth used
	Satellite news gathering	TV networks; cable news channels	Managed SCPC, with real-time access control
Part-time and occasional video services	Video backhaul, syndication, business TV, distance education	TV networks; cable channels; corporations and government agencies; service packagers; universities	Fractional bandwidth, based on power and bandwidth used; may include other services and real-time access control

operator will have a decided upper hand in the negotiation. On the other hand, if multiple sources exist, then there tends to be a significant degree of price competition. Other important aspects have to do with the term of the agreement and the nature of the backup service that the satellite operator can put on the table.

Another factor is the neighborhood principle that was introduced previously. If this factor is involved, then the pricing and availability of the transponder for the particular use could be severely impacted. For all of these reasons, there is no such thing as a standard price for a transponder, and prices can vary not by percentages but by integer factors. During the mid-1980s, for example, a standard C-band transponder on U.S. domestic satellite such as Westar 4 rented for $60,000 per month, while at the same time a cable TV network might be paying $250,000 for a transponder with the same technical performance on a cable satellite like Galaxy I. The difference in neighborhood value is the result of 80 million cable households that the Galaxy I transponder can reach.

13.2.2   Space Segment Provision

Having decided on the satellite with its associated coverage footprint, frequency band, bandwidth, and EIRP, a satellite operator still has a variety of ways to offer the full-time capacity of a transponder. Users are concerned about the type of transaction, the term of the contract, and the degree of backup (protection) that the operator is willing and able to provide. The alternatives that are commonly applied in the industry are identified in Table 13.2 and reviewed in the following sections.

Table 13.2   Typical Arrangements for the Provision of Space Segment Services

Arrangement	Term	Features
Transponder sales	Life of the satellite	Warranty payback provision for total transponder failure
Long-term leases	Life of the satellite	Comparable to transponder sale, without warranty (payments stop in event of failure to provide service)
	3 to 8 years	Not considered a sale, since transponder reverts to operator at end of term
Short-term leases	Fixed term (cannot be canceled)	Used for business startup and limited duration applications (to hold capacity for use)
	Month-to-month	Itinerant usage for short-term applications; extension may be by mutual agreement
	Ability to cancel on notice	Offers flexibility, but provides less security to the operator
Part-time and occasional usage	Operator reserves capacity for specific events	Most flexible service (discussed later)
Partial-transponder (SCPC)	Fractional power and bandwidth, fixed term or monthly	Comparable to short-term leases; needs operator care to control user power and intermodulation

Each has technical, operational, financial, and legal implications. Because of the wide variety of user needs and the competitive nature of the market, there are other combinations and variations that are applied from time to time.

13.2.2.1   Transponder Access Guarantees and Backup

Full-time or part-time services are tied to a particular satellite and frequency. Customers assume that the capacity is provided as specified in their contract. However, there are situations where a particular transponder does not perform as required or fails, and action must be taken to restore service as soon as possible. In an extreme and infrequent case, the total failure of a satellite could seriously impact the viability of services without a contingency plan. We consider below the standard provisions for guaranteeing backup by the satellite operator. In any of these situations, it is assumed that the operator would have already attempted to employ a spare amplifier switched into the same transponder channel (such redundancy would be employed to maintain service and therefore revenue).

The transponder guarantees are:

- *Protected service:* The satellite operator will provide backup in the event of transponder failure, using capacity drawn from reserve transponders or by preempting service to a lower priority (preemptible) customer. The replacement capacity can come from the same satellite (intrasatellite protection), another satellite in the operator's fleet (intersatellite protection), or an in-orbit fleet spare satellite that would be drifted into place (total replacement).
- *Unprotected service:* No backup is guaranteed, but this customer cannot be preempted to restore a protected service. This is the most popular form of

service as it balances risk versus cost. The fact that the service cannot normally be terminated gives the user good confidence.

- *Preemptible service:* No backup is provided and service can be preempted to restore a protected customer. An order of preemption should be defined in each customer's contract.

Customers who need the guarantee of protected service should expect to pay a premium, whereas others who accept preemptable transponders could justify a discount. For most customers, the nonpreemptable class has the best balance of price and security since access to the transponder cannot be interrupted unless the operator cannot bring the service back to life.

Whenever service is stopped, the corresponding action will depend on the nature of the contract. A sale of a transponder will come to an end if there is a total failure; the customer usually is entitled to a rebate of a portion of the purchase price in proportion to the unused remaining life of the satellite (i.e., warranty payback). This would be based on an assumed minimum lifetime, perhaps 12 years of the expected lifetime of 15 years. A more harsh arrangement would be that the customer simply loses out and no refund is forthcoming. The case of a transponder lease or rental is more straightforward: if the transponder cannot be restored, then the customer simply stops paying.

13.2.2.2   Transponders Sales and Ownership

Subject to local regulatory control and availability, users may be able to purchase transponders on a condominium basis. This was pioneered in the United States by Hughes Communications with the sale of all 24 transponders on Galaxy I to cable TV networks. Some of the advantages of such ownership are:

- Access for the life of the satellite, which is subject to preemption conditions, if any; also, backup may be provided (see Section 13.2.2.1);
- Control of all transmissions through the transponder and greater influence over the operation of the satellite, which will depend on the provisions of the purchase agreement;
- Ability to finance with leveraged leasing or other innovative financing means;
- Knowledge of who your neighbors are—this may be stipulated in the purchase agreement.

Purchase agreements are usually private transactions that allow the buyer and seller to negotiate the terms of service, guarantees, and financing (if applicable). Contracts of this type seldom become public, except if the buyer is a publicly traded company in the United States, in which case it must reveal any major transactions and commitments. Satellite operators prefer long-term commitments and generally would like to sell out the satellite prior to launch. From that point, there is no further marketing for the particular satellite, and the primary responsibilities relate to the proper operation and control of harmful interference. Early failure of a transponder will cause termination of the agreement with specified rights and remedies, unless alternate transponders are included in the deal.

13.2.3   Selling Occasional Video Service

Many of the video users identified in Table 13.1 need only a few hours of satellite transmission at one time. The events could be of a one-time nature or could be repeating on a weekly or monthly schedule. Owners of unused transponders find that occasional video is a way to get revenue without tying up the capacity long term. Generally speaking, the revenue from occasional video, while important, is less than can be had from a full-time user. Some users only require occasional services and hence can save considerable money by not committing to full-time transponders. The carrier may wish to restrict occasional video capacity since it can cannibalize long-term deals with captive TV networks.

The market for occasional video service in the United States is the most developed in the world and covers many applications and types of customers. Among the most established are:

- Entertainment program syndication—the delivery of specific TV programs that are sold individually or as a series to local TV stations;
- News backhaul, such as bureaus and on location—mostly ad hoc links for specific events from fixed or temporary locations, using SNG vehicles and fly away terminals;
- Sports backhaul—from stadiums, race tracks, and Olympic venues;
- Private broadcasts (business TV)—produced by corporations to deliver a professionally produced program to inform employees, announce new products, or communicate other critical information from the top of the organization;
- Distance education—a potentially very valuable use of FSS satellite capacity to conduct formal education, professional instruction, and a wide variety of training; used by universities, government agencies, and corporations.

Many of these applications are not large enough to justify a full-time transponder. This means that without the availability of the service, many uses would not exist to a significant degree. Of course, there could be an organization that is large enough and has sufficient resources to make the commitment. Digital compression increases the number of TV channels per transponder and therefore has reduced the effective cost of space segment even further.

A satellite operator or other service provider that wishes to enter the occasional video market will need to put into place an adequate service infrastructure. This consists of, as a minimum, an order desk to receive requests from customers, a network operation center to coordinate access to the transponders and record usage, and a billing organization to invoice customers and verify payment. As the pool of transponders grows along with the customer base, the operator will find it necessary to automate some of these activities through a computerized reservation and billing system. Virtually all of these systems that have been created are customized jobs for the specific operator. This is the best way to ensure that the computer software and the business work effectively hand in hand. A system of this type can be extended to include reserving outside facilities like terrestrial links and uplink Earth stations, along with studio facilities that might be required to complete the end-to-end service.

Arranging part-time and occasional services has become a bit of a cottage industry, with businesses in this area requiring little capital commitment and relatively few employees. A leader in this field is The Space Connection, based in Burbank, California. This company sells satellite capacity on most of the satellites covering the United States. They do this from capacity in their own inventory as well as transponders booked from the inventory of others. They are able to quickly arrange space and ground segment elements, confirm all of the bookings with the appropriate suppliers, and bill customers for the entire package. In 2003, The Space Connection entered the SNG field by becoming the exclusive agent for PanAmSat's SBS 6 satellite. This required the addition of an uplink management center to control access to the satellite.

13.2.4 Partial Transponder and SCPC Services

The satellite operator can extend its market to smaller customers by offering fractional transponder power and bandwidth on a full-time basis. However, there are several issues that should be addressed.

- A transponder will be committed to this service even if there is one carrier operating. Other customers will have to be added as rapidly as possible to achieve an adequate fill factor.
- Each user will pay a relatively small fraction of the required revenue; revenue will increase over time according to the fill and churn (e.g., customers who terminate their service and are subsequently replaced by new customers).
- Many customers must be marketed to (each one perhaps infrequently).
- Partial transponder services are susceptible to interference from high-power carriers in cross-polarized transponders and on adjacent satellites. Assignments should be made with this in mind, and actions must be taken rapidly when interference occurs.

Users can also interfere with each other since they share power and bandwidth. For this reason, carrier power levels and frequencies must be precisely and continuously monitored by the satellite operator. Any significant deviations by users must be addressed quickly and thoroughly to deal with financial and performance issues.

Buyers of partial transponder capacity usually do not care about the neighborhood for which the satellite provides. The FDMA technique that is part and parcel to partial transponder service allows for individual customers to operate almost independently of each other. VSAT networks, data broadcasters, and private audio networks that do not connect with one another can share the same transponder. Therefore, there is no synergy among these users, and price is a major differentiating factor in satellite selection.

An exception to this rule would apply to commercial radio networks that deliver radio programming to local radio stations. Users prefer either full-time channels or, at a minimum, fixed time blocks for the duration of daily programs. Radio networks can also be supported through a value-added network that is operated by a third party. The manner in which radio, data, video, and telephony networks can be delivered on a value-added basis is covered in the next section.

13.3   Value-Added Service Offerings

Satellite communication was originally seen as a domestic telecommunications business that competed for the long distance telephone subscriber. Rather than being a competitor, what has evolved is that satellite communication is a mutual effort of the satellite operator, the user, and conventional telecommunications companies in the fiber and wireless markets. Terrestrial carriers have an important role since they provide PSTN access, backhaul connections, the last mile between satellite link and user, and backup services over the public network. PanAmSat, for example, entered into such a cooperative agreement with Level (3) Communications to create what they call the Virtual Teleport. The idea is that a customer would not need to arrange a backhaul circuit from their location in one city, say Houston, TX, to a PanAmSat teleport in Atlanta, GA. Using the nationwide fiber network of Level (3), PanAmSat will pick up the traffic at the customer's location, transfer it via fiber to the teleport, and uplink it to the appropriate satellite. There are physical facilities involved, but the teleport itself need not be located in the same city as the customer. This concept was proposed in the early 1990s by Hong Kong Telecom (now Reach Networks) to offer MSS land Earth station services to Asian countries not wishing to invest in their own land Earth station.

Value-added services cover a wide range of capabilities and offerings, making it difficult to generalize. Table 13.3 provides a way of viewing the scope of the industry, extending from the simple case of maintaining a customer's equipment, through more complex arrangements like sharing an uplinking Earth station, to the most extensive offerings of complete integrated networks. The degree of value added at the top indicates how the scope of service increases from left to right.

13.3.1   Entering the Competitive End-to-End Services Business

Developed economies offer an environment where it is relatively easy to piece together a complete service from the available elements. In other parts of the world,

Table 13.3   Satellite Communication Industry Scope, Showing the Range from the Space Segment Only Through the Full Range of Value-Added Services

| Application Segment | Degree of Value Added → | | | |
	Space Segment Only	Customer Assistance	Systems Integration	End-to-End Service
Video	Full transponder	Event arrangement, maintenance, shared uplink	Uplink Earth station, equipment manufacture, network control	Video distribution network, DTH systems
Data	Full and partial transponder	Shared hub, maintenance	Hub implementation, equipment manufacture, network control	Data distribution, Internet and intranet access, backhaul, videoconferencing, VoIP, strategic IT applications
Telephony	Full and partial	Shared hub, maintenance	Hub implementation, equipment manufacture, network control	Rural telephony, MSS systems, telephone trunks

users will find this more of a challenge and therefore may need to engineer and construct new fiber or wireless facilities. Satellite operators and other companies in the industry attempt to satisfy this need by establishing a version of one-stop shopping for satellite communication. Companies that offer these value-added services may form joint ventures to minimize the risk to the provider. The challenge of this business is for the provider to gain control of all aspects of the service. Furthermore, operating and marketing expenses can be very high since a working network capability is usually a prerequisite.

There are many examples of end-to-end service offerings and companies built upon them. Successful ones run the risk of declining into unprofitability as the technology is diffused and low-cost providers enter the market. Also, new technology can make the service concept obsolete. Some examples of vertically integrated service offerings, representing the right-hand column of Table 13.3, are:

- Packaged video transmission services, from end-to-end including the uplinks and downlinks, and terrestrial connections using fiber and microwave—considering the video standards and interfaces required to properly connect the two ends together (discussed in Chapter 4);
- VSAT networks via shared hubs and terminals provided under contract—with appropriate network management and on-call maintenance service (discussed in Chapters 8 and 9) ;
- Application systems that deliver a certain end-to-end capability for a specific industry could be bundled in with the network—to provide an enterprise solution that gains competitive advantage;
- Video teleconferencing networks and meeting room services;
- DTH systems, including hardware, software, and possibly programming (discussed in Chapter 6);
- Mobile satellite communications through GEO or non-GEO constellations (discussed in Chapter 11).

The dilemma of this market is that customers seem to demand a complete solution; however, the cost and risk associated with having the capability could be uneconomic. One must take the plunge in financial and operational terms to be able to address such customers, and make a commitment to stay with the business through lean startup years. DTH and MSS systems cost in the billions of dollars—and can keep costing in high operating expenses. VSAT networks are perhaps less of a financial bet, but their upside is also less. Some organizations begin with a network to serve internal users and then become service providers to share their operating costs. Among these are the costs associated with the space segment, which must be paid regardless if anyone is actually transmitting information. Having others contribute is one option to offset these expenses while maintaining the needed internal capability.

13.3.2   Selling Value-Added Services as a Systems Integrator

A company wishing to be involved with value-added services may enter the market as a systems integrator. In this line of business, the provider takes a fixed price

contract to install a working system for the end user. The project may take the form of a time-and-materials agreement where the buyer assumes the risk of a cost over-run; this eliminates much of the provider's risk associated with building the network and offering services to the market. Traditionally, systems integration was the domain of aerospace companies and defense contractors that performed under contract. Philco Ford built a tropospheric scatter backbone throughout the former Republic of Vietnam during the 1960s. Hughes Aircraft integrated a satellite communication network in Indonesia in the 1970s; Aerospatiale did likewise for the members of the Arab League in the 1980s. EDS and IBM extended the concept to corporations wishing to contain data processing and internal telecommunications expenses.

Satellite communication offers systems integration opportunities on a smaller scale because it takes less in the way of equipment and operational capability to create a wide area network. The satellite is usually in service and able to provide the critical linkages. Earth stations are installed at the prescribed locations using substantially off-the-shelf components. User terminals are distributed to end users in much the same way as cellular phones and DTH dishes. The systems integrator manages the design and implementation according to the contracted cost and schedule. Companies that perform this integration role on the ground include GlobeCom Systems, Inc., Alcatel Space, Singapore Technologies, Allen Communications, and ND SatCom. In some cases, the relationship could extend to operating and maintaining portions or even the entire system. The latter is discussed in the next section.

13.3.3   Maintenance Services

Once the ground segment is placed into service, the operator or user is faced with the costly and difficult question of providing adequate maintenance for the life of the system. Anyone who has worked this side of the business knows the following:

- Systems and facilities, like an automobile, must receive proper maintenance if quality service is expected; backup equipment must likewise be maintained to meet the availability requirement.
- Engineers who design and install the system seldom think about the O&M phase. As a consequence, maintenance people have to reverse engineer many aspects of the system in order to figure out (1) how it works and (2) how to fix it when a failure occurs. It is better to involve maintenance engineers in some phase of system implementation, and engineers in the operating phase.
- If everything is working properly, no one says anything good (or bad); this is exactly the situation we experience with our telephone and cable TV service (who has ever called their service provider to thank them for providing what's paid for?).
- If there is a failure that affects service, everyone is upset (customers, managers, and marketing people). It is best to isolate and correct problems without it ever reaching the end user.

The last lesson learned is that there is never enough money available for (1) adequate staff or services to maintain the system, (2) adequate test equipment, (3) enough spare parts and units, and (4) training of maintenance engineers and technicians.

Maintenance support may be obtained from the original vendors or third-party maintenance organizations. Otherwise, adequately skilled staff will take on the responsibility for troubleshooting and repair down to the circuit board level. This is where the service provider can step in to help the customer and keep the network or facility in good operating condition. It turns out that the number of people who perform this task is much less than it might appear. The key to success as a maintenance service provider is to have a motivated team of professionals with a thorough understanding of every element of the system. This sounds much tougher than it is. The principle of specialization can be used to deal with a complex system such as a DTH broadcast center or VSAT shared hub. Also, if the Earth stations are dispersed over a wide area, then the maintenance staff will need to be dispersed as well. There is a vital need for good communication among the maintenance team members, who should use wireless devices like pagers and cellular phones. Communication in remote areas is now possible with MSS networks.

This brings our discussion to another area of maintenance where the traditional concepts often cannot cope. Software is now a central element in every modern telecommunications and video network. Introduction and maintenance of sophisticated software is a major concern, particularly in satellite control systems and data networks. Internal software development is often undertaken out of necessity; however, there are substantial risks whenever a major software development project is undertaken [1]. Mark Norris has suggested the following simple process to encourage success.

- *Articulate what you want:* Write down your requirements and review them internally to obtain buy-in from all parties.
- *Get the right supplier for the job:* Conduct a proper procurement process and do not base the selection on Web site information and vendor hype.
- *Ensure that you keep control of the design:* Put a competent project manager in charge and give this person the needed authority.
- *Keep tabs on the supplier:* Never leave them alone and expect them to deliver on the final date.

In many organizations, there is a tendency to build up staff to minimize risk of service outage. Such an approach is greatly appreciated by users but can cause costs to escalate out of control. Therefore, it is important to minimize the required number of positions for management reasons. If there are too few people, then there is no time to rest and errors will occur more frequently. Too many people is also bad because of the cost and the boredom that result.

The proper operation of a satellite application system depends on the quality of the O&M. If this is the responsibility of the service provider, then there should be a process for measuring the effectiveness of the O&M function and how well it fulfills its obligations. The following are some suggested metrics that this author has found useful:

- *Availability:* This is the percentage of time, typically a month, that the service is working properly.
- *Mean time to repair:* This is the classic maintenance metric, used also in availability calculations.
- *Number of trouble calls and average length of an open "trouble ticket":* A trouble ticket is simply the record of a complaint from a user, indicating the nature of the problem, who reported it, and the actions taken to resolve it. The ticket is closed when the problem is resolved or the user declares the issue moot (because the user closed down the facility, left, or canceled the request for some other reason). Managing trouble calls and tickets is complex and can be supported by a computerized system.
- *Customer satisfaction, measured through a survey:* The customer is asked to rate the service quality, the people, and the company on a scale of, say, 1 to 5, 5 being the best. These numbers are collected monthly and provided to all participants. The most effective survey is taken personally either over the telephone or face-to-face. Simply e-mailing the survey usually is not enough. Or, if the survey is mailed, it should be followed up with a telephone call or visit.

13.3.4 The Services Contract and Service Level Agreement

Our approach to the business of satellite communications naturally takes us to the topic of contracts and the terms and conditions that are included. A properly crafted contract will protect both sides, being specific where needed so that no confusion will result. According to R. Wikanto, former director of engineering of the Indonesian Telecommunications Agency (now PT Telkom), the contract, after negotiation, can be put into a drawer and hopefully not referred to again. However, if there are problems with the project, then the document comes out and provides the basis to resolve differences.

The following is a framework for a typical services contract, including the basic terms of an SLA. The latter is a document that defines the required service characteristics, how they will be measured, and the actions to be taken when needed to either restore or modify services. In modern IT and telecom provisioning, the SLA specifies the standards to be followed and the consequences if they are not met. The contract has the added feature of defining what both the buyer and supplier are required to do.

Typical Content of a Satellite Application Contract

A   Objectives of contract

The contract for voice, video conferencing, and data services for the prescribed services (the "Contract") shall provide a turn-key solution for the purchasing organization (the "Purchaser") that includes installed and tested equipment along with satellite capacity and ground-based telecommunications services. The supplier of these elements (the "Contractor") has full and complete responsibility for meeting the requirements, for the initial installation and throughout the life of the Contract.

This applies to those elements directly provided by the Contractor as well as those supplied by third-party organizations in the satellite and terrestrial communications field. Except as provided for under Section G, Excusable Delays and Outages, there shall be no departing from this obligation.

B Purchaser responsibilities

Purchaser undertakes to provide space and other resources to support the efforts of the Contractor. These are reviewed below.

B.1 Remote site

Purchaser will make available to Contractor the space and facility support for the installation and operation of remote site equipment installed by Contractor. Purchaser will facilitate Contractor's access as necessary to complete the work tasks leading to verifying proper operation of equipment and services.

B.2 Licenses and approvals

Purchaser will assist Contractor in applying for and obtaining any licenses or approvals for installation and operation of ground and satellite communications equipment, including licenses to operated the remote site equipment to access the satellites. This is viewed as a cooperative effort, requiring both parties to aggressively seek these permissions on a timely basis.

B.3 Access and security

Purchaser will use reasonable best efforts to allow Contractor access to the appropriate space. Purchaser will provide physical security in accordance with its current practice for owned equipment, facilities, and personnel.

C Contractor responsibilities

C.1 Licenses and permits

As the carrier for these services, Contractor shall be required to obtain any government licenses necessary to provide the services defined in the Contract. Other licenses and permits needed to perform the work and deliver services between remote sites and centralized offices shall be the sole responsibility of the Contractor.

C.2 Equipment purchase, delivery, installation and test

Contractor is responsible to design the satellite network composed of remote site, teleport, and centralized office elements, and procure the equipment necessary to provide the specified services. Contractor will install all equipment according to Section E and conduct verification testing that demonstrates that functional and performance requirements are met. Purchaser shall be invited to witness any verification testing.

C.3   Service support and monitoring

Contractor shall maintain a central Network Operations Center on a 24-hour, 365-day per year basis. The NOC shall be in continuous contact with the remote site, teleports, and equipment installed at the centralized offices. The objective of this requirement is to assure that difficulties with equipment and services are detected before they cause service outages. Calls from the remote sites for help should be limited to those times when independent means of communication with the NOC fail.

C.4   Uninterrupted supply of services

Contractor shall supply the specified services between the remote sites and purchaser's centralized offices without interruption for the term of this contract. Any interruption will be subject to significant penalties and ultimate default of this Contract. Contractor shall not intentionally interrupt the services nor allow its third party suppliers to interrupt the service except in extreme and unusual circumstances, as covered in Section G.

D   Service Level Agreement

Purchaser and Contractor will enter into a Service Level Agreement (SLA) as part of the commitment of the Contractor to provide services during the term. The SLA is a contract that specifies, in measurable terms, what services the network will furnish and that they will meet contracted service standards.

Some metrics that the SLA will specify for each satellite service include:

- What percentage of the time services will be available
- The number of users that can be served simultaneously
- Specific performance benchmarks to which actual performance will be periodically compared
- The schedule for notification in advance of network changes that may affect users
- Help desk response time for various classes of problems
- Remote access capability from various locations
- Usage statistics that will be provided
- Satellite capacity usage

D.1   Satellite service

The satellite service SLA will cover use of satellite links for broadband communications between remote sites and any Purchaser centralized office.

D.2   Remote site installation

The remote site installation SLA will cover the equipment, including antennas, RF, and baseband equipment, and their proper operation as a system with the satellite links.

D.3   Teleport

It is the responsibility of the Contractor to enter into an approved SLA with teleport operator(s) so that the requirements of the SLA with Purchaser are satisfied.

D.4   Terrestrial connection services

It is the responsibility of the Contractor to enter into an approved SLA with providers of terrestrial connection services so that the requirements of the SLA with Purchaser are satisfied.

E   Work to be performed

Contractor is responsible for developing and maintaining all project work in accordance with schedules and plans to be reviewed and approved by Purchaser. The specific content in terms of equipment to be supplied, installation services to be performed, and test and verification of the installed equipment will be provided in a detailed accounting, broken down by item, cost, and duration (if appropriate). Once agreed, this listing cannot be changed by Contractor without the written approval of Purchaser.

Purchaser must be kept well informed of progress and potential difficulties during the development of the satellite system. Contractor will conduct design reviews with Purchaser at critical points in the project, at a location appropriate to the work being performed. The following design reviews will be held as a minimum:

Preliminary, within 30 days of Contract signing

Final, at least 30 days before scheduled delivery date

Acceptance, after completion of acceptance testing and before services can commence. Purchaser will be invited and may witness key tests, according to an agreed test plan

At the conclusion of a review, managers from Purchaser and Contractor shall summarize and agree upon action items to be tracked. The review will not have been deemed completed until all action items are closed by mutual agreement.

F   Preemption of Purchaser access to satellite capacity and teleports

Contractor or its subcontractors and agents shall not preempt Purchaser access to and use of satellite capacity or teleports, except in extreme, unusual, and unexpected circumstances, as covered in Section G. Preemption shall not occur for the convenience of the satellite operator in assigning transponders or satellites in order to reallocate their capacity or accommodate another customer.

G   Excusable delays and outages

Service is expected to be provided on a 24-hour, 365-day per year basis, without disruption or delay in provision. The Contractor shall take all means available to assure that this need is met. However, in certain cases, it may be impossible to satisfy this need, resulting in delay or outage. Excusable delays and outages are limited to the following:

- Hurricane
- Civil unrest or war
- Earthquake of 8 or greater magnitude in vicinity of critical facility used to provide service or equipment
- As directed by the U.S. government (FCC or other federal government agency) or competent foreign government pertaining to the location.

H   Payment terms and billing

Contractor and Purchaser will agree upon terms of payment during the installation and operation of the system. Contractor will invoice Purchaser, indicating full billing detail, and Purchaser will make payment in accordance with the Contract. Contractor and Purchaser will agree on the format of invoices.

I   Taxes and jurisdiction

The Contract shall be executed in the State of California, USA, and will be subject to California law. Taxes will be imposed and collected in accordance with California and U.S. federal law.

J   Deliverables

The Contract will identify all deliverable equipment, spare parts, documentation, training, and services throughout the term. The following delineates the general requirements and process for the Contractor to follow in making deliverables readily available to Purchaser.

J.1   Equipment and spare parts

Contractor will pre-assemble and test equipment at its factory or staging facilities before shipment to and installation at remote sites, teleports, or Purchaser centralized offices. This is intended to assure proper operation before committing to permanent installation on site.

Equipment will be new and free of defects after installation. Contractor is responsible for any warrantees, including their own as well as those of the original manufacturer.

Contractor shall provide sufficient spare parts for the duration of the Contract.

J.2   Documentation

Contractor will supply complete documentation for all equipment and in sufficient quantities to allow Purchaser to operate and maintain equipment at all locations. Manuals will be clear and readable, with sufficient drawing and numerical detail to facilitate their use in the field. Contact information, including current telephone numbers and e-mail addresses, shall be indicated; updates will be provided as soon as they become known to Contractor.

J.3   Training

Contractor is responsible to arrange for training of Purchaser staff as required to support the operation and maintenance of supplied equipment. This will include a 3-day overview of the system and its operation. Detailed courses by subcontractors and suppliers shall also be offered.

The above abstract is not actually a legal contract as this must be drafted by competent legal counsel and negotiated between the contractor and purchaser. However, it provides the key provisions that the contract would have to address. All of these factors have a direct impact on the quality and reliability of the network being provided under the contract.

13.4   The Marketing Organization

The fundamental role of marketing is to create the business and technical capabilities that attract good customers. From this, a sustainable business strategy is developed. As discussed in the next section, building a new market is difficult and expensive. Even if you are fortunate to have a solid business, you still must protect that position through continued marketing activity.

A successful marketing strategy is built on people who are familiar with the important elements of the business. Marketing teams should consist of professionals from different functional areas, such as sales, engineering, finance, operations, and promotion/public relations. For a satellite operator, the cost of creating and supporting such a team is small compared to the investment in space segment assets. Therefore, resources spent on excellence in this area will usually pay for itself in increased revenues. Moving toward value-added services, the cost of marketing and adding customers can exceed the value. For this reason, careful study of the total expense should be undertaken before commitments are made.

The following is a suggested list of the more significant costs of a strong marketing organization.

- *Marketing staff:* the multidisciplinary team that analyzes the markets and customer needs, identifies the right type of offering, and creates the infrastructure and systems needed to enter the particular market or segment.
- *Sales staff (may be part of marketing):* the high-energy people who take what the marketers have developed and expose it to live customers. In satellite communication, sales people may take on the principal customer-relations role but leave the technical and business dealing to other professionals with greater content knowledge.
- *Service demonstration capabilities:* this consists of temporary and permanent facilities that allow customers to experience the service capabilities without making a major commitment.
- *Travel and related expenses:* members of the marketing team must travel a significant percentage of time to maintain contact with customers, strategic partners, and regulatory bodies.

- *Advertising and supporting materials:* this is the promotion campaign to inform prospective customers of the availability of the service capability and to generate a positive impression of the organization.
- *Subcontractor arrangements:* any service or offering will involve components supplied by outside organizations.
- *Legal expenses:* these are to prepare and negotiate contracts and service agreements.
- *Tracking of revenue and costs:* success in the business is not guaranteed. The business activity must know where it is from a financial standpoint at all times.

The particular makeup of the marketing organization will depend on the special needs of the business and the structure of the organization. Due to the complexity of satellite communications and the unfamiliarity of many potential customers, members of the team must be conversant in the technology, economics, and potential customer base. By monitoring progress and making the appropriate adjustments, the best match to the market will be obtained.

13.5   Financing a Satellite System

A satellite system is a major capital commitment, representing a fixed asset with a given lifetime once the satellite is launched. Fortunately, a satellite has wide area coverage and is very versatile if it is not tied to a particular network configuration. The operating cost of the space segment of a satellite system is low relative to the capital. Almost any organization can acquire satellite technology because of the diffusion of expertise that has occurred over the decades since Early Bird was launched.

The cost of the ground network can equal or exceed that of the space segment. Capital and operations expenses are more in balance; that is, the total capital of a ground station project is approximately equal to the annual cost of operation, including return on investment. Fortunately, the ground segment operator can repair and upgrade ground facilities; the satellite, once launched, cannot effectively be modified. Users generally operate their own ground stations, reducing the capital costs (and risk) of satellite operators. Recently, global operators like Intelsat, NewSkies, and PanAmSat have developed networks of teleports to deliver end-to-end solutions and value-added services.

13.5.1   Elements of Capital Budgeting Analysis

The cost of implementing a satellite system is relatively easy to estimate because there are only a few items to consider, each of which is quite expensive. It is also relatively straightforward to estimate the annual financing and operating expenses. In some satellite systems, some revenue comes in lump sums from transponder purchases or prepaid leases, while other revenue comes over time from service charges and lease payments. The profit projection of the system is obtained by taking the difference between the revenue and the cost at each point in time. These can be

converted back into a present value of the profit by discounting the profit with a cost of money factor (typically in the range of 10% to 20% when considering U.S. dollars).

The following are some typical examples of each type of cost.

- Capital assets, also the sunk cost of the system, determined by adding the purchase price of equipment and facilities along with the cost of initiating the business. They include:
  - *First cost of system hardware:* this is usually specified in the satellite purchase contract and launch services agreement;
  - *Cost of facility construction or improvement:* the satellite operator may run this project internally or include it with the satellite purchase;
  - *Management and other expenses, tied to initiation of the service or business:* the preoperations period along with that immediately following launch maybe considered part of system construction.
- Annual cost factors are costs that will be expended each year in the running of the business. They include:
  - *Recovery of investment:* converting the capital cost into an annual cost by using the depreciation factor, which provides for the recovery of the original investment by the source of capital;
  - *Return on investment:* the cost of money and profit on the operation (e.g., 10% to 20%);
  - *Internal operating costs:* routine and nonroutine expenses incurred each year in the running of the service business;
  - *External operating costs:* money paid to outside entities, including other service providers for PSTN connectivity, maintenance of hardware and software, rent, and utilities;
  - *Marketing expense:* the money expended to acquire customers (discussed in Section 13.4);
  - *Overhead, general, and administrative expense:* the sum of expenses for running the overall business, which includes office rentals, secretarial support, human resources, procurement support, and corporate officers' salaries.

This background can be made clearer through an example. Say that we are considering investing in a Ku-band satellite to serve cable TV systems in a developing market. The cost of the satellite and launch are assumed to be $150 million, and the premium to insure it until on-station, at 20%, is $30 million. The investment in TT&C ground equipment might be another $10 million. For the sake of simplicity, we assume that this total investment of $190 million is all made in the year prior to launch. Say that we can operate the satellite for 1 year at a cost of $5 million (this could either be by our own staff or under an outsourcing agreement with another operator). The rest of our organization that gets involved with marketing the services and running the administrative part of the company might cost another $3 million per year. To make these annual expenses more conservative, we can assume that there is inflation of 5% per year.

On the revenue side of this example, let us assume that our satellite has 32 transponders and that we have been able to sell six of them for a total of $50 million at the

time of launch. From that point, our annual revenue from transponder leasing and other services (occasional and SCPC) add up to $15 million in the first year. Our owners have informed us that we must give them a return of 15% before tax on their initial investment (assuming that we have no debt).

Putting all of this together, we obtain the spreadsheet in Figure 13.3. To satisfy the 15% return requirement, the revenue growth rate from year to year must be at least 25% for there to be a positive net present value (NPV). This means that by the 10th year of operation, the annual revenue has grown from $15 million (not counting the initial transponder sales) to $112 million, increasing by a factor of 7.5.

This simple example is not far from the truth. The cost of ownership and operation of a satellite is easy to determine. The harder part is finding a satisfactory market and assuring adequate growth of revenues. It would be nice if our little satellite company could sell out the transponders in the first year, in which case we could keep our marketing expenses to a minimum. A real company might incur greater costs to find customers and to attract them over time. This is highly dependent on the degree of competition and the obstacles that customers face in using our services.

13.5.2 Sources of Capital for New Satellite Systems

The money for financing a satellite communications business can come from a wide variety of sources, both internal and external. Internal sources are within the corporate or governmental structure of the overall enterprise. If funding from internal sources is adequate, then there is no real need to approach capital markets. Obtaining these types of funds requires an understanding of the internal capital budgeting process and little else. Therefore, this section concentrates on external sources.

		Year:	1	2	3	4	5	6	7	8	9	10
Investment												
Satellite		80										
Launch		70										
Insurance		30										
TT&C ground		10										
TOTAL INVESTMENT		190										
Annual expenses												
Inflation rate	5%											
Satellite operation			5	5	6	6	6	6	7	7	7	8
Marketing			1	1	1	1	1	1	1	1	1	2
Administrative			2	2	2	2	2	3	3	3	3	3
TOTAL ANNUAL COST			8	8	9	9	10	10	11	11	12	12
Revenue												
Growth rate	25%											
Transponder sales			50									
Leases and services			15	19	23	29	37	46	57	72	89	112
TOTAL REVENUE			65	19	23	29	37	46	57	72	89	112
NET CASH			57	10	15	20	27	36	46	60	78	99
Discount rate	15%											
Discounted cash flow			50	8	10	11	13	15	17	20	22	25
Present value		191										
Net present value		1										

Figure 13.3 Example of a financial spreadsheet analysis for a hypothetical space segment consisting of one operating satellite.

The two possible external sources are debt and equity. Debt is nothing more than a direct loan from a bank or other institution that must be repaid with interest. Typical forms of debt include direct loans and bonds. The holder of the debt (the lender) is usually not an active participant in the business. Equity, on the other hand, represents a purchase of some percentage of ownership in the business, usually without a payment guarantee like the interest that applies in the case of debt. Preferred stock is a form of equity that includes a payment guarantee. Equity can be defined by privately held shares of stock in the corporation (or simply a percentage of the company's ownership) or publicly traded shares that can be resold on a public stock exchange somewhere in the world. Public shares can be bought and sold more or less freely, while private shares usually cannot.

Between the basic categories of debt and equity, there is a wide variety of external sources. Listed next are some of the more popular ones that have been used or attempted in the satellite communication industry from time to time. For any particular service business, the sources will be used or combined in innovative ways to optimize the cost of funds. Another important factor is the desire of the initial owners of the business to maintain adequate control, which is an issue with both categories of funds. With equity, there is a direct loss of control (called dilution) as new investors are granted their respective share of control in exchange for capital. The control that remains for the founders could be driven to zero. In the case of debt, the lender will almost always demand covenants in the debt instrument, the purpose of which is to preclude the borrower from weakening the financial position of the lender. These covenants would reduce the flexibility of future sales of debt or equity and could actually force the business to be run in a more conservative manner. Convertible debt, which is popular in the high-tech field, allows the lender to exchange debt for equity in the business. For a publicly traded satellite business, the lender may liquidate the debt by first converting debt to equity and selling the stock for cash. This may or may not be allowed under the terms of the loan agreement.

Forms of debt include:

- Bank loans;
- Government loans or loan support;
- Commercial bonds, low-yield, high-quality;
- Commercial bonds, high-yield, junk bonds;
- Convertible bonds and warrants;
- Project finance;
- Leases, sale/leaseback;
- Vendor financing, debt.

Forms of equity include:

- Stock, private placement;
- Stock, initial public offering;
- Limited partnership and limited liability company;
- Joint ventures, strategic partnership;
- Vendor financing, equity.

The most popular sources of funding for satellite ventures at the time of this writing are (1) bank loans with government guarantees; (2) initial public offerings (considerably reduced since the tech wreck of 2000); (3) vendor financing, both debt and equity based; (4) junk bonds; and (5) private placements from strategic partners. Large private financings of the 1990s include AMSC, a publicly traded company; Globalstar, both a limited partnership and publicly traded company; Iridium, AsiaSat, and XM Satellite Radio. Each arrived at their respective positions through a multistep process of using more than one of the available financing means. In some cases, it was intentional, and in others, it was a case of the best option for the next round. The attraction of going public—allowing investors to exit the business by cashing out—has waned of late and many service providers have had to operate as privately owned firms that rely on loans and private invested capital.

One way to approach the capital markets is to locate an experienced financial advisor. Many of the firms in this field are subsidiaries of major international banks, like Bank of America, Chase Manhattan, JP Morgan, and ING Bank. These companies may carry out a portion of the financing itself in return for a percentage fee, or they may charge directly for their services. It is a good idea to investigate several of these firms before making a selection.

13.5.3   Evaluating Venture Viability

A satellite communication venture must stand up to the same scrutiny applied to other business opportunities. The following are several of the more critical factors:

- Where will the revenue come from (who are the customers)? The research to obtain a credible market forecast can be substantial. Surveys, focus groups, pilot tests, and other market trials are good ways to get a handle on this challenge to the business.
- How will customers get their money? Demand is measured in buying power (cash and solid credit), not just interest. The Internet boom proved that giving a service away to gain "eyeballs" is not enough to establish a customer base.
- What services are you offering that they will find attractive? The service mix must be specific and something users understand and see of value.
- What are the blockages that will prevent your customers from using your service? Obstacles exist in the regulatory sphere as well as in distribution and gaining brand recognition.
- Who will provide parts of the service that you cannot? No provider can expect to be self-sufficient as many of the ingredients would be better provided by other companies. Still, it can be a challenge to find the right business associates and to tie them up adequately.
- How will you sell your service? This will depend on whether it is business-to-consumer (B2C) or business-to-business (B2B). These overworked terms nevertheless communicate the distinction between consumer markets and trade business. The former represents the largest demand in terms of an organization and cost of doing business. Companies like DIRECTV, EchoStar, Sky, and XM Satellite Radio have demonstrated good prowess in this area. The

B2B area requires much less in the way of investment and expense to reach customers; purchases tend to be larger and buyer-seller bonding stronger. Companies like PanAmSat and Stratos Global have built successful businesses by giving the customer what they want and making it affordable at the same time.

• If your market does not support the system financially, what is (are) your backup marketing plan(s)? Every attempt at new market entry is a learning experience—sellers profit from this learning by making the right changes in business strategy. Iridium would not or could not change from a strategy of focusing on the international business traveler to other segments in time to prevent bankruptcy. Inmarsat, on the other hand, found ways to grow out of the limited maritime mobile market to add business in the data communications field.

The technical and operational side of the system must also be investigated before giving approval. The following aspects should be well thought out ahead of time:

• A solid technical concept for the network with contained risk;
• Presale commitments for transponders or a credible marketing concept;
• Orbital positions or coordination of them;
• Startup money to carry through the first year;
• A contract or commitment from a spacecraft manufacturer or other integrated supplier (for a turn-key delivery);
• A launch commitment or reservation;
• An experienced organization to manage the construction and operation of the system.

In the 1970s, entering satellite communications meant buying an entire satellite system, including spacecraft, TT&C, and all of the Earth stations. Over the years, smaller network opportunities opened up through cable TV programming, video networks, syndication, data broadcasting, radio networks and music distribution, teleports, and private networks (business video and VSATs). These networks each require space segment capacity to function.

Satellite system operation is not typically a startup business for entrepreneurs (although the late Rene Anselmo, founder of PanAmSat, and EchoStar founder, Charlie Ergen, are notable exceptions). You need to have a solid business strategy in order to survive on a long-term basis. In addition, there must be an adequate source of funds over this term. Currently, gaining access to funds is an extreme challenge, even though new entrants have traditionally appeared from literally any quarter of the field.

13.6   Trends in Satellite Communications Business and Applications

On a global basis, the satellite communications industry brings in between $50 and $70 billion from the mix of satellite and ground equipment sales, communications and broadcasting services via satellite, and associated services for systems integration, content creation and management, and consulting. The industry only

represents a narrow slice of the overall telecommunications and mass media pie. There are physical limits on the number of satellites in orbit and the power they can generate in space. The resulting RF links cannot practically deliver services as pervasively as fiber and land-based wireless systems. The complementary nature of satellite and terrestrial services will, however, provide many areas where the relative merits can be blended very effectively. This evolution is suggested in the timeline in Figure 13.4. We conclude the chapter and this book with some comments on how satellite communications applications will develop further in response to needs and competition.

13.6.1   Broadband Applications to Mobile and Fixed Locations

The strategies and technologies covered throughout this book clearly demonstrate the versatility of satellite communications in the context of the Internet and digital video. The higher bandwidths now available from more powerful satellites and better performing VSATs and digital receivers are giving the industry a new boost. This, in turn, motivates the larger base of terrestrial operators to take steps to improve their offerings. The result is a push-pull effect on the satellite industry, which must continue to innovate to maintain its niche. Thus, VSAT terminals that only delivered hundreds of kilobits now routinely offer megabits per second of throughput. Direct integration with the Internet Protocol and MPEG make satellite data networks more effective than the current crop of terrestrial alternatives. The challenge remains to do this at a competitive cost, something the DTH TV has clearly demonstrated but where VSATs are still lagging. One hope, now largely lost (often due to low availabilities and abysmal system planning that ignored market realities) was that Ka-band satellites with multiple spot beams and onboard processing would so reduce the cost of bandwidth that a broadband service could compete on equal footing with DSL and cable modems. The two remaining developers of such networks in the United States, Wild Blue and Hughes SpaceWay, may produce working Ka-band systems and crack this particular nut. In the meantime, broadband VSATs at C- and Ku-bands ply the seas, fly in the air, and land anywhere with

1970s
Long distance
 telephone
Video backhaul

1980s
Video and audio
 distribution
Point-to-point data
Broadcast data
Private networks

1990s
Video broadcast
Interactive data
 and voice
Mobile
 communications
Alternate routes
 and backup

2000s
DTH services
Mobile (to all
 platforms for all
 applications)
Interactive
 multimedia
Emergency and
 temporary
 service

Figure 13.4   Evolutionary trends in the commercial satellite communications market.

high-speed access and backbone services that work well and overcome the digital divide.

13.6.2 Focus on Valuable Segments

Like any successful business, satellite applications need to focus on those segments where the technology provides a special fit. This principally is due to the lack of other telecommunication or broadcast infrastructure (such as postwar Iraq and the skies over the Atlantic Ocean). Mobility is also the key—an area where direct entry by fiber is unlikely. Wireless networks based on the GSM, WCDMA, and 802.11 standards fill vital needs, but do not reach everywhere on the planet. Small but valuable niches exist wherever people and organizations with money are found.

13.6.3 Satellites and the Digital Divide

Satellite applications excel in crossing the digital divide, as we define it. Rather than the social question of who can afford broadband Internet access, we concern ourselves with reaching areas and locations where terrestrial networks fail to reach at all. Running fiber or microwave links to thousands of islands in an archipelago nation like Indonesia or throughout mountainous territory of China, Afghanistan, and Colombia is simply impractical—no commercial organization is going to do this. Likewise, a small population on a Pacific or Indian Ocean island cannot expect to receive service from a traditional telephone company or TV broadcaster. For them, the satellite is the best means of delivering services. The future portends exciting times with new applications and new businesses. Perhaps you, the reader, will start one.

Reference

[1]   Norris, M., *Survival in the Software Jungle*, Norwood, MA: Artech House, 1995.

About the Author

Bruce R. Elbert is a leading authority on satellites and telecommunications, with more than 30 years of working experience as a consultant and executive. In addition to his international industrial background, Mr. Elbert is a prolific writer of books and articles and has been featured as a speaker at numerous conferences and seminars. He is the editor of two Artech House series: Space Technology and Applications, and Technology Management and Professional Development. Mr. Elbert is an adjunct professor at the University of Wisconsin–Madison and an instructor at UCLA Extension, specializing in satellite communications and information technology. He has a B.S. and an M.S. in electrical engineering from City College of New York and the University of Maryland, respectively, and an M.B.A. from Pepperdine University. He may be contacted by e-mail at bruce@applicationstrategy.com or through his company's Web site at http://www.applicationstrategy.com.

Index