TECHNICKÁ UNIVERZITA
V KOŠICIACH

- VÝZVA pre tvoju BUDÚCNOSŤ

Moderné vzdelávanie pre vedomostnú spoločnosť/
Projekt je spolufinancovaný zo zdrojov EÚ

SATELITNÉ TECHNOLÓGIE A SLUŽBY
SATELLITE TECHNOLOGY AND SERVICES

Fakulta elektrotechniky a informatiky

Stanislav Marchevský

Európska únia
Európsky sociálny fond

Agentúra
Ministerstva školstva, vedy, výskumu a športu SR
pre štrukturálne fondy EÚ

Operačný program
VZDELÁVANIE

Táto publikácia vznikla za finančnej podpory z **Európskeho sociálneho fondu** v rámci Operačného programu **VZDELÁVANIE.**

Prioritná os 1 Reforma vzdelávania a odbornej prípravy

Opatrenie 1.2 Vysoké školy a výskum a vývoj ako motory rozvoja vedomostnej spoločnosti.

NÁZOV: Satelitné technológie a služby
AUTORI: prof. Ing. Stanislav Marchevský,CSc.
RECENZENTI: prof. Ing. Dušan Repčík, CSc., doc. Ing. Václav Moucha, CSc.
VYDAVATEĽ: Technická univerzita v Košiciach
ROK: 2015
ROZSAH: 235 strán
NÁKLAD: 50 ks
VYDANIE: prvé
ISBN: 978-80-553-1983-4

Rukopis neprešiel jazykovou úpravou.

Za odbornú a obsahovú stránku zodpovedajú autori.

Abstrakt v SJ

Publikácia s názvom Satelitné technológie a služby: je stručná ale pritom komplexná e-learningová učebnica s témou satelitných technológií a vybraných aplikácií. Venuje sa pomerne veľkému rozsahu problémov satelitných komunikačných technológií a satelitnej navigácie. Cieľom publikácie bolo predstaviť predmet jasne a výstižne s kľúčovými názornými príkladmi.

Abstrakt v AJ

Satellite Technology and Services: is a concise and yet comprehensive e-learning book on the subject of satellite technology and its applications, covering the relaively large area of satelite communications and navigation technology. This book spans the entire breadth of satellite-enabled end user applications. The aim has been to present the subject matter in a clear and concise manner with key illustrative examples.

Foreword

One of the important components of a broadband communication system is the satellite link, the other being optical. Satellite communication has become the backbone of long distance communication irrespective of geographical conditions. Satellites have passed the age when their use was restricted to outer space experiments and remote sensing. Today many satellites are multipurpose satellites which are used forcommunication, meteorological data collection, search and rescue, global positioning systems, mineral and oil exploration etc. Satellite communication has transformed the world into a "global village".

A satellite system represents one of the most sophisticated and intriguing systems to design. Engineers need to consider almost all aspects of applied sciences, engineering and technology: They apply the principles of a variety of scientific disciplines such as physics of materials, sensor technology, virtual instrumentation, communication engineering, automatic control systems, mechanics of structures and so on.

.After teaching this subject for nearly a decade I found that there is a need for a structured textbook which would cover the subject comprehensiyely so as to im part to the reader, full command of the basie concepts and enough of application-oriented knowledge to design satelite links. Though there are a number of books on satellite communications with excellent contents (as referred here in the bibliography) the exact requirement of the students remains to be fully met. The objective of this book is to fulfill their need. In this book I have tried my best to present the subject in a simple way, yet convey all the important aspects of space and satellite communication. This book is suitable for a one-semester course „Satellite technology and services“ and is written keeping in mind both, undergraduate and postgraduate students. Practising engineers in this field can also refer this book. This book covers, besides the basie concepts of satellite system; important parameter calculations and design concepts. The emphasis is on geostationary satellites. Beginning with orbiting parameters, the book gradually progresses to the more advanced coricepts finally detailing the design of a complete multiple access links and earth station.

Obsah

List of figures

Figure 8.36 System to resolve ambiguity in phase comparison monopulse tracking technique

List of tables

List of Acronyms

A	
ACM	adaptive coded modulation
ACTS	Advanced Communications Technology Satellite
A/D	analog to digital converter
ADM	adaptive delta modulation
ADPCM	adaptive differential pulse code modulation
AGARD	Advisory Group for Aeronautical Research and Development (NATO)
AIAA	American Institute of Aeronautics and Astronautics
AM	amplitude modulation
AMI	alternate mark inversion
AMSS	aeronautical mobile satellite service
AOCS	Attitude and Orbit Control System
ATS-	Applications Technology Satellite-
AWGN	additive white Gaussian noise
Az	azimuth (angle)
B	
BB	baseband
BER	bit error rate
BFSK	binary frequency shift keying
BO	backoff
BOL	beginning of life
BPF	band pass filter
BPSK	binary phase shift keying
BSS	broadcast satellite service
C	
CBR	carrier and bit-timing recovery
CDC	coordination and delay channel

CDF	cumulative distribution function
CDMA	code division multiple access
CEPIT	Coordinamento Esperimento Propagazione Italsat
CEPT	European Conference of Postal and Telecommunications Administrations
C/I	carrier to interference ratio
CLW	cloud liquid water
cm	centimeters
C/N	carrier-to-noise ratio
C/No	carrier-to-noise density
COMSAT	Communications Satellite Corporation
CONUS	continental United States
CPA	copolar attenuation
CRC	cyclic redundancy check
CSC	common signaling channel
CTS	Communications Technology Satellite
CVSD	continuously variable slope delta modulation
D	
D.C.	down converter
DA	demand assignment
DAH	Dissanayake, Allnutt, and Haidara (rain attenuation model)
DAMA	demand assigned multiple access
dB	decibel
dBHz	decibel-Hz
dbi	decibels above isotropic
dBK	decibel-Kelvin
dBm	decibel-milliwatts
dBW	decibel-watt
DEM	demodulator
DOS	United States Department of State

DS	digital signaling (also known as T-carrier TDM signaling)
DSB/SC	double sideband suppressed carrier
DSI	digital speech interpolation
DS-SS	direct sequence spread spectrum
E	
Eb/No	energy per bit to noise density
EHF	extremely high frequency
EIRP	effective isotropic radiated power
El	elevation angle
EOL	end of life
erf	error function
erfc	complimentary error function
ERS	empirical roadside shadowing
ES	earth station
ESA	European Space Agency
E-W	east-west station keeping
F	
FA	fixed access
FCC	Federal Communications Commission
FDM	frequency division multiplex
FDMA	frequency division multiple access
FEC	forward error correction
FET	field effect transistor
FH-SS	frequency hopping spread spectrum
FM	frequency modulation
FSK	frequency shift keying
FSS	fixed satellite service
FT	frequency translation transponder

G	
GEO	geostationary satellite orbits
GHz	gigahertz
GSO	geosynchronous satellite
G/T	receiver figure of merit
H	
HEO	high elliptical earth orbit, high earth orbit
HEW	Health Education Experiment
HF	high frequency
HP	horizontal polarization
hPa	hectopascal (unit for air pressure, equal to 1 cm H2O)
HPA	high power amplifier
Hz	hertz
I	
IEE	Institute of Electrical Engineers
IEEE	Institute of Electrical and Electronics Engineers
IF	intermediate frequency
INTELSAT	International Satellite Organization
ISI	intersymbol interference
ITU	International Telecommunications Union
ITU-D	International Telecommunications Union, Development Sector
ITU-R	International Telecommunications Union, Radiocommunications Sector
ITU-T	International Telecommunications Union, Telecommunication Standards Sector
J	
K	
K	degrees Kelvin
Kbps	kilobits per second
kg	kilogram

KHz	kilohertz
km	kilometers
L	
LEO	low earth orbit
LF	low-frequency
LHCP	left hand circular polarization
LMSS	land mobile satellite service
LNA	low noise amplifier
LNB	low noise block
LO	local oscillator
LPF	low pass filter
M	
m	meters
MA	multiple access
MAC	medium access control
Mbps	megabits per second
MCPC	multiple channel per carrier
MEO	medium earth orbit
MF	medium frequency
MF-TDMA	multi-frequency time division multiple access
MHz	megahertz
MKF	street masking function
MMSS	maritime mobile satellite service
MOD	modulator
MODEM	modulator/demodulator
MSK	minimum shift keying
MSS	mobile satellite, service
MUX	multiplexer
N	

NASA	National Aeronautics and Space Administration
NF	noise figure (or noise factor)
NGSO	non geosynchronous (or geostationary) satellite orbits
NIC	nearly instantaneous companding
NRZ	non return to zero
N-S	north-south station keeping
NTIA	National Telecommunications and Information Agency
NTSC	National Television System Committee
O	
OBP	on-board processing transponder
OFDM	orthogonal frequency division multiplexing
OOK	on/off keying
P	
PA	pre-assigned access
PAL	phase alternation line
PAM	pulse amplitude modulation
PCM	pulse code modulation
PFD	power flux density
PLACE	Position Location and Aircraft Communication Experiment
PN	pseudorandom sequence
PSK	phase shift keying
PSTN	public switched telephone network
Q	
QAM	quadrature amplitude modulation
QPSK	quadrature phase shift keying
R	
REC	receiver
RF	radio frequency
RFI	radio frequency interference

RHCP	right hand circular polarization
RZ	return to zero
S	
SC	service channel
SCORE	Signal Communications Orbiting Relay Experiment
SCPC	single channel per carrier
SDMA	space division multiple access
SECAM	SEquential Couleur Avec Memoire
SGN	satellite news gathering
SHF	super high frequency
SITE	satellite instructional television experiment
S/N	signal-to-noise ratio
SS	subsatellite point
SS/TDMA	time division multiple access, satellite switched
SSB/SC	single sideband suppressed carrier
SSPA	solid state amplifier
SYNC	synchronization
T	
TDM	time division multiplex(ing)
TDMA	time division multiple access
TDRS	Tracking and Data Relay Satellite
TEC	total electron content
T-R	transmitter-receiver
TRANS	transmitter
TRUST	Television Relay Using Small Terminals
TT&C	tracking, telemetry and command
TTC&M	tracking, telemetry, command and monitoring
TTY	teletype
TWT	traveling wave tube

TWTA	traveling wave tube amplifier
U	
UHF	ultra high frequency
USSR	Union of Soviet Socialist Republics
UW	unique word
V	
VA	voice activation (factor)
VF	voice frequency (channel)
VHF	very high frequency
VLF	very low frequency
VOW	voice order wire
VP	vertical polarization
VPI&SU	Virginia Polytechnic Institute and State University
VSAT	very small antenna (aperture) terminal
W	
WVD	water vapor density
X	
XPD	cross-polarization discrimination
Y	
Z	

Glossary

Apogee: Point on the satellite orbit farthest from the centre of the Earth. The apogee distance is the distance of the apogee point from the centre of the Earth

Argument of perigee: This parameter defines the location of the major axis of the satellite orbit. It is measured as the angle between the line joining the perigee and the centre of the Earth and the line of nodes from the ascending to the descending node in the same direction as that of the satellite orbit

Ascending node: The point where the satellite orbit cuts the Earth's equatorial plane, when it passes from the southern hemisphere to the northern hemisphere

Azimuth angle – Earth station: The azimuth angle of an Earth station is the angle produced by the line of intersection of the local horizontal plane and the plane passing through the satellite, Earth station and centre of the Earth with the true north

De-spun Antenna: An antenna system placed on a platform that is spun in a direction opposite to the direction of spin of the satellite body. This ensures a constant pointing direction for the satellite antenna system

Earth coverage: Surface area of the Earth that can possibly be covered by a satellite

Eclipse: An eclipse is said to occur when sunlight fails to reach the satellite's solar panel due to an obstruction from a celestial body. The major and most frequent source of an eclipse is due to the satelite coming in the shadow of Earth, known as the solar eclipse. Another type of eclipse known as the lunar eclipse occurs when the moon's shadow passes across the satellite

Elevation angle – Earth station: The elevation angle of an Earth station is the angle produced by the line of intersection of the local horizontal plane and the plane passing through the satellite, Earth station and centre of the Earth with the line joining the Earth station and the satellite

Footprint: Same as Earth coverage Ground track: This is an imaginary line formed by the locus of the lowest point on the surface of the Earth. The lowest point is the point formed by the projection of the line joining the satellite with the centre of the Earth on the surface of the Earth

Slant range: The line-of-sight distance between the satellite and the Earth station

Spin stabilization: A technique for stabilizing the attitude of a satellite in which the satellite body is spun around an axis perpendicular to the orbital plane. Like a spinning top, the spinning satellite body offers inertial stiffness, thus preventing the satellite from drifting from its desired orientation

Station keeping: Station keeping is the process of maintenance of the satellite's attitude against different factors that cause temporal drift

Three-axis stabilization: Also known as body stabilization, a technique for stabilizing the attitude of a satellite in which stabilization is achieved by controlling the movement of the satellite along the three axes, i.e. yaw, pitch and roll, with respect to a reference

Apogee: Point on the satellite orbit farthest from the centre of the Earth. The apogee distance is the distance of the apogee point from the centre of the Earth

Argument of perigee: This parameter defines the location of the major axis of the satellite orbit. It is measured as the angle between the line joining the perigee and the centre of the Earth and the line of nodes from the ascending to the descending node in the same direction as that of the satellite orbit

Ascending node: The point where the satellite orbit cuts the Earth's equatorial plane, when it passes from the southern hemisphere to the northern hemisphere

Centrifugal force: The force acting outwards from the centre of the Earth on any body orbiting it

Centripetal force: A force that is directed towards the centre of the Earth due to the gravitational force of attraction of Earth

Descending node: The point where the satellite orbit cuts the Earth's equatorial plane, when it passes from the northern hemisphere to the southern hemisphere

Eccentricity: Referring to an elliptical orbit, it is the ratio of the distance between the centre of the Earth and the centre of the ellipse to the semi-major axis of the ellipse. It is zero for a circular orbit and between 0 and 1 for an elliptical orbit

Equatorial orbit: An orbit in which the satellite's orbital plane coincides with the Earth's equatorial plane

Equinox: An equinox is said to occur when the angle of inclination of the Earth's equatorial plane with respect to the direction of the sun is zero. Such a situation occurs twice a year, one on 21 March called the spring equinox and the other on 21 September called the autumn equinox

First cosmic velocity: This is the injection velocity at which the apogee and perigee distances are equal, with the result that the satellite orbit is circular

Geostationary Earth orbit (GEO): A satellite orbit with an orbit height at 35 786 km above the surface of the Earth. This height makes the satellite's orbital velocity equal to the speed of rotation of Earth, thus making the satellite look stationary from a given point on the surface of the Earth

Inclination: Inclination is the angle that the orbital plane of the satellite makes with the Earth's equatorial plane

Inclined orbit: An orbit having an angle of inclination between 0◦ and 180◦

Injection velocity: This is the horizontal velocity with which a satellite is injected into space by the launch vehicle with the intention of imparting a specific trajectory to the satellite

Kepler's first law: The orbit of an artificial satellite around Earth is elliptical with the centre of the Earth lying at one of its foci

Kepler's second law: The line joining the satellite and the centre of the Earth sweeps out equal areas in the plane of the orbit in equal times

Kepler's third law: The square of the time period of any satellite is proportional to the cube of the semi-major axis of its elliptical orbit

Low Earth orbit (LEO): A satellite orbit with an orbital height of around 150 km to 500 km above the surface of Earth. These orbits have lower orbital periods, shorter propagation delays and lower propagation losses

Medium Earth orbit (MEO): A satellite orbit with an orbital height around 10 000 km to 20 000 km above the surface of the Earth

Molniya orbit: A highly inclined and eccentric orbit used by Russia and other countries of the erstwhile Soviet Union for providing communication services

Orbit: A trajectory that is periodically repeated

Perigee: A point on a satellite orbit closest to the centre of the Earth. The perigee distance is the distance of the perigee point from the centre of the Earth

Polar orbit: An orbit having an angle of inclination equal to 90◦

Prograde orbit: Also called a direct orbit, an orbit where the satellite travels in the same direction as the direction of rotation of Earth. This orbit has an angle of inclination between 0◦ and 90◦

Project Iridium: Project Iridium is a global communication system conceived by Motorola that makes use of satellites in low Earth orbits. A total of 66 satellites are arranged in a distributed architecture with each satellite carrying 1/66 of the total system capacity

Retrograde orbit: An orbit where the satellite travels in a direction opposite to the direction of rotation of Earth. This orbit has an angle of inclination between 90◦ and 180◦

Right ascension of the ascending node: The right ascension of the ascending node indicates the orientation of the line of nodes, which is the line joining the ascending and descending nodes, with respect to the direction of the vernal equinox. It is expressed as an angle (□) measured from the vernal equinox towards the line of nodes in the direction of rotation of Earth. The angle could be anywhere from 0◦ to 360◦

Second cosmic velocity: This is the injection velocity at which the apogee distance becomes infinite and the orbit takes the shape of a parabola. It equals√2 times the first cosmic velocity

Solstices: Solstices are said to occur when the angle of inclination of the Earth's equatorial plane with respect to the direction of the sun is at its maximum, i.e. 23.4◦. These are like equinoxes and also occur twice during the year, one on 21 June called the summer solstice and the other on 21 December called the winter solstice

Sun-synchronous orbit: A sun-synchronous orbit, also known as a helio-synchronous orbit, is one that lies in a plane that maintains a fixed angle with respect to the Earth–sun direction

Third cosmic velocity: This is the injection velocity at which the satellite succeeds in escaping from the solar system. It is related to the motion of Earth around the sun. For injection velocities beyond the third cosmic velocity, there is a region of hyperbolic flights outside the solar system

Trajectory: A path traced by a moving body

True anomaly of a satellite: This parameter is used to indicate the position of the satellite in its orbit. This is done by defining an angle (θ), called the true anomaly of the satellite, formed by the line joining the perigee and the centre of the Earth with the line joining the satellite and the centre of the Earth

Preface

The word ‘satellite’ is a household name today. It sounds very familiar to all of us irrespective of our educational and professional background. It is no longer the prerogative of a few select nations and is not a topic of research and discussion that is confined to the premises of big academic institutes and research organizations. Today, it is not only one of the main subjects taught at undergraduate, graduate and postgraduate level; it is the bread and butter for a large percentage of electronics, communications and IT professionals working for academic institutes, science and technology organizations and industry. Most of the books on satellite technology and its applications cover only communications-related applications of satellites, with either occasional or no reference to other important applications, which include remote sensing, weather forecasting, scientific, navigational and military applications. Also, space encyclopedias mainly cover the satellite missions and their applications with not much information on the technological aspects.

Satellite Technology and Services: is a concise and yet comprehensive e-learning book on the subject of satellite technology and its applications, covering the relaively large area of satelite communications and navigation technology. This book spans the entire breadth of satellite-enabled enduser applications. The aim has been to present the subject matter in a clear and concise manner with key illustrative examples.

After an introductory chapter, the book presents fundamental concepts applicable generally across all the systems. Subsequent chapters delve into techniques and examples of specific systems and services available directly from personal satellite terminals. Such applications encompass broadcasting, communications (narrowband and wideband, commercial, military and amateur), navigation and satellite-based distress services.

1 Introduction to satelite communications

A communications satellite is an orbiting artificial earth satellite that receives a communications signal from a transmitting ground station, amplifies and possibly processes it, then transmits it back to the earth for reception by one or more receiving ground stations. Communications information neither originates nor terminates at the satellite itself. The satellite is an active transmission relay, similar in function to relay towers used in terrestrial microwave communications.

The commercial satellite communications industry has its beginnings in the mid-1960s, and in less than 50 years has progressed from an alternative exotic technology to a mainstream transmission technology, which is pervasive in all elements of the global telecommunications infrastructure. Today's communications satellites offer extensive capabilities in applications involving data, voice, and video, with services provided to fixed, broadcast, mobile, personal communications, and private networks users. Satellite communications are now an accepted fact of everyday life, as evidenced by the antennas or 'dishes' that dot city and country horizons, or the nearly instantaneous global news coverage that is taken for granted, particularly in times of international crises. The communications satellite is a critical element in the overall telecommunications infrastructure, as represented by Figure 1.1, which highlights, by the shaded area, the communications satellite component as related to the transmission of information. Electronic information in the form of voice, data, video, imaging, etc., is generated in a user environment on or near the earth's surface. The information's first node is often a terrestrial interface, which then directs the information to a satellite uplink, which generates an RF (radio frequency) radiowave that propagates through the air link to an orbiting satellite (or satellites). The information bearing radiowave is amplified and possibly processed at the satellite, then reformatted and transmitted back to a receiving ground station through a second RF radiowave propagating through the air link. Mobile users, indicated by the vehicle and handheld phone on the figure, generally bypass the terrestrial interface only for direct mobile-to-mobile communications.

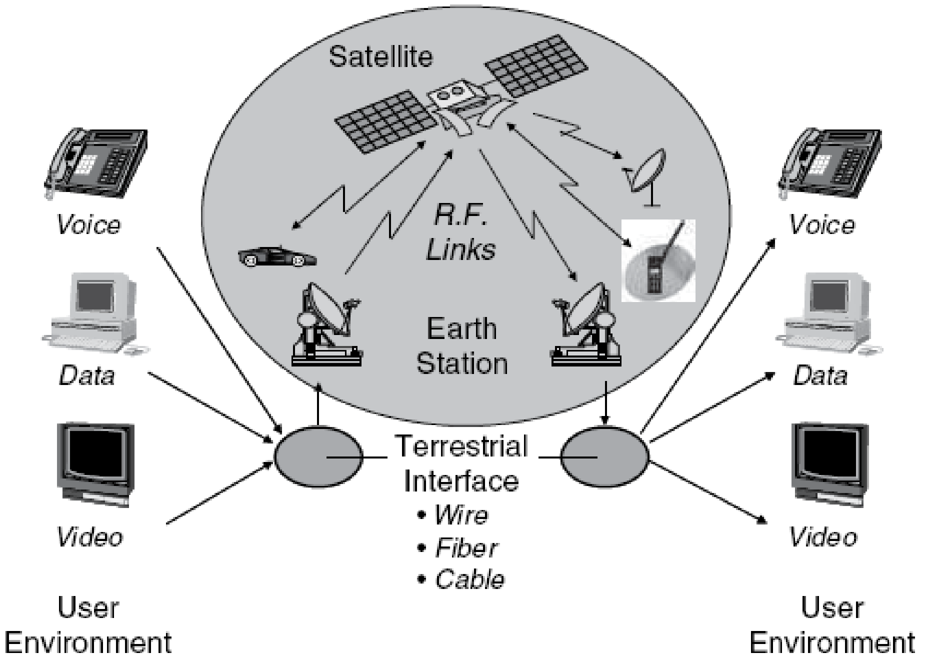


Figure 1.1 Communications via satellite in the telecommunications infrastructure

Communications by satellite offers a number of features that are not readily available with alternative modes of transmission, such as terrestrial microwave, cable, or fiber networks. Some of the advantages of satellite communications are:

• *Distance Independent Costs.* The cost of satellite transmission is basically the same, regardless of the distance between the transmitting and receiving earth stations. Satellite based transmission costs tend to be more stable, particularly for international or intercontinental communications over vast distances.

• *Fixed Broadcast Costs.* The cost of satellite broadcast transmission, that is, transmission from one transmit ground terminal to a number of receiving ground terminals, is independent of the number of ground terminals receiving the transmission.

• *High Capacity. Satellite* communications links involve high carrier frequencies, with large information bandwidths. Capacities of typical communications satellites range from 10s to 100s of Mbps (Mega-bits per second), and can provide services for several hundred video channels or several tens of thousands of voice or data links.

• *Low Error Rates.* Bit errors on a digital satellite link tend to be random, allowing statistical detection and error correction techniques to be used. Error rates of one bit error in 106 bits or better can be routinely achieved efficiently and reliably with standard equipment.

- *Diverse UserNetworks.* Large areas of the earth are visible from the typical communications satellite, allowing the satellite to link together many users simultaneously. Satellites are particularly useful for accessing remote areas or communities not otherwise accessible by terrestrial means. Satellite terminals can be on the surface, at sea, or in the air, and can be fixed or mobile.

The successful implementation of satellite wireless communications requires robust air links providing the uplink and downlink paths for the communications signal. Transmission through the atmosphere will degrade signal characteristics however, and under some conditions it can be the major impediment to successful system performance.Adetailed knowledge of the types of atmospheric effects that impact satellite communications and the means to predict and model them for application to communications link design and performance is essential for wireless satellite link engineering. The effects of the atmosphere are even more significant as current and planned satellites move up to higher operating frequencies, including the Ku-band (14 GHz uplink/12 GHz downlink), Ka-band (30 GHz/20 GHz), and V-band (50 GHz/40 GHz), where the effects of rain, gaseous attenuation, and other effects will increase.

1.1 Some Basic Communications Satellite System Definitions

This section provides some of the basic definitions and parameters used in the satellite communications industry, which will be used throughout the book in the evaluation and analysis of satellite communications systems design and performance

1.1.1 Satellite Communications Segments

We begin with the communications satellite portion of the communications infrastructure, shown by the shaded oval in Figure 1.1. The satellite communications portion is broken down into two areas or segments: the space segment and the ground (or earth) segment.

1.1.1.1 Space Segment

The elements of the space segment of a communications satellite system are shown on Figure 1.2. The space segment includes the satellite (or satellites) in orbit in the system, and the ground station that provides the operational control of the satellite(s) in orbit. The ground station is variously referred to as the Tracking, Telemetry, Command (TT&C) or the Tracking, Telemetry, Command and Monitoring (TTC&M) station. The TTC&M station provides essential spacecraft management and control functions to keep the satellite operating safely in orbit. The TTC&M links between the spacecraft and the ground are usually separate from the user communications links. TTC&M links may operate in the same frequency bands or in other bands. TTC&M is most often accomplished through a separate earth terminal facility specifically designed for the complex operations required to maintain a spacecraft in orbit.

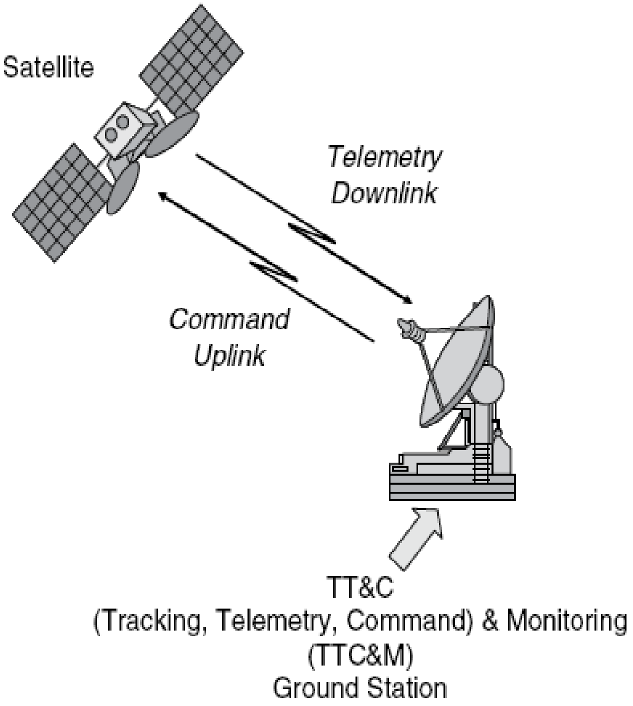


Figure 1.2 The space segment for a communications satellite network

1.1.1.2 Ground Segment

The ground segment of the communications satellite system consists of the earth surface area based terminals that utilize the communications capabilities of the Space Segment. TTC&M ground stations are not included in the ground segment. The ground segment terminals consist of three basic types:

- fixed (in-place) terminals;
- transportable terminals;
- mobile terminals.

Fixed terminals are designed to access the satellite while fixed in-place on the ground. They may be providing different types of services, but they are defined by the fact that they are not moving while communicating with the satellite. Examples of fixed terminals are small terminals used in private networks (VSATs), or terminals mounted on residence buildings used to receive broadcast satellite signals.

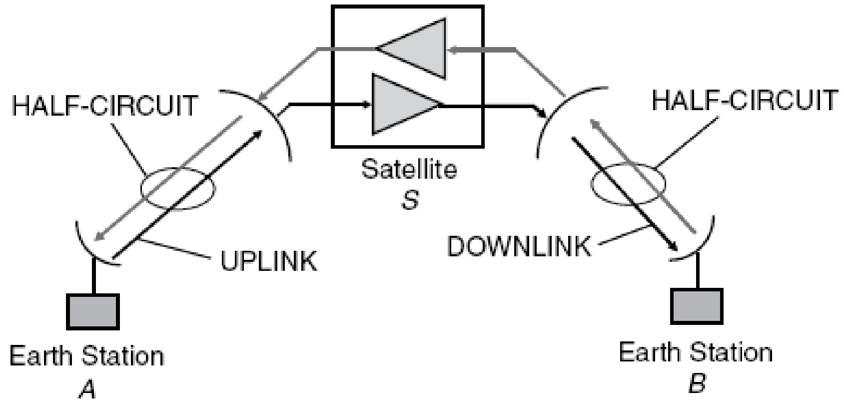
Transportable terminals are designed to be movable, but once on location remain fixed during transmissions to the satellite. Examples of the transportable terminal are satellite news gathering (SGN) trucks, which move to locations, stop in place, and then deploy an antenna to establish links to the satellite.

Mobile terminals are designed to communicate with the satellite while in motion. They are further defined as land mobile, aeronautical mobile, or maritime mobile, depending on their locations on or near the earth surface.

1.1.2 Satellite Link Parameters

The communications satellite link is defined by several basic parameters, some used in traditional communications system definitions, others unique to the satellite environment. Figure 1.3 summarizes the parameters used in the evaluation of satellite communications links. Two oneway free-space or air links between Earth Stations A and B are shown. The portion of the link from the earth station to the satellite is called the uplink, while the portion from the satellite to the ground is the down Note that either station has an uplink and a downlink. The electronics in the satellite that receives the uplink signal, amplifies and possibly processes the signal, and then reformats and transmits the signal back to the ground, is called the transponder, designated by the triangular amplifier symbol in the figure (the point of the triangle indicates the direction of signal transmission). Two transponders are required in the satellite for each twoway link between the two ground stations as shown. The antennas on the satellite that receive and transmit the signals are usually not included as a part of the transponder electronics – they are defined as a separate element of the satellite payload.

A channel is defined as the one-way total link from A-to-S-to-B, OR the link from B-to-S-to-A. The duplex (two-way) links A-to-S-to-B AND B-to-S-to-A establish a circuit between the two stations. A half-circuit is defined as the two links at one of the earth stations, that is A-to-S AND S-to-A; OR B-to-S AND S-to-B. The circuit designations are a carry-over from standard telephony definitions, which are applied to the satellite segment ofthe communications infrastructure.



CHANNEL – one way link from $A \rightarrow B$ or $B \rightarrow A$
CIRCUIT – full duplex link – $A \rightleftarrows B$
HALF CIRCUIT – two way link – $A \rightleftarrows S$ or $S \rightleftarrows B$
TRANSPONDER – ▷ basic satellite repeater electronics, usually one channel

Figure 1.3 Basic link parameters in the communications satellite link

1.1.3 Satellite Orbits

The characteristics of satellite orbits in common use for a vast array of satellite communications services and applications are discussed in detail in Chapter x. We introduce here the satelite orbit terms for the four most commonly used orbits in satellite communications, shown in Figure 1.4. The basic orbit altitude(s) and the one-way delay times are shown for each orbit,along with the common abbreviation designations.

1.1.3.1 Geosynchronous Orbit (GSO or GEO)

The GSO orbit is by far the most popular orbit used for communications satellites. A GSO satellite is located in a circular orbit in the equatorial plane, at a nominal distance of 36 000 km at a stable point, which maintains the satellite at a fixed location in the sky. This is a tremendous advantage for satellite communications, because the

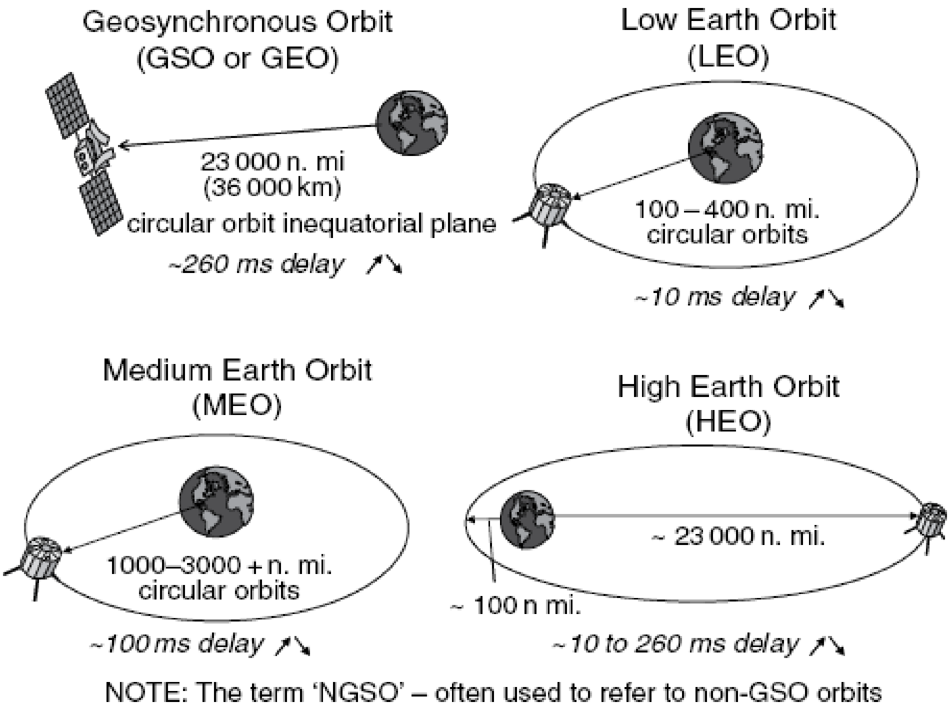


Figure 1.4 Satellite orbits

pointing direction remains fixed in space and the ground antenna does not need to track a moving satellite. A disadvantage of the GSO is the long delay time of ∼260 ms, which can affect network synchronization or impact voice communications.

1.1.3.2 Low Earth Orbit (LEO)

The second most common orbit is the low earth orbit (LEO), which is a circular orbit nominally 160 to 640 km above the earth. The delay is low, ∼10 ms, however the satellite moves across the sky, and the ground station must actively track the satellite to maintain communications.

1.1.3.3 Medium Earth Orbit (MEO)

The MEO is similar to the LEO, however the satellite is in a higher circular orbit – 1600 to 4200 km. It is a popular orbit for navigation satellites such as the GPS constellation

1.1.3.4 High Earth Orbit (HEO)

The HEO is the only non-circular orbit of the four types. It operates with an elliptical orbit, with a maximum altitude (apogee) similar to the GSO, and a minimum altitude (perigee) similar to the LEO. Satellite orbits that are not synchronous, such as the LEO, MEO, or HEO, are often referred to as non-geosynchronous orbit (NGSO) satellites.

1.1.4 Frequency Band Designations

The frequency of operation is perhaps the major determining factor in the design and performance of a satellite communications link. The wavelength of the free space path signal is the principal parameter that determines the interaction effects of the atmosphere, and the resulting link path degradations. Also, the satellite systems designer must operate within the constraints of international and domestic regulations related to choice of operating free space path frequency.

Two different methods of designation have come into common use to define radio frequency bands. Letter band designations, derived from radar applications in the 1940s, divide the spectrum from 1 to 300 GHz into eight bands with nominal frequency ranges, as shown on Figure 1.5.The K-band is further broken down intoKU-band (K-lower) andKA-band (K-above).

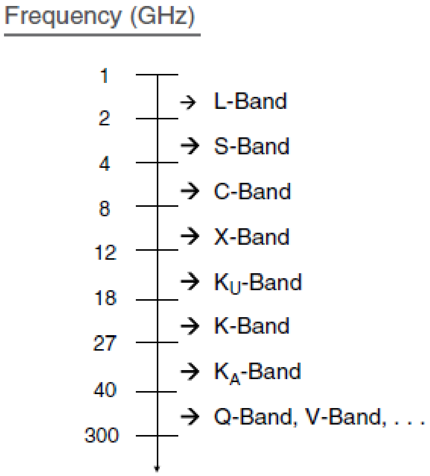


Figure 1.5 Letter band frequency designations

The boundaries of the bands are not always followed, and often some overlap is observed. For example, some references consider C-band as 3.7–6.5 GHz andKu-band as 10.9–12.5 GHz. The bands above 40 GHzhave seen several letter designations used, including Q-band,W-band, U-band, and W-band. The ambiguity in letter band designations suggests that they should be used with caution – particularly when the specific frequency is an important consideration.

A second designation divides the spectrum from 3 Hz to 300 GHz into bands based on decade steps of nominal wavelength, as shown in Figure 1.6. This designation is less ambiguous than the letter designation, however, as we shall see in later chapters, most satellite communications links operate within only three or four of the bands, VHF

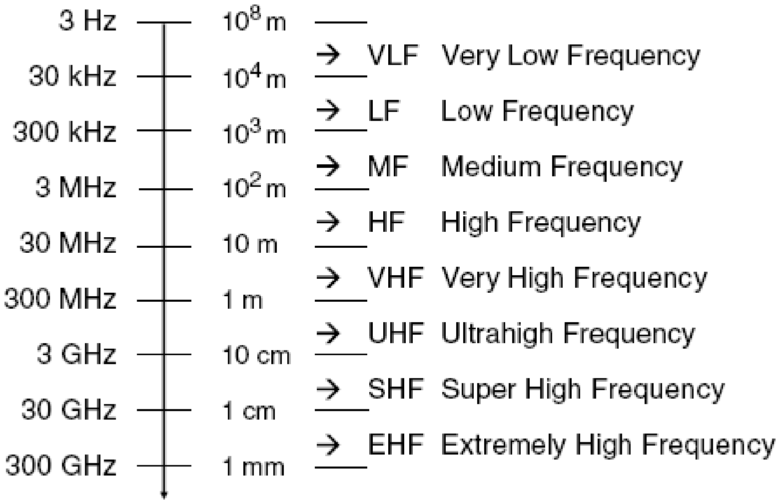


Figure 1.6 Frequency band designations by wavelength

Figure 1.6 Frequency band designations by wavelength

1.2 Regulatory Process for Satellite Communications

Satellite operators and owners must operate within constraints of regulations related to fundamental parameters and characteristics of the satellite communications system. The satelite communications system parameters that fall under the regulatory umbrella include:

- choice of radiating frequency;
- maximum allowable radiated power;
- orbit locations (slots) for GSO.

The purpose of the regulation is to minimize radio frequency interference and, to a lesser degree, physical interference between systems. Potential radio interference includes not only other operating satelite systems, but also terrestrial communications systems, and other systems emitting energy in the same frequency bands.

The discipline involved with the development of the technical, analytical, and institutional elements supporting the allocation and regulation of the frequency spectrum is usually referred to as spectrum management or frequency management. Most countries have active organizations, both in the government and the commercial sectors, involved with spectrum management, particularly those organizations responsible for the development of satellite systems or the provision of satellite based services.

There are two levels of regulation and allocation involved in the process: international and domestic. The primary organization responsible for international satellite communications systems regulation and allocation is the International Telecommunications Union (ITU), headquartered in Geneva, Switzerland.

The ITU was formed in 1932 from the International Telegraph Union, created in 1865. It is a United Nations Specialized Agency, currently with over 190 members. The ITU structure is similar to the United Nations, with a General Secretariat, elected Administrative Council, boards, and committees for the conduct of Technical and Administrative functions. The ITU has three primary functions:

- allocations and use of the radio-frequency spectrum;
- telecommunications standardization;
- development and expansion of worldwide telecommunications.

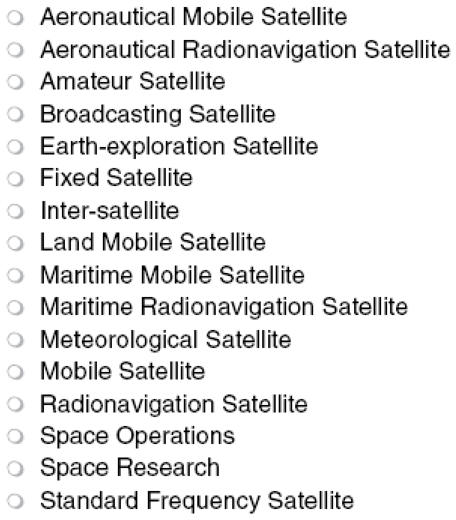
The three functions are accomplished through three sectors within the ITU organization: the Radiocommunications Sector (ITU-R), responsible for frequency allocations and use of the radio-frequency spectrum; the Telecommunications Standards Sector (ITU-T), responsible for telecommunications standards; and the Telecommunications Development Sector (ITU-D), responsible for the development and expansion of worldwide telecommunications.

The international regulations developed by the ITU are handed down and processed by each country, where the domestic level regulations are developed. The ITU does not have enforcement powers – each individual country is left to manage and enforce the regulations within its boundaries. The responsibility for managing regulations in the United States is with the Federal Communications Commission (FCC) and the National Telecommunications and Information Agency (NTIA). Satellite systems operated by the federal government operate through the NTIA, while all other systems, including commercial and local government systems, operate through the FCC. The US Department of State coordinates all the frequency and spectrum management activities and represents the US at the ITU and its related organizations. Other countries have their own mechanisms and organizations responsible for the spectrum management function – usually government agencies or bureaus working in close cooperation with satellite systems and services providers. Two attributes determine the specific frequency bands and other regulatory factors for a particular satellite system:

- service(s) to be provided by the satellite system/network; and
- location(s) of the satellite system/network ground terminals.

Both attributes together determine the frequency band, or bands, where the satellite system may operate.

Figure 1.7 lists the major services as designated by the ITU that are relevant to satelite systems. Some service areas are divided into several sub areas. The mobile satellite service (MSS) area, for example, is further broken down into the aeronautical mobile satelite service (AMSS), the land mobile satellite service (LMSS), and the maritime mobile satelite service (MMSS), depending on the physical locale of the ground based terminals. If the terminals are located on more than one locale, for example on land and sea, then the MSS would apply.



- Aeronautical Mobile Satellite
- Aeronautical Radionavigation Satellite
- Amateur Satellite
- Broadcasting Satellite
- Earth-exploration Satellite
- Fixed Satellite
- Inter-satellite
- Land Mobile Satellite
- Maritime Mobile Satellite
- Maritime Radionavigation Satellite
- Meteorological Satellite
- Mobile Satellite
- Radionavigation Satellite
- Space Operations
- Space Research
- Standard Frequency Satellite

Figure 1.7 Satellite services as designated by the International Telecommunications Union (ITU) *(source: ITU [15]; reproduced by permission of International Telecommunications Union)*

The second attribute, the location of the earth terminals, is determined by the appropriate service region. The ITU divides the globe into three telecommunications service regions, as is shown in Figure 1.8.

The three regions divide the earth land areas approximately into the major land masses – Europe and Africa (Region 1), the Americas (Region 2), and the Pacific Rim countries (Region 3). Each service region is treated as independent in terms of frequency allocations, because the general assumption is that systems operating in any one of the regions are protected by geographic separation from systems in the other service regions. International frequency allocations are provided for systems operating on a global basis.

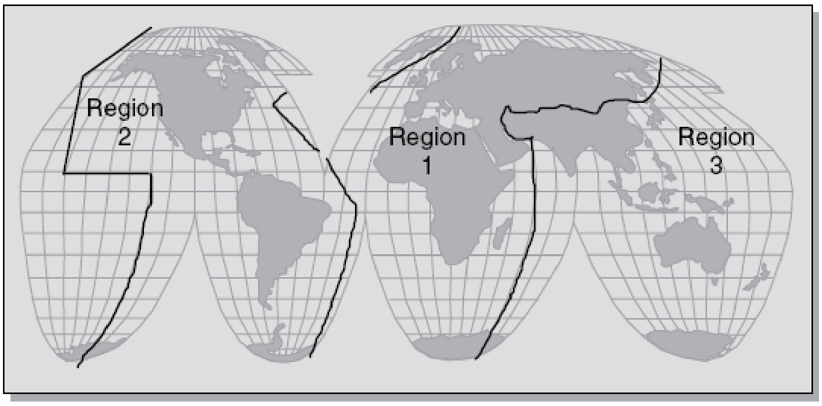


Figure 1.8 ITU telecommunications service regions *(source: ITU [15]; reproduced by permission of International Telecommunications Union)*

References

[1] Clarke, A.C.:‘Extraterrestrial Relays,’Wireless World, Vol. 51, pp. 305 308, October 1945.

[2] Davis, M.I. Krassner, G.N.: ‘SCORE First Communications Satellite,’ Journal of American Rocket Society,Vol. 4, May 1959.

[3] Jaffe,L.: ‘Project Echo Results,’ Astronautics, Vol. 6, No. 5, May 1961.

[4] E. Imboldi,E-Hershberg,D.: ‘Courier Satellite Communications System,’Advances in the Astronautical Sciences,Vol. 8, 1961.

[5] Special Issue on ProjectWest Ford, Proceedings of the IEEE, Vol. 52, No. 5, May 1964.

[6] K.W. Gatland, K.W.:Telecommunications Satellites, Prentice Hall, New York, 1964.

[7] D.H. Martin, D.H.: Communications Satellites 1958–1988, The Aerospace Corporation, December 31, 1986.

[8] J. Alper,J.-Pelton,J.N eds., ‘The INTELSATGlobal System, Progress in Astronautics andAeronautics,’Vol. 93, AIAA, New York, 1984.

[10] Almond, J. ‘Commercial Communications Satellite Systems in Canada,’ IEEE Communications Magazine, Vol. 19, No. 1, January 1981.

[11] W.N. Redisch and R.L. Hall, ‘ATS 6 Spacecraft Design/Performance,’ EASCON’74 Conference Record, October 1974.

[12] Special Issue onATS 6, IEEE Transactions on Aerospace and Electronics Systems,Vol. 11, No. 6, November 1975.

[13] Wright,D.L-Day,J.W.B.: ‘The Communications Technology Satellite and the Associated Ground Terminals for Experiments,’AIAA Conference on Communications Satellites for Health/Education Applications, July 1975.

[14] Ippolito,L.J.:‘Characterization of the CTS 12 and 14 GHz Communications Links Preliminary Measurements and Evaluation,’ International Conference on Communications: ICC’76, June 1976.

[15] International Telecommunications Union, www.itu.int.

[16] Ippolito, L.J.: Satellite Communications Systems Engineering. Atmospheric Effects, Satellite Link Design and System Performance, Wiley, Singapore,2008

2 Satellite Orbits

The orbital locations of the spacecraft in a communications satellite system play a major role in determining the coverage and operational characteristics of the services provided by that system. This chapter describes the general characteristics of satellite orbits and summarizes the characteristics of the most popular orbits for communications applications.

The same laws of motion that control the motions of the planets around the sun govern artificial earth satellites that orbit the earth. Satellite orbit determination is based on the Laws of Motion first developed by Johannes Kepler and later refined by Newton in 1665 from his own Laws of Mechanics and Gravitation. Competing forces act on the satellite; gravity tends to pull the satellite in towards the earth, while its orbital velocity tends to pull the satelite away from the earth. Figure 2.1 shows a simplified picture of the forces acting on an orbiting satellite.

The gravitational force, F_{in} , and the angular velocity force, F_{out} , can be represented as

$$F_{in} = m\left(\frac{\mu}{r^2}\right) \tag{2.1}$$

and

$$F_{out} = m\left(\frac{v^2}{r}\right) \tag{2.2}$$

Where m=satellite mass; v=satellite velocity in the plane of orbit; r =distance from the center of the earth (orbit radius); and_=Kepler’s Constant (or Geocentric Gravitational Constant)= 3.986004×105 km3/s2.

Note that for $F_{in} = F_{out}$

$$v = \left(\frac{\mu}{r}\right)^{\frac{1}{2}} \tag{2.3}$$

This result gives the velocity required to maintain a satellite at the orbit radius r. Note that for the discussion above all other forces acting on the satellite, such as the gravity forces from the moon, sun, and other bodies, are neglected.

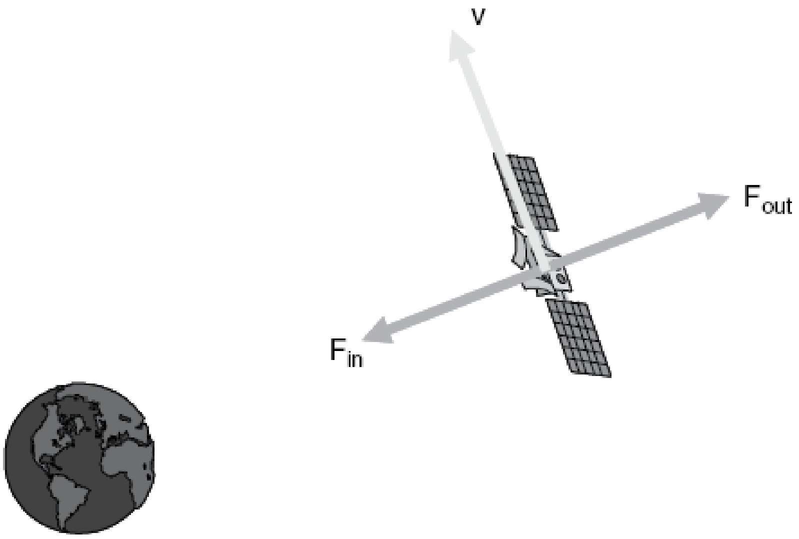


Figure 2.1 Forces acting on a satellite

2.1 Kepler's Laws

Kepler's laws of planetary motion apply to any two bodies in space that interact through gravitation. The laws of motion are described through three fundamental principles.

Kepler's First Law, as it applies to artificial satellite orbits, can be simply stated as follows: 'the path followed by a satellite around the earth will be an ellipse, with the center of mass of earth as one of the two foci of the ellipse.' This is shown in Figure 2.2.

F

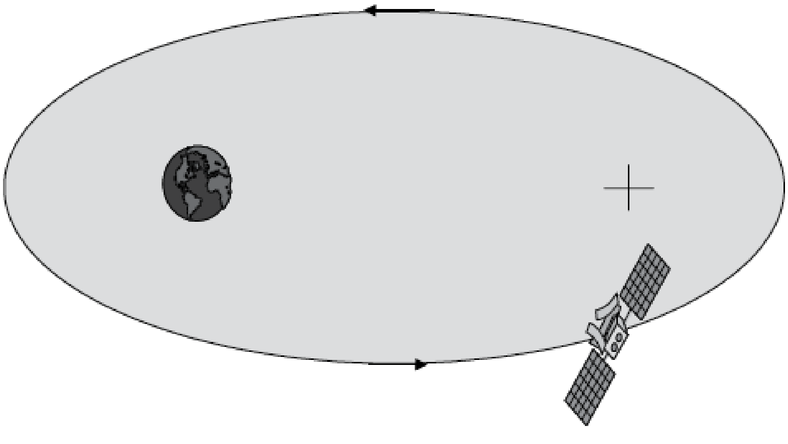


Figure 2.2 Kepler's First Law

Figure 2.2 Kepler's First Law

If no other forces are acting on the satellite, either intentionally by orbit control or unintentionally as in gravity forces from other bodies, the satellite will eventually settle in an ellipticalorbit, with the earth as one of the foci of the ellipse. The 'size' of the ellipse will depend on satellite mass and its angular velocity

Kepler's Second Law can likewise be simply stated as follows: 'for equal time intervals, the satellite sweeps out equal areas in the orbital plane.' Figure 2.3 demonstrates this concept.

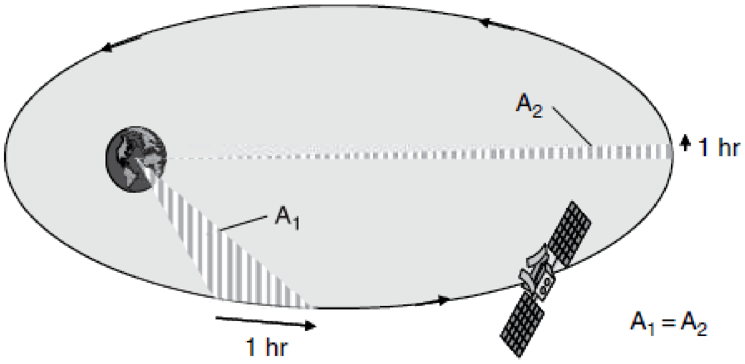


Figure 2.3 Kepler's Second Law

Figure 2.3 Kepler's Second Law

The shaded area A1 shows the area swept out in the orbital plane by the orbiting satelite in a one hour time period at a location near the earth. Kepler's second law states that the area swept out by any other one hour time period in the orbit will also sweep out an area equal to A1. For example, the area swept out by the satellite in a one hour period around the point farthest from the earth (the orbit's apogee), labeled A2 on the figure, will be equal to A1, i.e.: A1 =A2. This result also shows that the satellite orbital velocity is not constant; the satellite is moving much faster at locations near the earth, and slows down as it approaches apogee. This factor will be discussed in more detail later when specific satellite orbit types are introduced.

Stated simply, Kepler's Third Law is as follows: 'the square of the periodic time of orbit is proportional to the cube of the mean distance between the two bodies.' This is quantified as follows:

$$T^2 = \left[\frac{4\pi^2}{\mu}\right] a^3 \tag{2.4}$$

where T=orbital period in s; a=distance between the two bodies, in km; μ= Kepler's Constant =3.986004×105 km3/s2.

If the orbit is circular, then a=r, and

$$r = \left[\frac{\mu}{4\pi^2}\right]^{\frac{1}{3}} T^{\frac{2}{3}} \tag{2.5}$$

This demonstrates an important result:

$$\text{Orbit Radius} = [\mathit{Constant}] \times (\text{Orbit Period})^{\frac{2}{3}} \tag{2.6}$$

Under this condition, a specific orbit period is determined only by proper selection of the orbit radius. This allows the satellite designer to select orbit periods that best meet particular application requirements by locating the satellite at the proper orbit altitude. The altitudes required to obtain a specific number of repeatable ground traces with a circular orbit are listed in Table 2.1.

Table 2.1 Orbit altitudes for specified orbital periods

Revolutions/day	Nominal period (hours)	Nominal altitude (km)
1	24	36 000
2	12	20 200
3	8	13 900
4	6	10 400
6	4	6400
8	3	4200

2.2 Orbital Parameters

Figure 2.4 shows two perspectives useful in describing the important orbital parameters used to define earth-orbiting satellite characteristics. The parameters are:

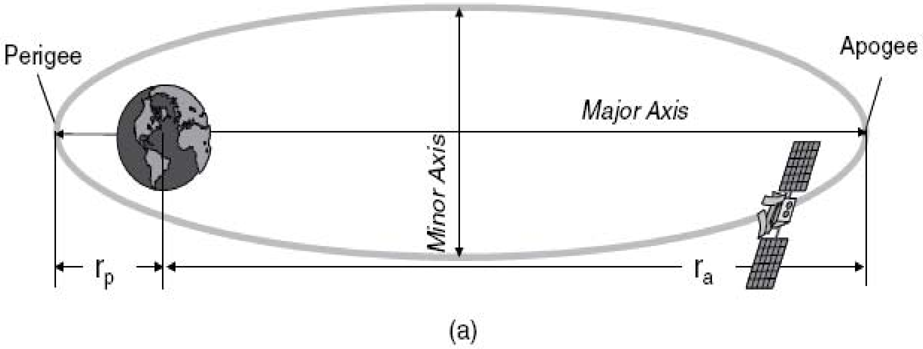
- Apogee – the point farthest from earth.
- Perigee – the point of closest approach to earth.
- Line of Apsides – the line joining the perigee and apogee through the center of the earth.

- Ascending Node – the point where the orbit crosses the equatorial plane, going from south to north.
- Descending Node – the point where the orbit crosses the equatorial plane, going from north to south.
- Line of Nodes – the line joining the ascending and descending nodes through the center of the earth.
- Argument of Perigee, ω–the angle from ascending node to perigee, measured in the orbital plane.
- Right Ascension of the Ascending Node, φ-the angle measured eastward, in the equatorial plane, from the line to the first point of Aries (Y) to the ascending node.

The eccentricity is a measure of the 'circularity' of the orbit. It is determined from

$$e = \frac{r_a - r_p}{r_a + r_p} \tag{2.7}$$

where e=the eccentricity of the orbit; r_a =the distance from the center of the earth to the apogee point; and r_p =the distance from the center of the earth to the perigee point.



(a)

Figure 2.4 Earth-orbiting satellite parameters

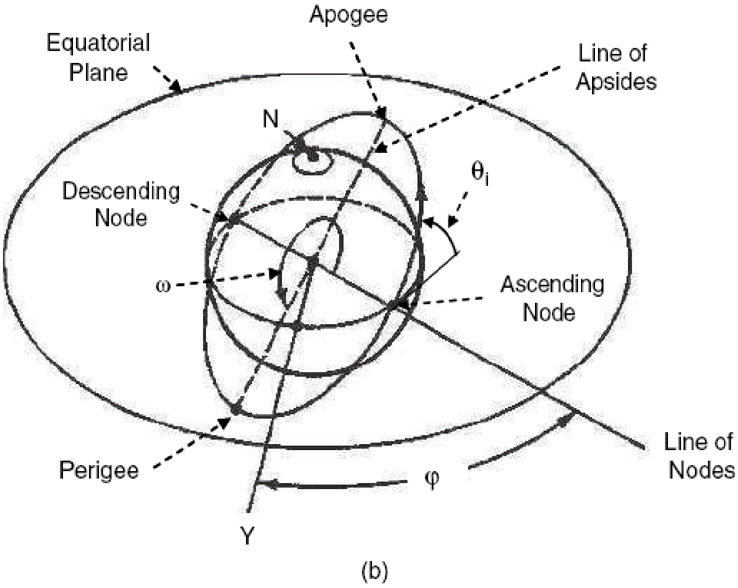


Figure 2.4 *(continued)*

The higher the eccentricity, the ‘flatter’ the ellipse. A circular orbit is the special case of an ellipse with equal major and minor axes (zero eccentricity). That is:

Elliptical Orbit 0<e<1

Circular Orbit e=0

The inclination angle, v_i, is the angle between the orbital plane and the earth’s equatorial plane. As atellite that is in an orbit with some inclination angle is in an inclined orbit.Asatellite that is in orbit in the equatorial plane (inclination angle=0◦) is in an equatorial orbit.Asatellite that has an inclination angle of 90◦ is in a polar orbit. The orbit may be elliptical or circular, depending on the orbital velocity and direction of motion imparted to the satellite on insertion into orbit.

Figure 2.5 shows another important characteristic of satellite orbits. An orbit in which the satellite moves in the same direction as the earth’s rotation is called a prograde orbit. The inclination angle of a prograde orbit is between 0◦ and 90◦. A satellite in a retrograde orbit moves in a direction opposite (counter to) the earth’s rotation, with an inclination angle between 90◦ and 180◦. Most satellites are launched in a prograde orbit, because the earth’s rotational velocity enhances the satellite’s orbital velocity, reducing the amount of energy required to launch and place the satellite in orbit. An almost endless number of combinations of orbital parameters are available for satelite orbits. Orbital elements defines the set of parameters needed to uniquely specify the location of an orbiting satellite. The minimum number of parameters required is six:

- Eccentricity;
- Semi-Major Axis;
- Time of Perigee;
- Right Ascension of Ascending Node;
- Inclination Angle;
- Argument of Perigee.

Figure

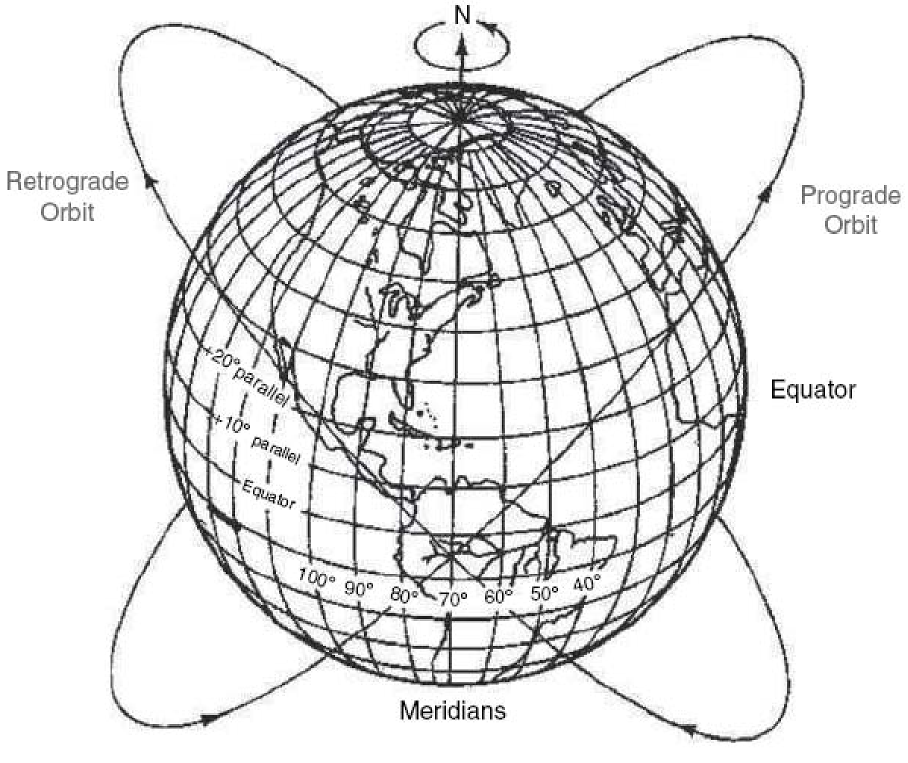


Figure 2.5 Prograde and retrograde orbits

These parameters will uniquely define the absolute (i.e., the inertial) coordinates of the satelite at any time t. They are used to determine the satellite track and provide a prediction of satelite location for extended periods beyond the current time.Satellite orbits coordinates are specified in sidereal time rather than in solar time. Solar time, which forms the basis of all global time standards, is based on one complete rotation of the earth relative to the sun. Sidereal time is based on one complete rotation of the earth relative to a fixed star reference, as shown in Figure 2.6.

Since Sidereal time is based on one complete rotation of the earth relative to a fixed star reference at essentially an infinite distance, rather than the sun, a mean Sidereal day is shorter than a mean Solar day by about 0.3 %, as indicated on Figure 2.6.

2.3 Orbits in Common Use

With all the possible combinations of orbit parameters available to the satellite designer, an almost endless list of possible orbits can be used. Experience has narrowed down the list of orbits in common use for communications, sensor, and scientific satellites, and they are introduced in the following sections.We begin with the most popular orbit used for communications satellites – the geostationary (or geosynchronous) orbit.

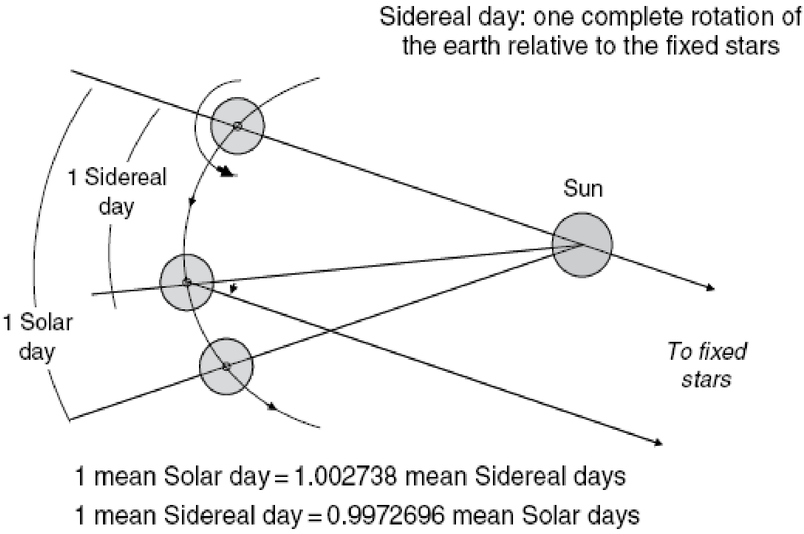


Figure 2.6 Sidereal time

2.3.1 Geostationary Orbit

Kepler’s third law demonstrated that there is a fixed relationship between orbit radius and the orbit period of revolution (see Equation (2.6)). Under this condition a specific orbit period can be determined by proper selection of the orbit radius.

If the orbit radius is chosen so that the period of revolution of the satellite is exactly set to the period of the earth’s rotation, one mean sidereal day, a unique satellite orbit is defined. In addition, if the orbit is circular (eccentricity=0), and the orbit is in the equatorial plane (inclination angle=0◦), the satellite will appear to hover motionless above the earth at the subsatellite point above the equator. This important special orbit is the geostationary earth orbit (GEO).

From Kepler's third law, the orbit radius for the GEO, rS, is found as

$$r_S = \left[\frac{\mu}{4\,\pi^2}\right]^{\frac{1}{3}} T^{\frac{2}{3}} = \left[\frac{3.986004 \times 10^5}{4\,\pi^2}\right]^{\frac{1}{3}} (86\,164.09)^{\frac{2}{3}} \qquad (2.8)$$
$$= 42\,164.17\,\text{km}$$

where T=1 mean sidereal day=86 164.09 s.

The geostationary height (altitude above the earth's surface), h_S, is then

$$\begin{aligned} h_S &= r_S - r_E \\ &= 42\,164 - 6378 \qquad (2.9) \\ &= 35\,786\,\text{km} \end{aligned}$$

where r_E = equatorial earth radius =6378 km.

The value of h_S is often rounded to 36 000 km for use in orbital calculations.The geostationary orbit is an ideal orbit that cannot be achieved for real artificial satellites because there are many other forces besides the earth's gravity acting on the satellite. A'perfect orbit', i.e., one with e exactly equal to zero and with _i exactly equal to 0◦, cannot be practically achieved without extensive station keeping and a vast amount of fuel to maintain the precise position required. A typical GEO orbit in use today would have an inclination angle slightly greater than 0 and possibly an eccentricity that also exceeds 0. The 'real world'GEO orbit that results is often referred to as a geosynchronous earth orbit (GSO) to differentiate it from the ideal geostationary orbit.1 Most current communications satellites operate in a geosynchronous earth orbit, which is ideally suited for the transfer of communications information between two or more points on the earth through a 'relay' that is fixed in space, relative to the earth. Figure 2.7 shows the basic elements of the geosynchronous earth orbit as it applies to satellite operations. The GSO location provides a fixed path from the ground to the satellite; therefore little or no ground tracking is required.A satellite in GSO sees about one-third of the earth’s surface, so three GSO satellites, placed 120◦ apart in the equatorial plane, could provide global coverage, except for the pole areas (to be discussed further later).

The period of revolution for the geostationary orbit is 23 hours, 56 minutes, which is the time for the earth to complete one revolution about its axis, measured relative to the star field reference (sidereal time). It is four minutes shorter than the 24-hour mean solar day because of the earth’s movement around the sun. The geosynchronous orbit does suffer from some disadvantages, even though it is the most heavily implemented orbit for current communications systems because of its fixed earthsatellite geometry and its large coverage area. The long path length produces a large path loss and a significant latency (time delay) for the radiowave signal propagating to and from the satellite. The two-way (up to the satellite and back) delay will be approximately 260 ms for a ground station located at a mid-latitude location. This could produce problems, particularly for voice communications or for certain protocols that cannot tolerate large latency.

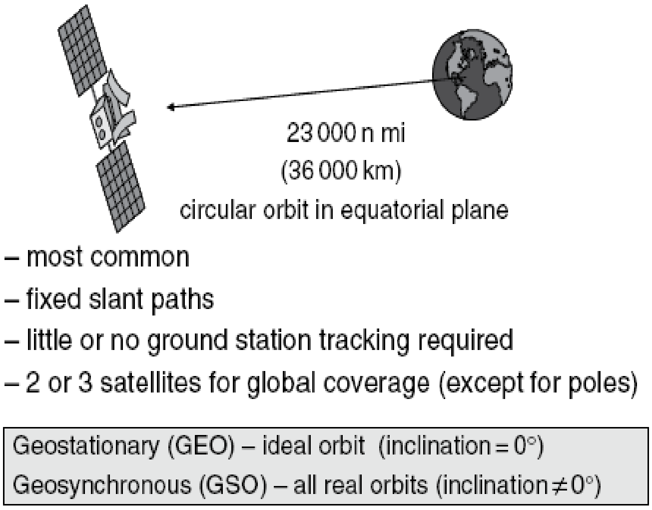


Figure 2.7 GSO – Geosynchronous earth orbit

The GSO cannot provide coverage to high latitude locations. The highest latitude, at which the GSO satellite is visible, with a 10◦ earth station elevation angle, is about 70◦, North or South latitude. Coverage can be increase somewhat by operation at higher inclination angles, but that produces other problems, such as the need for increased ground antenna tracking, which increases costs and system complexity.

The differentiation between the ideal and actual orbits by use of the terms ‘geostationary’ and ‘geosynchronous’ is by no means an accepted global standard.

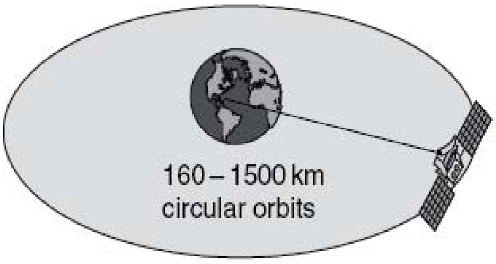
Often the terms are used interchangeably or all orbits may be defined by only one of the terms.We will maintain the definitions introduced above to avoid possible confusion.

The number of satellites that can operate in geostationary orbits is obviously limited, because there is only one equatorial plane, and the satellites must be spaced to avoid interference between each other. The allocation of geostationary orbital locations or slots is regulated by international treaties through the International Telecommunications Union, in close coordination with frequency band and service allocations, as discussed in Chapter 1. Current allocations place satellites in the range of 2–5◦ apart for each frequency band and service allocation, meaning that only 72–180 slots are available for global use, depending on the frequency band and service provided.

2.3.2 Low Earth Orbit

Earth satellites that operate well below the geostationary altitude, typically at altitudes from 160 to 2500 km, and in near circular orbits, are referred to as low earth orbit or LEO satellites.2 The low earth orbit satellite has several characteristics that can be advantageous for communications applications, as summarized on Figure 2.8.

FF



- requires earth terminal tracking
- approx. 8 to 10 minutes per pass for an earth terminal
- requires multiple satellites (12, 24, 66, . . .) for global coverage
- popular for mobile satellite communications applications

Figure 2.8 LEO – Low earth orbit

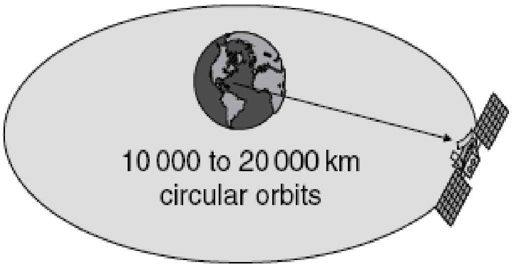
The earth-satellite links are much shorter, leading to lower path losses, which result in lower power, smaller antenna systems. Propagation delay is also less because of shorter path distances. LEO satellites, with the proper inclinations, can cover high latitude locations, including polar areas, which cannot be reached by GSO satellites. A major disadvantage of the LEO satellite is its restricted operations period, because the satellite is not at a fixed location in the sky, but instead sweeps across the sky for as little as 8 to 10 minutes from a fixed location on earth. If continuous global or wide area coverage is desired, a constellation of multiple LEO satellites is required, with links between the satellites to allow for point-to-point communications. Some current LEO satellite networks operate with 12, 24, and 66 satellites to achieve the desired coverage.

The oblateness (non-spherical shape) of the earth will cause two major perturbations to the LEO orbit. The point on the equator where the LEO satellite crosses from south to north (the ascending node) will drift westward several degrees per day.Asecond effect of the earth's oblateness is to rotate the orientation of the major axis in the plane of the orbit, either clockwise or counterclockwise. If the inclination is set to about 63°, however, the forces that induce the rotation will be balanced and the major axis direction remains fixed.

The LEO orbit has found serious consideration for mobile applications, because the small power and small antenna size of the earth terminals are a definite advantage. More LEO satellites are required to provide communications services comparable to the GSO case, but LEO satellites are much smaller and require significantly less energy to insert into orbit, hence total life cycle costs may be lower.

2.3.3 Medium Earth Orbit

Satellites that operate in the range between LEO and GSO, typically at altitudes of 10 000 to 20 000 km, are referred to as medium altitude orbit, or MEO satellites. The basic elements of the MEO are summarized on Figure 2.9



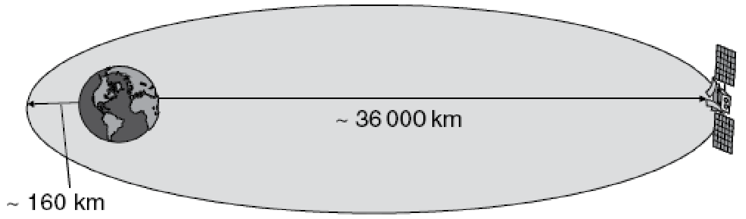
– similar to LEO, but at higher circular orbits
– 1 to 2 hours per pass for an earth terminal
– used for meteorological, remote sensing and position location applications

Figure 2.9 MEO – Medium earth orbit

The desirable features of the MEO include: repeatable ground traces for recurring ground coverage; selectable number of revolutions per day; and adequate relative satellite-earth motion to allow for accurate and precise position measurements. A typical MEO would provide one to two hours of observation time for an earth terminal at a fixed location. MEO satellites have characteristics that have been found useful for meteorological, remote sensing, navigation, and position determination applications. The Global Positioning System (GPS), for example, employs a constellation of up to 24 satellites operating in 12-hour circular orbits, at an altitude of 20 184 km.

2.3.4 Highly Elliptical Orbit

Satellites operating in high elliptical (high eccentricity) orbits (HEO) are used to provide coverage to high latitude areas not reachable by GSO, and those that require longer contact periods than available with LEO satellites. The orbital properties of the elliptical orbit defined by Kepler’s second law, as discussed previously, can be used to offer extended dwell time over areas near the apogee, when it is farthest from the earth but is moving the slowest in orbit (Figure 2.10).



– popular for high latitude or polar coverage
– often referred to as the ‘MOLNIYA’ orbit
– 8 to 10 hours of 12 hour HEO orbit available for communications from earth terminal, with ‘GSO like’ operations

Figure 2.10 HEO – Highly elliptical earth orbit

The most popular HEOorbit used for communications satellites is the Molniya orbit, named for the satellite system that serviced the (former) Soviet Union. The orbit is designed to provide extended coverage in the high northern latitudes that comprise most of the former Soviet Union’s land mass, where GSO satellites cannot provide coverage. A typical Molnyia orbit has a perigee altitude of about 1000 km, and an apogee altitude of nearly 40 000 km. This corresponds to an eccentricity of about 0.722. The inclination is chosen at 63.4◦ to prevent major axis rotation, as described in the previous section. The orbit has a nominal period of 12 hours, which means that it repeats the same ground trace twice each day. The highly elliptical orbit causes the satellite to spend nearly ten hours of each rotation over the northern hemisphere, and only two hours over the southern hemisphere. Two satellites in HEO Molniya orbits, properly phased, can provide nearly continuous coverage to high latitude locations in the northern hemisphere, because at least one of the satellites will be in view at any time during the day.

2.3.5 Polar Orbit

A circular orbit with an inclination near 90° is referred to as a polar orbit. Polar orbits are very useful for sensing and data gathering services, because their orbital characteristics can be selected to scan the entire globe on a periodic cycle. Landsat, for example, operated with an average altitude of 912 km, and an orbital period of 103 minutes, tracing out 14 revolutions each day. Each day the orbit shifted about 160 km west on the equator, returning to its original position after 18 days and 252 revolutions.

2.4 Geometry of GSO Links

The GSO is the dominant orbit used for communications satellites. In this section we develop the procedures to determine the parameters required to define the GSO parameters that are used to evaluate satellite link performance and design.

The three key parameters for the evaluation of the GSO link are:

d=range (distance) from the earth station (ES) to the satellite, in km

φ_z =azimuth angle from the ES to the satellite, in degrees

θ=elevation angle from the ES to the satellite, in degrees

The azimuth and elevation angles are referred to as the look angles for the ES to the satellite.

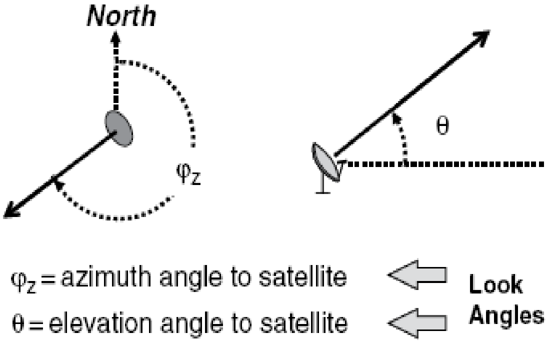


Figure 2.11 GSO look angles to satellite

Figure 2.11 shows the geometry and definitions of the look angles with respect to the earth station reference.

There are many sources available in the orbital mechanics and satellite literature that describe the detailed development of the calculations for the GSO parameters, range, elevation angle, and azimuth angle. Two good examples are provided in References 1 and 2.

The calculations involve spherical geometry derivations and evaluations requiring several stages of development. There are also several software packages available for the determination of orbital parameters, for both GSO and NGSO satellites networks. Our intent here is to summarize the final results of the various derivations and to allow us to apply the GSO parameters to the evaluation of free space links for communications satelite applications.

The input parameters required to determine the GSO parameters are:

l_E =earth station longitude, in degrees

l_S =satellite longitude, in degrees

L_E =earth station latitude, in degrees

L_S =satellite latitude in degrees (assumed to be 0, i.e., inclination angle=0)

H=earth station altitude above sea level, in km

The point on the earth's equator at the satellite longitude is called the subsatellite point (SS).

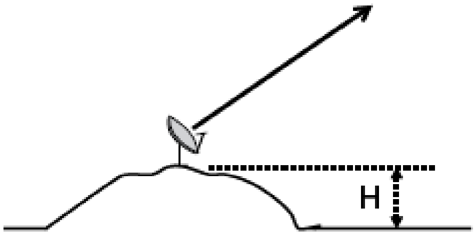


Figure 2.12 Earth station altitude

Figure 2.12 clarifies the definition of earth station altitude.

Longitude and latitude sign values are based on the sign convention shown in Figure 2.13. Longitudes east of the Greenwich Meridian and latitudes north of the equator are positive.

Additional parameters required for the calculations are:

Equatorial Radius: re =6378.14 km

Geostationary Radius: rS =42 164.17 km

Geostationary Height (Altitude): hGSO =rS −re =35 786 km

Eccentricity of the earth: ee =0.08182

An additional parameter required for the calculation of the GSO parameters is the differential longitude, B, defined as the difference between the earth station and satellite longitudes:

$$B = l_E - l_S \tag{2.10}$$

where the sign convention of Figure 2.13 is followed.

For example, for an earth station located in Washington, DC, at the longitude of 77◦W, and a satellite located at a longitude of 110◦W:

$$B = (-77) - (-110) = +33^\circ$$

Figure 2.13 Sign convention for longitude and latitude

2.3.6 Range to Satellite

The determination of the range to the satellite from the earth station requires the radius of the earth at the earth station latitude and longitude, R. It is found as

$$R = \sqrt{l^2 + z^2} \tag{2.11}$$

where

$$l = \left(\frac{r_e}{\sqrt{1 - e_e^2 \sin^2(L_E)}} + H \right) \cos(L_E) \tag{2.12}$$

and

$$z = \left(\frac{r_e\left(1 - e_e^2\right)}{\sqrt{1 - e_e^2 \sin^2(L_E)}} + H \right) \sin(L_E) \tag{2.13}$$

An intermediate angle, Ψ_E, is also defined:

$$\Phi_E = \tan^{-1}\left(\frac{z}{l}\right) \tag{2.14}$$

The range d is then found from

$$d = \sqrt{R^2 + r_s{}^2 - 2\, R\, r_s\, \cos(\Psi_E)\cos(B)} \tag{2.15}$$

This result will be used to determine several important parameters for satellite link analysis, including the free space path loss, which is directly dependent on the complete path length from the earth station antenna to the satellite antenna.

2.3.7 Elevation Angle to Satellite

The elevation angle from the earth station to the satellite, θ, is determined from

$$\theta = \cos^{-1}\left(\frac{r_e + h_{GSO}}{d} \sqrt{1 - \cos^2(B)\ \cos^2(L_E)} \right) \tag{2.16}$$

where r_e =equatorial radius=6378.14 km; h_{GSO} =geostationary altitude=35 786 km; d = range, in km; B = differential longitude, in degrees; and L_E =ES latitude, in degrees.

The elevation angle is important because it determines the slant path through the earth’s atmosphere, and will be the major parameter in evaluating atmospheric degradations such as rain attenuation, gaseous attenuation, and scintillation on the path. Generally, the lower the elevation angle, the more serious the atmospheric degradations will be, because more of the atmosphere will be present to interact with the radiowave on the path to the satellite.

2.3.8 Azimuth Angle to Satellite

The final parameter of interest is the earth station azimuth angle to the satellite. First, an intermediate angle A_i is found from

$$A_i = \sin^{-1}\left(\frac{\sin(|B|)}{\sin(\beta)}\right) \tag{2.17}$$

where |B| is the absolute value of the differential longitude

$$|B| = |l_E - l_S|$$

and

$$\beta = \cos^{-1}[\cos(B)\cos(L_E)]$$

The azimuth angle φ_z is determined from the intermediate angle Ai from one of four possible conditions, based on the relative location of the earth station and the subsatellite point on the earth’s surface. The condition is determined by standing at the earth station (ES) and looking in the direction of the subsatellite point (SS). That direction will be one of four possible general directions: northeast (NE), northwest (NW), southeast (SE), or southwest (SW), as shown in Figure 2.14. The resulting equation to determine φ_z for each of the four conditions is given in

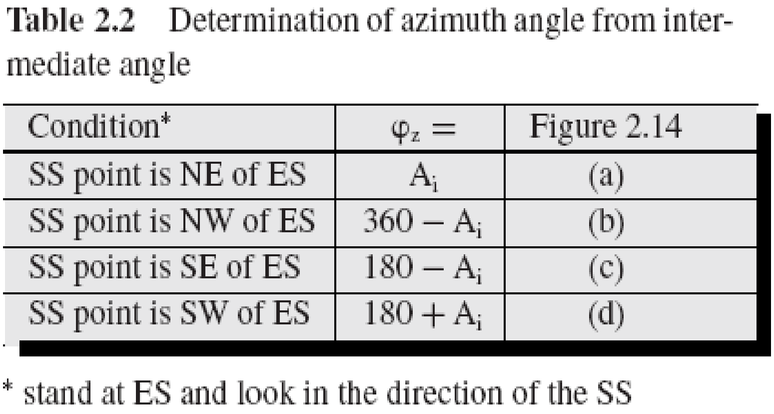


Table 2.2 Determination of azimuth angle from intermediate angle

Condition*	$\varphi_z =$	Figure 2.14
SS point is NE of ES	A_i	(a)
SS point is NW of ES	$360 - A_i$	(b)
SS point is SE of ES	$180 - A_i$	(c)
SS point is SW of ES	$180 + A_i$	(d)

* stand at ES and look in the direction of the SS

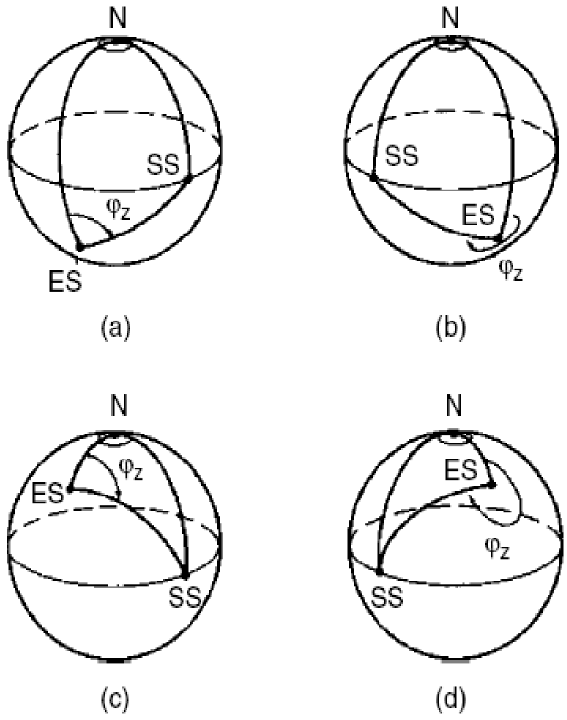


Figure 2.14 Determination of azimuth angle condition

Two special cases can occur where the azimuth angle can be directly observed.

- If the earth station is located at the same longitude as the subsatellite point, the azimuth angle will be 180° if the earth station is in the northern hemisphere and 0° if the earth station is in the southern hemisphere. This can be verified from the conditions of Table 2.2, with B=0.

- If the earth station is located on the equator, the azimuth angle will be 90° if the earth station is west of the subsatellite point and 270° if the earth station is east of the subsatellite point.

This can also be verified from the conditions of Table 2.2.

References

[1] Roddy,D.: Satellite Communications, Third Edition, McGraw-Hill, New York, 2001:

[2] Pratt, T. - Bostian, C.W. - Allnutt, J.E.: Satellite Communications, Second Edition, John Wiley & Sons, Inc., New York, 2003.

[3] Ippolito, L.J.: Satellite Communications Systems Engineering. Atmospheric Effects, Satellite Link Design and System Performance, Wiley, Singapore,2008

3 System Design

3.1 System Configuration

Figure 3.1 shows the basie configuration of mobile satellite communicationsystems. The system consists of three basie segments: a satellite, a gateway Earth station and a mobile Earth station (Figure 1.1). From the standpoint ofsystem design, a propagation path has to be added as the fourth segment. In mobile satellite communication systems, the propagation path is a very important factor that mainly affects the channel quality of the communication system. In land mobile satellite communications, the most serious propagation problem is the effect of blocking caused by buildings and surrounding objects which cause signals from the satellite to shut down completely. The second problem is shadowing caused by rrees and foliage, which results in signal attenuation. The third is multipath fading, which is mainly caused by buildings. However, this effect can usually be ignored because of the use of directionalantennas and the grear attenuation of reflected signals. In maritime satelite communications, fading caused by sea surface reflection is the most serious propagation problem. Rain attenuation has to be considered in higher frequency bands such as the Ka band and millimeter-wave band. However, it can be neglected in the L band.

A gateway and a mobile Earth station can be broken down into anantenna, a diplexer (DIP), a set of upconverters and downconverters (U/C and DfC), a high-power amplifier (HPA) and a low-noise arnplifier (LNA), and a set of a modulator (MOD) and a demodulator (DEM). The configuration for a satellite is almost the same as for gateway and mobile Earth stations and can be broken down into an antenna and a set of upconverters and downconverters, and a set of this on board equipment is called a transponder. Almost all of the present commercial satellites do not have a set of modulators and demodulators.

They can only transmit a signal after convening the frequency and amplify received weak signals. This type of transponder used on the present commercial satellites is called a bent pipe transponder or a transparent transponder. The main parameters that characterize the performance of the three segments-namely, satellites and gateway and mobile Earth stations-are GIT (the ratio of antenna gain to system noise ternperature or a figure of merit), effective isotropically radiated power (EIRP) and C!No (the ratio of carrier power to noise power density). The GIT and EIRP are

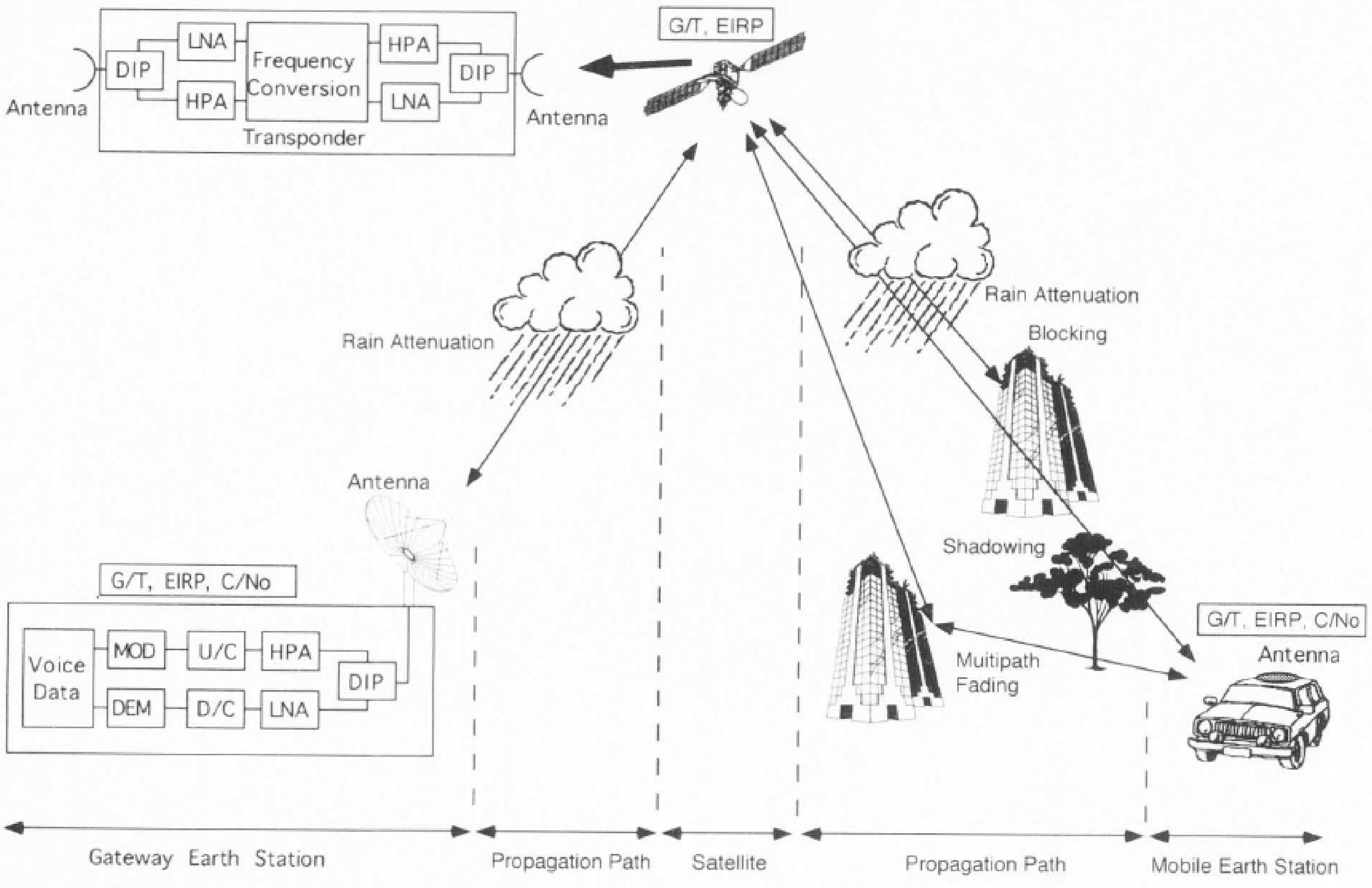


Figure 3.1 System configuration of mobile satellite communications.

frequendy used concepts in satellite communications, and they denote the receiving and transmitting capabilities, respectively, of a satellite, a gateway Earth station, and a mobile terminal. The C!No denotes the quality of the communication channel. These parameters will be described in detail in the following sections.

3.2 Main Parameters in Link Budget

3.2.1 Terminal Noise

The noise performance of communication systems can be described by a te rmfor noise temperature. The use of a fictitious temperature stems from the fact that the basie source of noise in electrical circuits is the thermal agitation of electrons in resistive circuit components. The open-circuit root mean square (rms) noise voltage Vn generated in a resistance of value R ohms (fl) at an absolute temperature T of kelvin (K) is given by Nyquist [l]

$$V_n = \sqrt{4\kappa TRB} \qquad \text{(volts)} \tag{3.1}$$

where K is a Boltzman's constant (1.38 x 10^{-23} W/s/K), and B is the frequency bandwidth (Hz) in which the noise voltage is measured. It is well known that maximum power can be delivered to an external load by a generator with a given internal load when the impedance of the external load is a complex conjugate of the source impedance. From this, it can easily be shown that thermal noise power P_n delivered to this optimum load by the thermal noise source of resistance R at temperature T is given by

$$P_n = \frac{V_n^2}{4R} = \kappa TB \qquad \text{(watts)} \tag{3.2}$$

It must be noted that noise power does not depend on a particular value of resistance but only on absolute temperature T and frequency bandwidth B.

Hence, noise power density N_0 per unit frequency bandwidth (l Hz) is given by

$$N_0 = kT \quad \text{(watts/Hz)} \tag{3.3}$$

It is convenient to use decibel (dB) expressions in calculating the parameters of mobile satellite communications such as antenna gain, noise power, free-space propagation attennuation, and so on. In rhis book, log(A) is denoted by the symbol [A] . Hence, N_0 is denored in decibels as follows:

$$[N_0] = [k] + [T]$$

$$= 10 \cdot \log(k) + 10\log(T)$$

$$= 10\log(1.38 \times 10^{-23}) + 10\log(T)$$

$$= -228.6 + 10\log(T) \qquad \text{(dBW/Hz)} \qquad (3.4)$$

Example 3.1

Noise power density generated by a resistor at a temperature of 27°C is calcullated by (3.4) as

$$[N_0] = -228.6 + 10\log(273 + 27)$$

$$= -228.6 + 24.8$$

$$= -203.8 \qquad \text{(dBW/Hz)}$$

3.2.2 Noise Figure

The performance of electrical circuits or components are evaluated by the parameters of a noise figure (NF), which is defined by

$$NF = \frac{\dfrac{S_{\text{in}}}{N_{\text{in}}}}{\dfrac{S_{\text{out}}}{N_{\text{out}}}} = \frac{\dfrac{S_{\text{in}}}{\kappa T_0 B}}{\dfrac{GS_{\text{in}}}{G(\kappa T_0 B + \kappa T_{\text{in}} B)}} = 1 + \frac{T_{\text{in}}}{T_0} \qquad (3.5)$$

where S_{in} and N_{in} denote the power of signal and noise, respectively, at the input port of the circuit and S_{out} and N_{out} denote the same at the output port. The letters G and B denote the gain and frequency bandwidth of a circuit, as shown in Figure 3 .2.

The T_o denotes the physical temperature of circumstances in which the circuit is immersed, and T_{in} denotes the equivalent input noise temperature, which is an equivalent value of noise temperature at the input port of thermal noise generated in the circuit. The noise figure is frequently described in decibel units as follows:

$$[NF] = 10 \cdot \log(1 + T_{in}/T_o) \qquad \text{(dB)} \qquad 3.6$$

When a noise figure is given, T_{in} is calculated as

$$[NF] = To(l0\ 10^{NF/10} - l) \qquad (K) \qquad 3.7$$

Example 3.2

l. When T_{in} = 400K, and T_0 = 300K, NF is given by (3.6) as follows:

[NF] = 10 log(l + 400/300) = 3.7 (dB)

2. When NF= 4 dB, and T_0 = 290K, T_{in} is given by (3.7) as fellows:

$T_{in} = 290(10^{4/10}-1) = 438.4$ (K)

Figure 3.3 shows the relation between noise figure in decibels and equivalent noise temperatures when T_0 = 300K.

In the same manner as the noise figure, if a circuir or a feed line has loss L_f as shown in Figure 3.4, L_f can be defined by

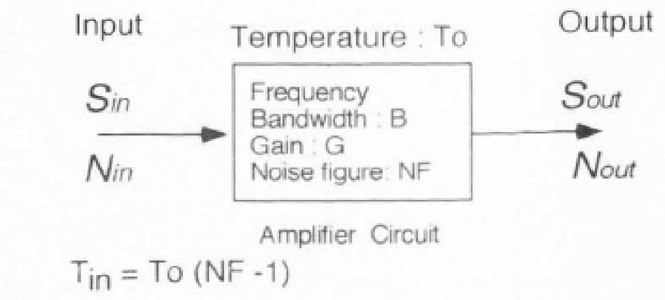


Figure 3.2 Noise figure of an amplifier circuit.

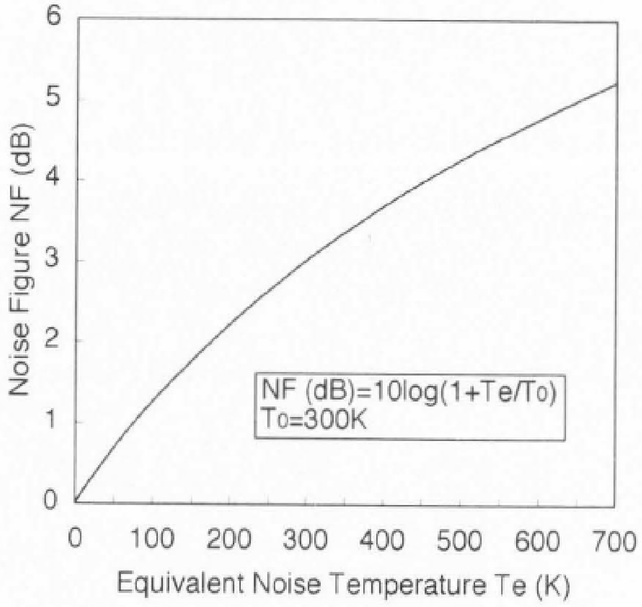


Figure 3.3 Noise figure and equivalent noise temperature.

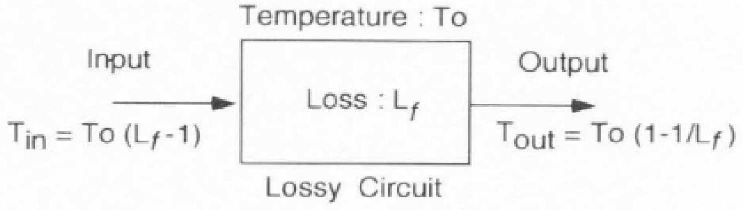


Figure 3.4 Equivalent noise temperature for a loss circuit.

$$L_f = \frac{\dfrac{S_{\text{in}}}{N_{\text{in}}}}{\dfrac{S_{\text{out}}}{N_{\text{out}}}}$$

$$= \frac{\dfrac{S_{\text{in}}}{\kappa T_0 B}}{\dfrac{\dfrac{1}{L_f} S_{\text{in}}}{\dfrac{1}{L_f}(\kappa T_0 B + \kappa T_{\text{in}} B)}}$$

$$= \frac{T_0 + T_{\text{in}}}{T_0} \tag{3.8}$$

$$\therefore \; T_{\text{in}} = T_0(L_f - 1) \tag{3.9}$$

Therefore, an equivalent noise remperature ar output port Tour can be obtained by dividing lin by L_f as follows:

$$T_{\text{out}} = T_0\left(1 - \frac{1}{L_f}\right) \tag{3.10}$$

Example 3.3

When a circuit has a loss of 3 dB and T0 = 300K, equivalent input and output noise temperatures can be given by (3 .9) and (3.10), respectively, as follows:

3.2.3 Noise Temperature of a Receiver

In general, a receiving system has cascade connection of loss and amplification circuits as shown in Figure 3.5. A signal from a satellite is received by an antenna with a gain of

$$T_{\text{in}} = 300 \cdot \left(10^{\frac{3}{10}} - 1\right) = 298.6 \text{ (K)}$$

$$T_{\text{out}} = 300 \cdot \left(1 - \frac{1}{10^{\frac{3}{10}}}\right) = 149.6 \text{ (K)}$$

G, and the equivalent antenna noise temperature at an output port of the antena

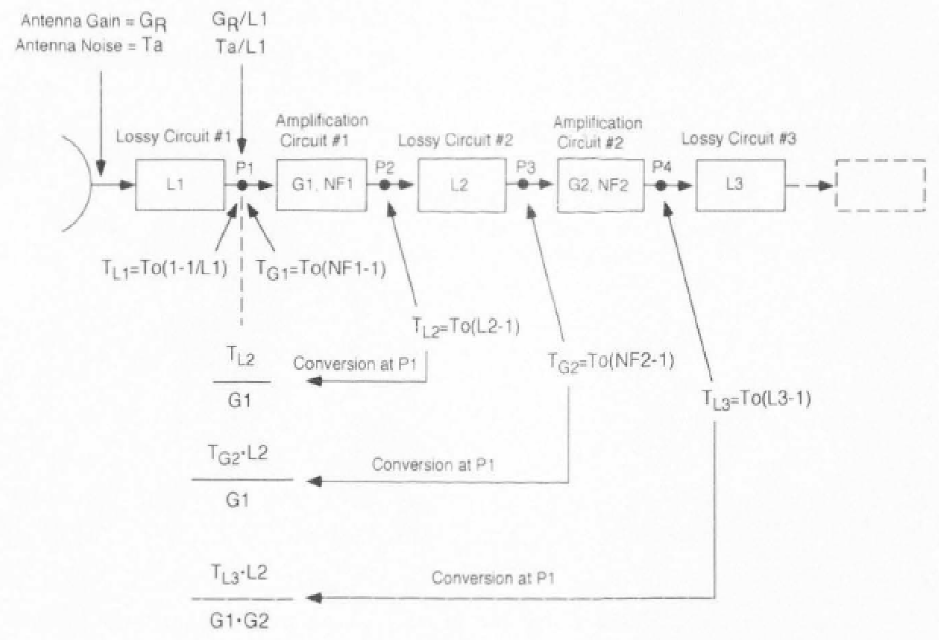


Figure 3.5 Cascade connection of loss and amplifying circuits.

output port of the an ten na is denoted as T_a. Losses in loss circuits #n (n is l, 2, 3, ...) are denoted as L_n, and gains and noise figures of amplification circuits #n are denoted as Gn and NF,1' respectively. In Figure 3.5, equivalent noise temperature TLI at output port P1 of loss circuit #l is described as $T_{L1} = T_0(l – l/ L_l)$. When amplification circuit #l is concerned, equivalent noise temperature T_{Gl} at input port Pl of circuit #l can be described as $T_{Gl} = T_0(NF- 1)$. Next, when loss circuit #2 is concerned, equivalent noise temperature T_{L2} at input port P2 of circuit #2 can be described as $T_{L2} = T_0(L2 - 1)$. In the same manner, equivalent noise temperature at the input port of each circuit can be described as shown in Figure 3.5. If equivalent noise temperature T_{L2} is measured at input port Pl of amplification circuit #l, this can be expressed by dividing it by gain Gl of circuit #l. In the same manner, equivalent noise temperarure at the input port of each circuit can be converted to the equivalent input noise temperature at the input port of amplification circuit #l as shown in Figure 3.5.

Hence, the equivalent input noise temperature T_{in} of the whole circuit (the receiver) at input port Pl to amplifier #l can be described as follows:

$$T_i = T_{L1} + T_{G2} + \frac{T_{L2}}{G_1} + \frac{T_{G2} \cdot L_2}{G_1} + \frac{T_{L3} \cdot L_2}{G_1 \cdot G_2} + \ldots . \qquad (3.11)$$

If G_1»1 all terms after the third term can be neglected compared to the first and second terms of (3.11). Therefore, the noise performance of the first-stage amplifier and loss circuit is found to dominate the performance of the receiver.

Overall equivalent input noise temperature T_s at the input port of the receiver can be expressed as

$$\begin{aligned} T_s &= \frac{T_a}{L_1} + T_{L1} + T_{G1} \\ &= \frac{T_a}{L_f} + T_0\left(1 - \frac{1}{L_f}\right) + T_R \qquad (3.12) \end{aligned}$$

where TR denotes the equivalent input noise temperature of the first-stage amplifier of the receiver, which is usually called a low-noise amplifier (LNA), and L_f denotes the loss of a feed line between the anten na and the LNA. The value T_a denotes the equivalent antenna noise temperature and T_s is the system noise temperature. It must be noted that T_s depends on the measured point, and it can be usually expressed at the input port to the LNA.

3.2.4 Figure of Merit (GIT)

Regarding the antenna, gain G_R ar receiving frequencies and equivalent input noise ternperature T_a can be denoted a G_R /L1 (= G_s) and Ta/ L_1, respectively, at the input port of amplification circuit #l. The value G_s means system gain at the input port to the LNA. Consequently, the ratio of antenna gain to noise temperature at the input port to the LNA can be written as

$$\begin{aligned} \frac{G_s}{T_s} &= \frac{\dfrac{G_R}{L_f}}{\dfrac{T_a}{L_f} + T_0\left(1 - \dfrac{1}{L_f}\right) + T_R} \\ &= \frac{G_R}{T_a + T_0(L_f - 1) + T_R L_f} \qquad (3.13) \end{aligned}$$

where G_s/ T_s is sometimes simply described as G/T (G over T). The G/T is an essential parameter of a receiver.

Figure 3.6 shows the relation between G/T and feeder loss in a 15-dBi antenna, which presents typical antenna gain in upcoming mobile satelite communication systems. Although T_a depends on factors such as frequency and beamwidth, a typical value is about 80K to l00K in the L band. The value L_F is a total loss of feed lines and components such as diplexers, cables, and phase shifters if a phased array antenna is used.

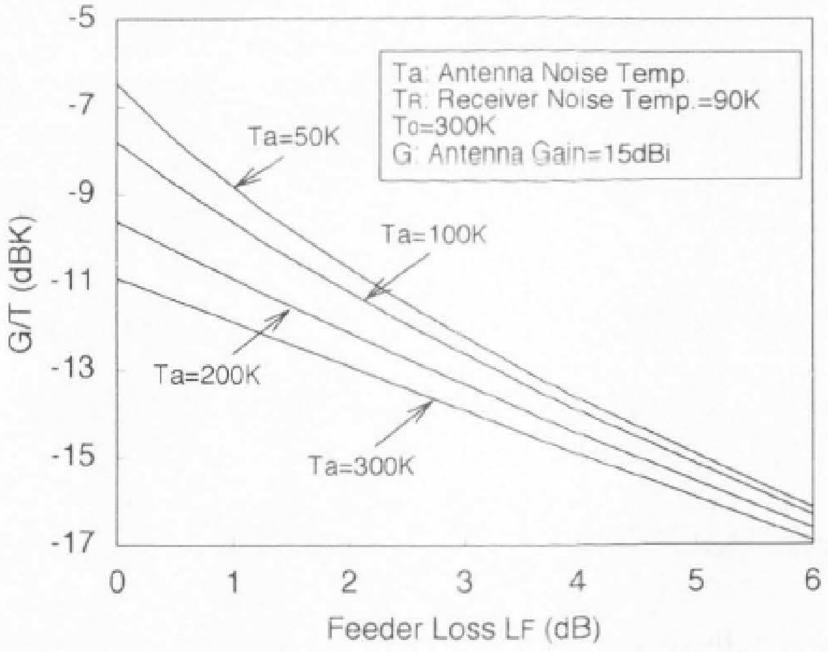


Figure 3.6 Relation between *G/T* and feeder loss. Antenna gain is 15 dBi.

3.3 Relation Between Transmitted and Received Power

The previous section showed that the sensitivity of a receiver is determined by G/T Next, we will consider what amount of power is available at the receiver.

Figure 3.7 shows the relation between transmitted and received power.

Although a perfect omnidirectional pattern in three dimensions can never be achieved, the concept of such an idea! antenna is very useful in theoretical analysis. If a transmitting antenna has an ideal isotropic radiation pattern in three dimensions, the power density on the spherical surface is

$$P_D = \frac{P_T}{4\pi d^2} \qquad (\text{watts/m}^2) \qquad (3.14)$$

where Pr and d denore the transmitted power and the distance between the transmitting and receiving antennas. If the transmitting anrenna has a gain of G_r, the power density of (3.14) can be written as

$$P_D = \frac{G_T \cdot P_T}{4\pi d^2} \qquad (\text{watts/m}^2) \qquad (3.15)$$

where G_T/P_T are considered to be the radiation power transmitted by an ideal omnidirectiona1 antenna. Therefore, this term is considered as an effective (or equivalent) isotropically radiated power (EIRP), and it is expressed as follows in antilogarithm and decibel expressions, respectively:

$$\text{EIRP} = G_T \cdot P_T \qquad (\text{watts}) \qquad (3.16)$$

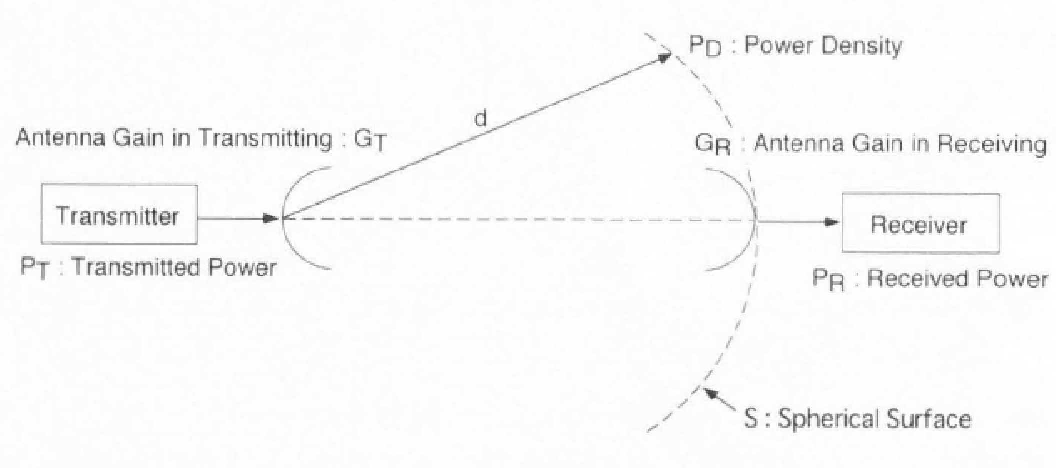


Figure 3.7 Concept of relation between transmitted and received power.

$$[\text{EIRP}] = [G_T] + [P_T] \qquad (\text{dBW}) \qquad (3.17)$$

EIRP is frequently used, and is an important concept in sate!lite communication systems to show the capab ilities of transmission.

Then, power P_R received by the receiving amenna, which has physical aperture area A and aperture efficiency η, is

$$\begin{aligned}P_R &= \frac{G_T \cdot P_T}{4\pi d^2} \cdot A \cdot \eta \\ &= \frac{\lambda^2}{(4\pi d)^2} \cdot (G_T \cdot P_T) \cdot G_R \\ &= \frac{(G_T \cdot P_T) \cdot G_R}{\left(\frac{4\pi d}{\lambda}\right)^2} \\ &= \frac{\mathrm{EIRP} \cdot G_R}{L_P} \qquad (3.18)\end{aligned}$$

Consequemly, $A\eta$ denotes the effective aperture area of the antenna, which is related to G_R and wavelength λ of the frequency obtained by the following equation [2]:

$$A \cdot \eta = \frac{\lambda^2}{4\pi} \cdot G_R \qquad (3.19)$$

Example 3.4

The frequency is 1 500 MHz. The gain of an antenna whose diameter is 100 cm and aperture efficiency is 0.6 is given by (3. 19) as follows:

$$\begin{aligned}G_R &= \frac{4\pi}{\lambda^2} A \cdot \eta \\ &= 4\pi \times \left(\frac{1500 \times 10^6}{3 \times 10^8}\right)^2 \times \pi\left(\frac{1}{2}\right)^2 \times 0.6 = 148.0 \\ [G_R] &= 10\log(148.0) = 21.7\ (\mathrm{dB})\end{aligned}$$

Free-space propagation loss L_p is caused by geometrical attenuation in propagation from the transmitter to the receiver. Figure 3.8 shows free-space propagation loss in the decibel scale at 1.5 GHz (L band), 4 GHz (C band),

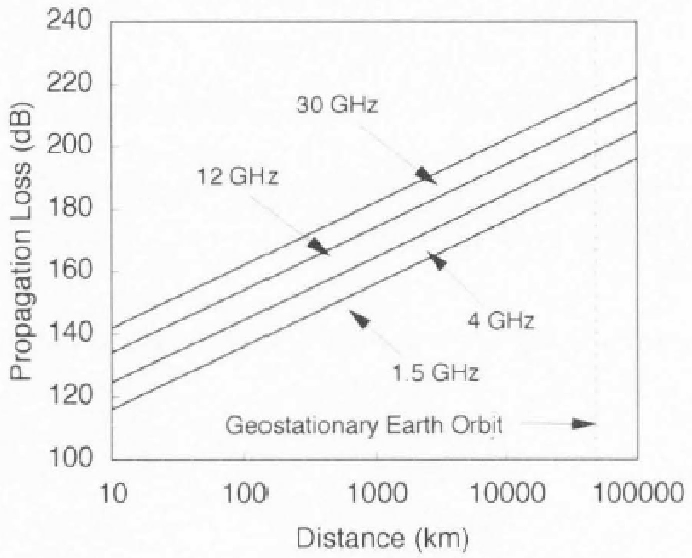


Figure 3.8 Free-space propagation loss.

12 GHz (Ku band), and 30 GHz (Ka band) . The GEO has a geostationary orbit, which is about 36,000 km above the equator,

$$L_P = \left(\frac{4\pi d}{\lambda}\right)^2 \tag{3.20}$$

Example 3.5

Free-space propagation loss at l 500 MHz from the geostationary satellite to the equator immediately below is calculated as follows:

$$L_P = \left(\frac{4\pi \times 36500 \times 10^3}{0.2}\right)^2 = (7.3\pi \times 10^8)^2$$

$$[L_P] = 10\ \log(7.3\pi \times 10^8)^2 = 20\log(7.3\pi \times 10^8) = 20(8 + 1.36) = 187.2\ (\text{dB})$$

3.4 Signal-to-Noise Ratio (C/N_0) in Satellite Communication Links

The radio frequency stages of an Earth station and a satellite, in general, consist of an antenna, a feed line, a diplexer, a high-power amplifier (HPA), and allow-noise amplifier (LNA), as shown in Figure 3.9. In Figure 3.9, G_T and G_R denote antenna gains in transmitting and receiving, respectively, and P_{out} and P_T denote the output power of an LNA and input power to an antenna, respectively. T he other notations are the same as those shown in the previous section.

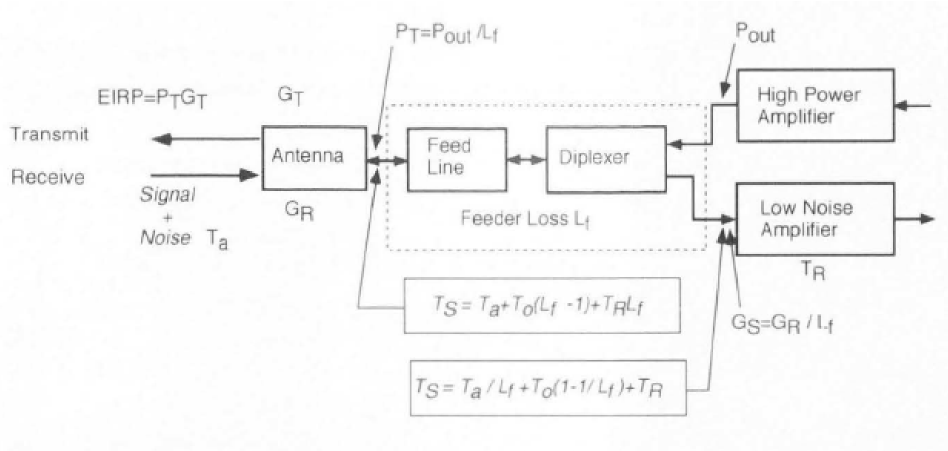


Figure 3.9 Block diagram of the RF stage of an Earth station.

The ratio of input signal power (C) to noise power (N) at the input point to the antenna can be written as follows using (3.2) and (3.16):

$$\begin{aligned}\frac{C}{N} &= \frac{\dfrac{\left(\dfrac{P_{\text{out}}}{L_f}\cdot G_T\right)\cdot G_R}{L_P}}{\kappa T_S B}\\ &= \frac{\dfrac{(P_T\cdot G_T)\cdot G_R}{L_P}}{\kappa T_S B}\\ &= \frac{\dfrac{\text{EIRP}\cdot G_R}{L_P}}{\kappa T_S B}\\ &= \frac{\text{EIRP}}{L_P}\left(\frac{G_R}{T_S}\right)\frac{1}{\kappa B} \qquad (3.21)\end{aligned}$$

When noise power density (C/N$_0$) is considered, (3. 19) can be written as

$$\frac{C}{N_0} = \frac{\text{EIRP}}{L_P}\left(\frac{G_R}{T_S}\right)\frac{1}{\kappa} \qquad (3.22)$$

Equations (3.21) and (3.22) are basie equations that show the quality of receiving signals from a satellite ro an Earth station, whose path is called a downlink.

Equation (3.20) can be written as follows in decibel expression:

$$\left[\frac{C}{N_0}\right] = [P_{\text{out}}] - [L_f] + [G_T] - [L_P] + [G_R] - [T_S] - [\kappa]$$

$$= [\mathrm{EIRP}] - [L_P] + [G_R] - [T_S] + 228.6 \qquad (3.23)$$

$$= [\mathrm{EIRP}] - [L_P] + \left[\frac{G_R}{T_S}\right] + 228.6 \text{ (dBHz)}$$

Equation (3.23) gives us an insight into the channel quality of a down link. The transmitted power (EIRP) is attenuated by free-space propagation (Lp) from a satellite to the Earth, amplified by receiving antenna gain (G_R), and attenuated by system noise (Ts). The channel quality of an uplink from the Earth to a satellite is expressed the same as (3.21-23).

In digital communications, the required C/No is determined by the bit error rate (BER) of the required quality of communication channels.

Example 3.6

A geostationary satellite transmits a signal at 1500 MHz to a mobile Earth station on the equator immediately under it. The parameters are as follows:

Satellite transmission power (l W)	0	dBW
Satellite antenna gain (D = 1m)	21.7 dBi	Example 3.4
Propagarion loss (d = 36 000 km)	187.2 dB	Example 3.5
Mobile antenna gain (D = 40 cm, η = 0.8)	15.0 dBi	
System noise temperature of a mobile Earth station (about 300K)	24.8 dBK	

Here, C/N_0 can be given by (3.21) as follows:

C/N_0 =0+ 21 7 - 187.2 + 15- 24.8 + 228.6 = 53.3 (dBHz)

In the above, we have considered channel quality in a downlink and an uplink separately. What about total channel quality from the base station to the mobile Earth station through the satellite? This can be easily understood by knowing that thermal noise, which is generated in an uplink and a downlink, is linearly added step by step. In general, interference noise, which is generated in the system from other systems, is added to the thermal noise. The total ($C/No)_T$ is given by

$$\left(\frac{C}{N_0}\right)_T = \frac{C}{(N_0)_U + (N_0)_D + I_0} \qquad (3.24)$$

$$= \left\{ \frac{1}{\left(\frac{C}{N_0}\right)_U} + \frac{1}{\left(\frac{C}{N_0}\right)_D} + \frac{1}{\left(\frac{C}{I_0}\right)} \right\}^{-1}$$

where I_0 denotes the power density of inrerference noise and U and D denote the uplink and downlink, respectively.

In (3.24), if value (C/N_0) for one of the links is sufficiently small compared to the other values, for example $(C/N_0)_D$ << $(C/No)_U$ and $(C/No)_D$ << (C/I_0) ,total quality $(C/No)_T$ can be approximately be given by ($C/No)_D$· This means that the total quality of a communication channel will be dominated by the poorest quality communication link. Figure 3.10 shows a calculated example for the total (C!No) depending on the uplink (C/N_0) 5• The paramerers are downlink $(C/N_0)_s$ of 50 dB Hz, 55 dB Hz, and 60 dBHz. It can easily be understood in this case that the total channel quality is dominated by the poor downlink and the total channel quality never exceeds the downlink quality no matter how much uplink quality is increased.

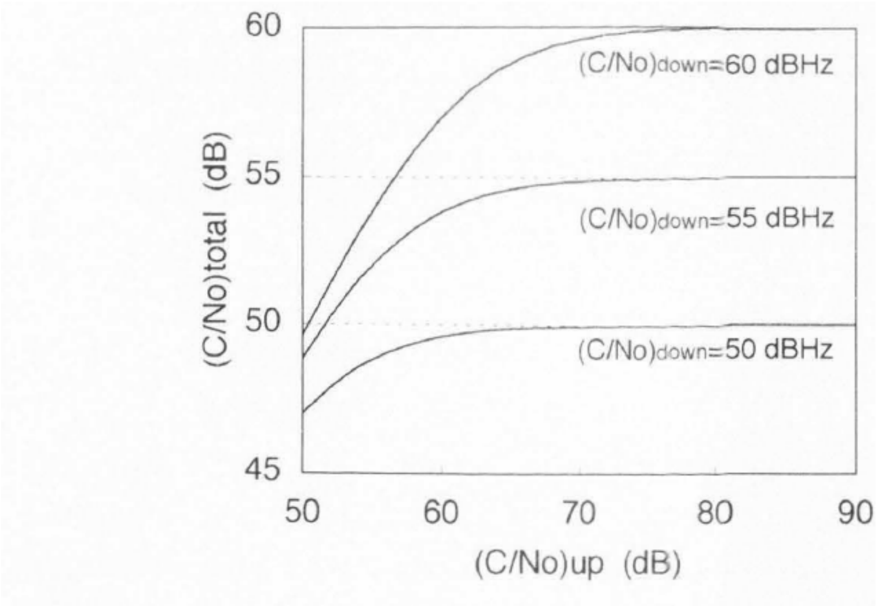


Figure 3.10 The relation between total (C/N_0) and uplink/downlink $(C/N_0)_S$.

Example 3.7

Consider the case of aeronautical satellite communication experiments using the ETS-V satellite [3]. The total C/N_0 for the forward cornmunication link from a gateway Earth station (GES at Kashirna) to an aircraft Earth station (AES at Anchorage) via satellite can be calculated by (3. 24). Interference noise has not been considered. Frequencies of 6 GHz and 1.5 GHz were used between the base Earth station and the satellite, and the satellite and the aircraft, respectively.

From a GES to a satellite (uplink)

GES EIRP	60.7 dBW
Propagation loss (6 GHz)	199.4 dB (d = 37,270 km)
Satellite antenna gain	21.7 dBi
Feeder loss	3.0 dB

Uplink total C = 60.7- 199.4 + 21.7- 3.0 = -120.0 (dBW)

$[N_0]$ = -228.6 + 10 log(300) = -203.8 (dBHz)

:. uplink (Cl No) u= -120.0 + 203.8 = 83.8 (dBHz)

From a satellite to an AES (downlink)

Satellite EIRP	30.5 dBW
Propagation loss (1.5 GHz)	88.5 dB (d = 41,097 km)
AES antenna gain	14.0 dBi
Antenna tracking error	0.5 dB
Feeder loss	3.0 dB

Downlink total C= 30.5- 188.5 + 14.0-0.5-3.0 = -147.5 (dBW)

$[N_0]$ = -228.6 + 10 log(300) = 203.8 (dBHz)

:. downlink $(C/N_0)_D$ = -147.5 + 203.8 = 56.3 (dBHz

Therefore, total $(C/N_0)_T$ is calculated as

$$\therefore \left(\frac{C}{N_0}\right)_T = \frac{1}{\dfrac{1}{\left(\dfrac{C}{N_0}\right)_U} + \dfrac{1}{\left(\dfrac{C}{N_0}\right)_D}} = \frac{1}{\dfrac{1}{10^{8.38}} + \dfrac{1}{10^{5.63}}} = 425{,}822.28$$

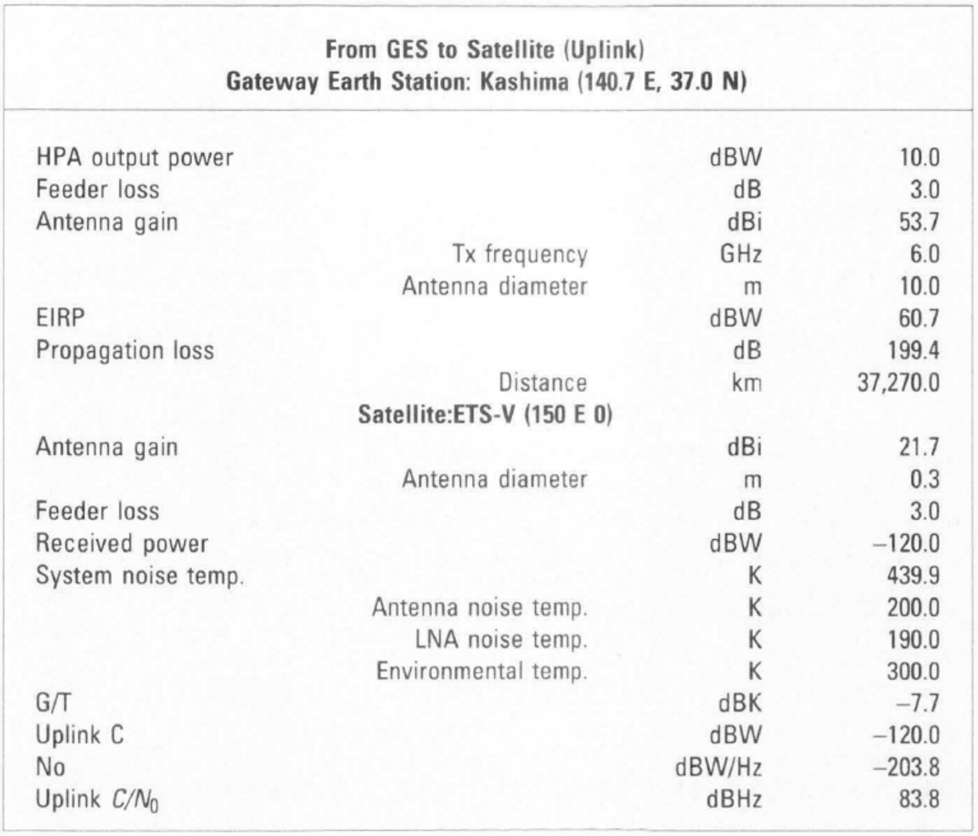
$$\therefore \left[\left(\frac{C}{N_0}\right)_T\right] = 10\,\log(425{,}822.28) = 56.3\ (\text{dBHz})$$

Ir is confirmed that the total quality of communication channels is dominated by the poorest communication link, which is a downlink in this example.

A more detailed calculation of the example link budget is in Table 3.1.

Table 3.1

Example of a Forward-Link Budget for Aeronautical Satellite Communications [3]



From GES to Satellite (Uplink) **Gateway Earth Station: Kashima (140.7 E, 37.0 N)**			
HPA output power		dBW	10.0
Feeder loss		dB	3.0
Antenna gain		dBi	53.7
	Tx frequency	GHz	6.0
	Antenna diameter	m	10.0
EIRP		dBW	60.7
Propagation loss		dB	199.4
	Distance	km	37,270.0
	Satellite:ETS-V (150 E 0)		
Antenna gain		dBi	21.7
	Antenna diameter	m	0.3
Feeder loss		dB	3.0
Received power		dBW	–120.0
System noise temp.		K	439.9
	Antenna noise temp.	K	200.0
	LNA noise temp.	K	190.0
	Environmental temp.	K	300.0
G/T		dBK	–7.7
Uplink C		dBW	–120.0
No		dBW/Hz	–203.8
Uplink C/N_0		dBHz	83.8

References

[l] Nyquist, H. , "Thermal Agitation of Electric Charge in Conductor," Physical Review, Vol. 32, July 1928, pp. 110- 11 3.

[2] Stutzman, W. L. and G. A. Thiele, Antenna Theory and Design, New York, NY: John Wiley & Sons, 198 1, p. 59.

[3] Radio Regulations, Articles ll , 27 and 28.

[4] Ippolito, L.J.: Satellite Communications Systems Engineering. Atmospheric Effects, Satellite Link Design and System Performance, Wiley, Singapore,2008

[5] Ohmori,S,-Wakana,H.- Kawase,S.:Mobile Satellite Communications, Artech House, Boston,London, 1998

4 VSAT technology

Introduction Satellites for communication services have evolved quite significantly in size and power since the launch of the first commercial satellites in 1965. This has permitted a consequent reduction in the size of earth stations, and hence their cost, with a consequent increase in number. Small stations, with antennas in the order of 1.2–1.8 m, have become very popular under the acronym VSAT, which stands for ’Very Small Aperture Terminals’. Such stations can easily be installed at the customer’s premises and, considering the inherent capability of a satellite to collect and broadcast signals over large areas, are being widely used to support a large range of services. Examples are broadcast and distribution services for data, image, audio and video, collection and monitoring for data, image and video, two-way interactive services for computer transactions, data base inquiry, internet access and voice communications.

The trend towards deregulation, which started in the United States, and progressed in other regions of the world, has triggered the success of VSAT networks for corporate applications. This illustrates that technology is not the only key to success. Indeed, VSAT networks have been installed and operated only in those regions of the world where demand existed for the kind of services that VSAT technology could support in a cost effective way, and also where the regulatory framework was supportive. This chapter aims to provide the framework of VSAT technology in the evolving context of satellite communications in terms of network configuration, services, economics, operational and regu- latory aspects.

4.1 VSAT network definition

VSAT, now a well established acronym for Very Small Aperture Terminal, was initially a trademark for a small earth station marketed in the 1980s by Telcom General in the USA. Its success as a generic name probably comes from the appealing association of its first letter V, which establishes a ‘victorious’ context, or may be perceived as a friendly sign of participation, and SAT which definitely establishes some reference to satellite communications.

In this chapter, the use of the word ‘terminal’ which appears in the clarification of the acronym will be replaced by ‘earth station’, or station for short, which is the more common designation in the field of satellite communications for the equipment assembly allowing reception from or transmission to a satellite. The word terminal will be used to designate the end user equipment (telephone set, facsimile machine, television set, computer, etc.) which generates or accepts the traffic that is conveyed within VSAT networks. This complies with regulatory texts, such as those of the International Telecommunications Union (ITU), where for instance equipment generating data traffic, such as computers, are named ‘Data Terminal Equipment’ (DTE). VSATs are one of the intermediary steps of the general trend in earth station size reduction that has been observed in satellite communications since the launch of the first communication satellites in the mid 1960s. Indeed, earth stations have evolved from the large INTELSAT Standard A earth stations equipped with antennas 30 m wide, to today’s receive-only stations with antennas as small as 60 cm for direct reception of television transmitted by broadcasting satellites, or handheld terminals for radiolocation such as the Global Postioning System (GPS) receivers. Present day handheld satellite phones (IRIDIUM, GLOBALSTAR) are pocket size. Figure 4.1 illustrates this trend. Therefore, VSATs are at the lower end of a product line which offers a large variety of communication services; at the upper end are large stations (often called trunking stations) which support large capacity satellite links. They are mainly used within international switching networks to support trunk telephony services between countries, possibly on different continents. Figure 4.2 illustrates how such stations collect traffic from end users via terrestrial links that are part of the public switched network of a given country. These stations are quite expensive, with costs in the range of $10 million, and require important civil works for their installation. Link capacities are in the range of a few thousand telephone channels, or equivalently about one hundred Mbs^{-1}. They are owned and operated by national telecom operators, such as the PTTs, or large private telecom companies. At the lower end are VSATs. These are small stations with antenna diameters less than 2.4 m, hence the name ‘small aperture’ which refers to the area of the antenna. Such stations cannot support satellite links with large capacities, but they are cheap, with manufacturing costs in the range of $1000 to $5000, and easy to install any where, on the roof of a building or on a parking lot. Installation costs are usually less than $2000. Therefore, VSATs are within the financial capabilities of small corporate companies, and can be used to set up rapidly small capacity satellite links in a flexible way. Capacities are of the order ofa few tens of kbs^{-1}, typically 56 or 64 kbs^{-1}.

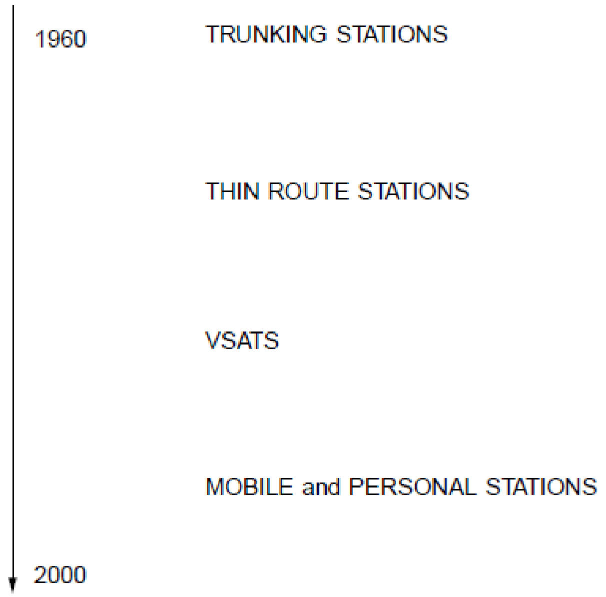


Figure 4.1 VSAT: a step towards earth station size reduction

The low cost of VSATs has made these very popular,with a market growth of the order of 20–25% per year in the nineties. There were about 50 000 VSATs in operation worldwide in 1990, and more than 600 000 twelve years later. This trend is likely to continue. Referring to transportation, VSATs are for information transport, the equivalent of personal cars for human transport, while the large earth stations mentioned earlier are like public buses or trains. At this point it is worth noting that VSATs, like personal cars, are available at one's premises. This avoids the need for using any public network links to access the earth station. Indeed, the user can directly plug into the VSAT equipment his own communication terminals such as a telephone or video set, personal computer, printer, etc. Therefore, VSATs appear as natural means to bypass public network operators by directly accessing satellite capacity. They are flexible tools for establishing private networks, for instance between the different sites of a company. Figure 4.3 illustrates this aspect by emphasising the positioning of VSATs near the user compared to trunking stations, which are located at the top level of the switching hierarchy of a switched public network.

The bypass opportunity offered by VSAT networks has not always been well accepted by national telecom operators as it could mean loss of revenue, as a result of business traffic being diverted from the public network. This has initiated conservative policies by national telecom operators opposing the deregulation of the communications sector. In some regions of the world, and particularly in Europe, this has been a strong restraint to the development of VSAT networks.

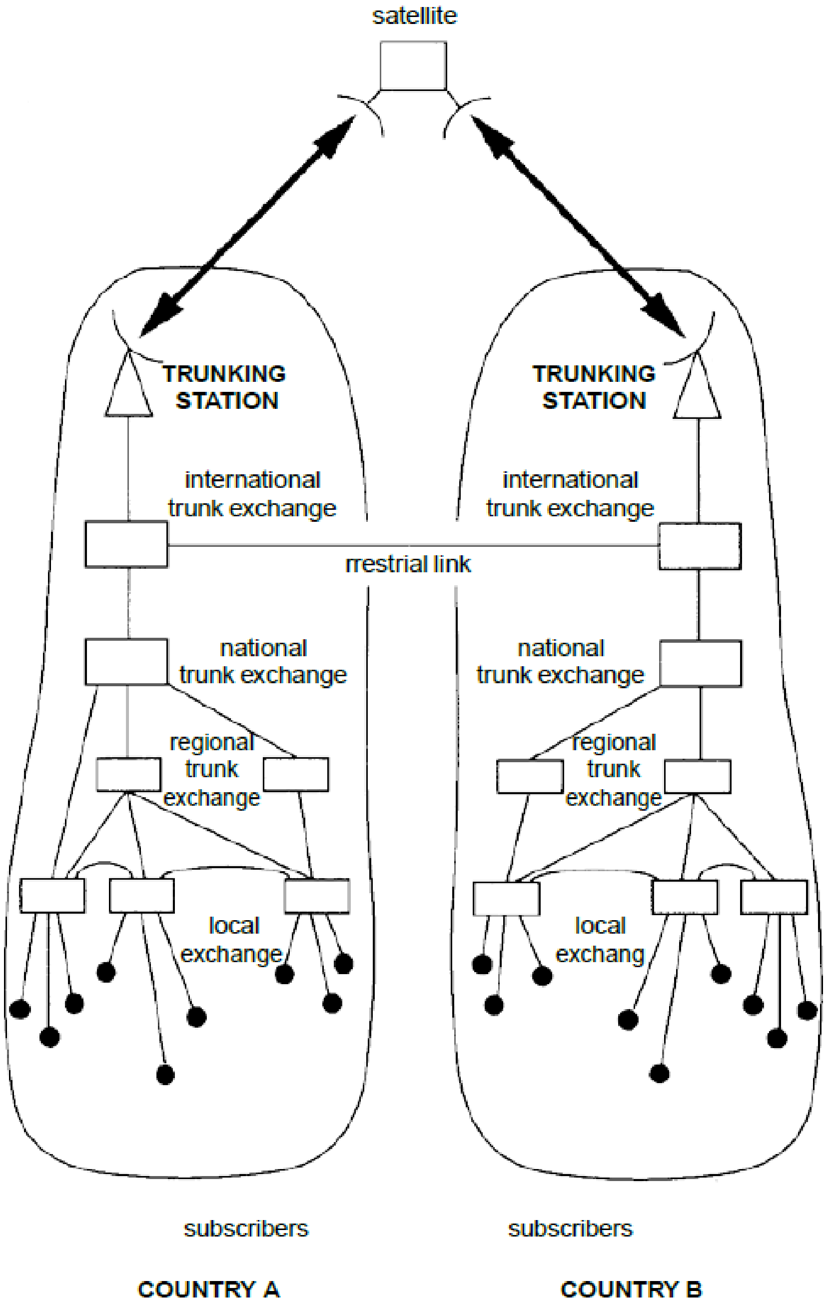


Figure 4.2 Trunking stations

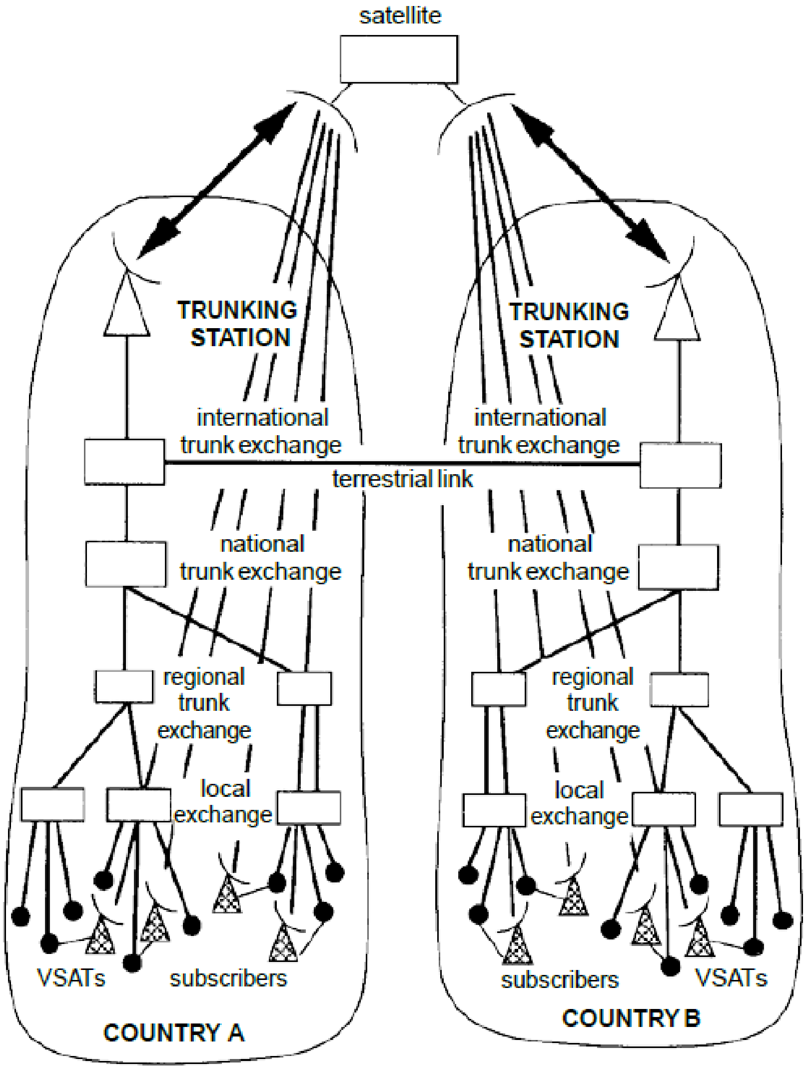


Figure 4.3 From trunking stations to VSATs

4.2 Configurations of VSAT Network

As illustrated in Figure 4.3, VSATs are connected by radio frequency (RF) links via a satellite, with a so-called uplink from the station to the satellite and a so-called downlink from the satellite to the station (Figure 4.4). The overall link from station to station, sometimes called hop, consists of an uplink and a downlink. A radio frequency link is a modulated carrier conveying information. Basically the satellite receives the uplinked carriers from the transmitting earth stations within the field of view of its receiving antenna, amplifies those carriers, translates their frequency to a lower band in order to avoid possible output/input interference, and transmits the amplified carriers to the stations located within the field of view of its transmitting antenna. Present VSAT networks use geostationary satellites, which are satellites orbiting in the equatorial plane of the earth at an altitude above the earth surface of 35 786 km. It has been shown in Chapter 2 that the orbit period at this altitude is equal to that of the rotation of the earth. As the satellite moves in its circular orbit in the same direction as the earth rotates, the satellite appears from any station on the ground as a fixed relay in the sky. Figure 4.5 illustrates this geometry. It should be noted that the distance from an earth station to the geostationary satellite induces a radio frequency carrier power attenuation of typically 200 dB on both uplink and downlink, and a propagation delay from earth station to earth station (hop delay) of about 0.25 s.

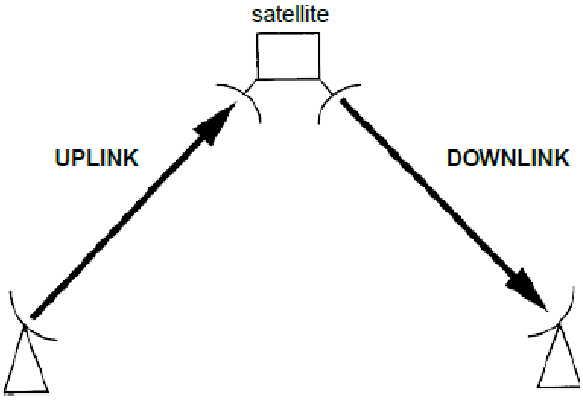


Figure 4.4 Definition of uplink and downlink

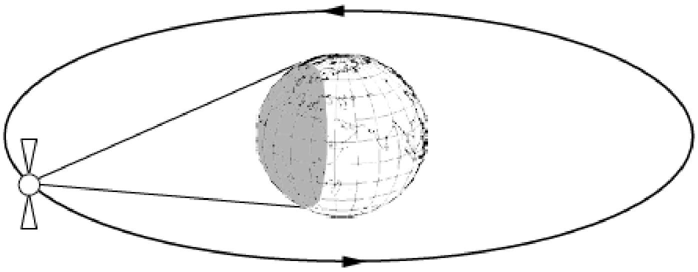


Figure 4.5 Geostationary satellite

As all VSATs are visible from the satellite, carriers can be relayed by the satellite from any VSAT to any other VSAT in the network, as illustrated by Figure 4.6. Regarding meshed VSAT networks, as shown in Figure 1.6, one must take into account the following limitations:

– typically 200 dB carrier power attenuation on the uplink and the downlink as a result of the distance to and from a geostation- ary satellite;

– limited satellite transponder radio frequency power, typically a few tens of watts;

– small size of the VSAT, which limits its transmitted power and its receiving sensitivity.

As a result of the above, it may well be that the demodulated signals at the receiving VSAT do not match the quality requested by the user terminals. Therefore direct links from VSAT to VSAT may not be acceptable.

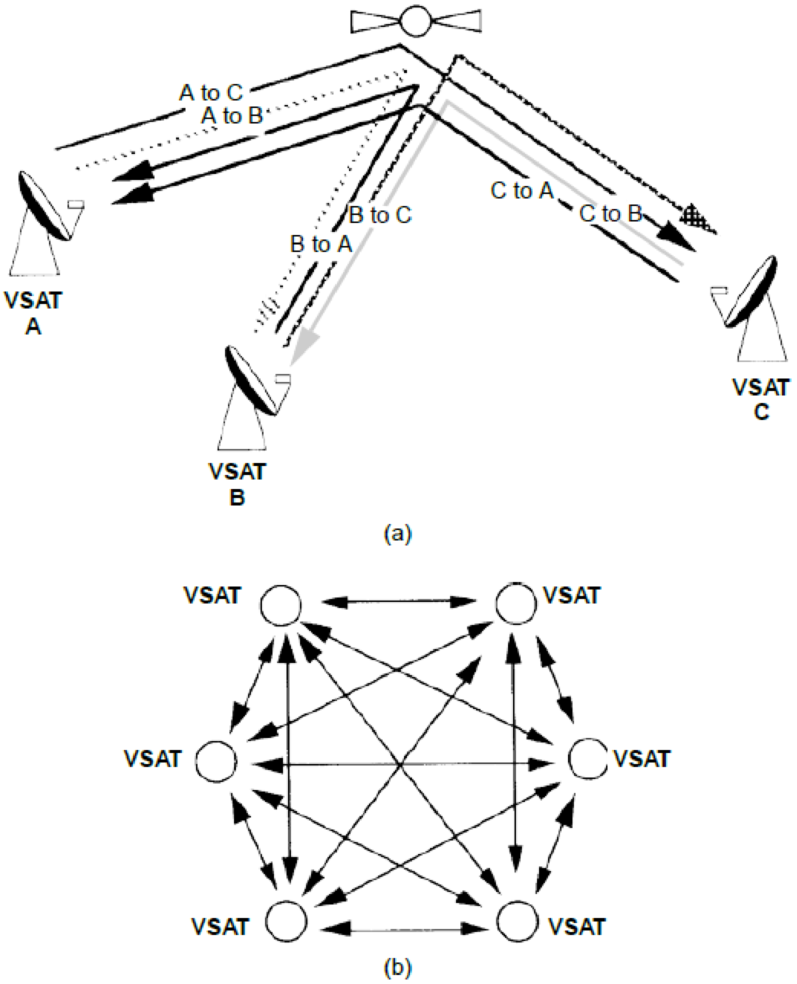


Figure 4.6 Meshed VSAT network. (a) Example with three VSATs (arrows represent information flow as conveyed by the carriers relayed by the satellite); (b) simplified representation for a larger number of VSATs (arrows represent bidirectional links made of two carriers travelling in opposite directions)

The solution then is to install in the network a station larger than a VSAT, called the hub. The hub station has a larger antenna size than that of a VSAT, say 4 m to 11 m, resulting in a higher gain than that of a typical VSAT antenna, and is equipped with a more powerful transmitter. As a result of its improved capability, the hub station is able to receive adequately all carriers transmitted by the VSATs, and to convey the desired information to all VSATs by means of its own transmitted carriers. The architecture of the network becomes star-shaped as shown in Figures 4.7 and 4.8. The links from the hub to the VSAT are named outbound links. Those from the VSAT to the hub are named inbound links. Both inbound and outbound links consist of two links, uplink and downlink, to and from the satellite, as illustrated in Figure 4.4. hub is located to remote sites where the receive-only VSATs are installed.

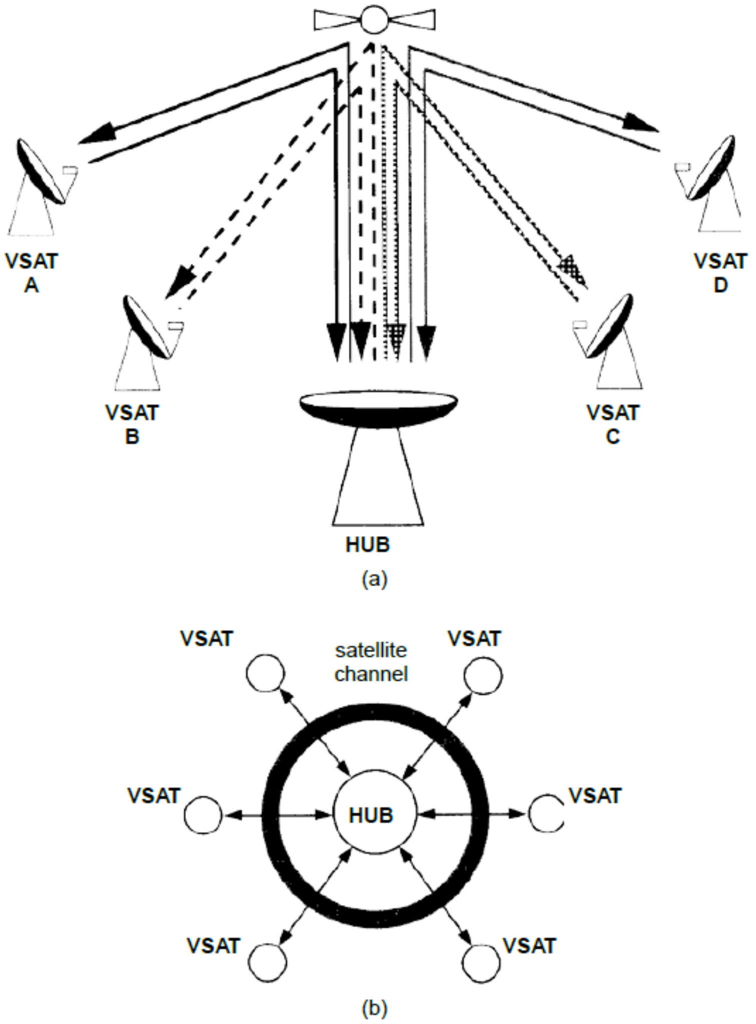


Figure 4.7 Two-way star-shaped VSAT network. (a) Example with four VSATs (arrows represent information flow as conveyed by the carriers relayed by the satellite); (b) simplified representation for a larger number of VSATs (arrows represent bidirectional links made of two carriers travelling in opposite directions)

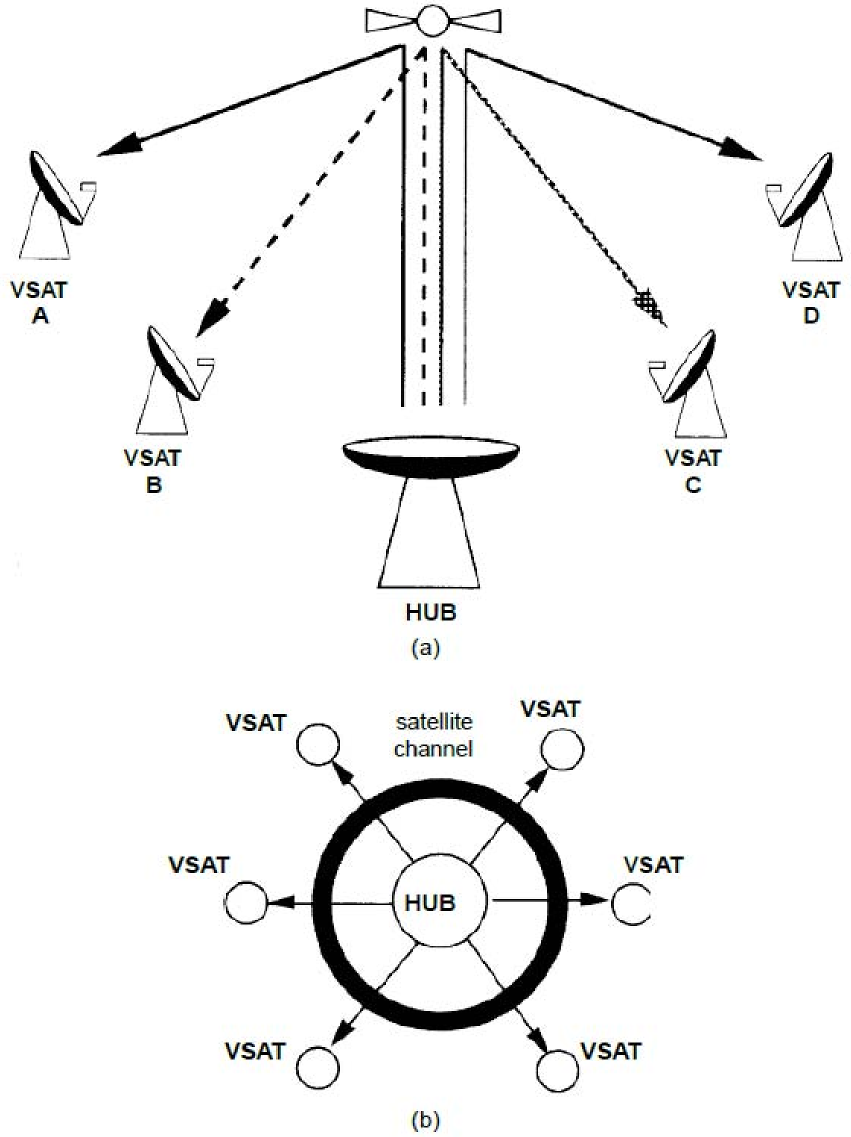


Figure 4.8 One-way star-shaped VSAT network. (a) Example with four VSATs (arrows represent information flow as conveyed by the outbound carriers relayed by the satellite); (b) simplified representation for a larger number of VSATs (arrows represent unidirectionnal links)

4.3 User Terminal Connectivity

User terminals are connected to VSATs and may be expected to communicate with one another thanks to the VSAT network. The two-way connectivity between user terminals can be achieved in two ways, depending on the VSAT network configuration:

– either thanks to direct links from VSAT to VSAT via satellite, should the link performance meet the requested quality. This applies in particular to the mesh configuration illustrated in Figure 4.6. The user terminal connectivity is illustrated in Figure 4.9;

– or by double hop links via satellite in a star-shaped network, with a first hop from VSAT to hub and then a second hop using the hub as a relay to the destination VSAT (as illustrated in Figure 4.10).

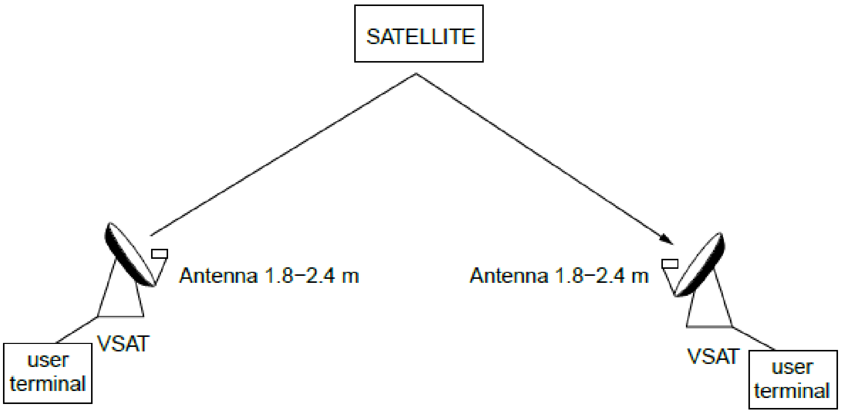


Figure 4.9 User terminal connectivity within meshed VSAT networks

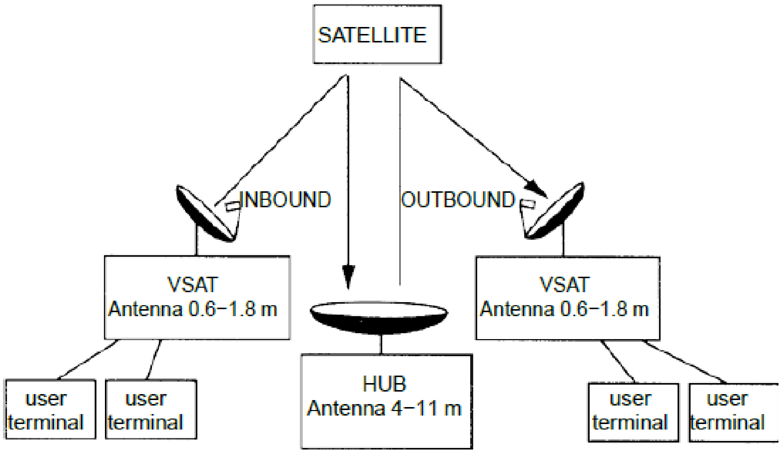


Figure 4.10 User terminal connectivity using the hub as a relay in star-shaped networks

– Comparing Figure 4.9 and 4.10 indicates a smaller antenna for VSATs within a star configuration than for VSATs in a meshed network. This is due to the linkage to a hub for VSATs in a star- shaped network, which provides more power on the outbound link and an improved ability to receive carriers transmitted by VSATs on the inbound link, compared to VSATs in a meshed network, as a result of the larger size of the hub.

– In conclusion, star-shaped networks are imposed by power limitations resulting from the small size of the VSAT earth stations, in conjunction with power limitation of the satellite transponder. This is particularly true when low cost VSATs are desired. Meshed networks are considered whenever such limitations do not hold. Meshed networks have the advantage of a reduced propagation delay (single hop delay is 0.25 s instead of 0.5 s for double hop) which is especially of interest for telephony services.

4.4 VSAT Network Applications and Types of Traffic

VSAT networks have both civilian and military applications. These will now be presented.

4.4.1 Civilian VSAT networks

4.4.1.1 Types of service

– As mentioned in the previous section, VSAT networks can be configured as one-way or two-way networks. Table 4.1 gives examples of services supported by VSAT networks according to these two classes.

– It can be noted that most of the services supported by two-way VSAT networks deal with interactive data traffic, where the user terminals are most often personal computers. The most notable exceptions are voice communications and satellite news gathering. Voice communications on VSAT network means telephony with possibly longer delays than those incurred on terrestrial lines, as a result of the long satellite path. Telephony services imply full connectivity, and delays are typically 0.25 s or 0.50 s depending on the selected network configuration, as mentioned above.

– Satellite news gathering (SNG) can be viewed as a temporary network using transportable VSATs, sometimes called ‘fly-away’stations, which are transported by car or aircraft and set up at alocation where news reporters can transmit video signals to a hub located near

– the company’s studio. Of course the service could be considered as inbound only, if it were not for the need to check the uplink from the remote site, and to be in touch by telephone with the staff at the studio. As fly-away VSATs are constantly transported, assembled and disassembled, they must be robust, lightweight and easy to install. Today they weigh typically 100 kg and can be installed in less than 20 minutes.

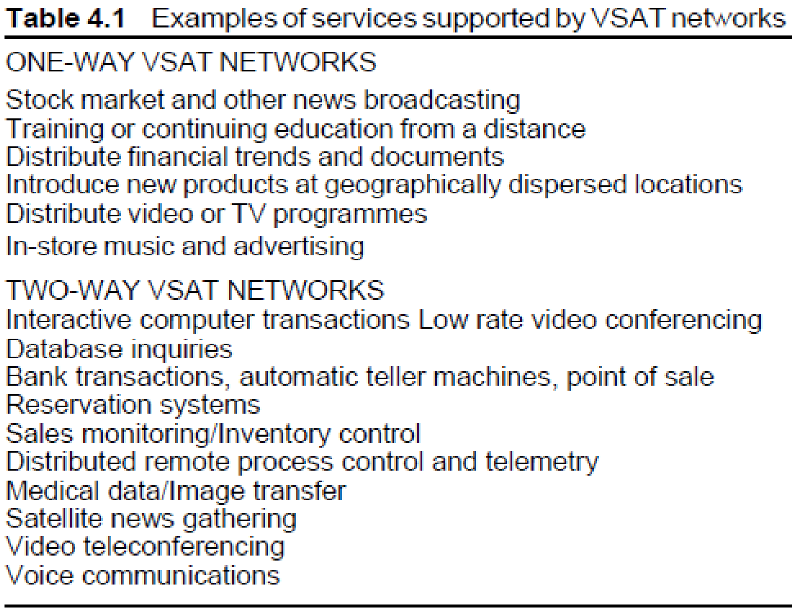


Table 4.1 Examples of services supported by VSAT networks

ONE-WAY VSAT NETWORKS

Stock market and other news broadcasting
Training or continuing education from a distance
Distribute financial trends and documents
Introduce new products at geographically dispersed locations
Distribute video or TV programmes
In-store music and advertising

TWO-WAY VSAT NETWORKS
Interactive computer transactions Low rate video conferencing
Database inquiries
Bank transactions, automatic teller machines, point of sale
Reservation systems
Sales monitoring/Inventory control
Distributed remote process control and telemetry
Medical data/Image transfer
Satellite news gathering
Video teleconferencing
Voice communications

4.4.1.2 Types of traffic

– Depending on the service, the traffic flow between the hub and the VSATs may have different characteristics and requirements.

– Data transfer or broadcasting, which belongs to the category of one- way services, typically displays file transfers of one to one hundred megabytes of data. This kind of service is not delay sensitive, but requires a high integrity of the data which are transferred. Examples of applications are computer download and distribution of data to remote sites.

– Interactive data is a two-way service corresponding to several transactions per minute and per terminal of single packets 50 to 250 bytes long on both inbound and outbound links. The required response time is typically a few seconds. Examples of applications are bank transactions and electronic funds transfer at point of sale.

Inquiry/response is a two-way service corresponding to several transactions per minute and terminal. Inbound packets (typically 30–100 bytes) are shorter than outbound packets (typically 500–2000 bytes). The required response time is typically a few seconds. Examples of applications are airline or hotel reservations and database enquiries. Supervisory control and data acquisition (SCADA) is a two-way service corresponding to one transaction per second or minute per terminal. Inbound packets (typically 100 bytes) are longer than outbound packets (typically 10 bytes). The required response time ranges from a few seconds to a few minutes. What is most important is the high data security level and the low power consumption of the terminal.

Table 4.2 Types of traffic

Type of traffic	Packet length		Required response time	Usage mode	Examples
	Inbound	Outbound			
Data transfer or broadcasting	not relevant	1–100 Mbytes	not delay sensitive	Usually during low traffic load periods (night time)	Computer download, distribution of data to remote sites
Interactive data	50–250 bytes	50–250 bytes	a few seconds	several transactions per minute per terminal	Bank transactions, electronic funds transfer at point of sale
Inquiry/response	30–100 bytes	500–2000 bytes	a few seconds	several transactions per minute per terminal	Airline reservations, database enquiries
Supervisory control and data acquisition (SCADA)	100 bytes	10 bytes	a few seconds/ minutes	one transaction per second/minute per terminal	Control/monitoring of pipelines and offshore platforms, electric utilities and water resources

Examples of applications are control and monitoring of pipelines, offshore platforms, electric utilities and water resources. Table 4.2 summarises the above discussion.

4.4.2 Military VSAT networks

VSAT networks have been adopted by many military forces in the world. Indeed the inherent flexibility in the deployment of VSATs makes them a valuable means of installing temporary communications links between small units in the battlefield and headquarters located near the hub. Moreover, the topology of a star-shaped network fits well into the natural information flow between field units and command base. Frequency bands are at X-band, with uplinks in the 7.9–8.4 GHz band and downlinks in the 7.25–7.75 GHz band. The military use VSAT must be a small, low weight, low power station that is easy to operate under battlefield conditions. As an example, the manpack station developed by the UK Defence Research Agency (DRA) for its Milpico VSAT military network is equipped with a 45 cm antenna, weighs less than 17 kg and can be set up within 90 seconds. It supports data and vocoded voice at 2.4 kbs−1. In order to do so, the hub stations need to be equipped with antennas as large as 14 m. Another key requirement is low probability of detection by hostile interceptors. Spread spectrum techniques are largely used.

4.5 VSAT NETWORKS: INVOLVED PARTIES

The applications of VSAT networks identified in the previous section clearly indicate that VSAT technology is appropriate to business or military applications. Reasons for this are the inherent flexibility of VSAT terminal technology. Which are the involved parties as far as corporate communications are concerned?

– The user is most often a company employee using office communication terminals such as personal computers, telephone sets or fax machines. On other occasions the terminal is transportable, as with satellite news gathering (SNG). Here the user is mostly interested in transmitting video to the company studio. The terminal may be fixed but not located in an office, as with supervisory control and data acquisition (SCADA) applications.

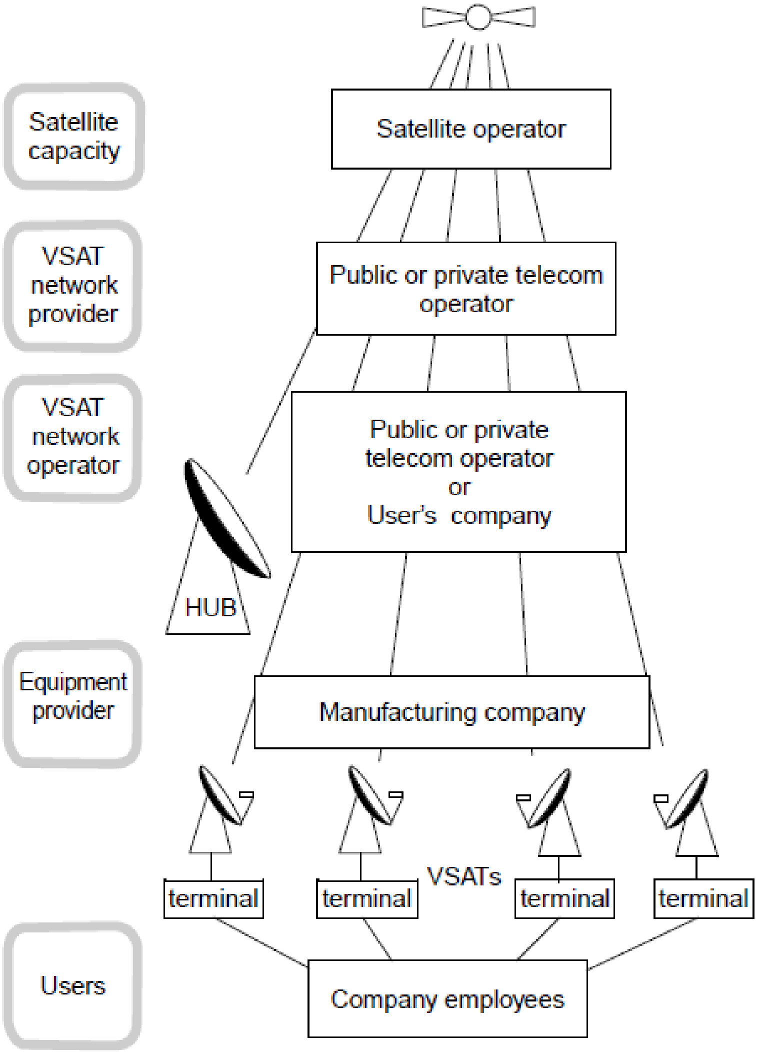


Figure 4.11 VSAT networks: involved parties

– The VSAT network operator may be the user's company itself, if the company owns the network, or it may be a telecom company (in many countries it is the national public telecom operator) who then leases the service. The VSAT network operator is then a customer to the network provider and/or the equipment provider. The VSAT network provider has the technical ability to dimension and install the network. It elaborates the network management system (NMS) and designs the corresponding software. Its inputs are the customer's needs, and its customers are network operators. The network provider may be a private company or a national telecom operator. The equipment provider sells the VSATs and/or the hub which it manufactures. It may be the network provider or a different party. For the VSAT network to work, some satellite capacity must be provided. The satellite may be owned by the user's company but this is a rare example of 'vertical integration', and most often the satellite is operated by a different party. This party may be a national or international private satellite operator. Figure 4.11 summarises the above discussion and it can serve as a convenient reference.

4.6 VSAT Network Options

4.6.1 Star or mesh

Section 4.2 introduced the two main architectures of a VSAT network: star and mesh. The question now is one architecture more appropriate than the other? The answer depends on three factors:

– the structure of information flow within the network;

– the requested link quality and capacity;

– the transmission delay.

These three aspects will now be discussed.

4.6.1.1 Structure of information flow

VSAT networks can support different types of application, and each has an optimum network configuration:

– Broadcasting: a central site distributes information to many remote sites with no back flow of information. Hence a star- shaped one-way network supports the service at the lowest cost.

– Corporate network: most often companies have a centralised structure with administration and management performed at a central site, and manufacturing or sales performed at sites scattered over a geographical area. Information from the remote sites needs to be gathered at the central site for decision making, and information from the central site (for example, relating to task sharing) has to be distributed to the remote ones. Such an information flow can be supported partially by a star-shaped one- way VSAT network, for instance for information distribution, or supported totally by a two-way star-shaped VSAT network. In the first case, VSATs need to be receive-only and are less expensive than in the latter case where interactivity is required, as this implies VSATs equipped with both transmit and receive equipment. Typically the cost of the transmitting equipment is two-thirds that of an interactive VSAT.

– Interactivity between distributed sites: other companies or organisations with a decentralised structure are more likely to comprise many sites interacting with one another. A meshed VSAT network using direct single hop connections from VSAT to VSAT is hence most desirable. The other option is a two-way star-shaped network with double hop connections from VSAT to VSAT via the hub.

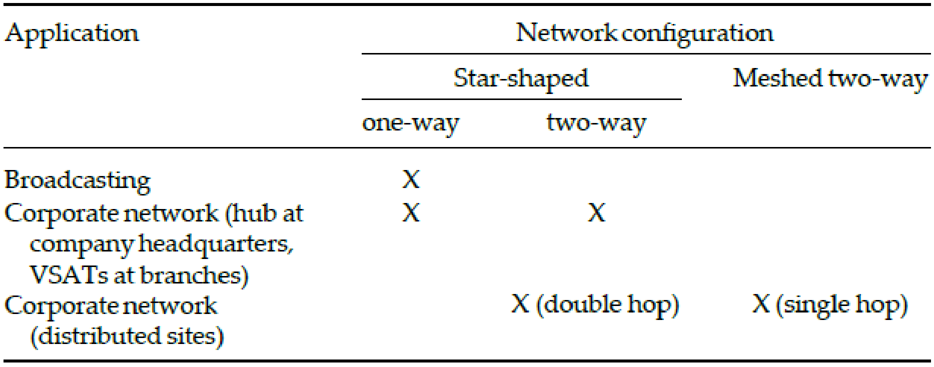
Table 4.3 summarises the above discussion.

4.6.1.2 Link quality and capacity

The link considered here is the link from the transmitting station to the receiving one. Such a link may comprise several parts. For instance a single hop link would comprise an uplink and a downlink (Figure 4.4), a double hop link would comprise two single hop links, one being inbound and the other outbound (Figure 4.10).

When dealing with link quality, one must refer to the quality of a given signal. Actually, two types of signal are involved: the modulated carrier at the input to the receiver and the baseband signals delivered to the user terminal once the carrier has been demodulated (Figure 4.13). The input to the receiver terminates the overall radio frequency link from the transmitting station to the receiving one, with its two link components, the uplink and the downlink. The earth station interface to the user terminal terminates the user-to-user baseband link from the output of the device generating bits (message source) to the input of the device to which those bits are transmitted (message sink).

Table 4.3 VSAT network configuration appropriate to a specific application



Application	Network configuration		
	Star-shaped		Meshed two-way
	one-way	two-way	
Broadcasting	X		
Corporate network (hub at company headquarters, VSATs at branches)	X	X	
Corporate network (distributed sites)		X (double hop)	X (single hop)

The link quality of the radio frequency link is measured by the $(C/N_0)_T$ ratio at the station receiver input, where C is the received carrier power and N_0 the power spectral density of noise.

The baseband link quality is measured by the information bit error rate (BER). It is conditioned by the E_b/N_0 value at the receiver input, where E_b(J) is the energy per information bit and N_0 (WHz^{-1}) is the noise power spectral density. As indicated in

Chapter 3, the E_b/N_0 ratio depends on the overall radio frequency link quality $(C/N_0)_T$ and the capacity of the link, measured by its information bit rate R_b (bs^{-1}):

$$\frac{E_b}{N_0} = \frac{(C/N_0)_T}{R_b} \quad (4.1)$$

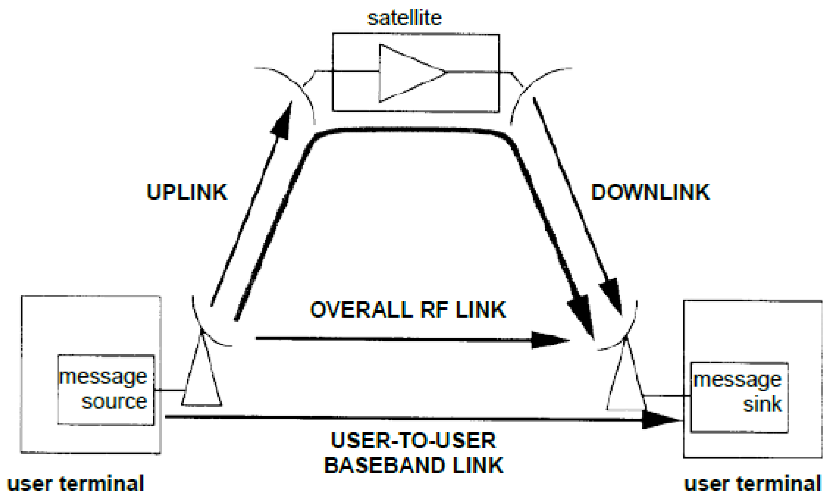


Figure 4.12 Overall radio frequency (RF) link and user-to-user baseband link

Figure 4.13 indicates the general trend which relates EIRP to G/T in a VSAT network, considering a given baseband signal quality in terms of constant BER. EIRP designates the effective isotropic radiated power of the transmitting equipment and G/T is the figure of merit of the receiving equipment. As can be seen from Figure 4.13, the double hop from VSAT to VSAT via the hub, when compared to a single hop, allows an increased link capacity without modifying the size of the VSATs. This option also involves a larger transmission delay.

4.6.1.3 Transmission delay

With a single hop link from VSAT to VSAT in a meshed network, the propagation delay is about 0.25 s. With a double hop from VSAT to VSAT via the hub, the propagation delay is twice as much, i.e. about 0.5 s. Double hop may be a problem for voice communications. How- ever it is not a severe problem for video or data transmission.

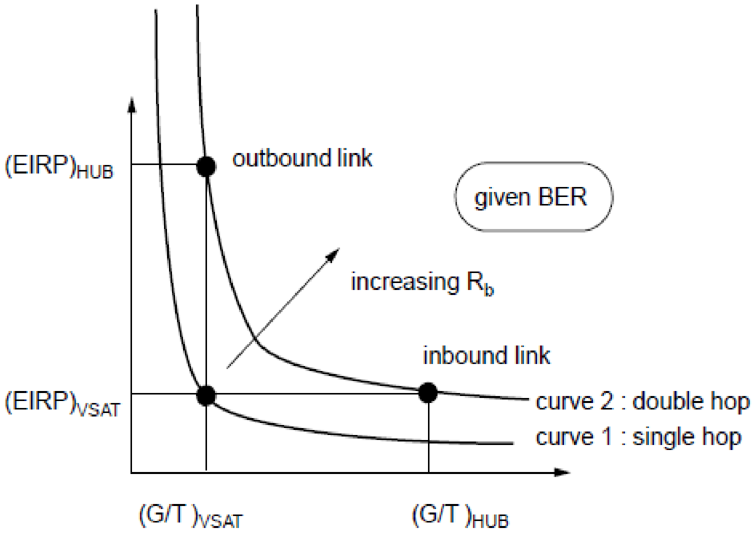


Figure 4.13 EIRP versus G/T in a VSAT network. Curve 1: single hop from VSAT to VSAT in a meshed network; Curve 2: double hop from VSAT to VSAT via the hub. Increased R_b means increased link capacity

Table 4.4 summarises the above discussion. Given the EIRP and G/T values for a VSAT, the designer can decide upon either a large delay from VSAT to VSAT and a larger capacity or a small delay and a lower capacity, by implementing either a star-shaped network, or a meshed one.

Table 4.4 Characteristics of star and mesh network configurations

	Network configuration	
	star-shaped(double hop)	meshed(single)
Capacity (given VSAT EIRP and *G/T*)	large	small
Delay (from VSAT to VSAT)	0.5 s	0.25 s

4.6.2 Frequency bands

VSAT networks are supposed to operate within the so-called 'fixed satellite service' (FSS) defined within the International Telecommu- nication Union (ITU). The only exception is when data is broadcast in association with broadcasting of television or audio programmes, within the so-called 'broadcasting satellite service' (BSS). The FSS covers all satellite communications between stations located while operating at given 'specified fixed points' of the earth.Transportable stations belong to this category, and hence the so-called 'fly-away' stations should use the same frequency bands as

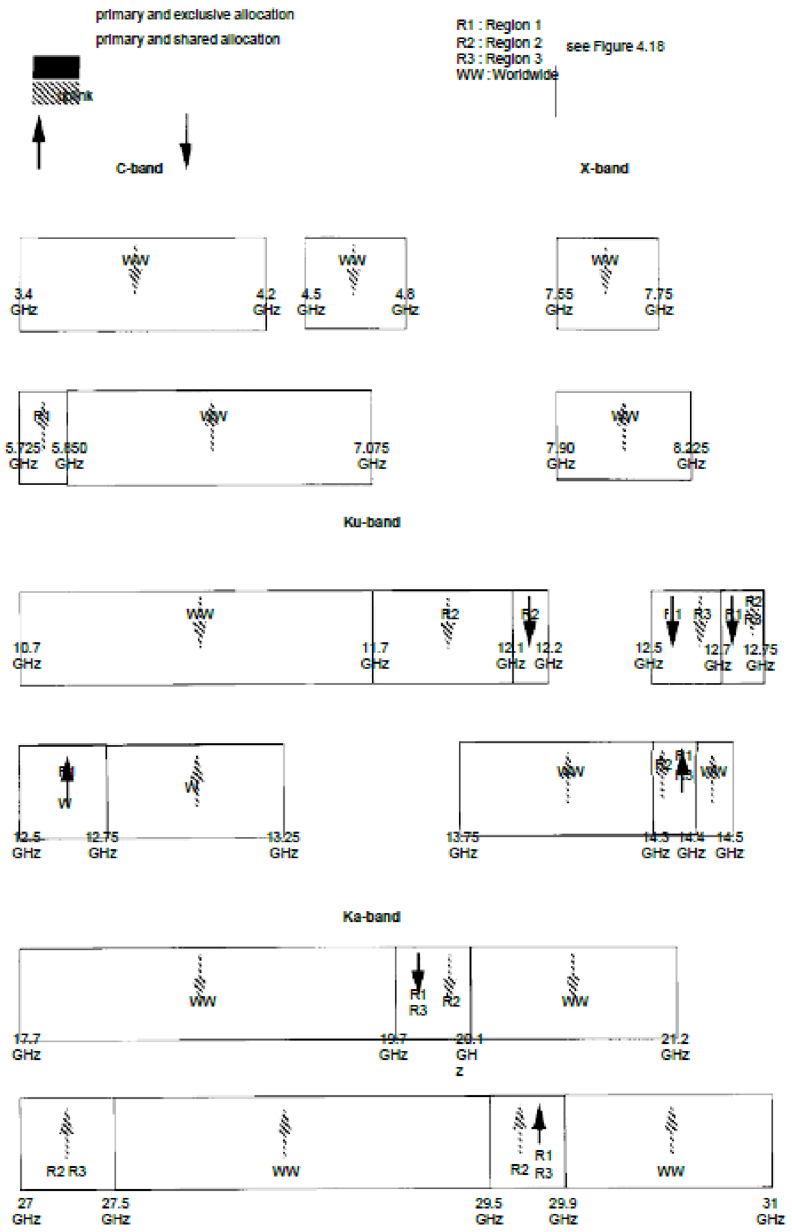


Figure 4.14 Frequency bands allocated to the fixed satellite service (FS
and usable for VSAT networks [ITU00]

fixed VSATs.The most commonly used bands for commercial applications are those allocated to the FSS at C-band and Ku-band. X-band is used by military systems. Few VSAT networks at Ka-band are commercial, most are experimental. Figure 4.14 gives the extension of these bands and provides some regulatory information. The figure displays uplinks and downlinks by means of arrows oriented upwards or downwards. The black arrows indicate a primary and exclusive allocation for FSS, which means in short that the FSS is protected against interference from any other service, which is then considered secondary. The striped arrows indicate a primary but shared allocation, which means that the allocated frequency bands can also be used by services other than

FSS with the same rights. Coordination is then mandatory, according to the procedure described in the ITU Radio Regulations. Figure 4.15 displays the geographical limits of regions R1, R2 and R3. As mentioned above, data may be carried in association with video signals within the frequency band allocated to the broadcasting satellite service. Possible bands are 11.7–12.5 GHz in regions 1 and 3, and 12.2–12.7 GHz in region 2, filling in the gaps of the bands represented in Figure 4.14 which deals with the fixed satellite service only.

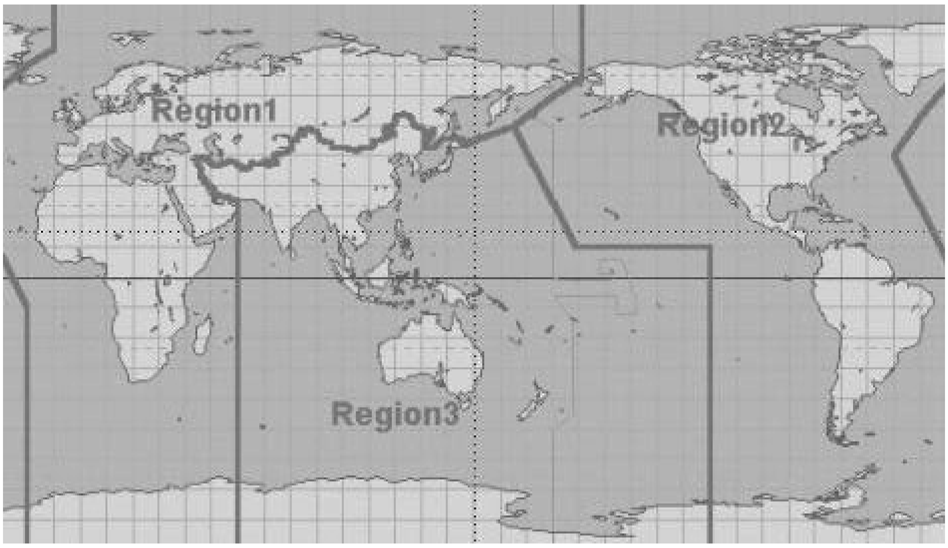


Figure 4.15 Regions 1, 2 and 3 in the world

The selection of a frequency band for operating a VSAT network depends first on the availability of satellites covering the region where the VSAT network is to be installed. To be considered next is the potential problem of interference. Interference designates unwanted carriers entering the receiving equipment along with the wanted ones. The unwanted carriers perturb the demodulator by acting as noise, adding to the natural thermal noise. Interference is a problem with VSATs because the small size of the antenna (small aperture) translates into a radiation pattern with a large beamwidth. Indeed as shown by equation (4.2) the half power beamwidth $\theta 3dB$ of an antenna relates to the product of its diameter by frequency as follows:

$$\theta_{3\mathrm{dB}} = \frac{70\ c}{D} \qquad \text{(degrees)} \qquad (4.2)$$

where D (m) is the diameter of the antenna, f (Hz) is the frequency, and c = 3 × 108 ms^{-1} is the velocity of light.

Therefore, the smaller the antenna diameter, the larger the beam- width, and the off-axis interfering carriers are more likely to be emitted or received with high antenna gain.

At this point it suffices to mention that interference is more likely to be a problem at C-band than at higher frequencies. There are two reasons for this: first, there is no primary and exclusive allocation to FSS at C-band. Second, given the earth station antenna diameter, interference is more important at C-band than at Ku-band, as the beamwidth is inversely proportional to the frequency, and thus is larger at C-band than at higher frequencies. To put this in perspective, equation (4.2) indicates, for a 1.8 m antenna, a beamwidth angle of 3◦ at 4 GHz, and only 1◦ at 12 GHz. This means that the receiving antenna is more likely to pick up carriers downlinked from satellites adjacent to the desired one at C-band than at Ku-band, especially as C-band satellites are many and hence nearer to each other. A typical angular separation for C-band satellites is 3◦, and is therefore comparable to VSAT antenna beamwidth. The same problem occurs on the uplink, where a small VSAT antenna projects carrier power in a larger angle at C-band than at Ku-band, and hence generates more interference on the uplink of adjacent satellite systems. However this is not a major issue as the transmit power of VSATs is weak. Finally it should be understood that C-band and parts of Ku- band are shared by terrestrial microwave relays, and this may be another source of interference. Ku-band offers a dedicated band free from any terrestrial microwave transmission (see black arrows in Figure 4.14), which is not the case for C-band. This simplifies the positioning of the VSAT and hub station as no coordination is implied. Figure 4.15 summarises the various interfering paths mentioned above. Where the small size of the antenna is at a premium, and should interference be too large, interference can be combated by using a modulation technique named spread spectrum, which consists of spreading the carrier in

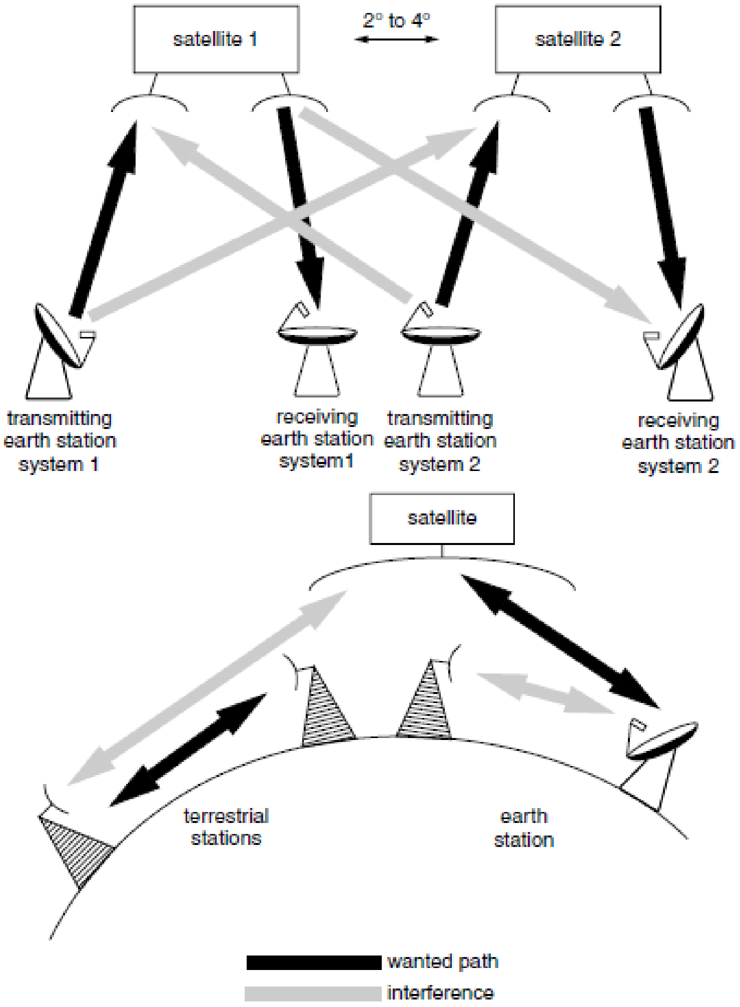
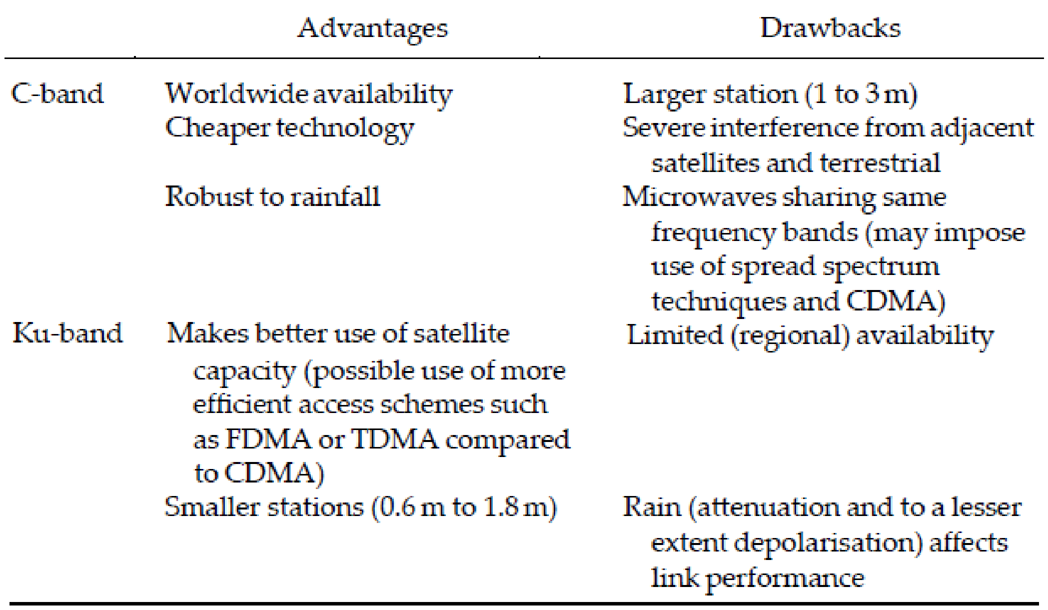


Figure 4.16 Interference paths

a much larger bandwidth than strictly required to transmit the information. This is an interesting technique as it provides not only interference protection but also potential for Code Division Multiple Access (CDMA) to a satellite channel. However, as a result of the greater utilised bandwidth, it is less bandwidth efficient compared to alternative multiple access techniques such as Frequency Division Multiple Access (FDMA) or Time Division Multiple Access (TDMA), which can be used where interference is not too severe.

One of the characteristics that must be taken into consideration is rain attenuation. Some power margin has to be incorporated into the network design to allow for the amount of power reduction of received carriers due to rain. This margin increases the

Table 4.5 Advantages and drawbacks of the most commonly available frequency bands



	Advantages	Drawbacks
C-band	Worldwide availability	Larger station (1 to 3 m)
	Cheaper technology	Severe interference from adjacent satellites and terrestrial
	Robust to rainfall	Microwaves sharing same frequency bands (may impose use of spread spectrum techniques and CDMA)
Ku-band	Makes better use of satellite capacity (possible use of more efficient access schemes such as FDMA or TDMA compared to CDMA)	Limited (regional) availability
	Smaller stations (0.6 m to 1.8 m)	Rain (attenuation and to a lesser extent depolarisation) affects link performance

cost of earth stations, and makes it prohibitive to provide enough carrier power during a large thunderstorm/downpour. At Ku-band, short (5–15 min) outages should be expected. Rain attenuation is higher at Ka-band and outages are likely to be longer when such systems develop. The great advantage of C-band is that it is not impaired by rain attenuation at all.

Finally, the cost of the equipment is another driving factor for choosing between C-band and Ku-band. Although C-band technology is cheaper, the larger size of the VSAT antenna for a similar performance makes the VSAT more expensive than at Ku-band. Table 4.5 summarises the advantages and drawbacks of the most commonly available frequency bands.

4.6.3 Huboptions

4.6.3.1 Dedicated large hub

A dedicated large hub (with antenna size in the range of 8–10 m) supports a full single network with possibly thousands of VSATs connected to it. The hub may be located at the customer’s organ- isation central site, with the host computer directly connected to it. It offers the customer full control of the network. In periods of expansion, changes in the network, or problems, this option may simplify the customer’s life. However, a dedicated hub represents the most expensive option and is only justified if its cost can be amortised over a sufficiently large number of VSATs in the network. The typical cost of a dedicated hub is in the region of $1 million.

4.6.3.2 Shared hub

Several separate networks may share a unique hub. With this option, hub services are leased to VSAT network operators. Hence the network operators are faced with minimum capital investment and this favours the initial implementation of a VSAT network. Therefore, shared hubs are most suitable for the smaller networks (less than 50 VSATs). However, sharing a hub has a number of drawbacks: Need for a connection from hub to host A shared hub facility is generally not colocated with the customer's host computer. Hence a backhaul circuit is needed to connect the hub to the host. The circuit may be a leased line or one provided by a terrestrial switched telephone network. This adds an extra cost to the VSAT network operation. Moreover, operational experience has shown the backhaul circuit to be the weakest link in the chain. Therefore this option means an increased failure risk. A possible way to mitigate this potential problem area would be using route diversity: for instance a microwave or satellite link could be used as a back-up for this interconnection. Possible limitation in future expansion A shared hub may impose an unforeseen capacity limitation, as the available capacity may be leased without notice to the other network operators sharing the hub. Guarantees should contractually be asked for by any network operator in this regard. 4.6.3.3 Mini-hub The mini-hub is a small hub (with antenna size in the range of 2–5 m) and a typical cost in the region of $100 000. It appeared as a result of the increased power from satellites and the improved performance of low-noise receiving equipment. The mini-hub has proved to be an attractive solution, as it retains the advantages of a dedicated hub at a reduced cost. It also eases possible installation problems in connection with downtown areas or communities with zoning restrictions, as a mini-hub entails a smaller antenna size and less rack mounted equipment than a large dedicated hub or even a shared hub. A typical mini-hub can support 300 to 400 remote VSATs.

4.7 VSAT Network Earth Stations

4.7.1 VSAT Station

Figure 4.17 illustrates the architecture of a VSAT station. As shown in the figure, a VSAT station is made of two separate sets of equipment:

the outdoor unit (ODU) and the indoor unit (IDU). The outdoor unit is the VSAT interface to the satellite, while the IDU is the interface to the customer's terminals or local area network (LAN).

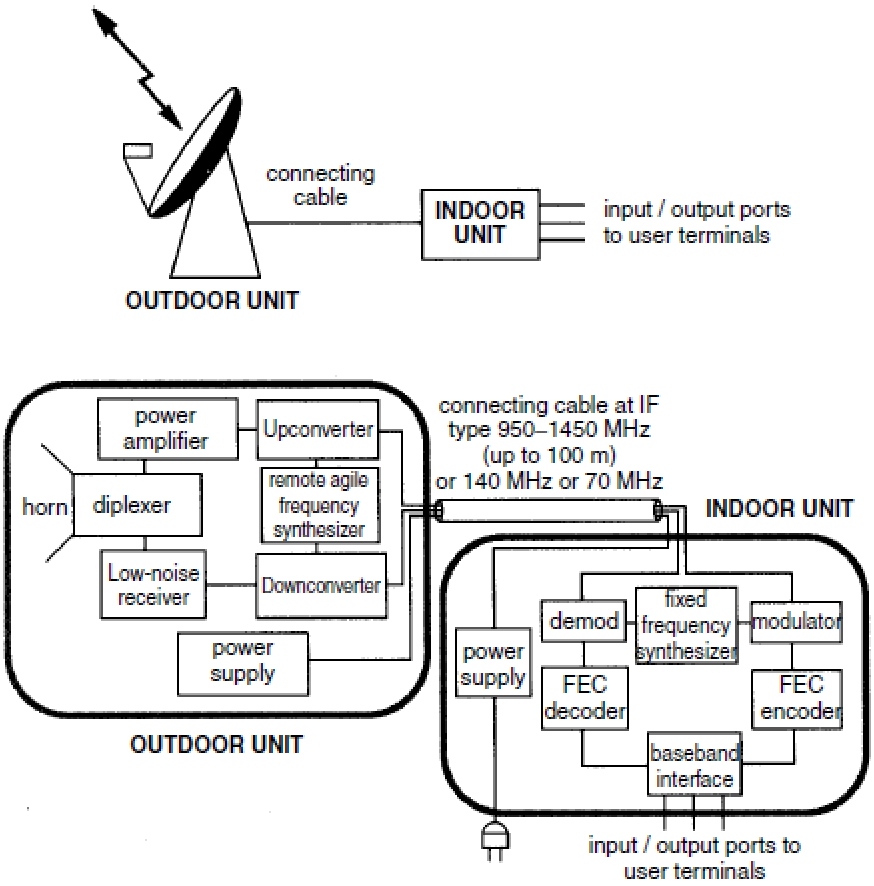


Figure 4.17 VSAT station equipment

4.7.1.1 The outdoor unit (ODU)

Figure 4.18 shows a photograph of an outdoor unit, with its antenna and the electronics package containing the transmitting amplifier, the low-noise receiver, the up- and down-converters and the frequency synthesiser. The photograph in Figure 4.19 provides a closer look at the electronics container. For a proper specification of the ODU, as an interface to the satellite, the following parameters are of importance:

– the transmit and receive frequency bands;

– the transmit and receive step size for adjusting the frequency of the transmitted carrier or for tuning to the received carrier frequency;

– the equivalent isotropic radiated power (EIRP), which determines the performance of the radio frequency uplink. The EIRP depends on the value of the antenna gain, and hence its size and transmit frequency, and on the transmitting amplifier output power ;

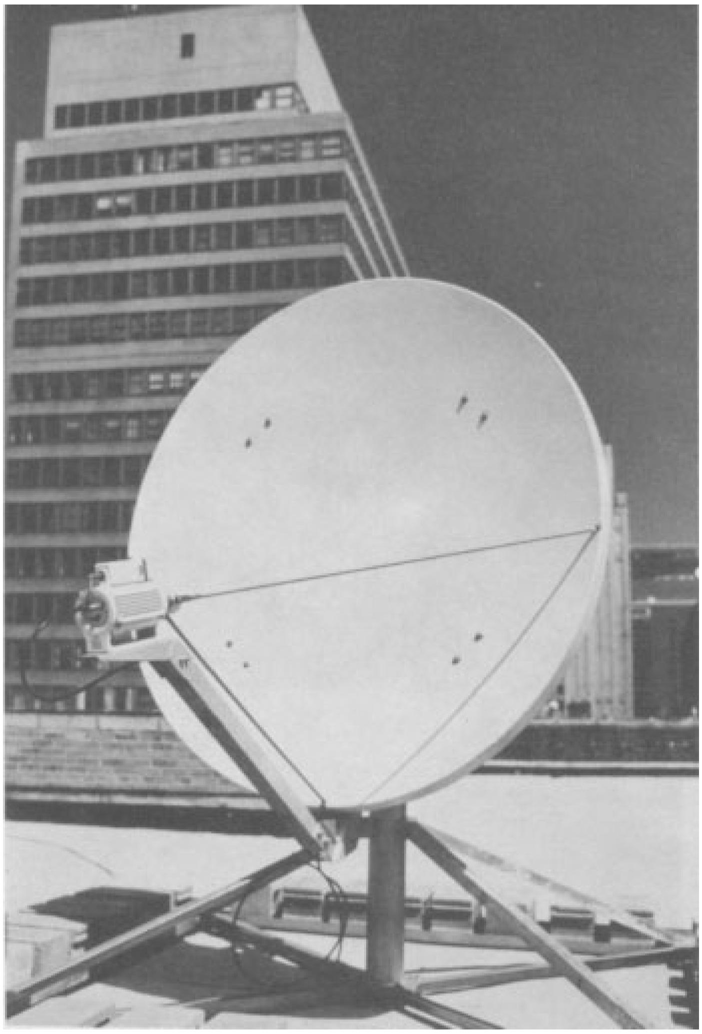


Figure 4.18 Photograph of the outdoor unit of a VSAT station.

– the figure of merit G/T, which determines the performance of the radio frequency downlink. The G/T ratio depends on the value of the antenna gain, and hence its size and receive frequency, and on the noise temperature of the receiver;

– the antenna sidelobe gain variation with off-axis angle which controls the off-axis EIRP and G/T, hence determining the levels of produced and received interference.

Operating temperature range, wind loading under operational and survival conditions, rain and humidity are also to be considered.Table 4.6 displays typical values for the ODU of a VSAT. LNA typical noise temperature of today’s VSAT receiver is 50 K at C- band and 120 K at Ku-band. Advances in HEMT FET technology now make possible uncooled LNAs having noise temperatures of 35 K at C-band and 80 K at Ku-band.

4.7.1.2 The indoor unit (IDU)

The indoor unit installed at the user's facility is shown in Figure 4.17. In order to connect his terminals to the VSAT, the user must access the ports installed on the rear panel of the outdoor unit, shown in the photograph in Figure 4.18. For a proper specification of the IDU, as an interface to the user's terminals or to a local area network (LAN), the following parameters are of importance:

– number of ports;

– type of ports: mechanical, electrical, functional and procedural interface;

– port speed: this is the maximum bit rate at which data can be exchanged between the user terminal and the VSAT indoor unit on a given port. The actual data rate can be lower.

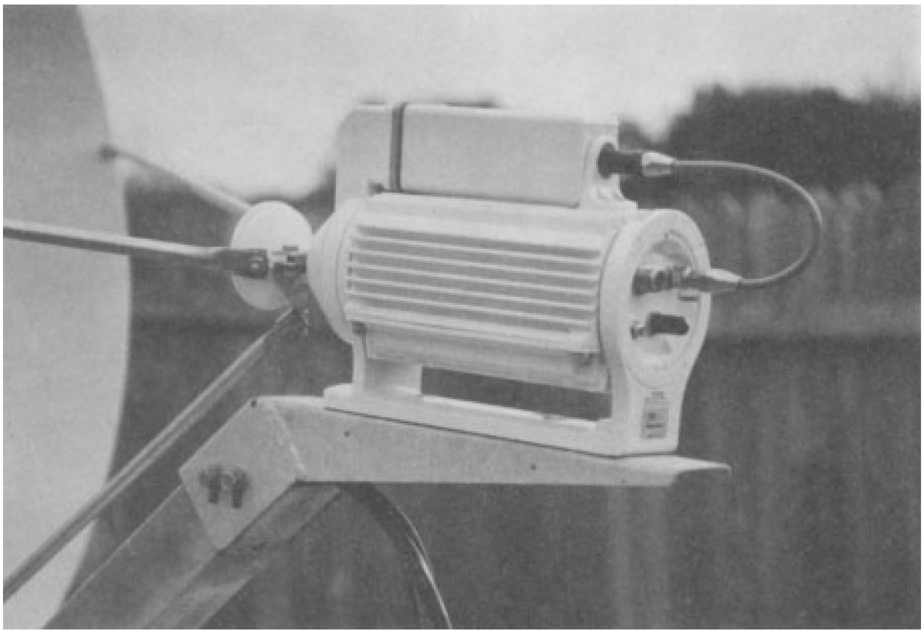
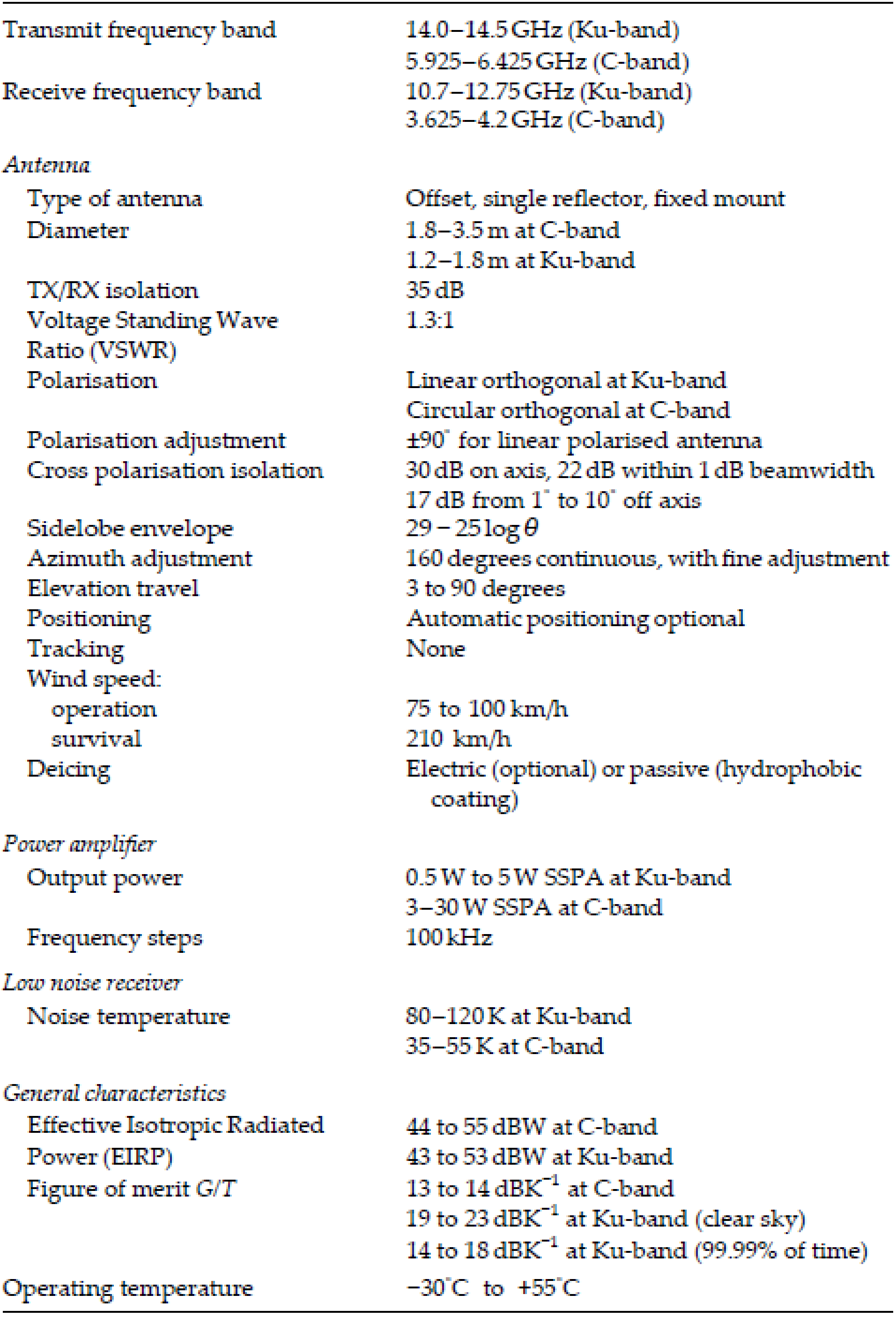


Figure 4.19 Photograph of the electronics container of the outdoor unit

Coherent modulation schemes such as biphase shift keying (BPSK) or quadrature phase shift keying (QPSK) are used. For acceptable performance transmission rate on the carrier should be higher than 2.4 kbs−1, otherwise phase noise becomes a problem. For lower data transmission rate values, phase shift keying is avoided and frequency shift keying (FSK) is used instead.

Table 4.6 Typical values for the ODU parts of a VSAT station



Transmit frequency band	14.0–14.5 GHz (Ku-band) 5.925–6.425 GHz (C-band)
Receive frequency band	10.7–12.75 GHz (Ku-band) 3.625–4.2 GHz (C-band)
Antenna	
Type of antenna	Offset, single reflector, fixed mount
Diameter	1.8–3.5 m at C-band 1.2–1.8 m at Ku-band
TX/RX isolation	35 dB
Voltage Standing Wave Ratio (VSWR)	1.3:1
Polarisation	Linear orthogonal at Ku-band Circular orthogonal at C-band
Polarisation adjustment	±90° for linear polarised antenna
Cross polarisation isolation	30 dB on axis, 22 dB within 1 dB beamwidth 17 dB from 1° to 10° off axis
Sidelobe envelope	$29 - 25 \log \theta$
Azimuth adjustment	160 degrees continuous, with fine adjustment
Elevation travel	3 to 90 degrees
Positioning	Automatic positioning optional
Tracking	None
Wind speed:	
operation	75 to 100 km/h
survival	210 km/h
Deicing	Electric (optional) or passive (hydrophobic coating)
Power amplifier	
Output power	0.5 W to 5 W SSPA at Ku-band 3–30 W SSPA at C-band
Frequency steps	100 kHz
Low noise receiver	
Noise temperature	80–120 K at Ku-band 35–55 K at C-band
General characteristics	
Effective Isotropic Radiated Power (EIRP)	44 to 55 dBW at C-band 43 to 53 dBW at Ku-band
Figure of merit G/T	13 to 14 dBK^{-1} at C-band 19 to 23 dBK^{-1} at Ku-band (clear sky) 14 to 18 dBK^{-1} at Ku-band (99.99% of time)
Operating temperature	−30°C to +55°C

4.8 Conclusions

This conclusion summarises the perceived advantages and draw- backs of VSAT networks.

4.8.1 Advantages

4.8.1.1 Point-to-multipoint and point-to-point communications

A VSAT network offers communications between remote terminals. As a result of the power limitation resulting from the imposed small size and low cost of the remote station, a VSAT network is most often star-shaped with remotes linked to a larger station called a hub. This star configuration often well reflects the structure of information flow within most large organisations which have a point of central control where the hub can be installed. The star configuration itself is not a severe limitation to the effectiveness of a VSAT network as point-to-point communications, which would conveniently be supported by a meshed network, can still be achieved via a double hop, using the hub as a central switch to the network.

4.8.1.2 Asymmetry of data transfer

As a result of its asymmetric configuration, a star-shaped network displays different capacities on the inbound link and on the out- bound link. This may be an advantage considering the customer need for asymmetric capacities in most of his applications. Should he use leased terrestrial lines which are inherently symmetric, i.e. offering equal capacity in both directions, the customer would have to pay for unused capacity.

4.8.1.3 Flexibility

A VSAT network inherently provides a quick response time for network additions and reconfigurations (one or two days) as a result of the easy displacement and installation of a remote station.

4.8.1.4 Private corporate networks

A VSAT network offers its operator end-to-end control over trans- mission quality and reliability. It also protects him from possible and unexpected tariff fluctuations, by offering price stability and the possibility to forecast its communication expenses. Therefore it is an adequate support to private corporate networks.

4.8.1.5 Low bit error rate

The bit error rate usually encountered on VSAT links is typically 10^{-7}.

4.8.1.6 Distance-insensitive cost

The cost of a link in a VSAT network is not sensitive to distance. Hence, cost savings are expected if the network displays a large number of sites and a high geographical dispersion.

4.8.2 Drawbacks

4.8.2.1 Interference sensitivity

A radio frequency link in a VSAT network is subject to interference as a result of the small earth station antenna size.

4.8.2.2 Eavesdropping

As a result of the large coverage of a geostationary satellite, it may be easy for an eavesdropper to receive a downlink carrier and access the information content by demodulating the carrier. Therefore, to prevent unauthorised use of the information conveyed on the carrier, encryption may be mandatory.

4.8.2.3 Loss of transponder may lead to loss of network

The satellite is a single point failure. Should the transponder that relays the carrier fail, then the complete VSAT network is out of order. Communication links can be restored by using a spare transponder. With a spare colocated on the same satellite, a mere change in frequency or polarisation puts the network back in operation. However, should this operation. However, should this transponder be located on another satellite, this may mean intervening on each site to repoint the antenna, and this may take some time.

4.8.2.4 Propagation delay (double hop = 0.5 s)

The propagation time from remote to remote in a star-shaped network imposes a double hop with its associated delay of about half a second. This may prevent the use of voice communications, at least with commercial standards.

References

[1] Abramson, N. (1992) Fundamentals of packet multiple access for satellite networks, IEEE Journal on Selected Areas in Communication, 10(2), 309–316.

[2] Chitre, D.M. and McCoskey, J.S. (1988) VSAT networks: architecture, protocols and management, IEEE Communications Magazine, 26(7), 28–38.

[3] Evans, B.G. (Ed.) (1999) Satellite Communications Systems, 3rd edition, IEE, London.

[4] Everett, J. (Ed.) (1992) VSATs: Very Small Aperture Terminals, IEE Telecommunications Series 28, Peter Pelegrinus.

[5] Ferrari, D. (1990) Client requirements for real time communications services, IEEE Communications Magazine, 11, 65–72.

[6] Fujii, A., Teshigawara, Y., Tejima, S. and Matsumoto, Y. (1986) AA/TDMA–Adaptive satelite access method for mini earth station networks, GLOBECOM 86, paper 42.4, Houston.

[7] Ha, Tri T. (1986) Digital Satellite Communications, MacMillan.

[8] (1988) Handbook on Satellite Communications,International Telecommunication Union, Geneva.

[9] Maral,G.: VSAT Networks. Second edition. Wiley, 2003

5 Global Navigation Satellite System (GNSS)

Recently, there is an increase interest in positioning techniques based on Global Navigation Satellite Systems (GNSS) such as Global Positioning System (GPS), cellular network infrastructure or on the integration of the two technologies for a wide spread of applications such as Automatic Vehicle Location (AVL), tracking systems, navigation, Pedestrian Navigation Systems (PNSs), intelligent transportation Systems, precise positioning and emergency callers. During the last 20 years there are many important events in the field of satellite navigation systems such as: (a) the full operational GPS in 1993, when 24 GPS satellites were operating in their assigned orbits, available for navigation use and providing Standard Positioning Services (SPS), (b) the new European satellite system Galileo, (c) the modernized of US satellite system GPS, and (d) the reconstruction of Russian satellite system Glonass. LBS projects aim to improve user-friendly info-mobility services for position determination by combining wireless communications, satellite navigation (GNSS) and geographic information systems (GIS), based on a mobile client/server architecture

The meaning of GNSS is the technical interoperability and compatibility between various satellite navigation systems such as modernized GPS, Galileo, reconstructed GLONASS to be used by civilian users without considering the nationalities of each system in order to promote the safety and convenience of life .

Our interest here is to outline the new technologies and applications evolved and appeared from the integration between the GNSS, GIS and wireless communications. We will give an introduction of GNSS by introducing the characteristic of the three satellite systems (GPS, GLONASS and Galileo). Satellite navigation systems has become integral part of all applications where mobility plays a important role. These functions will be at the heart of the mobile phone third-generation (3G) and fourth-generation (4G) networks such as the UMTS and LTE. In transportation systems, the presence of receivers will become as common as seat belts or airbags, with all car manufacturers equipping their entry-level vehicles with these devices., products and, consequently, applications and services. The milestone of satellite navigation is the real time positioning and time synchronization. For that reason the implementation of wide-area augmentation systems should be highlighted, because they allow a significant improvement of accuracy and integrity performance. WAAS, EGNOS and MSAS provide over US, Europe, Japan a useful augmentation to GPS, GLONASS and Galileo services. GNSS development has an interesting aspect due to its sensitive nature. Considerable events or developments are always subject to a couple of differentiators: technological developments and political decisions.

GPS and Glonass in all stages of improvements are strictly related to those differentiators. The approval and startup of the European Galileo program is considered by far the most real innovation. Technological and political decisions in Galileo substantiate that interoperability and compatibility must be reached in the forthcoming years. Such issues are the true GNSS improvement for the benefit of institutions and organizations. GNSS applications in all fields will play a key role, moving its use from the transportation domain to multimodal use, outdoors and indoors. It is expected that GNSS will increase significantly the precision in position domain.

The concept of reference system for navigation is essential since all the applications of GNSS are related to the coordinate system used. The main application of GNSS is focused on the potential of to TLF eBOOK determine the position in the Global reference system any where any time on the Globe in a simple, fast and cost-effective manner.

The integration between GNSS and other related technologies such as telecommunications (GSM, GPRS, UMTS, LTE), the Geographic Information Systems (GIS) and Inertial Navigation System (INS), has created numerous applications that needs more time to be discussed in details. Many research efforts have been exerted in order to find each new applications to promote the quality of our life using the GNSS.

The GNSS consist of three main satellite technologies: GPS, Glonass and Galileo. Each of them consists mainly of three segments: (a) space segment, (b) control segment and (c) user segment. These segments are almost similar in the three satellite technologies, which are all together make up the GNSS. As of today, the complete satellite technology is the GPS technology and most of the existing worldwide applications related to the GPS technology. The GNSS technology will become clearer after the operation of Galileo and the reconstruction of Glonass in the next few years.

5.1 GPS

The Global Positioning System (GPS) is a satellite-based navigation system that was developed by the U.S. Department of Defense (DoD) in the early 1970s. Initially, GPS was developed as a military system to fulfill U.S. military needs. However, it was later made available to civilians, and is now a dual-use system that can be accessed by both military and civilian users .

GPS provides continuous positioning and timing information, anywhere in the world under any weather conditions. Because it serves an unlimited number of users as well as being used for security reasons, GPS is a one-way-ranging (passive) system. That is, users can only receive the satellite signals. This chapter introduces the GPS system, its components, and its basic idea.

5.1.1 Overview of GPS

GPS consists, nominally, of a constellation of 24 operational satellites. This constellation, known as the initial operational capability (IOC), was completed in July 1993. The official IOC announcement, however, was made on December 8, 1993. To ensure continuous worldwide coverage, GPS satellites are arranged so that four satellites are placed in each of six orbital planes (Figure 5.1). With this constellation geometry, four to ten GPS satellites will be visible anywhere in the world, if an elevation angle of 10° is considered. As discussed later, only four satellites are needed to provide the positioning, or location, information.

GPS satellite orbits are nearly circular (an elliptical shape with a maximum eccentricity is about 0.01), with an inclination of about 55° to the equator. The semimajor axis of a GPS orbit is about 26,560 km (i.e., the satellite altitude of about 20,200 km above the Earth's surface). The corresponding GPS orbital period is about 12 sidereal hours (~11 hours, 58 minutes). The GPS system was officially declared to have achieved full operational capability (FOC) on July 17, 1995, ensuring the availability of at least 24 operational, nonexperimental, GPS satellites. In fact, since GPS achieved its FOC, the number of satellites in the GPS constellation has always been more than 24 operational satellites.

5.1.2 GPS segments

GPS consists of three segments: the space segment, the control segment, and the user segment (Figure 5.2). The space segment consists of the 24-satellite constellation introduced in the previous section. Each GPS satellite transmits a signal, which has a number of components: two sine waves (also known as carrier frequencies), two digital codes, and a navigation message. The codes and the navigation message are added to the carriers as binary biphase modulations. The carriers and the codes are used mainly to determine the distance from the user's receiver to the GPS satellites.

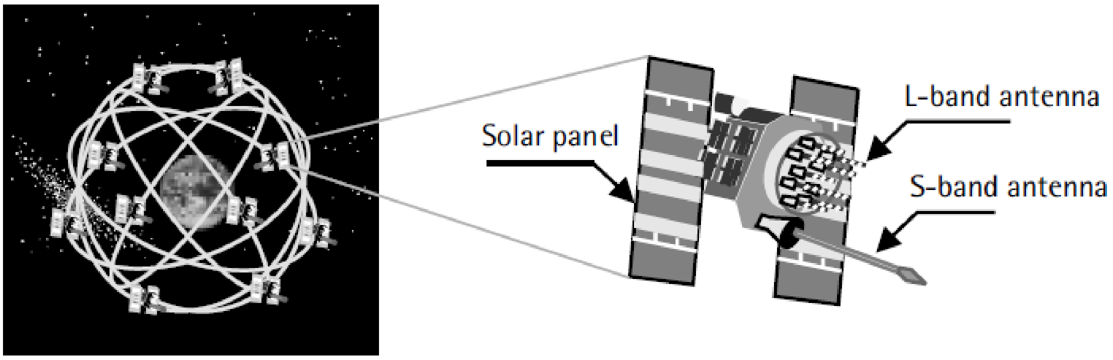


Figure 5.1 GPS constellation

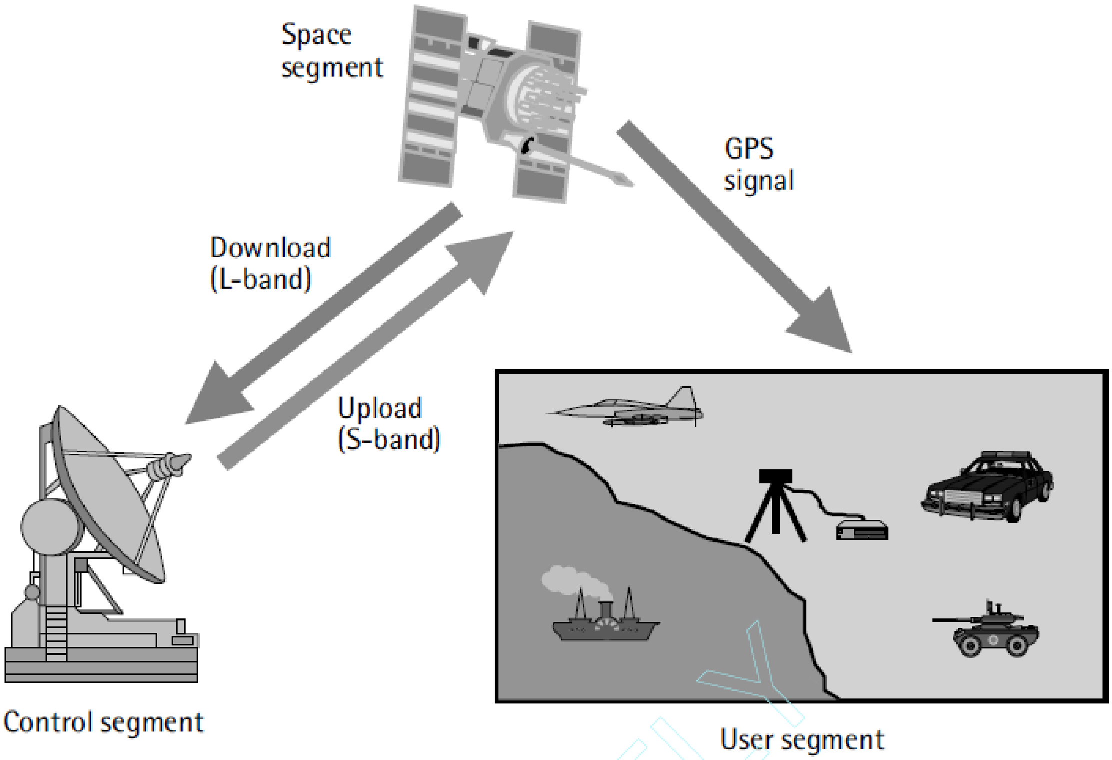


Figure 5.2 GPS segments.

The navigation message contains, along with other information, the coordinates (the location) of the satellites as a function of time. The transmitted signals are controlled by highly accurate atomic clocks onboard the satellites.

The control segment of the GPS system consists of a worldwide network of tracking stations, with a master control station (MCS) located in the United States at Colorado Springs, Colorado. The primary task of the operational control segment is tracking the GPS satellites in order to determine and predict satellite locations, system integrity, behavior of the satellite atomic clocks, atmospheric data, the satellite almanac, and other considerations. This information is then packed and uploaded into the GPS satellites through the S-band link.

The user segment includes all military and civilian users. With a GPS receiver connected to a GPS antenna, a user can receive the GPS signals, which can be used to determine his or her position anywhere in the world. GPS is currently available to all users worldwide at no direct charge.

5.1.3 GPS satellite generations

GPS satellite constellation buildup started with a series of 11 satellites known as Block I satellites (Figure 5.3). The first satellite in this series (and in the GPS system) was launched on February 22, 1978; the last was launched on October 9,1985. Block I satellites were built mainly for experimental purposes. The inclination angle of the orbital planes of these satellites, with respect to the equator, was 63°, which was modified in the following satellite generations . Although the design lifetime of Block I satellites was 4.5 years, some remained in service for more than 10 years. The last Block I satellite was taken out of service on November 18, 1995.

The second generation of the GPS satellites is known as Block II/IIA satellites (Figure 5.3). Block IIA is an advanced version of Block II, with an increase in the navigation message data storage capability from 14 days for Block II to 180 days for Block IIA. This means that Block II and Block IIA satelite scan function continuously, without ground support, for periods of 14 and 180 days, respectively. A total of 28 Block II/IIA satellites were launched during the period from February 1989 to November 1997. Of these, 23 are currently in service. Unlike Block I, the orbital plane of Block II/IIA satellites are inclined by 55° with respect to the equator. The design lifetime of a Block II/IIA satellite is 7.5 years, which was exceeded by most Block II/IIA satellites. To ensure national security, some security features, known as selective availability (SA) and antispoofing, were added to Block II/IIA satellites.

A new generation of GPS satellites, known as Block IIR, is currently being launched (Figure 5.3). These replenishment satellites will be backward compatible with Block II/IIA, which means that the changes are transparent to the users. Block IIR consists of 21 satellites with a design life of 10 years. In addition to the expected higher accuracy, Block IIR satellites have the capability of operating autonomously for at least 180 days without ground corrections or accuracy degradation. The autonomous navigation capability of this satellite generation is achieved in part through mutual satellite ranging capabilities. In addition, predicted ephemeris and clock data for a period of 210 days are uploaded by the ground control segment to support the autonomous navigation. More features will be added to the last 12 Block IIR satellites under the GPS modernization program, which will be launched at the beginning [7]. As of July 2001, six Block IIR satellites have been successfully launched.

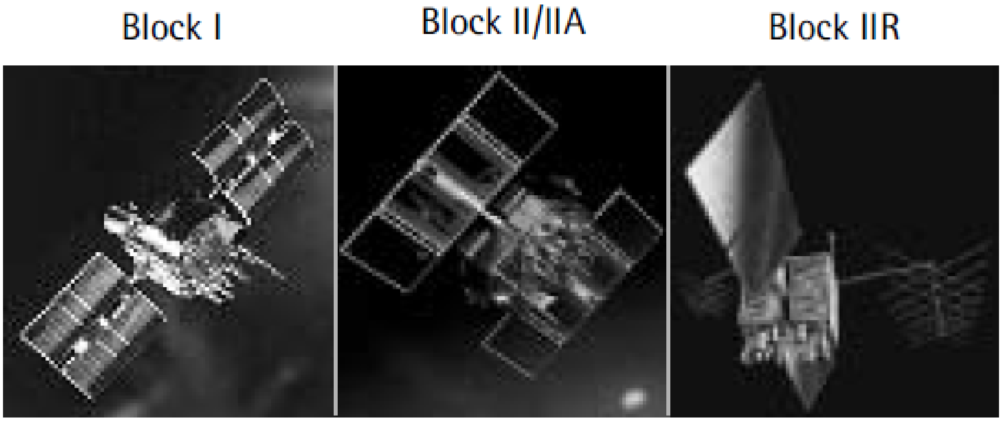


Figure 5.3 GPS satellite generations

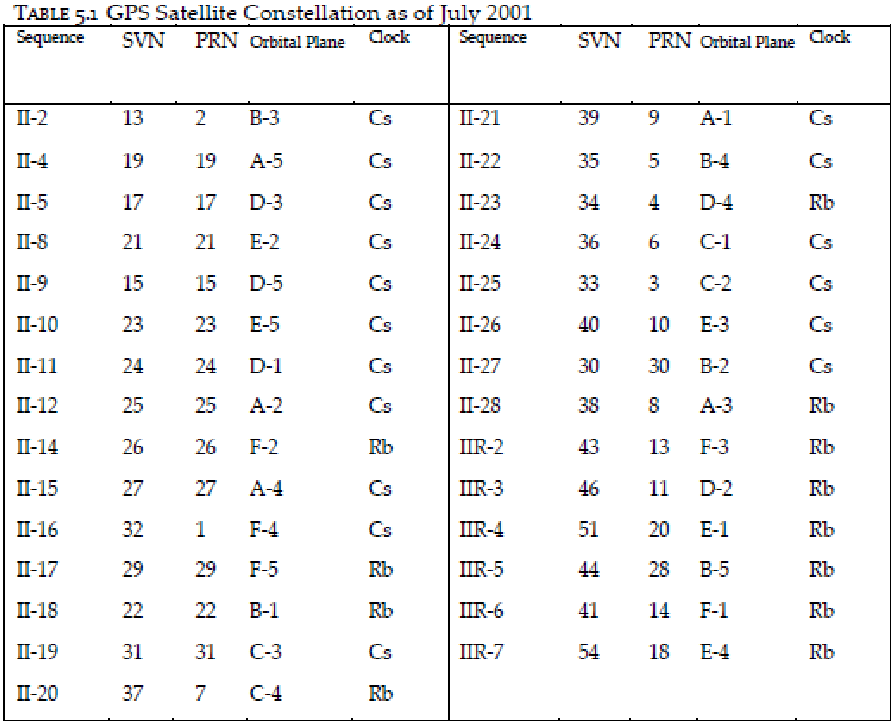
Block IIR will be followed by another system, called Block IIF (for "follow-on"), consisting of 33 satellites. The satellite life span will be 15 years. Block IIF satellites will have new capabilities under the GPS modernization program that will dramatically improve the autonomous GPS positioning accuracy). The first Block IIF satellite was scheduled to be launched in 2005 or shortly after that date.

5.1.4 Current GPS satellite constellation

The current GPS constellation (as of July 2001) contains five Block II, 18 Block IIA, and six Block IIR satellites (see Table 5.1). This makes the total number of GPS satellites in the constellation to be 29, which exceeds the nominal 24-satellite constellation by five satellites. All Block I satellites are no longer operational.

The GPS satellites are placed in six orbital planes, which are labeled A through F. Since more satellites are currently available than the nominal 24-satellite constellation, an orbital plane may contain four or five satellites. As shown in Table 5.1, all of the orbital planes have five satellites, except for orbital plane C, which has only four. The satellites can be identified by various systems. The most popular identification systems within the GPS user community are the space vehicle number (SVN) and the pseudorandom noise (PRN); the PRN number will be defined later. Block II/IIA satellites are equipped with four onboard atomic clocks: two cesium (Cs) and two rubidium (Rb). The cesium clock is used as the primary timing source to control the GPS signal. Block IIR satellites, however, use rubidium clocks only.

TABLE 5.1 GPS Satellite Constellation as of July 2001



Sequence	SVN	PRN	Orbital Plane	Clock	Sequence	SVN	PRN	Orbital Plane	Clock
II-2	13	2	B-3	Cs	II-21	39	9	A-1	Cs
II-4	19	19	A-5	Cs	II-22	35	5	B-4	Cs
II-5	17	17	D-3	Cs	II-23	34	4	D-4	Rb
II-8	21	21	E-2	Cs	II-24	36	6	C-1	Cs
II-9	15	15	D-5	Cs	II-25	33	3	C-2	Cs
II-10	23	23	E-5	Cs	II-26	40	10	E-3	Cs
II-11	24	24	D-1	Cs	II-27	30	30	B-2	Cs
II-12	25	25	A-2	Cs	II-28	38	8	A-3	Rb
II-14	26	26	F-2	Rb	IIR-2	43	13	F-3	Rb
II-15	27	27	A-4	Cs	IIR-3	46	11	D-2	Rb
II-16	32	1	F-4	Cs	IIR-4	51	20	E-1	Rb
II-17	29	29	F-5	Rb	IIR-5	44	28	B-5	Rb
II-18	22	22	B-1	Rb	IIR-6	41	14	F-1	Rb
II-19	31	31	C-3	Cs	IIR-7	54	18	E-4	Rb
II-20	37	7	C-4	Rb					

It should be pointed out that two satellites, PRN05 and PRN06, are equipped with corner cube reflectors to be tracked by laser ranging (Table 5.1).

5.1.5 Control sites

The control segment of GPS consists of a master control station (MCS), a worldwide network of monitor stations, and ground control stations (Figure 5.4). The MCS, located near Colorado Springs, Colorado, is the central processing facility of the control segment and is manned at all times.

There are five monitor stations, located in Colorado Springs (with the MCS), Hawaii, Kwajalein, Diego Garcia, and Ascension Island. The positions (or coordinates) of these monitor stations are known very precisely.

Each monitor station is equipped with high-quality GPS receivers and a cesium oscillator for the purpose of continuous tracking of all the GPS satellites in view. Three of the monitor stations (Kwajalein, Diego Garcia, and Ascension Island) are also equipped with ground antennas for uploading the information to the GPS satellites. All of the monitor stations and the ground control stations are unmanned and operated remotely from the MCS.

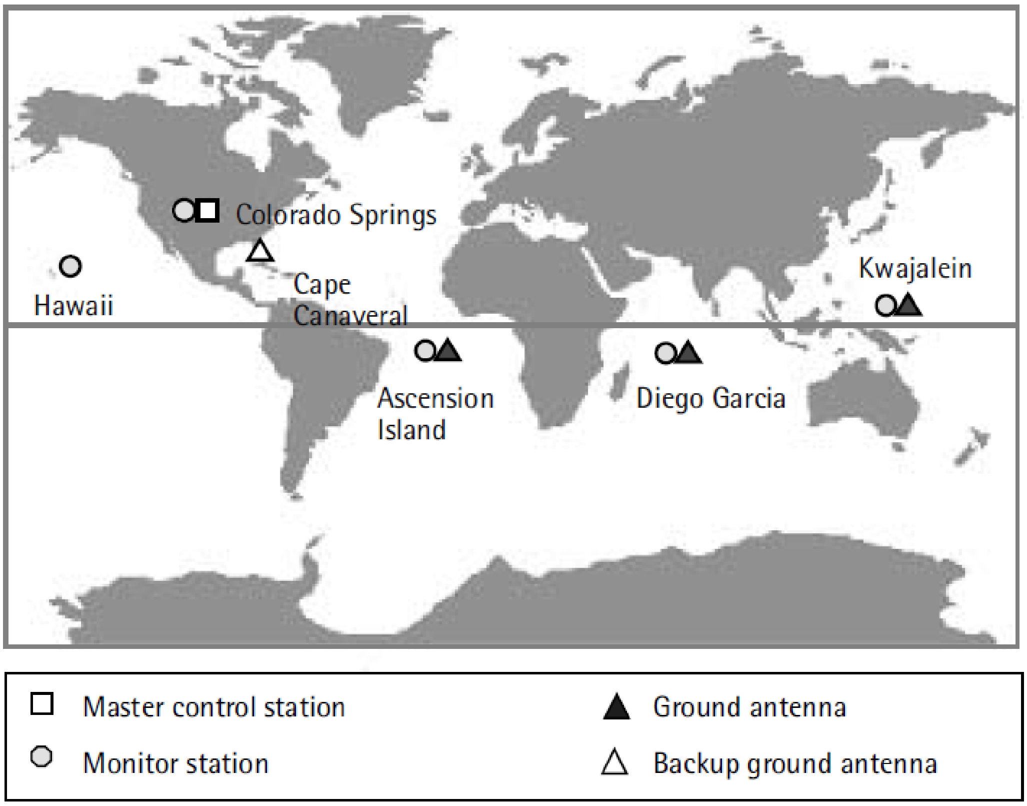


Figure 5.4 GPS control sites.

The GPS observations collected at the monitor stations are transmitted to the MCS for processing. The outcome of the processing is predicted satellite navigation data that includes, along with other information, the satellite positions as a function of time, the satellite clock parameters, atmospheric data, satellite almanac, and others. This fresh navigation data is sent to one of the ground control stations to upload it to the GPS satellites through the S-band link.

Monitoring the GPS system integrity is also one of the tasks of the MCS. The status of a satellite is set to unhealthy condition by the MCS during satellite maintenance or outages. This satellite health condition appears as a part of the satellite navigation message on a near real-time basis. Scheduled satellite maintenance or outage is reported in a message called Notice Advisory to Navstar Users (NANU), which is available to the public through, for example, the U.S. Coast Guard Navigation Center.

5.1.6 The basic idea

The idea behind GPS is rather simple. If the distances from a point on the Earth (a GPS receiver) to three GPS satellites are known along with the satellite locations, then the location of the point (or receiver) can be determined by simply applying the well-known concept of resection. That is all! But how can we get the distances to the satellites as well as the satellite locations?

As mentioned before, each GPS satellite continuously transmits a microwave radio signal composed of two carriers, two codes, and a navigation message. When a GPS receiver is switched on, it will pick up the GPS signal through the receiver antenna. Once the receiver acquires the GPS signal, it will process it using its built-in software. The partial outcome of the signal processing consists of the distances to the GPS satellites through the digital codes (known as the pseudoranges) and the satellite coordinates through the navigation message.

Theoretically, only three distances to three simultaneously tracked satellites are needed. In this case, the receiver would be located at the intersection of three spheres; each has a radius of one receiver-satellite distance and is centered on that particular satellite (Figure 5.5). From the practical point of view, however, a fourth satellite is needed to account for the receiver clock offset. The accuracy obtained with the method described earlier was until recently limited to 100m for the horizontal component, 156m for the vertical component, and 340 ns for the time component, all at the 95% probability level. This low accuracy level was due to the effect of the so-called selective availability, a technique used to intentionally degrade the autonomous real-time positioning accuracy to unauthorized users. With the recent presidential decision of terminating the selective availability, the obtained horizontal accuracy is expected to improve to about 22m (95%) probability level). To further improve the GPS positioning accuracy, the so-called differential method,

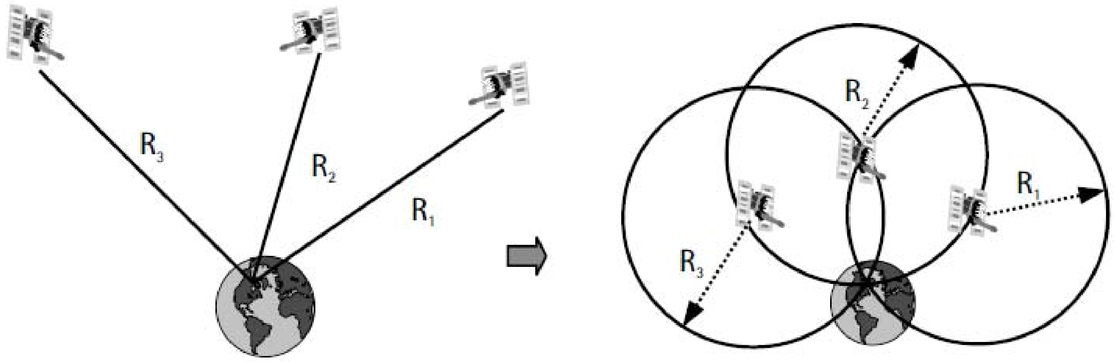


Figure 5.5 Basic idea of GPS positioning.

which employs two receivers simultaneously tracking the same GPS satellites, is used. In this case, positioning accuracy level of the order of a subcentimeter to a few meters can be obtained.

Other uses of GPS include the determination of the user's velocity, which could be determined by several methods. The most widely used method is based on estimating the Doppler frequency of the received GPS signal. It is known that the Doppler shift occurs as a result of the relative satellite-receiver motion. GPS may also be used in determining the attitude of a rigid body, such as an aircraft or a marine vessel. The word "attitude" means the orientation, or the direction, of the rigid body, which can be described by the three rotation angles of the three axes of the rigid body with respect to a reference system. Attitude is determined by equipping the body with a minimum of three GPS receivers (or one special receiver) connected to three antennas, which are arranged in a nonstraight line. Data collected at the receivers are then processed to obtain the attitude of the rigid body.

5.1.7 GPS positioning service

As stated earlier, GPS was originally developed as a military system, but was later made available to civilians as well. However, to keep the military advantage, the U.S. DoD provides two levels of GPS positioning and timing services: the Precise Positioning Service (PPS) and the Standard Positioning Service (SPS).

PPS is the most precise autonomous positioning and timing service. It uses one of the transmitted GPS codes, known as P(Y)-code, which is accessible by authorized users only. These users include U.S. military forces. The expected positioning accuracy provided by the PPS is 16m for the horizontal component and 23m for the vertical component (95% probability level).

SPS, however, is less precise than PPS. It uses the second transmitted GPS code, known as the C/A-code, which is available free of charge to all users worldwide, authorized and unauthorized. Originally, SPS provided positioning accuracy of the order of 100m for the horizontal component and 156m for the vertical component (95% probability level). This was achieved under the effect ofselective availability. With the recent presidential decision of discontinuing the SA, the SPS autonomous positioning accuracy is presently at a comparable level to that of the PPS.

5.1.8 GPS signals

The generated signals on board the satellites are based or derived from generation of a fundamental frequency ƒo=10.23 MHz. The signal is controlled by atomic clock and has stability in the range of 10−13 over one day. Two carrier signals in the L-band, denoted L1 and L2, are generated by integer multiplications of ƒo. The carriers L1 and L2 are biphase modulated by codes to provide satellite clock readings to the receiver and transmit information such as the orbital parameters. The codes consist of a sequence with the states +1 or -1, corresponding to the binary values 0 or 1. The biphase modulation is performed by a 180° shift in the carrier phase whenever a change in the code state occurs; see Figure 5.6. The clear/access code (C/A-code) and precision code (P-code) are used for the satellite clock reading, both are characterized by a pseudorandom noise (PRN) sequence. The W-code is employed to encrypt the P-code to the Y-code when Anti Spoofing (A-S) is applied. The navigation message is modulated using the two carriers (L1 and L2) at a chipping rate of 50 bps.

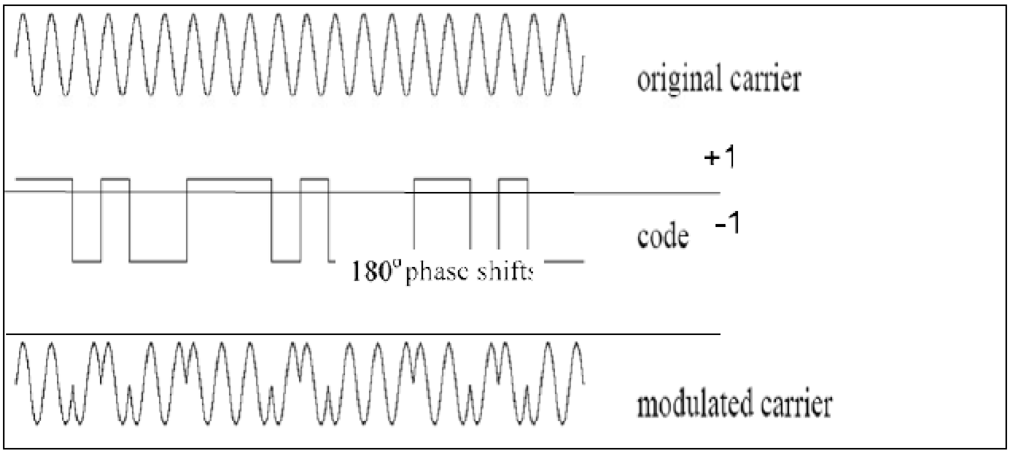


Figure 5.6. Biphase modulation of carrier

It contains information on the satellite orbits, orbit perturbations, GPS time, satellite clock, ionospheric parameters, and system status messages (Leick, 2003). The modulation of L1 by P- code, C/A-code and navigation message (D), is done using the quadrature phase shift keying (QPSK) scheme. The C/A-code is placed on the LI carrier with 90° offset from the P-code since they have the same bit transition epochs. For the L1 and L2 we have:

$$L1(t) = a1\,P(t)W(t)\cos(2\pi f1t) + a1C/A(t)D(t)\sin(2\pi f1t)$$

$$L2(t) = a2\,P(t)W(t)\cos(2\pi f\,2\,t)\ (1) \qquad (5.1)$$

The signal broadcast by the satellite is a spread spectrum signal, which makes it less prone to jamming. The basic concept of spread spectrum technique is that the information waveform with small bandwidth is converted by modulating it with a large-bandwidth. The generation of pseudo random sequence (PRN) in the code is based on the use of an electronic hardware device called tapped feed back shift register (FBSR). This device can generate a large variety of pseudo random codes, but in this way the generated code repeat it self after some very long time. The receiver could distinguish the signals coming from different satellites because the receiving C/A code (the Gold code), has low cross-correlation and is unique for each satellite (Leick, 2003). The navigation message consists of 25 frames with each frame containing 1500 bit and each frame is subdivided into 5 sub-frames with 300 bit. The information transmitted by the navigation message is periodically updated by the control segment.

5.2 GLONASS

The GLONASS (GLObal NAvigation Satellite System or "GLObalnaya NAvigatsionnaya Sputnikovaya Sistema") is nearly identical to GPS. Glonass satellite-based radio-navigationsystem provides the positioning and timing information to users. It is operated by the Ministry of Defense of the Russian Federation (GLONASS-ICD, 2002).Glonass space segment is consist of 24 satellites, equally distributed in 3 orbit separated by 120° in the equatorial plane. Satellite orbital altitude is about 19,130 km above the ground surface. This results in an orbital period of 11:15:44 corresponding to 8/17 of a sidereal day. The future of GLONASS seems uncertain due to economic problems facing the Russian Federation. The number of operational satellites was steadily decreasing over the past few years. The launch of three new GLONASS satellites in December 1998 was the first launch after a lapse of 3 years. As of January 2006, a total of 10 GLONASS satellites are operational. The oldest of the still active satellites was launched in October, 2000. According to Russian officials the GLONASS system shall again be restored by 2008.

The use of GLONASS in addition to GPS provides very significant advantages:

- increased satellite signal observations
- markedly increased spatial distribution of visible satellites
- reduced Horizontal and Vertical Dilution of Precision (DOP) factors
- decreased occupation times means faster RTK results

In order to determine a position in GPS-only mode the receiver must track a minimum of four satellites, representing the four unknowns of 3-D position and time. In combined GPS/GLONASS mode, the receiver must track five satellites, representing the same four previous unknowns and at least one GLONASS satellite to determine the GPS/GLONASS time offset.

With the availability of combined GPS/GLONASS receivers, users have access to a potential 48+ satellite-combined system. With 48+ satellites, performance in urban canyons and other locations with restricted visibility, such as forested areas improves, as more satellites are visible in the non-blocked portions of the sky. A larger satellite constellation also improves real-time carrier phase differential positioning performance.

Russia has committed itself to bringing the system up to the required minimum of 18 active satellites by the end of 2007, and signed an agreement with India that provides for the launches of GLONASS satellites on Indian launch vehicles. At the time of publication, April 2007, there are 12 operational GLONASS satellites and one newly launched GLONASS satellite at its commissioning phase. The Russian Government have set 2009 as the full deployment date of the 24-satellite constellation and ensured financial support to meet that date.

5.2.1 GLONASS System Design

As with GPS, the GLONASS system uses a satellite constellation to provide, ideally, a GLONASS receiver with six to twelve satellites at most times. A minimum of four satellites in view allows a GLONASS receiver to compute its position in three dimensions, as well as become synchronized to the system time.

The GLONASS system design consists of three parts:

- The Control segment
- The Space segment
- The User segment

All these parts operate together to provide accurate three-dimensional positioning, timing and velocity data to users worldwide.

5.2.1.1 The Control Segment

The Control Segment consists of the system control center and a network of command trackingstations across Russia. The GLONASS control segment, similar to GPS, must monitor the status of satellites, determine the ephemerides and satellite clock offsets with respect to GLONASS time and UTC (Coordinated Universal Time), and twice a day upload the navigation data to the satellites.

5.2.1.2 The Space Segment

The Space Segment is the portion of the GLONASS system that is located in space, that is, the GLONASS satellites that provide GLONASS ranging information. When complete, this segment will consist of 24 satellites in three orbital planes, with eight satellites per plane. Figure 1 on Page 3 shows a combined GPS and GLONASS satellite system.

5.2.1.3 The User Segment

The User Segment consists of equipment (such as a NovAtel OEMV family receiver) that tracks and receives the satellite signals. This equipment must be capable of simultaneously processing the signals from a minimum of four satellites to obtain accurate position, velocity and timing measurements. Like GPS, GLONASS is a dual military/civilian-use system. The system's potential civil applications are many and mirror those of GPS.

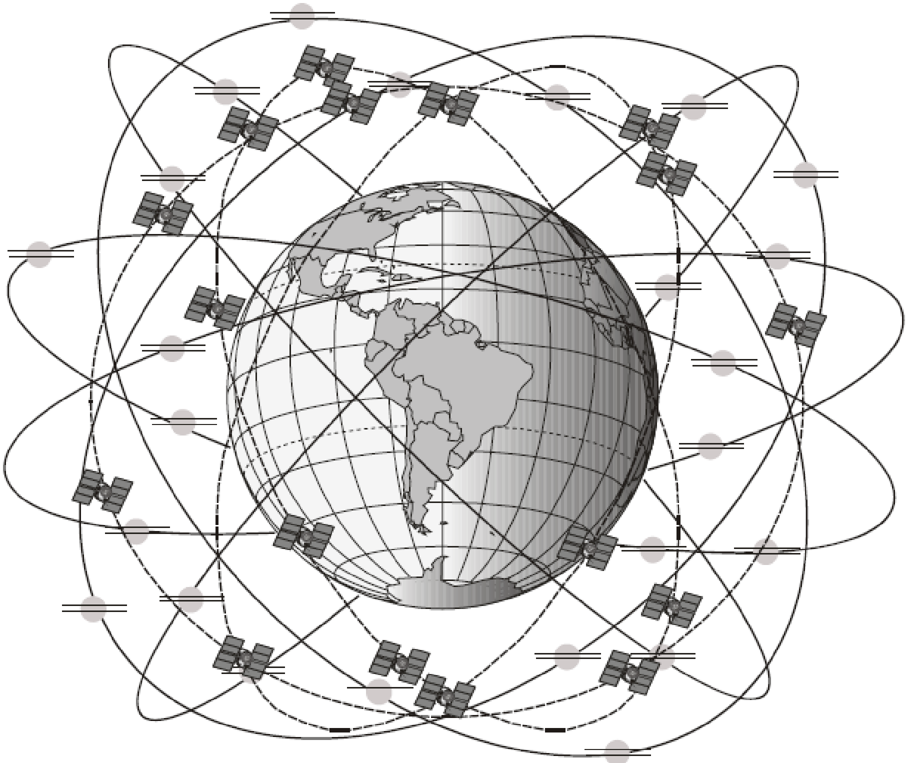


Figure 5.7 View of GPS and GLONASS Satellite Orbit Arrangement

Following are points about the GLONASS space segment:

- The geometry repeats about once every 8 days. The orbit period of each satellite is approximately 8/17 of a sidereal day such that, after eight sidereal days, the GLONASS satellites have completed exactly 17 orbital revolutions. A sidereal day is the rotation period of the Earth relative to the equinox and is equal to one calendar day (the mean solar day) minus approximately four minutes.

- Because each orbital plane contains eight equally spaced satellites, one of the satellites will be at the same spot in the sky at the same sidereal time each day.

- The satellites are placed into nominally circular orbits with target inclinations of 64.8 degrees and an orbital height of about 19,140 km, which is about 1,050 km lower than GPS satellites.

- Some of the GLONASS transmissions initially caused interference to radio astronomers and mobile communication service providers. The Russians consequently agreed to reduce the number of frequencies used by the satellites and to gradually change the L1 frequencies in the future to 1598.0625 – 1605.375 MHz. Eventually the system will only use 12 primary frequency channels (plus two additional channels for testing purposes).

- The GLONASS satellite signal identifies the satellite and provides:
  - position, velocity and acceleration vectors at a reference epoch to compute satellite locations
  - synchronization bits, data age and satellite health
  - offset of GLONASS time from UTC (SU) (formerly Soviet Union and now Russia)
  - almanacs of all other GLONASS satellites

5.2.2 GPS and GLONASS Satellite Identification

The GLONASS satellites each transmit on slightly different L1 and L2 frequencies, with P- code on both L1 and L2, and with C/A code, at present, only on L1. GLONASS-M satellites reportedly3 transmit the C/A code on L2.

Every GPS satellite transmits the L1 frequency centered at 1575.42 MHz. The GPS satellites are identifiable by their Pseudorandom Noise code number (PRN) with a NovAtel receiver.

Unlike GPS, all GLONASS satellites transmit the same code at different frequencies. They derive signal timing and frequencies from one of three on-board cesium atomic clocks operating at 5 MHz:

For example,

L1 = 1602 MHz + (n x 0.5625) MHz

where

n = the frequency channel number (n = 0, 1, 2 and so on)

It means that satellites transmit signals on their own frequency, separated by multiples of 0.5625 MHz or 562.5 kHz, from the frequency of other satellites; see Figure 5.8 below.

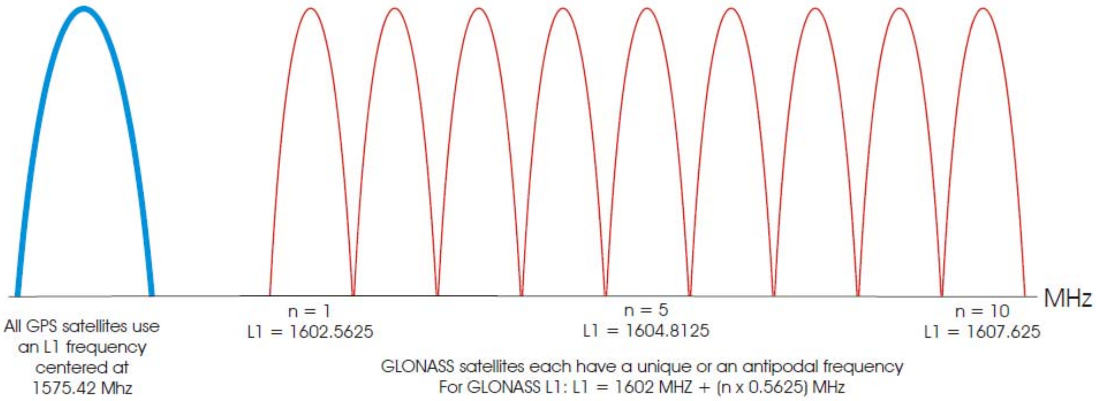


Figure 5.8 GPS and GLONASS L1 Frequencies

The signals are right-hand circularly polarized, like GPS signals, and have comparable signal strength. GLONASS accomplishes system operation (24 satellites and only 12 channels) by having antipodal satellites transmit on the same frequency. Antipodal satellites are in the same orbit plane separated by 180 degrees in argument of latitude. This is possible because the paired satellites will never appear at the same time in view of an operational receiver that is on the earth's surface, see Figure 3 below. At the time of publication, April 2007, four pairs of operational satellites share frequencies. Figure 3. A comparison of GPS with GLONASS satellites, signals, and messages is in Table 5.2.

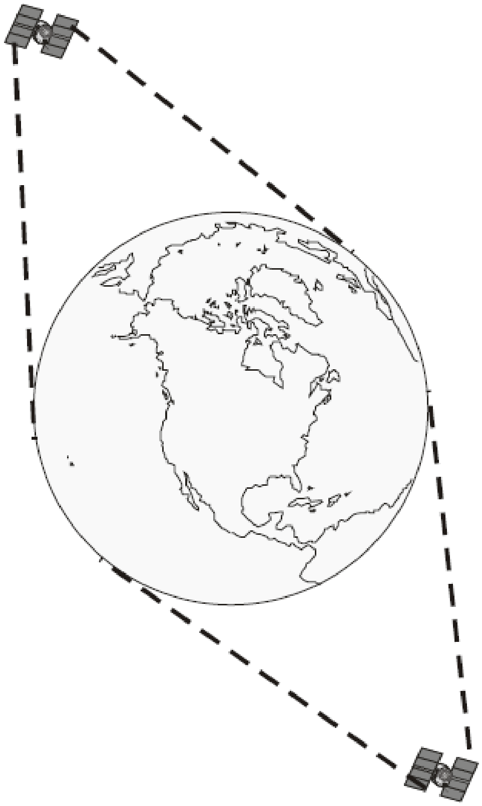


Figure 5.9 GLONASS Antipodal Satellites

5.2.2.1 Time

As stated earlier, both GPS and GLONASS satellites broadcast their time within their satelite messages. NovAtel's OEMV family of receivers are able to receive and record both time references as well as report the offset information between GPS and GLONASS time. Although similar, GPS and GLONASS have several differences in the way they record and report time.

5.2.2.2 GPS Time vs. Local Receiver Time

All logs output by the receiver report GPS Time expressed in GPS weeks and seconds into the week. The time reported is not corrected for local receiver clock error. To derive the closest GPS Time, you must subtract the clock offset shown in the TIME log from GPS Time reported. Refer also to the OEMV Family Firmware Reference Manual, available in PDF format from our website at http://www.novatel.com/support/docupdates.htm. GPS Time is based on an atomic time scale. Coordinated Universal Time as maintained by the U.S.

GPS Time is based on an atomic time scale. Coordinated Universal Time as maintained by the U.S. Naval Observatory (UTC (USNO) reported in NMEA logs) is also based on an atomic time scale, with an offset of an integer number of seconds with respect to GPS Time. GPS Time is designated as being coincident with UTC (USNO) at the start date of January 6, 1980 (00 hours). GPS Time does not count leap seconds, and therefore an offset exists between UTC (USNO) and GPS Time (at this date in April 2007: 14 seconds). The GPS week consists of 604800 seconds, where 000000 seconds is at Saturday/ Sunday midnight GPS Time. Each week at this time, the week number increments by one, and the seconds into the week resets to 0.

5.2.2.3 GLONASS Time vs. Local Receiver Time

GLONASS time is based on an atomic time scale similar to GPS. This time scale is UTC as maintained by Russia (UTC (SU)). Unlike GPS, the GLONASS time scale is not continuous and must be adjusted for periodic leap seconds. Leap seconds are applied to all UTC time references as specified by the International Earth Rotation and Reference System Service (IERS). Leap seconds are used to keep UTC close to mean solar time. Mean solar time, based on the spin of the Earth on its axis, is not uniform and its rate is gradually changing due to tidal friction and other factors such as motions of the Earth's fluid core.

GLONASS time is maintained within 1 ms, and typically better than 1 microsecond (µs), of UTC (SU) by the control segment with the remaining portion of the offset broadcast in the navigation message. As well, Moscow offsets GLONASS time from UTC (SU) by plus three hours. The GLOCLOCK log, refer to the OEMV Family Firmware Reference Manual, contains the offset information between GPS and GLONASS time.

5.2.2.4 Datum

A datum is a set of parameters (translations, rotations, and scale) used to establish the position of a reference ellipsoid with respect to points on the Earth’s crust. If not set, the receiver’s factory default value is the World Geodetic System 1984 (WGS84).

Table 5.2 Comparison of GLONASS and GPS Characteristics

Parameter	Detail		GLONASS	GPS
Satellites	Number of satellites		21 + 3 spares [a]	21 + 3 spares [a]
	Number of orbital planes		3	6
	Orbital plane inclination (degrees)		64.8	55
	Orbital radius (kilometers)		25 510	26 560
Signals	Fundamental clock frequency (MHz)		5.0	10.23
	Signal separation technique [b]		FDMA	CDMA
	Carrier frequencies (MHz)	L1	1598.0625 - 1609.3125 [c]	1575.42
		L2	1242.9375 - 1251.6875	1227.6
	Code clock rate (MHz)	C/A	0.511	1.023
		P	5.11	10.23
	Code length (chips)	C/A	511	1 023
		P	5.11 x 10^6	6.187104 x 10^{12}
C/A-code Navigation Message	Superframe duration (minutes)		2.5	12.5
	Superframe capacity (bits)		7 500	37 500
	Superframe reserve capacity (bits)		~620	~2 750
	Word duration (seconds)		2.0	0.6
	Word capacity (bits)		100	30
	Number of words within a frame		15	50
	Technique for specifying satellite ephemeris		Geocentric Cartesian coordinates and their derivatives	Keplarian orbital elements and perturbation factors
	Time reference [d]		UTC (SU)	UTC (USNO)
	Position reference (geodetic datum) [e]		PZ-90	WGS84

---

a At the time of publication, April 2007, there are 29 operational GPS satellites and 12 operational GLONASS satellites in orbit.

b Full GLONASS system operation will consist of 24 satellites and only 12 channels. Such a system of simultaneous multiple transmissions is known as frequency division multiple access (FDMA) and distinguishes GLONASS from GPS, which is a code division multiple access (CDMA) system. See also the *GPS and GLONASS Satellite Identification* section of this overview starting on *Page 4*.

c Refer to the GLONASS Interface Control Document (ICD), Version 5.0, Moscow, 2002 for more details. You can find GLONASS contact information on their website at http://www.glonass-ianc.rsa.ru.

d GLONASS and GPS use different time systems. GLONASS time is referenced to UTC (SU), the Russian National Etalon time scale, whereas, GPS Time is referenced to UTC as maintained by the U.S. Naval Observatory UTC (USNO). The GLONASS control segment periodically applies a time step to bring the system's time within several hundred nanoseconds of UTC.

e GLONASS ephemerides are referenced to the Parametry Zemli 1990 (PZ-90, or in English translation, Parameters of the Earth 1990, PE-90) reference frame. The realization of the PZ-90 frame through adopted reference station coordinates has resulted in offsets in origin and orientation as well as a difference in scale with respect to WGS84 used by GPS. Relationships between PZ-90 and WGS84 have now been established.

5.3 GALILEO

GALILEO is Europe's initiative for a state-of-the-art global navigation satellite system, providing a highly accurate, guaranteed global positioning service under civilian control. Galileo will be not too different from the other GNSS parts (modernized GPS and Glonass). It will provide autonomous navigation and positioning services, but at the same time will be interoperable with the two other global satellite navigation systems; the GPS and GLONASS. A user will be able to take a position with the same receiver from any of the satellites in any combination. By providing dual frequencies as standard, however, GALILEO will deliver real-time positioning accuracy down to the meter range. It will guarantee availability of the service under all, but the most extreme circumstances and will inform users within seconds of a failure of any satellite. This will make it appropriate for applications where safety is vital, such as running trains, guiding cars and landing aircraft. The combined use of GALILEO and other GNSS systems can offer much improved performance for all kinds of users worldwide. GALILEO is expected to be in operation by the year 2008. The first satellite of Galileo system (GIOVE A) has already been lunched in 27th December 2005.

5.3.1 Galileo segments

Galileo segments are almost similar to GPS, but with some modification. The main extension of Galileo compared to GPS is the implementation of a global/ regional segment for integrity monitoring. The objective is to assist the safety critical aircraft navigation and locate and guide railway trains.

5.3.1.1 Space Segment

The space segment or the constellation features consists of 30 Medium Earth Orbiting (MEO) satellites (27 and 3 active spare satellite), distributed evenly and regularly over three orbit planes. The projected altitude is slightly larger than for GPS 23,616 km and the inclination is 56°.

5.3.1.2 Groundsegment

The Galileo ground segment is responsible for managing the constellation of navigation satellites, controlling core functions of the navigation mission such as orbit determination of satellites, and clock synchronization, and determining and disseminating (via the MEO satellites) the integrity information, such as the warning alerts within time-to-alarm requirements, at global level. The Global ground segment will also provide interfaces with service centers. The Ground Control Segment will consist of about 12-15 reference stations, 5 up-link stations and two control centers. The ground segment also will include 16-20 monitor stations, three up-link stations for integrity data and two central stations for integrity computations.

5.3.1.3 User Segment

The user segment consists of different types of user receivers, with different capabilities related to the different GALILEO signals in order to fulfill the various GALILEO services Figure 5.7.

5.3.1.4 Galileo signals

The GALILEO frequency should respect the radio-regulations as they are discussed and agreed on at the International Telecommunications Union (ITU) forums such as the World RadioCommunication Conference (WRC). There were different studies that were conducted before the determination of the Galileo signal allocations in order to avoid interference with GPS and Glonass systems, which operate in the same portion of the RF spectrum. Galileo will provide several navigation signals in right-hand circular polarization (RHCP) in the frequency ranges of 1164-1215 MHz (E5a and E5b), 1260-1300 MHz (E6) and 1559-1592 MHz (E2-L1-E1) that are part of the Radio Navigation Satellite Service (RNSS) allocation (Hein et al., 2003). All Galileo satellites will share the same nominal frequency, making use of code division multiple access (CDMA) techniques. Galileo will use a different modulation scheme for its signals, the binary offset carrier (BOC) and quadrature phase skip keying (QPSK).

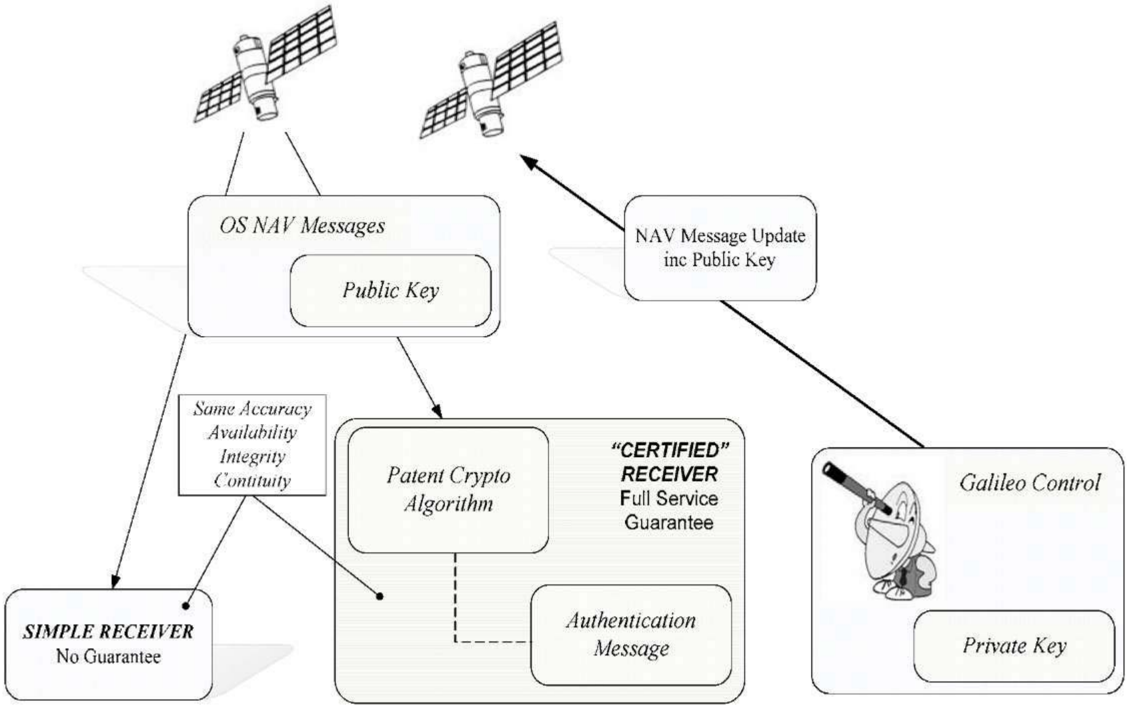


Figure 5.10. GALILEO System Architecture

5.3.1.5 Definition of Services

The Galileo constellation offers the capability of broadcasting globally a set of six signals supporting the open, commercial, safety-of-life and public regulated services. Each navigation signal is composed of one or two ranging codes and navigation data as well as, depending on the signal, integrity, commercial and search and rescue (SAR) data. Satellite-to- user distance measurements based on ranging codes and data are used in the GALILEO user receivers to fulfill the different GALILEO services). The main services are:

1.Open service (OS) data: These are transmitted on the E5a, E5b and E2-L1-E1 carrier frequencies. OSdata are available to all users and consist mainly of the navigation and SAR data. Open service offers positioning, navigation and timing signals, which can be accessed free of charge.

2.Commercial Service (CS), data: These are transmitted on the E5b, E6 and E2-L1-E1 carriers. All CSdata are encrypted and provided by service providers that interface with the Galileo Control Centre. Access to those commercial data is provided directly to the users by the service providers. The signal is designed to support very precise local differential applications (Sub- meter accuracy) using the open (option encrypted) signal overlaid with the PRS signal on E6 and also support the integration of GALILEO positioning applications and wireless communications networks.

3.Safety-of-life Services (SOL) data: These include mainly integrity and Signal in Space Accuracy (SISA)data. Combination of this Galileo services either with the current GPS as augmented by EGNOS corrections, or the future improved GPS and EGNOS integrity-only. Particularly, SOL is based on the satellite navigation signals without using added elements such as WAAS, and EGNOS. The accuracy required is about 4 meter over the Globe. This could be possible by introducing the ionospheric model based on multiple frequency measurements and modeling the other GNSS errors.

4.Public Regulated Service (PRS) data: These are transmitted on E6 and L1 carrier frequencies. ThePublic Regulated Service is provided on dedicated frequencies to provide the capability for greater continuity of service placed under EU Governments control for Public applications devoted to European and/or National Security, such as police, civil protection, law enforcement, civil protection such as some emergency services, as well as other governmental activities. The PRS is robust in order to be resistant to interference, jamming and other accidental or malicious aggressions.

5.4 GNSS SIGNALS

The overall of mentioned signals (Modernized GPS, Galileo and Glonass signals), make up the GNSS signals. Each satellite system has specific signal characteristics, but each system attempts to be compatible with the others in order to prevent the interferences and attenuation between the signals. It is important to consider that the processing of all signals should be performed using the same receiver, thus a complex receiver design is supposed to be designed and built. As mentioned above, The GNSS frequency plan shall respect the radio-regulations as they are discussed and agreed on at ITU forums. The available spectrum which can be used for the development of Radio-Navigation Satellite Systems (RNSS) is shown in Figure 5.8.

5.5 SIGNAL PROCESSING AND RECEIVER DESIGN

The main function of the signal processor in the receiver is the reconstruction of the carriers and extraction of codes and navigation messages. After this stage the receiver performs the Doppler shift measurement by comparing the received signal by a reference signal generated by the receiver. Due to the motion of satellite, the received signal is Doppler shifted. The code ranges are determined in the delay lock loop (DLL) by using code correlation. The correlation technique provides all components of bimodulated signals. The correlation technique is performed between the generated reference signal and the received one. The signals are shifted with respect to time so that they are optimally matched based on mathematical correlation. Currently some geodetic type receivers are available on the market tracking GPS and Glonass satellites simultaneously on both frequencies, in particular the Ashtech Z18 receiver and the TPS (Topcon Positioning Systems) Legacy receivers. The future GNSS receiver could be designed to track the different GNSS signals and could be of many types:

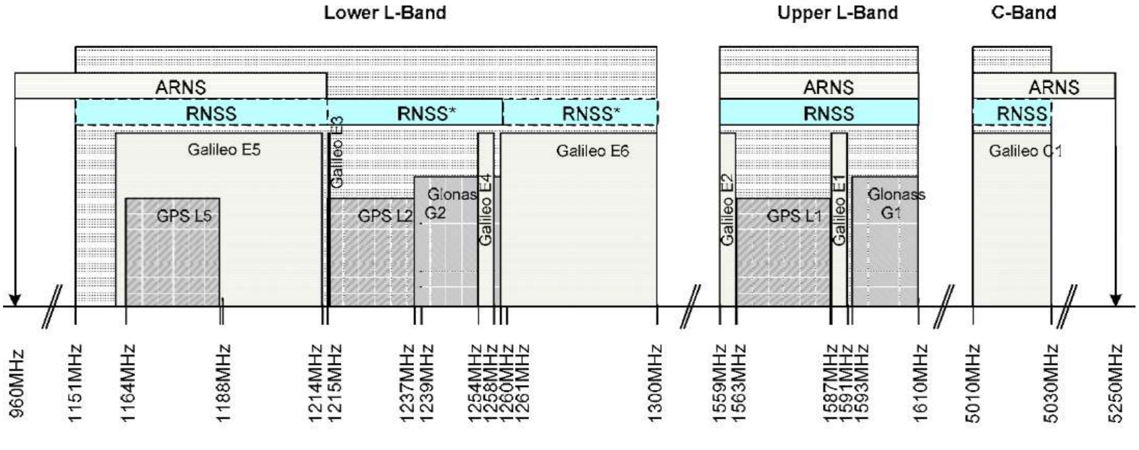


Figure 5.11. Radio-Navigation Satellite Systems (RNSS) frequency spectrum defined for GNSS signals

- The first type could process all GNSS signals GPS L1, L2, L5 and Galileo OS, CS using L1, E5 and E6 and also Glonass L1 and L2.

- The second type uses free signal and codes, GPS L1 and L2C and Galileo OS, on L1 and E5.
- The third type uses L1 and E5.
- Forth type uses GPS L1 and L2 (which are already in the market.
- Fifth type uses GPS and Glonass signals (which are already exist),

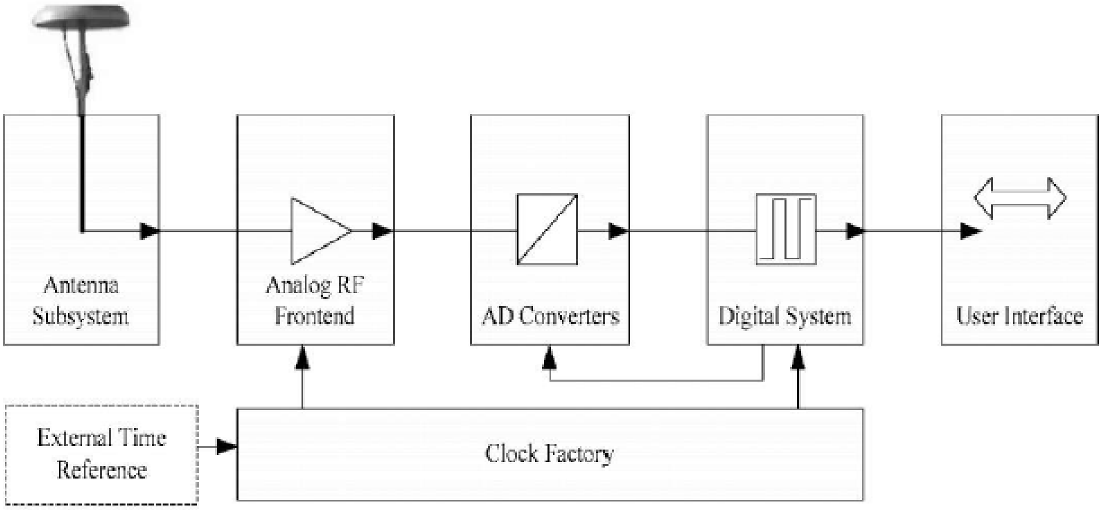


Figure. 5.12. Hybrid Galileo/GPS Receiver Concept

The most common receiver types are Intermediate Frequency receiver (IF) and the software defined radio receiver (SDR). In the RF front-end receiver the signal is down converted to an intermediate frequency and then sampled, but SDR uses direct digitization, or bandpass sampling. The main components of RF-FE combined GNSS receiver are shown in Figure 5.12. After sampling and analog to digital conversion (ADC) of the received signal, the receiver performs parallel de- spreading. The received base-band signal is multiplied in parallel with the spreading codes of all visible satellites. The received signal of each satellite is multiplied in parallel with different code delay offsets. These products are then accumulated to compute the cross-correlation function.

Because BOC signals are used in Galileo, supplementary measures are necessary due to the multiple correlation peaks of the auto-correlation function. Carrier tracking is performed using a phase-locked or frequency-locked loop (PLL or FLL). Coherent correlation combined with differential or non-coherent correlation can be done for the pilot and the data channel. Multiple signals will be available at L1 within the next few years. Galileo will use a different modulation scheme for its signals such as BOC and QPSK, while GPS uses binary phase shift keying (BPSK) modulation for the open signals at L1 and L2. The L5 signal that will appear with the Block IIF satellites in 2006, will have quadrature phase shift keying (QPSK). The binary offset carrier (BOC) modulation scheme of Galileo provides better multipath and receiver noise performance compared to the GPS binary phase shift keying (BPSK) modulation. More complex techniques are already developed for tracking BOC signal, such as bump jump and BPSK-like.

5.6 CONCLUSION

Global Navigation Satellite Systems (GNSS) technology has become vital to many applications that range from city planning engineering and zoning to military applications. It has been widely accepted globally by governments and organizations. That is why we expect to have very soon at least three GNSS systems: the USA GPS, European Galileo, and the Russian Glonass systems. There is a multibillion dollar investment in this field and intensive worldwide research activities. The impressive progress in wireless communications and networks has played a great role in increasing interest in GNSS and providing enabling methodologies and mechanisms. It is expected that all 3G (4G) and future generations of cellular phones will be equipped with GNSS chips. GNSS technology dominates the outdoor navigation, which provides accuracy to the range of few meters to 10 m in single point positioning technique or sub-meter to a few meter level in differential GNSS technique (DGNSS). Different techniques have been developed recently for indoor positioning. They offer either absolute or relative positioning capabilities with acceptable precision. Combining these technologies with GNSS allows to provide a more reliable and robust location solution. Most common implementation of Hybrid technology for GSM, GPRS and WCDMA is to combine A-GNSS with Cell-ID (Hybrid location technology).

References

[1] Kleusberg, A., and R. B. Langley, “The Limitations of GPS,” GPS World, Vol. 1, No. 2, March/April 1990, pp. 50–52.

[2] Shaw, M., K. Sandhoo, and D. Turner, “Modernization of the Global Positioning System,” GPS World, Vol. 11, No. 9, September 2000, pp. 36–44.

[3] Hoffmann-Wellenhof, B., H. Lichtenegger, and J. Collins, Global Positioning System: Theory and Practice, 3rd ed., New York: Springer-Verlag, 1994.

[4] El-Rabbany, A., The Effect of Physical Correlations on the Ambiguity Resolution and Accuracy Estimation in GPS Differential Positioning, Introduction to GPS Technical Report No. 170, Department of Geodesy and Geomatics Engineering, Fredericton, New Brunswick, Canada: University of New Brunswick, 1994.

[5] Wells, D. E., et al., Guide to GPS Positioning, Fredericton, New Brunswick: Canadian GPS Associates, 1987.

[6] Georgiadou, Y., and K. D. Doucet, “The Issue of Selective Availability,” GPS World, Vol. 1, No. 5, September/October 1990, pp. 53–56.

[7] Langley, R. B., “Time, Clocks, and GPS,” GPS World, Vol. 2, No. 10, November/December 1991, pp. 38–42.

[8] Kaplan, E., Understanding GPS: Principles and Applications, Norwood, MA: Artech House, 1996.

[9] Weill, L. R., “Conquering Multipath: The GPS Accuracy Battle,” GPS World, Vol. 8, No. 4, April 1997, pp. 59–66.

[10] Langley, R. B., “The GPS Receiver: An Introduction,” GPS World, Vol. 2, No. 1, January 1991, pp. 50–53.

[11] Schupler, B. R., and T. A. Clark, “How Different Antennas Effect the GPS Observable,” GPS World, Vol. 2, No. 10, November/December 1991, pp. 32–36.

[12] Nolan, J., S. Gourevitch, and J. Ladd, “Geodetic Processing Using Full Dual Band Observables,” Proc. ION GPS-92, 5th Intl. Technical Meeting, Satellite Div., Institute of Navigation, Albuquerque, NM, September 16–18, 1992, pp. 1033–1041.

[13] Klobuchar, J. A., “Ionospheric Effects on GPS,” GPS World, Vol. 2, No. 4, April 1991, pp. 48–51.

[14] Komjathy, A., Global Ionospheric Total Electron Content Mapping Using the Global Positioning System, Ph.D. dissertation, Department of Geodesy and Geomatics Engineering, Technical Report No. 188, University of New Brunswick, Fredericton, New Brunswick, Canada, 1997.

[15] Langley, R. B., “GPS, the Ionosphere, and the Solar Maximum,” GPS World, Vol. 11, No. 7, July 2000, pp. 44–49.

[16] Hay, C., and J. Wong, “Enhancing GPS: Tropospheric Delay Prediction at the Master Control Station,” GPS World, Vol. 11, No. 1, January 2000, pp. 56–62.

[17] Brunner, F. K., and W. M. Welsch, “Effect of the Troposphere on GPS Measurements,” GPS World, Vol. 4, No. 1, January 1993, pp. 42–51.

[18] Leick, A., GPS Satellite Surveying, 2nd ed., New York: Wiley, 1995.

[19] Kleusberg, A., “Comparing GPS and GLONASS,” GPS World, Vol. 1, No. 6, November/December 1990, pp. 52–54.

[20] Langley, R. B., “GLONASS: Review and Update,” GPS World, Vol. 8, No. 7, July 1997, pp. 46–51.

[21] Navtech Seminars and GPS Supply, GPS/GNSS newsletter, May 17, 2001.

[22] Johnson, N. L., “GLONASS Spacecraft,” GPS World, Vol. 5, No. 11, November 1994, pp. 51–58.

[23] Bazlof, Y. A., et al., “GLONASS to GPS: A New Coordinate Transformation,” GPS World, Vol. 10, No. 1, January 1999, pp. 54–58.

[24] CANSPACE, “China Puts Second Navigation Positioning Satellite into Orbit,” Canadian Space Geodesy Forum, December 21, 2000.

[25] CANSPACE, “Chinese Launch Navigation Satellite,” Canadian Space Geodesy Forum, October 31, 2000.

[26] CANSPACE, “Chinese Satellite Navigation System Update,” Canadian Space Geodesy Forum, January 9, 2001.

[27] Kayton, M., and W. R. Fried, Avionics Navigation Systems, 2nd ed., New York: Wiley, 1997.

[28] Commission of the European Communities, “Commission Communication to the European Parliament and the Council on GALILEO,” Brussels, Belgium, November 22, 2000.

6 Antennas for satellite applications

6.1 Introduction

Our principal aim is to provide the students with a basic understanding of the key system properties of antennas. Such an understanding allows us to determine the transmission (loss) between two antennas, such as might occur where a signal is broadcast from the satellite to a user terminal (or vice versa). This is an important precursor in the determination of received signal strength. There are a bewildering number of different antenna designs in use for satelite applications, and probably several times this number in use for other (i.e. terrestrial) applications. Our intention here is to give the students an appreciation of the generic antenna types used for various types of satellite application (fixed, mobile, handheld, etc.), together with some illustrative examples. In this respect we have categorized antennas according to their gain (their ability to focus signals in specified angular ranges) – low medium or high. In the beginning of radio development, mobile communication systems were conceived for the transmission and receiving of telegraphy and telephony signals via antenna from ships, cars, trains and aircraft. The consideration of antenna transmission is inevitable, especially in GMSC systems, where their propagation characteristics are much affected by different and changeable local environments during movement of mobile and differ greatly from those observed in fixed satelite systems. To create antenna hardware for mobile GMSC systems, engineers have to consider all related factors in order to realize full mechanical and transmission potentials.

This chapter describes antenna characteristics, requirements and basic relations of antenna systems; considerable antenna classification for maritime, land, aeronautical and personal satellite applications for ships, vehicles (tracks, buses, cars and trains), aircraft (airplanes and helicopters) and integrated solutions with GSM antenna on the top of handheld phones; retrospective of all antenna configurations for GMSC, such as low-gain omnidirectional antennas, three principal divisions of medium-gain directional antennas and three types of high-gain directional aperture antennas and finally, all type of antennas are presented for particular MSC systems and antenna mounting and tracking systems.

6.2 Antennas

Thus far, we have discussed the propagation of electromagnetic waves withoutconsideration of how we launch and recover unguided electromagnetic waves between satellites and the Earth. Antennas1 perform this vital function. They are the interface between unguided waves and the guided waves contained within a satellite or user terminal. Note that, although our primary focus in this section will be antennas for use with radio waves, this definition of an antenna is sufficiently general to allow the term to be used for other types of electromagnetic wave – including optical ‘antennas’ (more usually known as telescopes).

6.2.1 General concepts

Guided versus unguided Waves

Electromagnetic waves passing between satellite and user terminals on or near the ground are examples of unguided waves. Antennas effect the transition between these unguided electromagnetic waves and guided waves which can be routed within the satellite or user terminal to the appropriate electronics. Such guided waves are contained within waveguides. Familiar forms of waveguide include coaxial cable, rectangular and circular metal waveguide (a hollow metal tube) and microstrip – the latter being particularly suited to Printed Circuit Board (PCB) manufacturing techniques and the uses of surface-mount electronic components.

A transmitting antenna thus takes an output electrical signal from a transmitter waveguide and launches it as an unguided wave (into space or the atmosphere), while a receiving antenna captures some fraction of an incoming unguided wave and channels it into the receiver waveguide input.

6.2.2 Antenna Properties

A full exposition of the properties of antennas is beyond the scope of this chapter, and interested readers are directed to the relevant texts. The vast majority of antennas are passive, reciprocal devices – that is, they function equally when used to transmit (guided to unguided wave) and receive (unguided to guided wave) – exceptions to this generalization being specialized antennas that contain non-reciprocal active electronic elements (such as RF amplifiers, isolators, etc). It is convenient when introducing antenna properties to consider first the antenna as being transmitting (radiating) and there after to consider the antenna as receiving, and finally to demonstrate the relationship between these properties resulting from reciprocity. A fundamental attribute of a transmitting antenna is the degree to which the radiated unguided energy is distributed in angle. The equivalent property for receiving antennas would be the degree to which antenna sensitivity to incoming radiation is distributed in angle. In both cases we shall assume that we are sufficiently far away from the antenna to be in its far-field (the region where the electromagnetic field is effectively a spherical wave with field components normal to the direction of propagation).

Spherical Polar Coordinate System

When considering antenna radiation properties, where we are generally interested in the variation in radiation strength with angle and distance, it is usual practice to use the spherical polar coordinate system {R, θ , φ} rather than the normal Cartesian coordinates {x, y, z}. Here, R is the radial distance from the origin, θ is the zenith angle (‘elevation’ angle measured downwards from the z-axis – typically oriented along the antenna boresight) and φ is the azimuth angle (in the x – y plane).

6.2.2.1 Antenna Directivity

Consider first a transmitting antenna. We define the radiation intensity I (θ, φ) resulting from this antenna in a given direction defined by the spherical polar angles θ and φ as the power radiated in that direction per unit solid angle (IEEE, 1993). Much reference is made in the field of antennas to an isotropic radiator – a hypothetical antenna that radiates equally in all directions – as, for an isotropic radiator, the radiation intensity will be the same in all directions and equal to the power accepted by the antenna divided by 4π (since there are 4πsr in a sphere).We may quantify the degree to which a real antenna discriminates in favour of some directions at the expense of others in terms of its directivity.

Antenna directivity D (θ, φ) is defined as the ratio of the radiation intensity in a given direction divided by the radiation intensity averaged over all directions. In terms of the total radiated power, its directivity is thus

$$D(\theta, \phi) = 4\pi \frac{I(\theta, \phi)}{P_{\text{rad}}} \tag{6.1}$$

Clearly, for the hypothetical isotropic radiator mentioned earlier, the directivity is unity in all directions. Figure 6.1 illustrates antenna directivity polar plots for example low-gain (turnstile antenna), medium-gain (horn antenna) and high-gain reflector (prime focus reflector) antennas. Note that the radial scale is logarithmic (in dBi).

A directional antenna radiation pattern typically comprises a single main lobe, together with a number of lesser sidelobes (and/or backlobes). Much effort is expended in designing antennas to maximize the efficiency of the main lobe and minimize the average level of sidelobes – which affect received noise and interference from unwanted directions. We are typically interested in the maximum directivity D_{max} (for the antenna's preferred polarization), as this defines the maximum degree to which the

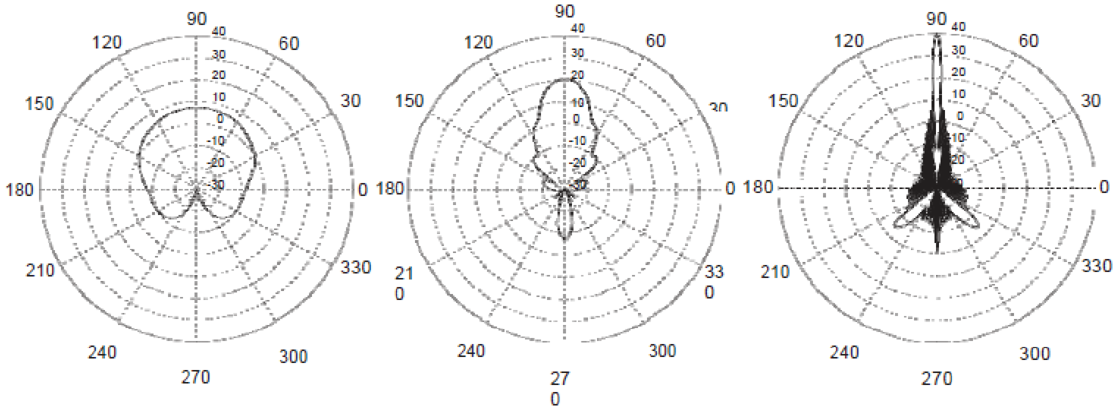


Figure 6.1 Example polar diagrams of antenna directivity (in dBi) for (left to right):

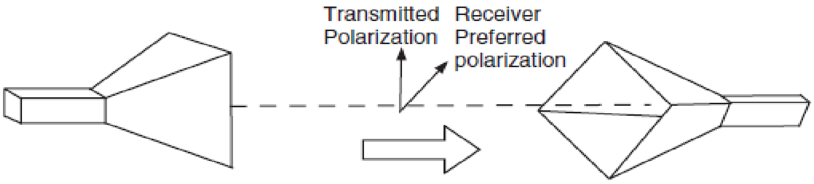


Figure 6.2 Antenna polarization mismatch.

crossed-dipole antenna can direct radiated power in a single direction (usually the antenna boresight in a unidirectional antenna).

6.2.2.2 Antenna Polarization

Antennas are designed to radiate waves with a particular polarization. This is usually either vertical or horizontal linear polarization or either left- or right-handed circular polarization. An antenna will therefore have preferred polarization, and the directivity may be resolved into directivities for two orthogonal polarizations (generally referred to as copolarization and cross-polarization). In general, directivities for the preferred and orthogonal polarizations are significantly different. The polarization mismatch loss factor resulting from non-optimum alignment of the receive antenna (refer to Figure 6.2) was defined previously in Chapter 3.

Although a fundamental antenna characteristic, the directivity of an antenna is rarely quoted, as it relates the radiation intensity to the radiated power rather than to the input power. The radiated power will always be less than the input power in practical antennas, owing to dissipative (ohmic) losses in the antenna construction and feed network, and a more useful property of antennas is antenna gain.

Antenna gain for a transmitting antenna is defined as the radiation intensity in a particular direction divided by the average radiation intensity that would result if all of the accepted input power were radiated, the latter being equal to the accepted input power P_{in} divided by 4π sr. Antenna gain is thus defined as

$$G(\theta,\varphi) \equiv 4\pi \frac{I(\theta,\varphi)}{P_{\text{in}}} \tag{6.2}$$

Gain and directivity may be related by introducing the antenna radiation efficiency η rad, defined as the ratio of the radiated power to the accepted input power, where upon

$$G = \eta_{\text{rad}} D \tag{6.3}$$

A typical value for the radiation efficiency for a commercial antenna is in the region of 60 – 80%. Antenna gain is usually quoted relative to the hypothetical lossless isotropic radiator, in dBi (meaning dB relative to isotropic). Gain of circular polarization antennas is quoted in dBiC (the C indicating circular polarization). The maximum gain Gmax (for the antenna's preferred polarization) is often of most interest.

6.2.2.3 Beamwidth

The Full-Width Half-Maximum (FWHM) beamwidth of an antenna is an important parameter that needs to be optimized for the particular application. The beamwidth of a satellite antenna defines the beam coverage area on the Earth. The beamwidth of a user terminal defines the degree to which the antenna needs to be pointed at the satellite, and the scope for interference to/from other satellites.

Relation Between Maximum Gain and Beamwidth

We have previously stated that antennas with high maximum gain have narrow beamwidths, and vice versa. As radiation intensity is defined as the power radiated per unit solid angle, we can relate the antenna maximum directivity to an equivalent beam solid angle ΩB, defined as the solid angle through which all the radiated power would stream if the power per unit solid angle were constant at the maximum value of the radiation intensity through this solid angle and zero elsewhere. Maximum directivity and beam solid angle are related via

$$D_{\mathrm{max}} = \frac{4\pi}{\Omega_{\mathrm{B}}} \qquad (6.4)$$

Gain Estimate

Equation (6.4) provides a useful way to estimate the maximum directivity for high-gain antennas. The beam solid angle for a high-gain ‘pencil beam’ antenna having antenna beamwiths vθx and vθy in the two principal axes can usefully be approximated by ΩB ≈ vθx vθy , from which the approximate maximum directivity (and potentially maximum gain) for specified antenna beamwidths may be estimated. Maximum antenna gain may thus be estimated using a suitable estimate for the radia- tion efficiency. Balanis (1997) gives the following approximate formula for maximum antenna gain ‘for many practical antennas’:

$$\Delta\theta = k\frac{\lambda}{D} \qquad (6.6)$$

where the value of k depends on the distribution of illumination across the radiating aperture. For a uniformly illuminated aperture, k → 1.02 rad (58.4◦). However, the aperture illumination of practical antennas is normally tapered towards the edges (from a maximum at the centre) in order to optimize efficiency while controlling sidelobe levels. A more typical illumination taper results in k → 1.22 rad (70◦).

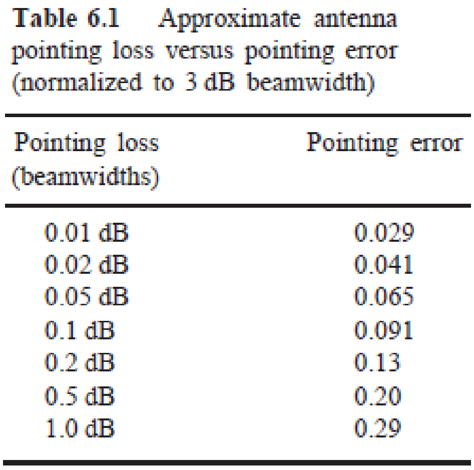


Table 6.1 Approximate antenna pointing loss versus pointing error (normalized to 3 dB beamwidth)

Pointing loss (beamwidths)	Pointing error
0.01 dB	0.029
0.02 dB	0.041
0.05 dB	0.065
0.1 dB	0.091
0.2 dB	0.13
0.5 dB	0.20
1.0 dB	0.29

Gain Approximation Near Boresight

The variation in antenna gain with offset angle for the main lobe near boresight is important when considering the accuracy required for antenna pointing and may be approximated (for each axis) by a parabola[3]

$$G(\theta) \approx G_{\max}\left(1 - 2\left(\frac{\delta\theta}{\Delta\theta}\right)^2\right) \tag{6.7}$$

where δθ is the mispointing angle – the angle offset from the peak gain direction (boresight) for the axis under consideration. We may use this approximation to estimate the antenna mispointing loss for different pointing errors, expressed as a fraction of the FWHM (3 dB) antenna beamwidth for that axis, as indicated in Table 6.1.

6.2.2.4 Impedance and Bandwidth

The impedance of an antenna is the ratio of the voltage across the antenna terminals to the current flowing into the antenna terminals. In general, the impedance of an antenna is complex (in both senses of the word), the voltage and current generally being offset in phase. The fraction of the real part of the antenna impedance that corresponds to power radiated from the antenna isknown as the radiation resistance. It must be noted that there is a fundamental relation between the volume of an antenna and its impedance bandwidth which ultimately limits their size reduction.

Typically, an antenna is only useful over a limited bandwidth, and the useful bandwidth of an antenna may be defined in a number of ways, but it is common to describe the bandwidth as the frequency range over which the antenna impedance presents an acceptable impedance mismatch loss .

Impedance Mismatch

Our definition of antenna gain relates the radiation efficiency to the accepted input power. Let us now explore the meaning of this phrase. In electronic circuits and transmission lines, efficient power transfer depends on matching the load impedance to that of the source. In general, a fraction of the input signal will be reflected at the antenna input terminals back along the transmission line towards the source (where it is dissipated) and the ratio $\Gamma^{\sim}$ of the amplitude of the reflected wave to that of the input wave (in complex phasor notation) is related to the mismatch in source and antenna impedances (Ramo, Whinnery and Van Duzer, 1965). The power reflection coefficient is $|\tilde{\Gamma}|^2$.

---

[3] When expressed in dBi, an alternative approximation for the main lobe gain is (for small angles)

$$(G(\theta))_{\mathrm{dB}} \approx (G_{\max})_{dB} - 12\left(\frac{\delta\theta}{\Delta\theta}\right)^2 \tag{6.8}$$

We define the accepted input power to be that which is not reflected back towards the source. The ratio of the accepted input power to the maximum available input (source) power (i.e. the input power if the antenna were properly matched) is known as the impedance mismatch loss (ratio) L_Z given by

$$\frac{1}{L_Z} = \left(1 - |\tilde{\Gamma}|^2\right) \qquad (6.9)$$

An equivalent situation occurs for the receiving antenna, in which case the ratio of the power delivered from the antenna to the load via the transmission line relative to the maximum output power that would be delivered into a matched load is also given by equation (6.9).

6.2.2.5 Transmitter Figure of Merit

The radiation intensity in a particular direction is proportional to the product of the accepted power and the antenna gain times. This product is the Effective Isotropic Radiated Power (EIRP), defined as

$$\mathrm{EIRP} \equiv P_t G_t \qquad (6.10)$$

where P_t is the accepted power in the transmitter antenna (after any transmission line losses), and G_t its gain. Physically, EIRP is the equivalent power that would have to be radiated by an, isotropic radiator to achieve the same intensity as the actual antenna in the direction of interest (usually in the direction of maximum antenna gain). The EIRP is thus an important figure of merit for a transmitting antenna. EIRP is typically given in dBW (dB relative to 1 W) or dBmW (dB relative to 1 mW).

6.2.2.6 Effective Aperture Area

Power Flux Density

We may relate the power flux density (PFD – the power per unit area) at some radius R from a transmitting antenna to the transmitted EIRP, ignoring any propagation losses, via

$$\mathrm{PFD} = \frac{\mathrm{EIRP}}{4\pi R^2} \qquad (6.11)$$

So far we have focused on the properties of a transmitting antenna in terms of the radiation intensity for a given input power. For a receiving antenna we may conveniently introduce an effective area (or effective aperture), defined as the ratio of the available power at the output terminals of a receiving antenna to the power flux density incident on the antenna from a specified direction (under the assumption that the input wave polarization is matched to that of the antenna). The concept of effective area is particularly intuitive for aperture antennas which have a defined physical aperture area, although, in general, the effective area is somewhat less than the physical area owing to inefficiencies. Consider the arrangement of Figure 6.3. Using the definition for effective area, the received power delivered to the antenna terminals is given by

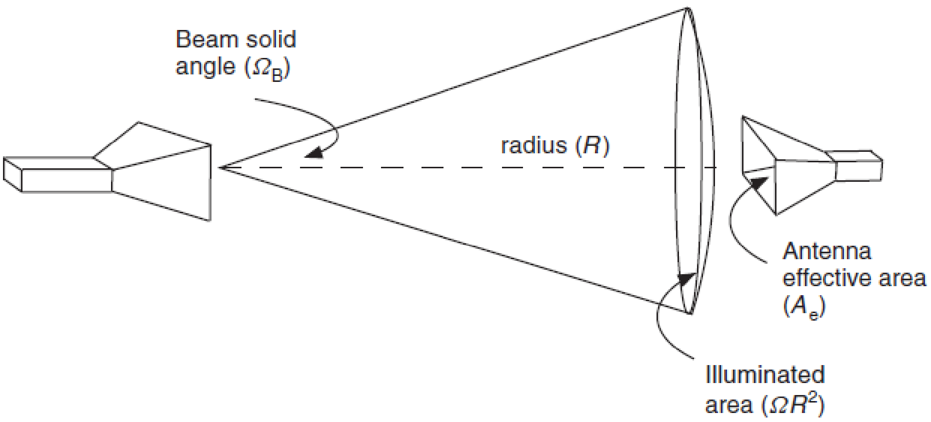


Figure 6.3 Beam solid angle, PFD and effective aperture.

where A_e is the effective aperture. As with directivity and gain, we are usually interested in the maximum effective area (the effective area in the direction of maximum antenna gain).

As a result of reciprocity, it should not matter whether the power flows from antenna 1 to antenna 2 or vice versa. As a consequence, we note that the product of the transmitter antenna gain and the receiver effective area is constant (i.e. $G_1 A_{e,2} = G_2 A_{e,1}$), which implies that, for any antenna, the ratio of effective area to maximum gain is a constant. Using this together with analytical expressions derived for the gain and effective area for a particular antenna – the short dipole – Friis (1946) derived a fundamental relationship between gain and effective aperture, valid for any antenna:

$$G = \frac{4\pi}{\lambda^2} A_e \tag{6.13}$$

Approximate Gain of Aperture Antennas

Equation (4.13) provides a useful means to relate antenna maximum gain of an aperture antenna to its physical aperture area. We shall define the antenna aperture efficiency ηa to be the ratio of the maximum effective aperture area to the physical aperture area. For well-designed reflector antennas, aperture efficiency is typically in the region 60–80%. Hence, for a circular aperture antenna (such as a circular reflector antenna) of diameter D, the maximum antenna gain is

$$G_{\max} \rightarrow \eta_a \left(\frac{\pi D}{\lambda}\right)^2 \tag{6.14}$$

6.2.2.7 Antenna Arrays

Individual antennas may be connected together to create antenna arrays with desirable properties. In antenna arrays, the component antenna elements are made to radiate together with specified amplitude and phase relationships (for each frequency). For this reason they are also referred to as phased arrays . Antenna arrays permit the construction of antenna systems with high maximum gains from low- or medium-gain element antennas. Such arrays can also offer advantages in terms of their physical geometry – for example, the antenna may take the form of a (thin) flat panel or be made conformal to some surface.

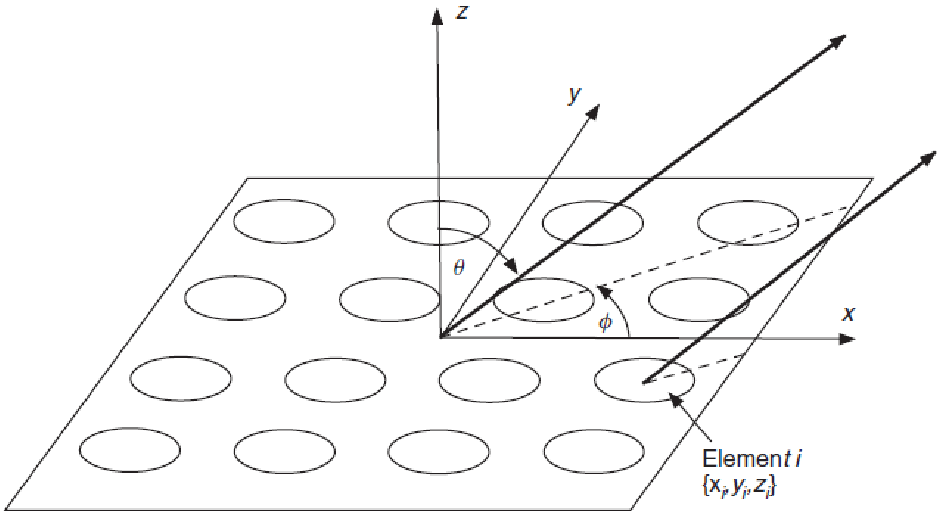


Figure 6.4 Planar phased array.

Phased arrays also provide additional degrees of flexibility with regard to control over radiation patterns – in principle, an array with N elements has (N − 1) degrees of freedom available in the control of its radiation pattern. Lastly, phased arrays may be constructed such that the phase relationship between the elements can be adjusted electronically, allowing the radiation pattern to be modified, at will – facilitating beam scanning and beam steering (although this flexibility usually comes at a significant additional cost, weight and power consumption). Although antenna arrays may in principle take almost any shape, linear, circular and planar geometry arrays are the most common forms, with linear and planar arrays being the most relevant array configurations for personal satellite applications. It will be useful to have a limited understanding of phased-array antennas; however, for a more detailed understanding of their behavior, the reader is directed to the many excellent textbooks in this field (Balanis, 1997; Kraus, 1988). With reference to the planar array in Figure 6.4, consider the contribution to the radiated far-field of an antenna array owing to the ith element of the array, located at {xi, y i, z i }. For a given far-field direction (θ, φ), the propagation phase advance (or delay) ψ_i of this element, relative to the array reference origin, is given by

$$\Delta\psi_i = \frac{2\pi}{\lambda}\left(x_i \cos(\phi)\sin(\theta) + y_i \sin(\phi)\sin(\theta) + z_i \cos(\theta)\right) \qquad (6.15)$$

Now, if ⍰$_i$ is the (complex phasor) electric field due to the ith element (referenced to the element coordinates), then the total farfield of the contributions from each array element (referenced to the array origin) is given by the sum of all field contributions

$$\bar{\mathcal{E}}_{\text{array}}(\theta, \phi) = \sum_1^N \tilde{w}_i \, \bar{\mathcal{E}}_i(\theta, \phi) \exp(\mathrm{i}\,\Delta\psi_i) \qquad (6.16)$$

where $\hat{w}_i$ is a (complex) weight for the ith element, introduced to control the radiation pattern. The phase and amplitude of the individual element weights may be used to control (or scan) the beam maximum gain direction and control beam shape and sidelobe levels.

Array Factor

Under the assumption that all of the antenna elements are identical and have the same relative orientation, and that electromagnetic coupling between elements can be neglected, the farfield patterns of all of the array elements (referenced to the elements themselves) will also be identical, and the total farfield may be expressed as the product of the element farfield and a (complex) array factor A¯F given by

$$\bar{\mathrm{AF}} \equiv \sum_1^N \tilde{w}_i \exp(\mathrm{i}\,\Delta\psi_i) \qquad (6.17)$$

The array factor is essentially the radiation pattern of an equivalent array of isotropic elements.The array antenna gain G_{array} may thus be simply expressed as the product of the element gain G_{elem} and the modulus squared of the array factor

$$G_{\text{array}}(\theta,\phi) = G_{\text{elem}}(\theta,\phi)\left|\bar{\text{AF}}(\theta,\phi)\right|^2 \tag{4.18}$$

Array Factor Under Cophasal Excitation

A particularly simple method to direct maximum array gain in a particular direction is to set the phase of each element weight equal and opposite to the relative phase delay resulting from the element position in the array. This is known as cophasal excitation. For an array with a fixed beam pattern, this may be implemented using a passive feed network.

As the array factor is a summation over N elements, and as the array gain is proportional to the square of the array factor, it may be inferred that the maximum array gain can be enhanced by a factor of up to N_2 compared with that for a single element. However, in a passively fed phased array there is a requirement to split (for a transmit array) or combine (for a receive array) the signal N ways, in order to form a single feed point for the antenna. This splitting/combining results in a power loss factor of 1/N . Consequently, in this type of phased array, the maximum gainenhancement is only a factor of N.

Linear Array

This result provides a useful first estimate of the number of elements needed to obtain a particular gain (or the gain given the number of elements). It may be used, for example, to estimate the maximum gain of a Yagi–Uda array antenna. The Yagi–Uda array is essentially a medium-gain linear array of elemental dipoles in which only one of the array elements is actually driven; the rest are excited parasitically.

Planar Array

The other common form of antenna array is the planar array. Typically, this comprises low-gain antenna elements (such as patch antennas) or sometimes medium-gain antennas (such as horn antennas) arranged in a rectangular or hexagonal grid. If the array comprises N closely packed aperture antenna elements, the maximum effective aperture of the array will tend to N times the effective aperture of a single element. Furthermore, if the effective area of the individual antenna elements is comparable with its physical aperture, then, for densely packed planar arrays, the maximum array gain will approach that of a single-aperture antenna with the equivalent total effective aperture.

Antenna array performance is sensitive to the interelement array spacing: too close and excessive mutual coupling degrades the beamwidth and impedance match; too far apart (more than half a wavelength) and a nominally unidirectional pattern will instead exhibit multiple (main) lobes – so called grating lobes.

6.2.3 Antennas for Personal Satellite Applications

6.2.3.1 Types of Antenna

Antennas come in a bewildering range of shapes and sizes, depending on their intended application and design heritage. Indeed, there are probably more different antenna designs than pages in this book. Nevertheless, most of these designs are adaptations of a relatively small number of generic types. Furthermore, antennas for use in satellite applications represent a relatively small subset of the available antenna types. To facilitate our discussion, we shall initially categorize antennas for use in satellite applications according to the general type of radiation pattern they produce. These patterns reflect the relative distribution of radiation as a function of angle for transmitting antennas and the relative distribution of sensitivity with angle for a receiving antenna. The following generic pattern types are considered here:

- torus-shaped, omnidirectional pattern (e.g. that for a common dipole antenna);
- low-gain (wide-beamwidth) unidirectional pattern (e.g. that for a patch antenna);
- medium-gain (moderate-beamwidth) unidirectional pattern (e.g. that for a horn antenna);
- high-gain (narrow-beamwidth) unidirectional pattern (e.g. that for a reflector antenna).

6.2.3.2 Torus-Shaped Omnidirectional Antennas

The torus-shaped omnidirectional antenna pattern is exemplified by the common dipole antenna and its cousin, the monopole (in effect, one-half of a dipole above a reflecting surface). This type of antenna, and its derivatives, is used in everyday devices such as mobile phones and wireless network devices. The familiar ‘doughnut’-shaped radiation pattern is particularly attractive for use in terrestrial mobile applications where transmissions emanate from near the horizon and where the azimuth angle of the signal with respect to the antenna is not predictable in advance (and where the cost of implementing a signal tracking scheme would be prohibitive). However, the usefulness of such antennas for satellite applications (for which the elevation angle may be significant) is limited by their toroidal radiation pattern which has maximum gain at the horizon (normal orientation) and zero gain at the zenith. Furthermore, their linear polarization would result in angle-dependent polarization loss. Antennas with torus-shaped omnidirectional patterns therefore find limited application in satellite applications.

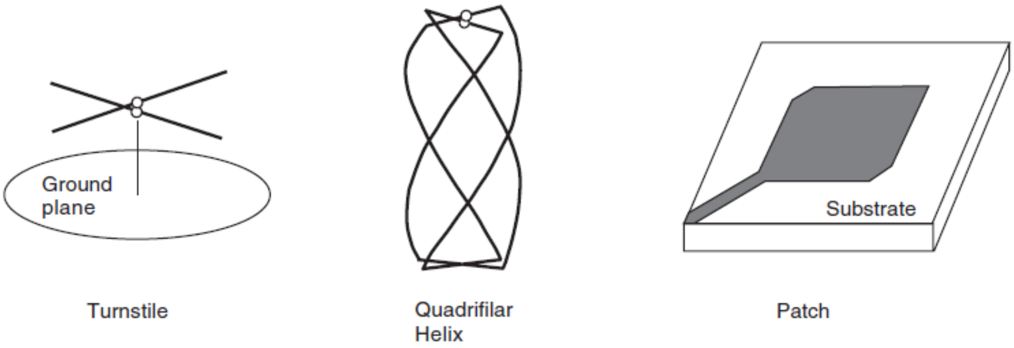


Figure 6.5 Low-gain antennas for circular polarization.

6.2.3.3 Low Gain (Near-Hemispherical-Coverage) Unidirectional Antennas

A more useful type of antenna pattern for use in mobile and handheld satellite applications is the low-gain unidirectional antenna with hemispherical (or near-hemispherical) coverage, as illustrated in Figure 6.5 (left). Correctly oriented, these antennas provide coverage of the upper hemisphere (where all visible satellites are located) and are useful where the azimuth of the satellite signal is unknown. No pointing of the antenna is needed in this case. Such antennas are widely used in mobile satellite applications, for example in GPS or other satellite navigation system receivers.

Examples of low-gain circular polarized antennas having nominally hemispherical radiation patterns are illustrated in Figure 6.1. These include: the crossed dipole or turnstile antenna (a pair of dipoles above a reflecting surface); a quadrifilar helix (a helix comprising four arms) and patch antenna.

6.2.3.4 Medium-Gain Directional Antennas

The nominally hemispherical radiation patterns of low-gain antennas, while attractive for mobile applications where antenna pointing is a problem, have limited application where increased sensitivity is required and some degree of antenna pointing is acceptable, or where coverage needs to be reduced. An example radiation pattern for a medium-gain antenna (a conical horn) is illustrated in Figure 6.1 (centre).

Medium-gain antennas, some examples of which are illustrated in Figure 6.6, provide more angular selectivity, which results in greater sensitivity. In the figure, examples are shown of a Yagi–Uda array of crossed dipoles (a variant of the familiar Yagi–Uda antenna4 used for terrestrial television reception), an axial-mode helix antenna and a horn antenna. The operation of the horn antenna (of which both pyramidal and conical variants exist) is perhaps easiest to understand; the antenna forms a taper between its aperture (opening) and the waveguide section at the other end; to a first approximation, the slower the taper, the narrower is the beam. Typical gains of these antennas are in the region of 10-20 dB better than for the low-gain antennas.

6.2.3.5 High-Gain Directional Antennas

Still higher EIRP and sensitivities (or smaller beamwidths) require still higher gain antennas. Typically, very high gains are produced by further focusing the patterns from a medium-gain antenna (the feed antenna) using techniques originally developed in optics, in particular using lenses and telescopes. Alternatively, high-gain antennas may be produced using two-dimensional arrays (grids) of lower medium-gain antennas (see Section 6.2.2.8).

.

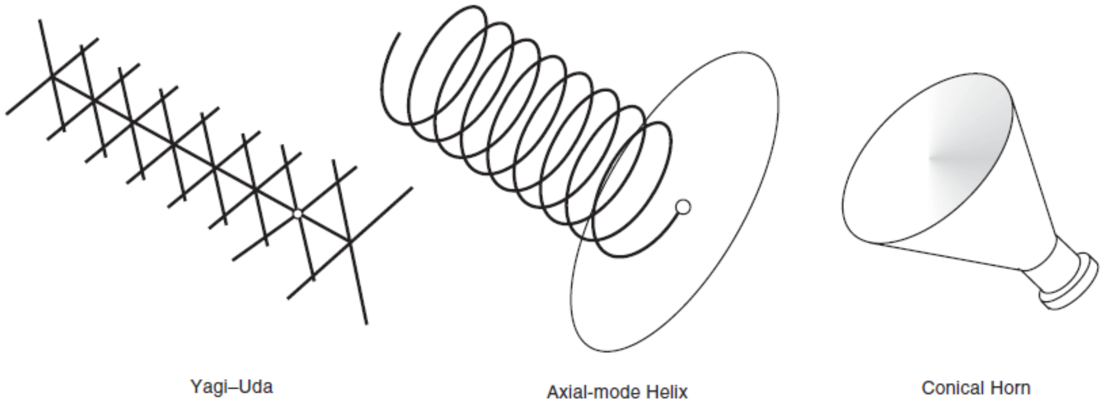


Figure 6.6 Medium-gain antennas.

High-gain reflector antennas transform the radiation patterns of a medium-gain antenna– such as a horn antenna – using reflecting lenses. Referring to Figure 6.7, the simplest reflector antenna has a feed antenna at the focus of a parabolic reflector (the point where parallel rays incident on the reflector converge). Such an antenna is known as a prime-focus antenna (top left). A drawback of the prime-focus antenna is that the feed (antenna) is located some distance in front of the reflector, and must be supported (and fed) with minimal obstruction to the main beam.

In the Cassegrain reflector antenna (named after the seventeenth-century telescope inventor), the feed is instead located at a small opening in the main reflector, and a small convex hyperbolic subreflector is used to deflect the beam from the feed onto the main parabolic reflector (Figure 6.7, top right). The use of a subreflector provides an additional degree of freedom in the antenna design, which allows Cassegrain antennas with higher gain than prime-focus reflectors, but, in order to be effective, the subreflector must itself be several wavelengths across (Balanis, 1997), which increases blockage. As a result, Cassegrain antennas are generally only used for very high-gain antennas. A third main catagory of reflector antenna is the Gregorian reflector (also named after a seventeenth-century telescope inventor), in which a concave elipsoid subreflector is used instead of the convex hyperberloid of the Cassegrain design. An advantage of the Gregorian reflector is that the separation of the feed and subreflector is increased and the feed antenna may be located behind the main reflector (Figure 6.7, bottom left). Again, Gregorian antennas are generally only used for very high-gain antennas.

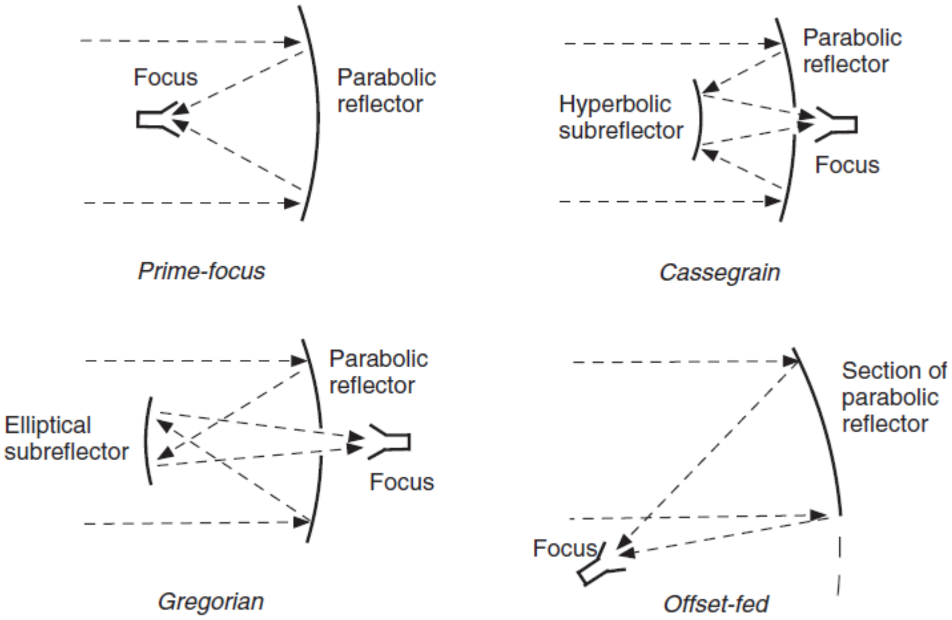


Figure 6.7 Reflector/telescope designs.

Feeding the antenna from behind the reflector is attractive in many situations, and, for relatively small reflector antennas, where the subreflector is small (in wavelengths) and almost flat, this type of antenna is referred to as a ‘splash-plate’ reflector.

A common drawback of all three reflector antenna types is that some part of the main beam is obstructed by the feed or subreflector and their support structures. This reduces antenna efficiency. In order to overcome this problem, the reflectors can be fed asymmetrically, as illustrated in Figure 6.7 (bottom right). In effect, these offset-fed reflector antennas have reflectors and subreflectors that are off-centre portions of larger reflectors. Such offset-fed reflectors can achieve higher efficiencies than centre-fed equivalents, albeit with slight degradation of polarization purity.

Maximum Gain of a Reflector Antenna

As the aperture size of an antenna increases (in wavelengths), the beamwidth decreases and the gain increases. In principle one might expect that the gain of an aperture antenna, and in particular that of a reflector antenna, can be increased ad infinitum simply by increasing the aperture size (for a given operating frequency). In practice, the maximum achievable gain will ultimately be limited by the departure of the antenna reflector geometry from the ideal. The effect of a given RMS surface roughness on surface efficiency has been estimated by Ruze (1966).

An RMS surface roughness of $\lambda/32$ will result in a surface efficiency of ∼85%, while an RMS roughness of $\lambda/16$ yields a surface efficiency of just ∼50%. Thus, at a frequency of 30 GHz (Kaband), we would require an RMS surface roughness better than 0.3mm in order to achieve a high antenna efficiency.

For mass-produced reflector antennas used in personal satellite applications, where the maximum antenna aperture is typically constrained by the required portability, pointing accuracy and/or mounting arrangements, this limit determines the required manufacturing tolerance and reflector mechanical strength (to avoid flexure) for these low-cost, mass-produced antennas. For the larger Earth station antennas used to anchor such services, additional strengthening is necessary, and ultimately the maximum achievable gain of extremely large-aperture (in wavelengths) Earth station antennas will be limited. For large satellite antennas, which must be very light and may be required to be unfurled in space, the impact of this limit on maximum antenna size can be significant, and is one limit on the exploitation of very narrow beamwidths (less than 1◦) in satellite applications.

6.2.3.6 Satellite Antennas

A variety of satellite antenna types are used to provide personal satellite services, depending on the satellite altitude, the desired beam coverage and the desired frequency and polarization. Intelsat, one of the largest fixed satellite service providers, which operates some 50 satellites spread worldwide, defines its satellite coverage by four types of beam:

- the global beam, which covers roughly one-third of the globe;
- the hemispherical beam, which covers roughly one-sixth of the disk;
- the zonal beam, which covers a large landmass like Europe;
- the spot beam, which covers a 'specific geographical area'.

Dual polarized antenna systems support two orthogonal polarizations – effectively doubling the satellite capacity

Antennas for Earth Cover Beam

Clearly, the beamwidth of a satellite antenna to provide coverage of the visible Earth will depend on satellite altitude. For a geostationary satellite, we find from Chapter 2 that the satellite nadir angle for zero minimum elevation angle is approximately 8.75◦ (i.e. a FWHM satellite antenna beamwidth of ∼17.5◦). The solid angle of a cone with this angle is ~0.073 sr. Therefore, the required antenna directivity is approximately 172 (22 dBi). In practice, the antenna pattern is not a simple conical shape, and the exact directivity of an Earth cover antenna will depend on the allowed gain reduction at the edge of cover. Earth coverage for a geostationary satellite is thus typically achieved by an antenna with a gain in the region of 13–22 dBi.

At microwave and millimetric frequencies, this value of gain is conveniently achieved using conical horn antennas, while at lower frequencies, for circular polarization, it is typically achieved using axial-mode helix antennas. Of course, where available, phased-array antennas may also be employed for Earth cover beams (using the appropriate phase weights).

Antennas for Zonal/Spot Beams

Spot beams – that is beams that are smaller than those that cover the visible Earth – are typically used to provide services requiring higher power/sensitivity. The shape and beamwidth of such spot beams depends on a number of factors and represent a compromise between maximum gain and coverage area. Radiation pattern for satellite antennas are often represented as a contour plot on the Earth to facilitate estimation of received signal quality.

The simplest spot beam antenna pattern is circular in shape, although the illuminated area appears non-circular on the ground owing to the Earth's curvature (essentially an elongation towards high latitude and/or longitude difference). To improve the transmission efficiency, radiation patterns (beams) may be shaped to best fit the service zone. Spot beams are typically provided using high-gain reflector antennas – typically, offset-fed reflector antennas. By way of example, Figure 6.8 shows the complement of reflector antennas on the Anik F-2 spacecraft, which provides a range of telecommunication services.

Phased-arrays antenna may also be used to provide spot beams for certain advanced applications, notably where the beam shape or location is altered dynamically.

At lower frequencies, the size of the satellite antenna reflector needed to obtain useful spot beams can present a problem, and typically these reflectors comprise a very light wire mesh, which is stowed away at launch and unfurled only when the satellite is in orbit. Notable examples of unfurlable reflectors are the Astromesh reflectors from Northrop Grumman (2010) used on Inmarsat- 4, Thuraya and MBSAT satellites. Figure 4.9 shows an artist's impression of the 9m diameter Astromesh reflector deployed on the Inmarsat-4 satellite. Another example of an unfurlable reflector is the Harris Folding Rib reflector, as deployed on the ACeS (Garuda) and ICO G1 satellites, which opens up rather like an umbrella (Harris Online, 2010). The maximum size of operational antenna at present is 18m launched recently by Terrestar in the United States to support personal applications and other services (Terrestar Online, 2010).

Antennas for Contoured Beams

For some applications it is desirable to synthesize the beam coverage to match particular geographic or political boundaries. There are two general approaches to forming complex contoured beams. One method is to use a standard reflector antenna but to deform the main reflector (or subreflector in a Cassegrain Antennas arrangement) so as to achieve the desired pattern. This approach is cost-effective but does not offer the flexibility to adjust the coverage once the satellite is launched.

Until recently, beam shapes of satellite antennas usually tended to remain fixed throughout a satellite's lifetime; more recently, and particularly for personal

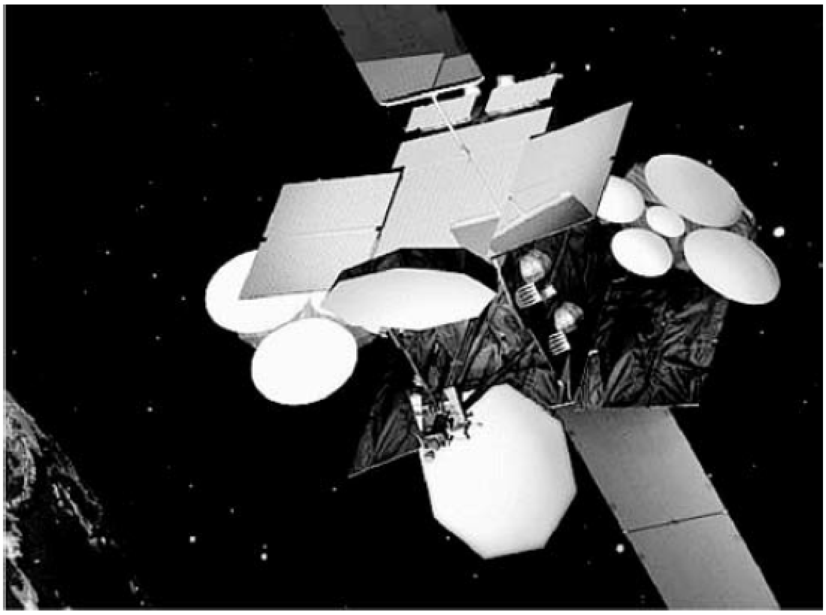


Figure 6.8 Anik F2 satellite antenna farm.

communication systems, satellites systems, have incorporated the capability to alter beams shapes, size or numbers, which allows for flexibility in allocation of satellite resources. For instance, an operator may decide to reduce or reshape beam size and illuminate uncovered areas in response to an event, or to introduce new products, or readjust coverage owing to a change in traffic pattern. The most flexible (but more expensive) approach is to use a phased array to feed a standard parabolic reflector. By adjusting the phase weights of the array elements, the beam shape can be optimized (and reoptimized) for the required coverage.

Figure 6.9 Artist's impression of the Inmarsat-4 satellite with its 9 m Astromesh reflector unfurled.

Figure 6.10 A 120-element Inmarsat-4 satellite L-band feed array being assembled

Antennas for Cellular, Multiple-Spot-Beam Coverage

It is often desirable to produce multiple spot beams, and it is increasingly common practice to maximize system capacity by synthesizing multiple beams in a hexagonal cellular-type pattern so as to maximize the potential for reuse of the limited available spectrum to a particular operator between non-adjacent 'cells', and to provide high EIRP. This type of cellular pattern of spot beams is generally formed by feeding a reflector antenna from a phased array at or near its focus. Figure 6.10 illustrates a 120-element L-band feed array being assembled for the Inmarsat-4 satellite. Inmarsat-4 (I-4) satellites can synthesize over 200 beams in the L-band service link, as illustrated in Figure 6.11. The satellite generates three types of beam – a global beam to cover the entire disc, 19 wide beams and 200 narrow spot beams to service portable user terminals. Each satellite deploys a 9 m furlable gold-plated molybdenum mesh reflector antenna which is offset fed by 120 helical

elements. Each beam is formed from about 20 feed elements fed with the appropriate amplitude and phase coefficients from an on-board digital beam former. Over 220 beams of different shapes and sizes can be formed by altering the beam coefficients. I-4 beam patterns have been altered on a number of occasions to satisfy the operational requirements.

6.2.3.7 Ground Station Antennas

Ground station antennas used as a hub to anchor satellite services will typically comprise high-gain reflector antennas at microwave frequencies. By way of illustration, Figure 6.12 shows an 9 m Cassegrain antenna from ViaSat. At lower frequencies, for circular polarization, arrays of axialmode helices or crossed-dipole Yagi–Uda antennas are often used.

6.2.3.8 User Equipment Antennas

Antennas for Fixed Satellite Applications

For applications where the orientation of the satellite is fixed with respect to the user, highgain antennas may be used. For wideband applications, offset-fed reflector antennas are generally employed. A typical example of a fixed reflector antenna from Kathrein is illustrated in Figure 6.13 (left). Planar arrays of patch antennas may also be used for medium-gain antennas for fixed services, and Figure 6.13 (right) shows a modern example of a mass-market square planar array mounted on a motorized turntable (auto-pointing antenna for use on recreational vehicles), also from Kathrein (Kathrein Online, 2010).

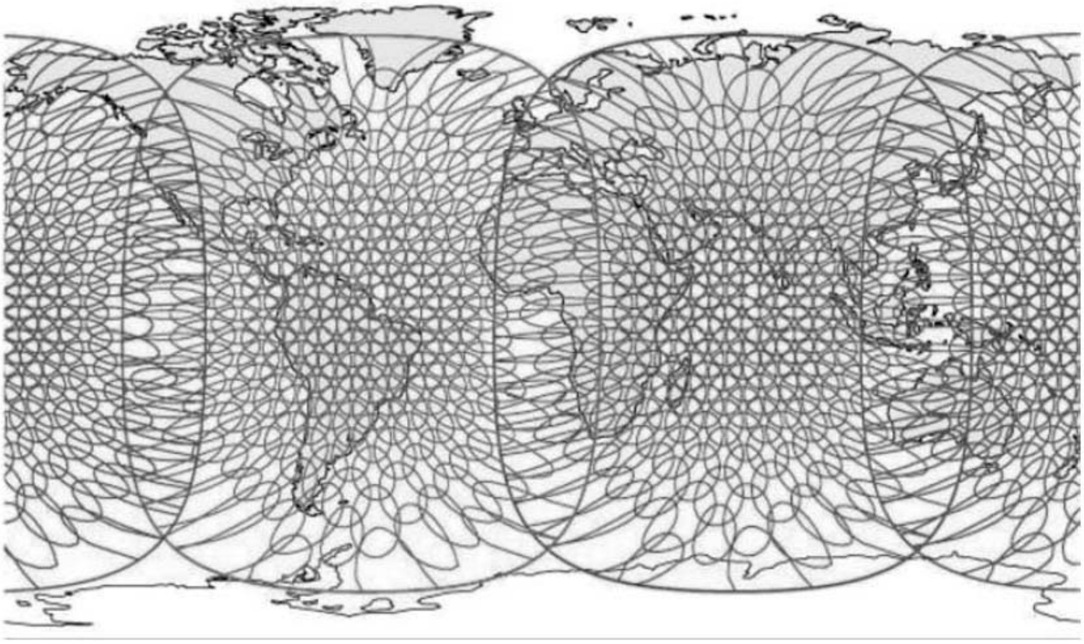


Figure 6.11 Example of cellular spot beam coverage produced by three Inmarsat fourth-generation satellites, located at 53° W, 64° E and 178° E.

Figure 6.12 An example 9 m cassegrain ground station antenna.

Figure 6.13 Example Ku-band satellite TV antennas: offset-fed reflector (left); motorized square planar antenna array (right).

Antennas for Portable Satellite Applications

A number of medium-gain directional antenna designs are available for portable applications where the user is required to erect the antenna and manually point it at the satellite. At UHF, foldable Yagi–Uda antennas are available that can be packed into a small carrying case, while medium-gain planar-phased arrays offer a compact form factor for medium-gain portable satellite applications.

Tracking Antennas for Mobile Satellite Applications

Antennas for applications where the user is mobile (for example, in a car, train or boat) generally fall into two categories. Less demanding applications may be satisfied using low-gain antennas with hemispherical radiation patterns, such as patches and quadrifilar helices. More demanding applications require a medium–high-gain directional antenna which needs continually to track the satellite as the platform moves.

The majority of tracking antennas employ an arrangement with the antenna (and often the RF electronics) mounted on a mechanical gimbal, which tracks (and compensates for) the motion of the platform. The antenna must track the satellite in at least two orthogonal axes: usually azimuth and elevation. In addition, those systems employing linear polarization generally require the antenna feed to track the polarization relative to the moving platform. The use of electronic beam steering using phased-array antennas instead of mechanical tracking is relatively rare owing to the generally higher cost and somewhat lower efficiency of electronic phased arrays – although a few designs employ electronically steered phased arrays for one axis (resulting in hybrid mechanical/electronic beam steering). Nevertheless, the use of mechanically-steered staring arrays (i.e. fixed boresight phased arrays) is increasingly common, as these can result in a reduced swept volume while tracking in comparison with normal reflectors.

The control mechanism for such tracking antennas may take one of two forms:

- Open loop. The direction of the satellite is predicted using only the motion of the platform.

Platform motion is typically obtained using electromechanical sensors. Typically, angular motion is obtained from gyroscopes, while heading is obtained from a flux gate compass (to compensate for gyroscope drift). A more sophisticated option is to detect platform motion electronically using an attitude sensor comprising an array of satellite navigation antennas.

- Closed loop. After the satellite has been found using open-loop control, the satellite signal is continuously monitored and the antenna is repointed so as to maximize the received signal. The tracking mechanism used will depend on the tracking speed required. A conical scan is typically employed, as it is relatively simple and cost effective to implement. In this scheme, the antenna continuously rotates around the nominal pointing direction, a fraction of a beamwidth away. Any signal variation is attributed to pointing error and is corrected, therefore conical scan can be confused by blockage/fading. The alternative is a monopulse arrangement, where the signal is received simultaneously in two or more directional beams with slightly different boresights, with the relative beam power used to monitor pointing.

Figure 6.14 Raysat T5 planar-phased array antenna.

High-gain tracking reflector antennas have been available for use on marine platforms for some time, and a splash-plate reflector antenna is often used on a small marine platform. The use of reduced-profile offset-fed high-gain antennas on Recreational Vehicles (RVs – motorized caravans) is a more recent development.

The height of tracking antennas on road vehicles is a significant issue. At Ku-band, a number of tracking antenna systems employ mechanically scanned planar arrays (so-called staring arrays). An example of such an antenna is that from Raysat, illustrated in Figure 6.14.

Although still mechanically scanned, these array antennas present a smaller swept volume than reflector antennas. Mechanically steered staring arrays are also used at L-band for mobile satelite services; typically, the array comprises a small number of axial-mode helices.

Figure 6.15 illustrates a design for an ultralow-profile tracking antenna for Ku-band TV reception, from RaySat, suited to mid-latitudes. This antenna employs a number of flat, planar, staring, phased arrays that are mechanically steered in both azimuth and elevation to point at the satellite. The aim is to limit the overall height of the antenna by using multiple subarrays, the outputs of which are combined electronically.

Ultimately, however, such low-profile antennas are limited by Lambert’s cosine law of illumination, with the result that many low-profile designs have lower limits on satellite elevation range Figure 6.14 because, in order to achieve a constant effective area (and hence constant antenna gain), as the satellite elevation decreases, the antenna will inevitably extend higher. For this reason, a common approach to low profile is to trade antenna height for width, thereby maintaining a constant area.

Figure 6.15 Raysat Speedway 1000 hybrid electronically/mechanically steered ultralow profile Ku-band antenna array for mobile TV reception.

Antennas for Handheld Satellite Applications

For handheld devices, satellite antennas must be compact and accommodate the unknown orientation of the satellite relative to the handset. The antenna pattern will therefore be omnidirectional in azimuth and preferably provide coverage of the whole upper hemisphere. Typical antenna types used for these applications are patch and quadrifilar helix antennas. In order to avoid the need to align the antenna polarization, circular polarization is almost universally employed for handheld services.

Clearly, the size of the antenna is a significant constraint in modern personal electronic devices. Increasingly high dielectric materials – such as ceramics – are used to reduce the antenna size by reducing the wavelength in the dielectric.

Figure 6.16 illustrates a range of GPS patch antenna from Cti International that use a ceramic dielectric insulator. The patch design to achieve circular polarization is a square with chamfered corners. For maximum sensitivity, GPS patch antennas are sometimes integrated with low-noise preamplifier to overcome the effects of cable loss.

Figure 6.17 shows a GeoHelix antenna designed for use in a variety of handheld satellite devices from Sarantel that employs a miniature quadrifilar helix wrapped around a high dielectric core (Sarantel Online, 2010). This type of antenna is employed in a number of handheld devices, including the latest Iridium mobile satellite handset.

Revision Questions

1. Estimate the beamwidth and maximum gain of a uniformly illuminated circular-aperture antenna of 1.2m diameter at a frequency of 12 GHz.

2. Which types of satellite antenna are typically used for the following coverage situations:

- earth cover beams;
- spot beams;
- contoured beams;
- multiple satellite spot beams?

3. Estimate the pointing loss for an 0.5m antenna at 12 GHz, for pointing errors of 0.5, 1 and 2◦.

What is the antenna beamwidth?

4. What types of antenna would you expect to find in a handset used for satellite applications at

L-band? Why is circular polarization used for most handheld terminal antennas?

5. Up to what frequency can the unmodified Rayleigh–Jeans formula be used at/near room temperature?

Estimate the effective noise temperature of an optical signal at λ = 1.06 μm.

6. What noise temperatures correspond to noise figures of 1, 2 and 5 dB?

7. Estimate the figure of merit for a low-gain antenna with hemispherical coverage, and explain your reasoning.

References

[1] Maini, A.K.- Agrawal, V.: Satellite Technology.Principles and Applications.Wiley 2011

[2] Richharia,M.-Westbrook,L.D.:Satellites Systems for Personal Applications,Willey 2010

[3] Balanis, C.A. (1997) Antenna Theory, Analysis and Design, 2nd edition, John Wiley & Sons, Inc., New York,NY.

[4] Barry, J.R. and Lee, E.A. (1990) Performance of coherent optical receivers. Proceedings of the IEEE, 78(8), 1369–1394.

[5] Boculic, R. (1991) Use basic concepts to determine antenna noise temperature. Microwaves and RF, March.

[6] Chu, L.J. (1948) Physical limitations in omnidirectional antennas. Journal of Applied Physics, 19, 1163–1175.

[7] Friis, H.T. (1944) Noise figures of radio receivers. Proceedings fo the IRE, July, 419–422.

[8] Friis, H.T. (1946) A note on a simple transmission formula. Proceedings fo the IRE, 34, 254.

[9] Harris online (2010) Commercial Deployable Antenna Reflectors. Available: http://download.harris.com/ app/public download.asp?fid=463 [accessed February 2010].

[10] Ho, C., Slobin, S., Kantak, A. and Asmar, S. (2008) Solar brightness temperature and corresponding antenna noise temperature at microwave frequencies. NASA InterPlanetary Network Progress

[11] Report 42-175, 15 November. Available: http://ipnpr.jpl.nasa.gov/ progress report/42-175/175E.pdf [accessed February 2010].

[12] IEEE (1993) IEEE Standard Definitions of Terms for Antennas, June.

[13] ITU-R P.618-8 (2003) Recommendation P.618-8, Propagation data and prediction methods required for the design of Earth–space telecommunication systems.

7 Satellite Multiple Access

Multiple Access (MA) refers to the general process used in communications systems in which system assets (circuits, channels, transponders, etc.) are allocated to users. The process, also called medium access control (MAC) for some wireless networks, is an important and some-times essential element in the communications system infrastructure, needed to ensure adequate capacity and link availability, particularly during times of heavy use of the communications system.

Satellite communications networks are particularly dependent on the inclusion of robust multiple access techniques, because satellite assets are usually limited in available power or available frequency spectrum and do not have the communications capacity to support all users at all times. Satellite links are designed to provide a desired link availability for average conditions, with some degradation expected during high demand times or during severe link outage periods. The goal of the MA process is to allow the communications network to respond to expected changes in user demand and adapt resources to provide the desired level of performance throughout high demand periods as well as average or limited demand conditions.

The primary assets available to the satellite communications systems designer to use in a multiple access process are satellite transponders and user ground terminals. Satellite MA techniques interconnect ground stations through multiple satellite transponders with the goal of optimizing several system attributes such as

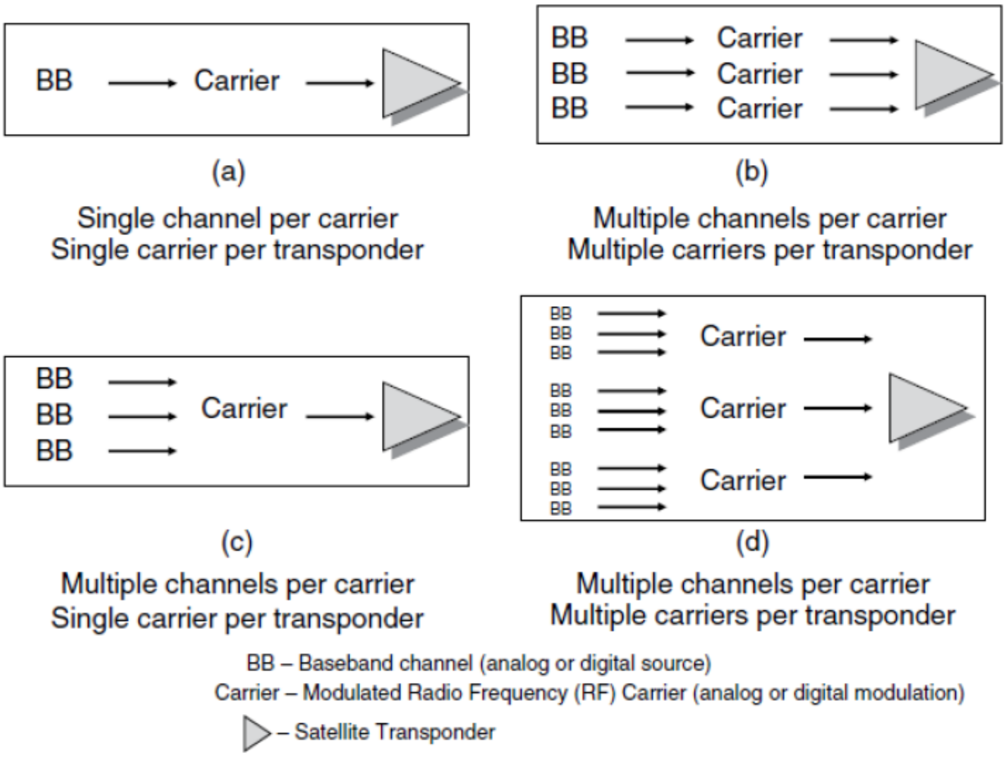
- spectral efficiency;
- power efficiency;
- reduced latency;
- increased throughput.

MA techniques are applicable to virtually all applications utilized by satellite systems, including both fixed and/or mobile users. Satellite systems often offer benefits over terrestrial transmission alternatives for implementation of efficient MA because the inherent ground/space link architecture allows network asset optimization without the need to add additional nodes or other components to the system.

The satellite transponder may be accessed in a number of different configurations, depending on the application and the satellite payload design. The frequency translation

(FT) transponder may be accessed by a single radio frequency (RF) carrier or by multiple carriers, with analog or digital modulation. Each carrier may be modulated by a single baseband (BB) channel or by multiple BB channels, from analog or digital sources.

Four basic multiple access configurations are identified in Figure 7.1. The simplest option,(a), consists of a single baseband channel modulating an RF carrier that feeds the satellite transponder. The baseband channel could be analog, such as analog voice or video, or a digital bit stream representing data, voice, or video. The modulation could be analog, such as amplitude modulation (AM) or frequency modulation (FM), or digital, such as frequency shift keying (FSK), or various forms of phase shift keying, such as BPSK or QPSK.



(a)
Single channel per carrier
Single carrier per transponder

(b)
Multiple channels per carrier
Multiple carriers per transponder

(c)
Multiple channels per carrier
Single carrier per transponder

(d)
Multiple channels per carrier
Multiple carriers per transponder

Figure 7.1 Access options in a satellite communications network

In the second option, (b), multiple single baseband/modulation chains are combined before feeding the transponder. In this case, the final amplifier in the transponder is usually operated in a power backoff mode to avoid intermodulation noise. Case (c) consists of a single modulated carrier; however, multiple baseband channels are multiplexed onto a single data stream before carrier modulation. Typical multiplexing formats include Frequency Division Multiplexing (FDM) for analog sources and Time Division Multiplexing (TDM) for digital sources. The most complex case, (d), consists of multiple multiplexed baseband channels modulating multiple RF carriers, with the multiple carriers all introduced to the single transponder. This option also requires power backoff to avoid intermodulation noise.

Cases (a) and (c) with a single carrier present in the transponder, are usually referred to as single channel per carrier (SCPC) operation. SCPC transponders can usually operate with input levels set to drive the final power amplifier to full saturation, providing high power efficiency. Cases (b) and (d) are multiple channel per carrier (MCPC) systems, which operate with input levels set well below the saturation level to avoid intermodulation that can cause crosstalk noise with analog data, or increased bit errors with digital data streams. This power backoff can be several dB, resulting in lower power efficiency for MCPC versus SCPC systems.

The MA methods available to the satellite system designer can be categorized into three fundamental techniques, differentiated primarily by the domain used in the process:

- Frequency Division Multiple Access (FDMA)
- Time Division Multiple Access (TDMA);
- Code Division Multiple Access (CDMA).

FDMA systems consist of multiple carriers that are separated by frequency in the transponder. The transmissions can be analog or digital, or combinations of both. In TDMA the multiple carriers are separated by TIME in the transponder, presenting only one carrier at any time to the transponder. TDMA is most practical for digital data only, because the transmissions are in a burst mode to provide the time division capability. CDMA is a combination of both frequency and time separation. It is the most complex technique, requiring several levels of synchronization at both the transmission and reception levels. CDMA is implemented for digital data only, and offers the highest power and spectral efficiency operation of the three fundamental techniques.

Multiple access options are further defined by secondary access techniques, which are usu-ally implemented within one or more of the three fundamental access technologies introduced above. These secondary techniques include:

- Demand Assigned Multiple Access (DAMA). Demand assigned networks change signal configuration dynamically to respond to changes in user demand. FDMA or

TDM Anetworks can be operated with pre-assigned channels, called fixed access (FA) or pre-assigned access (PA); or they can be operated as an assigned-on-demand DAMA network. CDMA is a random access system by its implementation, so it is a DAMA network by design.

- Space Division Multiple Access (SDMA). Space division multiple access refers to the capability to assign users to spatially separated physical links (different antenna beams, cells, sectored antennas, signal polarization, etc.), in addition to the MA inherent in the access method of implementation. It can be employed with any of the three basic MA techniques, and is an essential element of mobile satellite networks, which employ multibeam satellites, and may include frequency reuse and orthogonal polarized links to further increase network capacity.
- Satellite Switched TDMA (SS/TDMA). Satellite switched TDMA employs sequenced beam switching to add an additional level of multiple access in a frequency translation satellite. The switching is accomplished at RF or at an intermediate frequency (IF) and is unique to satellite based systems.
- Multi-frequency TDMA (MF-TDMA). This technique combines both FDMA and TDMA to improve capacity and performance for broadband satellite communications networks. The broadband baseband signal is divided up in frequency band and each segment drives a separate FDMA carrier. The received carriers are then recombined to produce the original broadband data.

The following sections discuss the three fundamental MA technologies further, including descriptions of the secondary access techniques as well. All of the fundamental MA technologies can be applied to terrestrial and space applications, but the focus here is on the specific implementations of MA for satellite-based systems and networks.

7.1 Frequency Division Multiple Access

Frequency division multiple access (FDMA) was the first MAtechnique to be implemented on satellite systems and is the simplest in principal and operation. Figure 7.2 shows a functional display of the FDMAprocess, featuring an example for three ground stations accessing a single frequency translation (FT) satellite transponder. Each station is assigned a specific frequency band for its uplink, f1, f2, and f3, respectively. The frequency/time plot of the figure shows that each ground station has exclusive use of its frequency band, or slot. The frequency slot is either pre-assigned or can be changed on demand. Frequency guard bands are used to avoid interference between the user slots.

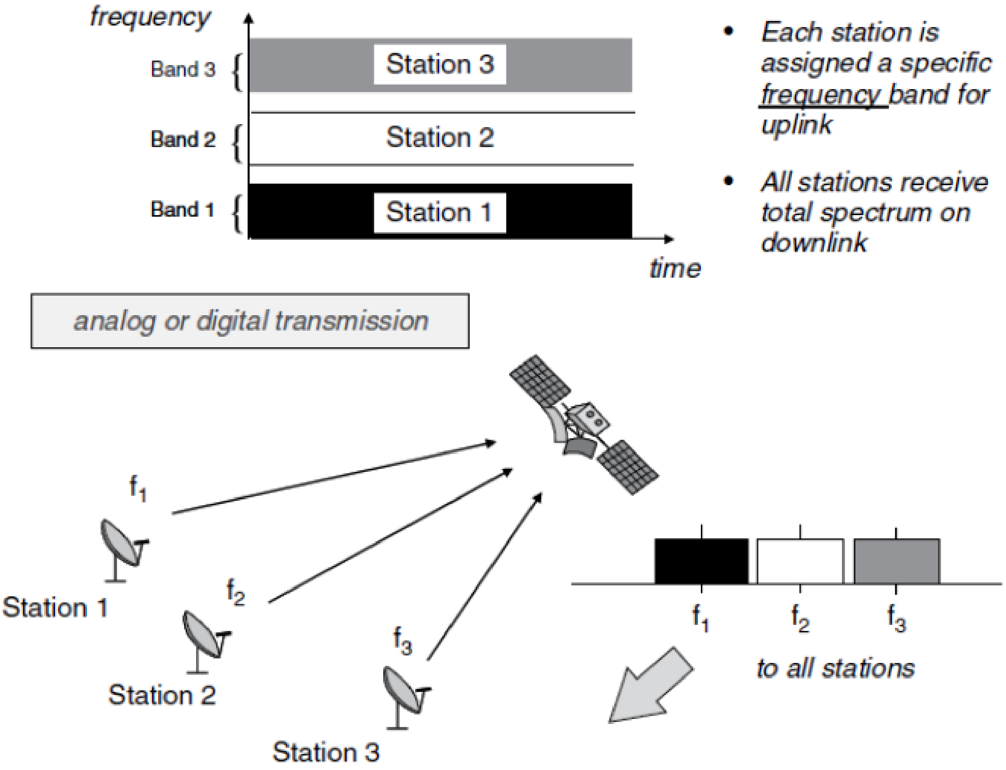


Figure 7.2 Frequency Division Multiple Access (FDMA)

The presence of the multiple carriers in the transponder final power amplifier requires power backoff operation to avoid intermodulation noise. The multiple carrier spectrum passes through the FT satellite and the full FDMA spectrum is transmitted on the downlink. The receiving sta-tion must be able to receive the full spectrum and can select the desired carrier for demodulation or detection.

FDMAtransmissions can be analog or digital, or combinations of both. FDMAis most useful for applications where a full time channel is desired – for example, video distribution. It is the least expensive to implement but has the potential to make inefficient use of the spectrum, because there can be 'dead times'on one or more channels when transmissions are not present.

Multiple access system performance must be analyzed by considered the specific pro-cessing elements used in the satellite communications information bearing signals. We select here two examples of typical transmission elements used in satellite FDMA applications for further analysis. The first example is PCM/TDM/PDK/FDMA.

7.1.1 PCM/TDM/PSK/FDMA

One of the most common MCPC FDMA systems employed in satellite communications is a PCM/TDM (pulse code modulation/time division multiplexed) application used for voice communications. The source data is PCM digitized voice, which is source combined using the DS-1 level TDM. The DS- 1 level consists of 24 64-kbps channels, multiplexed to a 1.544 Mbps TDM bit rate. The carrier modulation is phase shift keying, either BPSK or QPSK. The resulting MCPC signal structure, implemented with FDMA access, designated as PCM/TDM/PSK/FDMA, is the first MA system we consider for performance evaluation.

The capacity of the MA system is the most important parameter for evaluation. It determines the maximum number of users that can access the satellite and serves as the basis for decisions on demand access (DA) options on the link.

7.1.2 PCM/SCPC/PSK/FDMA

The second FDMA system to be considered is PCM/SCPC/PSK/FDMA, a popular digital baseband single carrier per channel (SCPC) system used for data and voice applications. No signal multiplexing is involved. Each incoming signal is A to D converted, and placed on a BPSK or QPSK modulated RF carrier for transmission over the satellite channel. A pair of channel frequencies is used for voice communications, one for each direction of transmission. One advantage of this SCPC FDMAapproach is that it can operate as a demand assignment access, where the carrier is turned off when not in use. The system can also use voice activation, which makes use of the statistics of voice conversations to share the SCPC carrier with multiple users.

A typical voice channel conversation is active only about 40% of the time in any one direction. A voice activation factor (VA) is used to quantify the improvement possible in the network. For example, a 36MHz transponder has a bandwidth limited capacity of 800 SCPC channels, using 45 KHz channel spacing, i.e.,

36MHz/45 KHz= 800

The 800 channels correspond to 400 simultaneous conversations. A typical voice activity specification based on modeling the channel as a sequence of Bernoulli trials with application of the binomial distribution would be: the probability that more than 175 of the 800 available channels will contain active speech in either direction is less the probability that more than 175 of the 800 available channels will contain active speech in either direction is less than 0.01 (< 1%). The voice activity factor is then

$$\mathrm{VA} = 10\log\left(\frac{800}{175 \times 2}\right) = 10\log\left(\frac{800}{350}\right) = 10\log(2.3) = 3.6\,\mathrm{dB} \qquad (7.1)$$

7.2 Time Division Multiple Access

The second multiple access technique to be used in satellite communications is time division multiple access (TDMA). With TDMA the multiple carriers are separated by TIME in the transponder, rather than by FREQUENCY as with FDMA, presenting only one carrier at any time to the transponder. This important factor allows the final amplifier in the satelite transponder to operate with a saturated power output, providing the most efficient use of the available power. Figure 7.3 shows a functional display of the TDMA process, fea-turing an example for three ground stations accessing a single frequency translation (FT) satelite transponder. Each station is assigned a specific time slot, t1, t2, and t3, respectively, for its uplink transmission of a burst (or packet) of data. The frequency/time plot of the figure shows that each ground station has exclusive use of the full transponder bandwidth during its time slot. The time slot is pre-assigned or can be changed on demand. Guard times are used between the time slots to avoid interference. TDMA is most practical for digital data only, because of the burst nature of the transmissions. Downlink transmission consists of interleaved set of packets from all the ground stations. A REFERENCE STATION, which could be one of the traffic stations or a separate ground location, is used to establish the synchronization reference clock and provide burst time operational data to the network.

TDMA offers a much more adaptive MA structure than FDMA regarding ease of reconfiguration for changing traffic demands. Figure 7.4 shows the signal structure for a typical TDMA network, consisting of N traffic stations. The total time period that includes all traffic station bursts and network information is called the TDMA frame. The frame repeats in time sequence and represents one complete transmission in the network. Typical frame times range from 1 to 20 ms. Each station burst contains a preamble and traffic data. The preamble contains synchronization and station identification data. The reference burst, from the reference station, is usually at the start of each frame, and provides the network synchronization and operational information.

Guard bands are included to prevent overlap and to account for different transmission times for each of the stations, based on their range to the satellite.

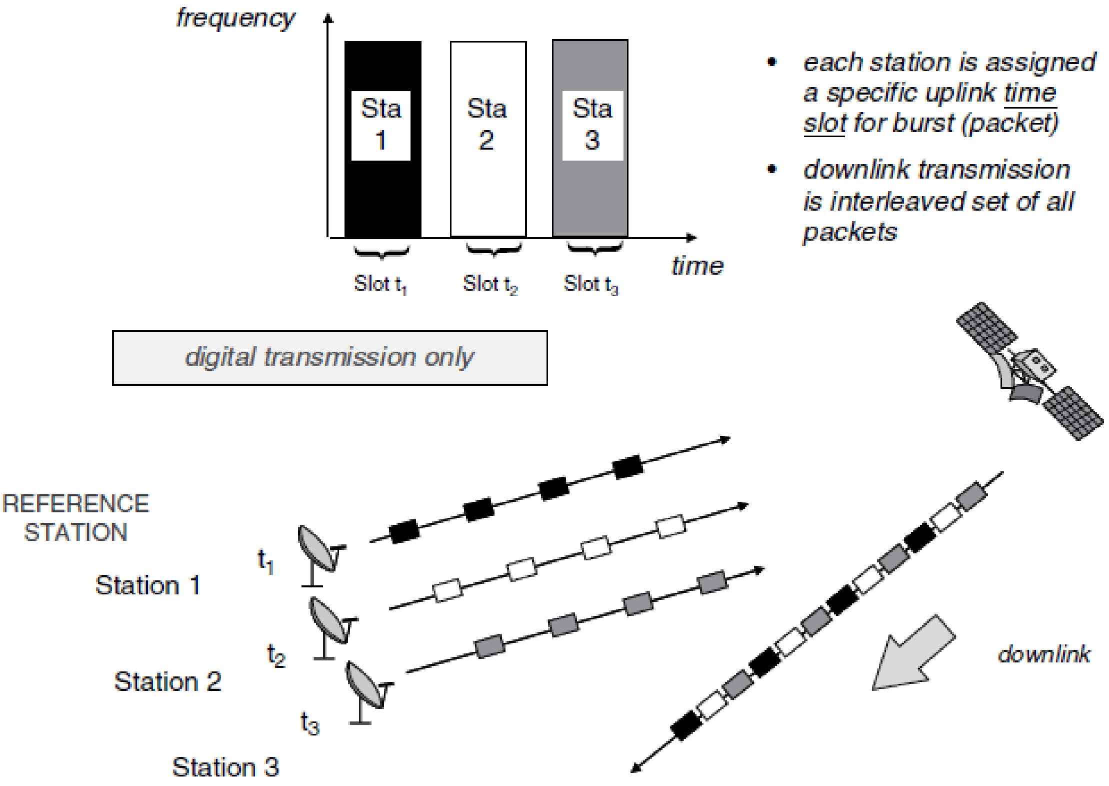


Figure 7.3 Time Division Multiple Access (TDMA)

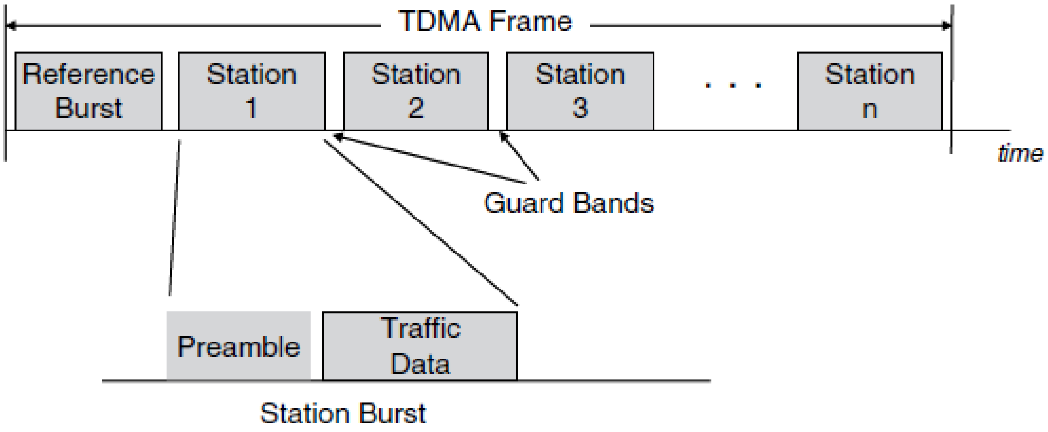


Figure 7.4 TDMA frame structure

Station bursts do not need to be identical in duration, and can be longer for heavier traffic stations or during higher use periods. The specific allocation of burst times for each of the stations within the frame is called the burst time plan. The burst time plan is dynamic and can be changed as often as each frame to adapt for changing traffic patterns.

7.2.1 PCM/TDM/PSK/TDMA

The most common TDMA network structure, popular in VSAT networks, consists of PCMbased baseband formatting, TDM source combining, and QPSK or BPSK modulation, designated as PCM/TDM/PSK/TDMA (see AppendixAfor a description of common communications signal processing elements). PCM/TDM/PSK/TDMA employs a single modulated carrier occupying the full transponder bandwidth. Typical TDMAdata rates for full transponder operation are

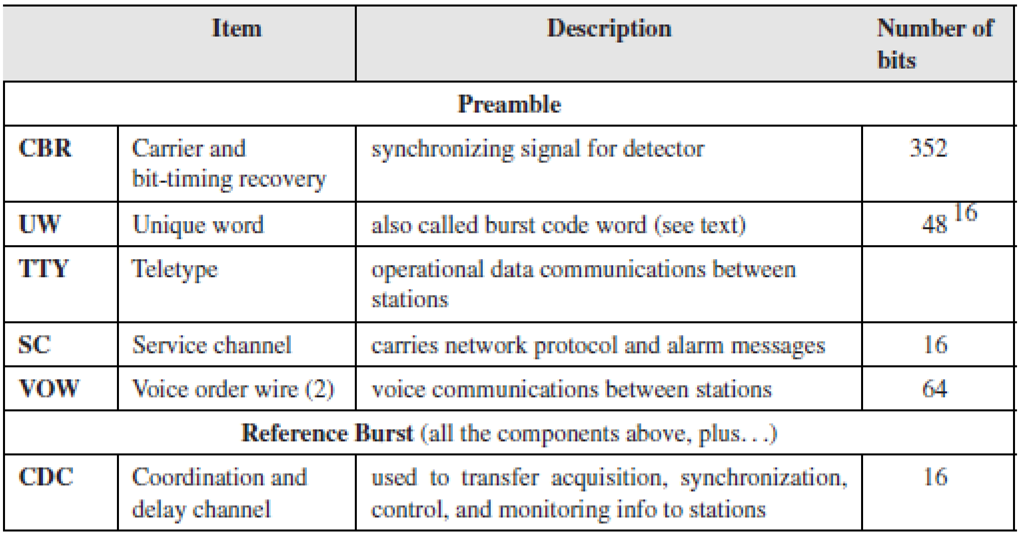
- 60 Mbps (36MHz transponder)
- 130 Mbps (72MHz transponder)

The TDMA frame period is chosen to be a multiple of 125µs, the standard PCM sampling period. Baseband formatting can be classical PCM or ADPCM. Most networks employ two reference bursts, from two reference stations, for redundancy, because a loss in carrier and bit timing recovery information results in complete network breakdown.

An important element of frame synchronization is to maintain a unique word (also called a burst code word) in the reference burst and in each station burst preamble. The unique word is typically a sequence of 24 to 48 bits, chosen for high probability of correct detection. It is the only repeating bit sequence in the frame. It is essential to maintain network synchronization and to accomplish carrier recovery in the PSK demodulation process.

The preamble consists of a number of elements, each with a specific purpose in the TDMA process. Typical components of the TDMA preamble and reference burst are summarized in Table 7.1 for a representative operational system, the INTELSAT TDMA system, deployed on many early INTELSAT satellites [1]. The INTELSAT TDMA system operates with a 2 ms frame period, and consists of two reference bursts per frame. The last column gives the number of bits allocated to each component in the INTELSAT structure. With QPSK modulation the transmission consists of 2 bits per symbol.

Table 7.1 INTELSAT TDMA preamble and reference burst structure *(source: Roddy [1]; reproduced by permission of Dennis Roddy, Satellite Communications, Third Edition, © 2001 The McGraw-Hill Companies)*



Item		Description	Number of bits
		Preamble	
CBR	Carrier and bit-timing recovery	synchronizing signal for detector	352
UW	Unique word	also called burst code word (see text)	48[16]
TTY	Teletype	operational data communications between stations	
SC	Service channel	carries network protocol and alarm messages	16
VOW	Voice order wire (2)	voice communications between stations	64
		Reference Burst (all the components above, plus. . .)	
CDC	Coordination and delay channel	used to transfer acquisition, synchronization, control, and monitoring info to stations	16

7.2.2 TDMA Frame Efficiency

The performance of a TDMA system can be evaluated by consideration of the TDMA frame efficiency, η_F, defined as

$$\begin{aligned}\eta_F &= \frac{\text{Number of bits available for traffic}}{\text{Total number of bits in frame}} \\ &= 1 - \frac{\text{Number of overhead bits}}{\text{Total number of bits in frame}}\end{aligned} \tag{7.2}$$

or, in terms of the TDMA frame elements,

$$\eta_F = 1 - \frac{n_r\, b_r + n_t\, b_p + (n_r + n_t) b_g}{r_T t_F} \tag{7.3}$$

where

t_F = TDMA frame time, in s

r_T = total TDMA bit rate, in bps

n_r = number of reference stations

n_t = number of traffic bursts

b_r = number of bits in reference burst

b_p = number of bits in traffic burst preamble

b_g = number of bits in guard band

Note that the frame efficiency is improved (increased) by: 1) a longer frame time, which increases the total number of bits; or 2) lowering the overhead (nontraffic bits) in the frame. The optimum operating structure occurs by providing the longest possible frame time with the lowest total number bits allocated to overhead functions.

7.2.3 TDMA Capacity

The network channel capacity for a TDMA network is most often evaluated in terms of an equivalent voice-channel capacity, nC. This allows evaluation of capacity for any type of data source bit stream: voice, video, data, or any combination of the three. The equivalent voice channel capacity is defined as

$$n_C = \frac{\text{Available information bit rate, } r_i}{\text{Equivalent voice channel bit rate, } r_C} = \frac{r_i}{r_C} \qquad (7.4)$$

$$r_C = 64 \text{ kbps} \qquad (7.5)$$

The channel capacity for a TDMA network is determined by the following steps.

Step 1

Determine the composite carrier-to-noise ratio available on the RF link, designated by an upper case T, as $(C/N)_T$. This value is developed from RF link evaluation as discussed in Chapter 3.

Step 2

Calculate the carrier-to-noise ratio required to achieve the threshold BER desired for the TDMA network, designated by a lower case t, as $(C/N)_t$, from

$$\left(\frac{C}{N}\right)_t = \left(\frac{E_b}{N_o}\right)_t - B_N + R_T + M_i + M_A \qquad (7.6)$$

where

$\left(\frac{E_b}{N_o}\right)_t$ = the $\left(\frac{E_b}{N_o}\right)$ required for the threshold BER
R_T = TDMA data rate at the desired frame efficiency, η_F, in dB
B_N = noise bandwidth of carrier, in dB
M_i = MODEM implementation margin, in dB (~1 to 3 dB)
M_A = Adjacent Channel Interference margin, (if any) (~1 to 2 dB)

Implementation margins are included to account for the deviation of modem performance from the ideal.

Step 3

The TDMA data rate RT is adjusted until

$$(C/N)_t > (C/N)_t \tag{7.7}$$

Step 4

The equivalent voice channel capacity, n_C, (see Equation 7.4), is now calculated. With the frame parameters for the reference bursts, traffic bursts, and guard bands defined as

t_F = TDMA frame time, in s

r_T = total TDMA bit rate, in bps

n_r = number of reference stations

n_t = number of traffic bursts

b_T = number of bits in total TDMA frame

b_r = number of bits in reference burst

b_p = number of bits in traffic burst preamble

b_g = number of bits in guard band

we can define the following bit rates, all in bps:

$$\text{Total TDMA Bit Rate}: \ r_T = \frac{b_T}{t_F} \tag{7.8}$$

$$\text{Reference Burst Bit Rate}: \ r_p = \frac{b_p}{t_F} \tag{7.9}$$

$$\text{Reference Burst Bit Rate}: \ r_r = \frac{b_r}{t_F} \tag{7.10}$$

$$\text{Guard Time Bit Rate}: \ r_g = \frac{b_g}{t_F} \tag{7.11}$$

The available bit rate for traffic (traffic bit rate) is then

$$r_i = r_T - n_r(r_r + r_g) - n_t(r_p + r_g) \tag{7.12}$$

The equivalent voice channel capacity is therefore

$$n_C = \frac{r_i}{r_C}$$

or

$$n_C = \frac{r_T}{r_C} - \frac{n_r(r_r + r_g)}{r_C} - \frac{n_t(r_p + r_g)}{r_C} \tag{7.13}$$

This result provides the number of equivalent voice channels that can be supported by the TDMA network for the specified TDMA bit rate, TDMA frame efficiency, and frame parameters.

Sample Calculation for Channel Capacity

The sample calculation for frame efficiency in the previous section found that a TDMA network operating with 12 traffic terminals could maintain a 95% frame efficiency with a TDMA data rate of 96.64 Mb/s. We wish to determine the channel capacity, expressed in terms of the equivalent voice-channel capacity, n_c, for this system. Assume the PCM data rate of r_c = 64 kbps for the evaluation.The element bit rates for the system, from Equations (7.16) through (7.19), are

$$\text{Total TDMA Bit Rate: } r_T = 96.64\ Mbps$$

$$\text{Preamble Bit Rate : } r_p = \frac{b_p}{t_F} = \frac{560}{0{,}002} = 280\ kbps$$

$$\text{Reference Burst Bit Rate : } r_r = \frac{b_r}{t_F} = \frac{576}{0.002} = 288\ kbps$$

$$\text{Guard Time Bit Rate :} r_g = \frac{b_g}{t_F} = \frac{128}{0.002} = 64\ kbps$$

The available bit rate for traffic is, from Equation (7.20),

$$r_i = r_T - n_r(r_r + r_g) - n_t(r_p + r_g) = 96{,}64 \times 10^3 - 12(280 + 64) \times 10^3$$
$$= 91{,}8\ Mbps$$

The equivalent voice-channel capacity is therefore

$$n_C = \frac{r_i}{r_c} = \frac{91{,}8 \times 10^6}{64 \times 10^3} = 1434{,}5 \text{ or } 1434 \text{ channels}$$

The result, rounded off to the next lowest integer, 1434, is the number of equivalent voice channels that can be supported to maintain a frame efficiency of 95% for the TDMA network.

7.2.4 Satellite Switched TDMA

The use of TDMA on a frequency translation satellite provides a high degree of robustness for efficient multiple access applications. TDMA also offers the possibility of extending the design options by adaption of additional capabilities in the network; a technique called satelite switchedTDMA, or SS/TDMA. SS/TDMA consists of a rapid reconfiguration of antenna beams on-board the satellite to provide an additional level of access capabilities over basic TDMA.

SS/TDMA adds antenna beam switching to provide additional MA capability to adapt to changing demand requirements. SS/TDMA utilized on a frequency translation satellite is not classified as an on-board processing technique, however, since there is no demodulation/ remodulation to baseband, as is done with on-board processing satellites. The on-board switching is accomplished at IF with an n × n switch matrix. Switching is done in synchronization with the TDMA bursts from the ground stations. Figure 7.5 shows the configuration o fa 3 × 3 satellite switched TDMA architecture.

The network shown on the figure consists of three regional beams, designated as West, Central, and East beams. The ground stations in the West beam are designated by upper case letters A, B, C; ground stations in the Central beam by numbers 1, 2, 3; and ground stations in the East beam by Roman numerals I, II, III. The switch matrix mode is shown on the right of the figure, labeled Mode 1, Mode 2, or Mode 3. The dashed lines across the three beams show the reconfiguration times between the switch modes. The blocks indicate the data packets being transmitted from each station, with the desired receive station designated by the number or letter on each.

Consider first Switch Mode 1, the lowest of the three in Figure 7.5. During the Switch Mode 1 time period, the West beam stations, A, B, and C, transmit data bursts addressed to stations in the Central beam only, i.e., 1, 2, or 3, as indicated by the labels on the individual bursts. Similarly, the Central beam stations, 1, 2, and 3, transmit bursts addressed only to the West beam stations, A, B, and C; and the East beam stations, I, II, and III, transmit bursts addressed only back to stations in the East beam, I, II, and III.

During the Switch Mode 2 time period, theWest beam stations transmit bursts addressed only to East beam stations; the Central beam stations transmit only back to Central beam stations; and the East beam stations transmit only toWest beam stations. In the Switch Mode 3 time period, the West beam stations transmit bursts addressed only back toWest beam stations; the Central beam stations transmit only to East beam stations; and the East beam stations transmit only to Central beam stations.

Figure 7.6(a), (b), and (c), shows the 3 × 3 satellite switch matrix settings for each of the three Switch Modes, 1, 2, and 3, shown on Figure 7.5, respectively. The switch matrix is maintained at the fixed switch locations throughout the switch mode period, and then

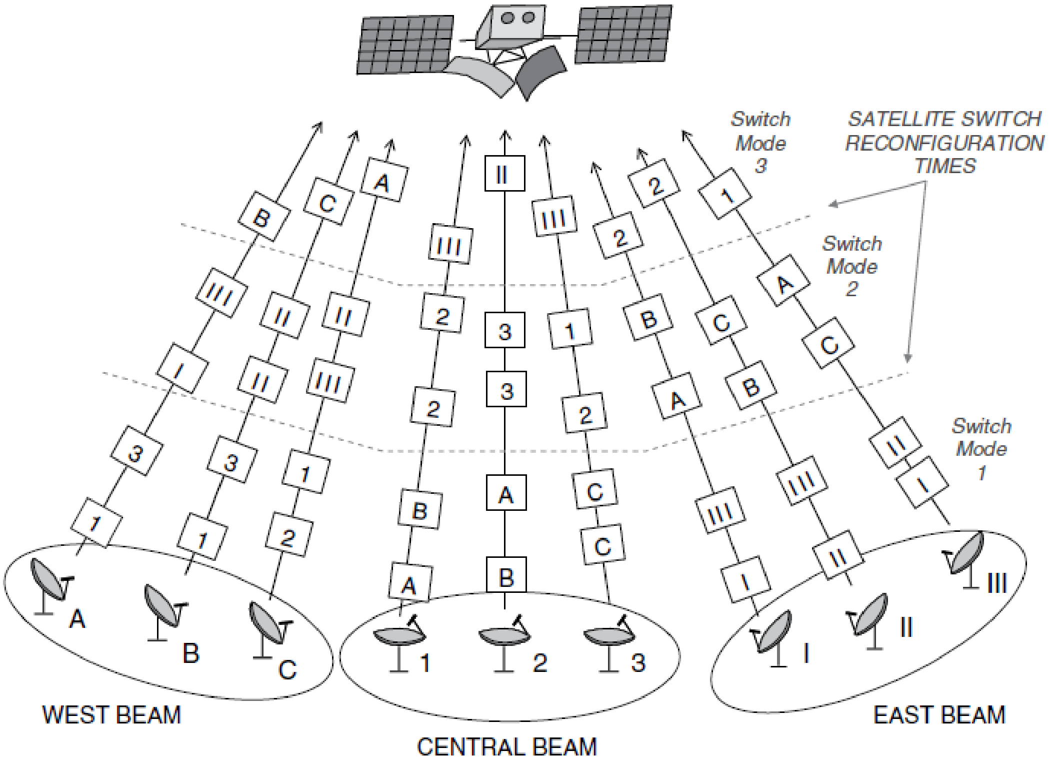


Figure 7.5 3 × 3 SS/TDMA network configuration

maintained at the fixed switch locations throughout the switch mode period, and then rapidly changed to the next configuration at the reconfiguration time, as shown by the dashed lines on Figure 7.5.

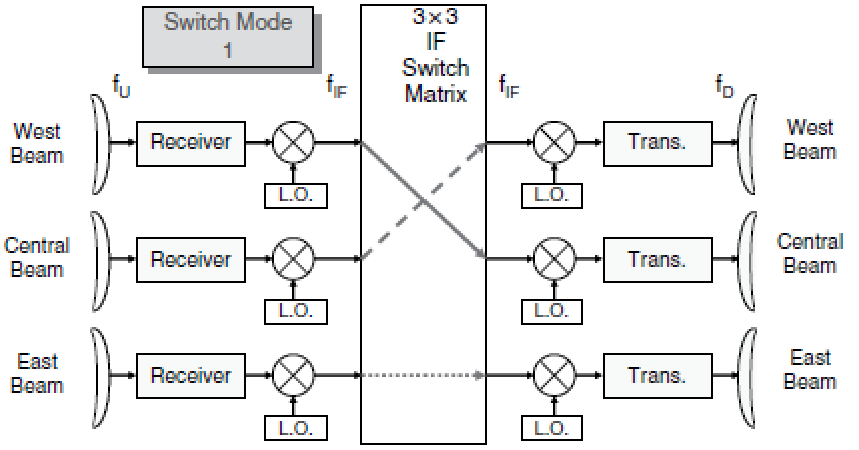
It should be pointed out that Figures 7.5 and 7.6 show only three of six possible switch modes for the 3 × 3 switch matrix. The remaining three switch modes are listed on Table 7.2, which shows all six of the possible switch locations possible with a 3 × 3 switch matrix. Data throughput is improved on a SS/TDMA implementation because receive stations only receive packets addressed to them, reducing processing time and station complexity. The number of switch positions, ns, required for full inter-connectivity of N beams, that is for an N × N switch matrix, is

$$n_S = N! \qquad (7.14$$

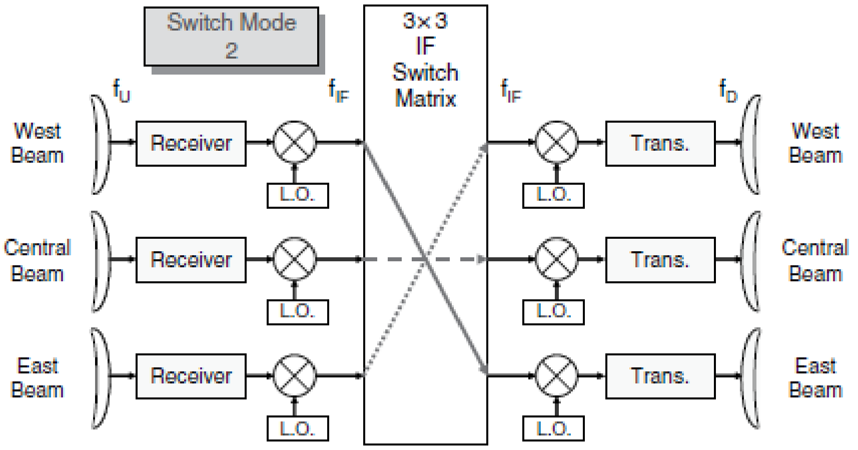
Full interconnectivity includes return back to the same beam. As N increases, the number of switch positions becomes quite large, i.e.,

N =3	n_s = 6 (3 × 3 network discussed above)
N =4	n_s = 24
N =5	n_s = 120
N =6	n_s = 720

Typical implementations consist of crossbar switch matrixes implemented in coax or waveguide. The switching elements are ferrite, diode, or FET (field effect transistor), with the dual-gate FET offering the best performance. In practice, only 3- or 4-beam SS/TDMA operation is practical because of the large size and weight required for larger IF switch matrices.



(a)



(b)

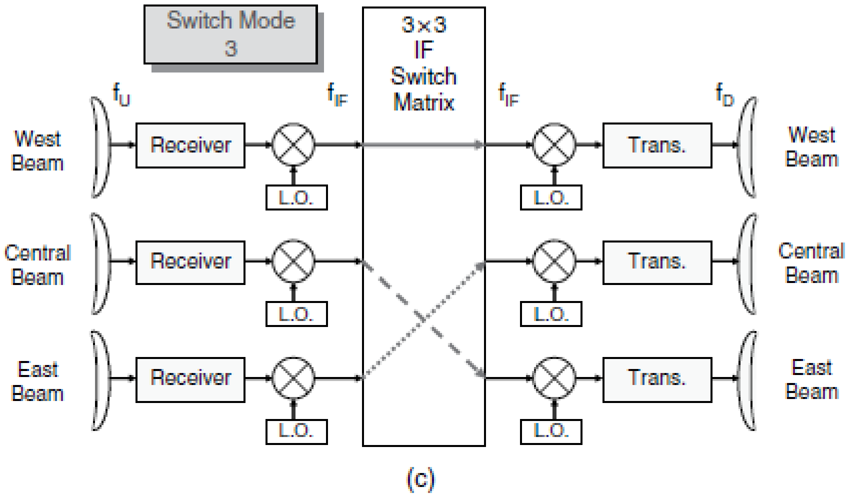


Figure 7.6 Switch matrix settings for 3 × 3 SS/TDMA network

Table 7.2 3 × 3 matrix switch modes

Uplink beam	**Downlink beam**					
	Switch position 1	**Switch position 2**	**Switch position 3**	**Switch position 4**	**Switch position 5**	**Switch position 6**
West	Central	East	West	West	Central	East
Central	West	Central	East	Central	East	West
East	East	West	Central	East	West	Central

SS/TDMA can also provide fixed switch positions that increase options and allow the network to support varied users. Two options are available: (a) if a switch mode is selected and remains fixed, the satellite operates as a traditional frequency reuse ‘bent pipe’ repeater; (b)the switch matrix could be set one-to-all, that is, in the 3×3 case above, one uplink beam, say the Central beam, could be fixed to all three downlink beams. This would provide a broadcast mode of operation, where the information from one ground station could be sent to all locations served by the satellite.

7.3 Code Division Multiple Access

The third fundamental multiple access (MA) technique, code division multiple access (CDMA) is a combination of both frequency and time separation. It is the most complex technique to implement, requiring several levels of synchronization at both the transmission and reception levels. CDMA is practical for digital formatted data only, and offers the highest power and spectral efficiency operation of the three fundamental techniques.

Figure 7.7 shows a functional display of the CDMAprocess, similar in presentation to those discussed previously for FDMA (Figure 7.2) and TDMA (Figure 7.3). Each uplink station is assigned a time slot and a frequency band in a coded sequence to transmit its station packets. The downlink transmission is an interleaved set of all the packets as shown in Figure 7.7.The downlink receive station must know the code of frequency and time locations in order to detect the complete data sequence. The receive station with knowledge of the code can recoup the signal from the noise-like signal that appears to a receiver that does not know the code.

Code Division Multiple Access is often referred to as Spread Spectrum or Spread Spectrum Multiple Access (SSMA) because of the signal spreading characteristics of the process CDMA offers several advantages over FDMA or TDMA due to its architecture.

- Privacy. The code is distributed only to authorized users, protecting the information from others.
- Spectrum Efficiency. Several CDMAnetworks can share the same frequency band, because the undetected signal behaves as Gaussian noise to all receivers without knowledge of the code sequence. This is particularity useful in applications such as NGSO Mobile Satellite Service systems, where bandwidth allocations are limited.
- Fading Channel Performance. Only a small portion of the signal energy is present in a given frequency band segment at any one time, therefore frequency selective fading or dispersion will have a limited effect on overall link performance.
- Jam Resistance. Again, because only a small portion of the signal energy is present in a given frequency band segment at any one time, the signal is more resistant to intentional or unintentional signals present in the frequency band, thereby reducing the effects on link performance.

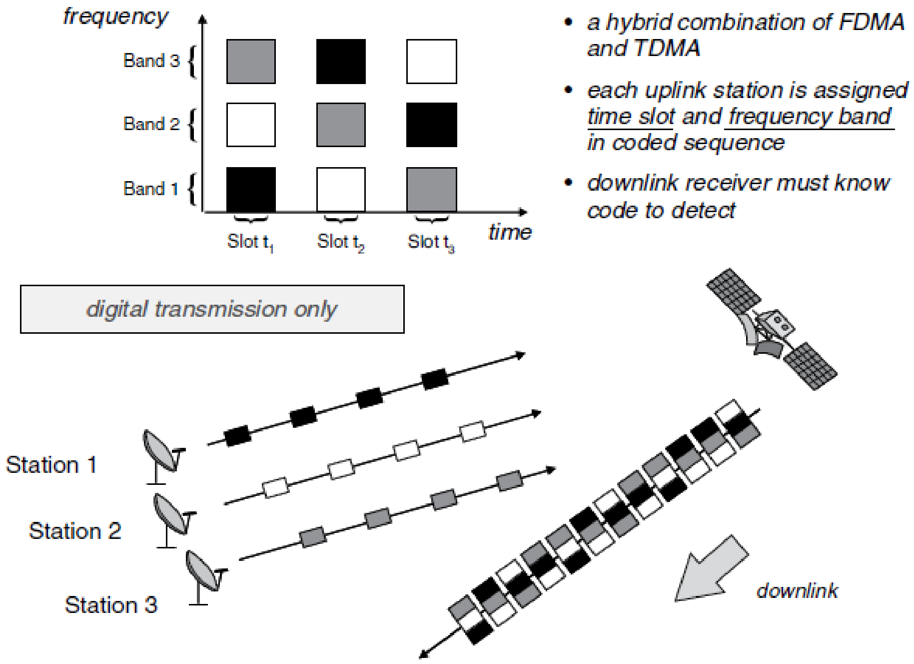


Figure 7.7 Code Division Multiple Access (CDMA)

Selection of the appropriate code sequence to use in the CDMA process is critical to its M successful implementation. The code sequence must be configured to avoid unauthorized decoding, yet short enough to allow efficient data transmissions without introducing latency or synchronization problems. The most successful type of code sequence for CDMA,which meets both of the above criteria, is the pseudorandom (PN) sequence. Pseudorandom means like random,' that is, appearing random but having certain non-random or deterministic features. The PN sequence used in CDMA systems is a finite length binary sequence, in which bits are randomly arranged. The autocorrelation of the PN sequence resembles the autocorrelation of band-limited white noise. The PN sequence used in CDMA systems is generated using sequential logic circuits and a feedback shift register. Figure 7.8 shows an example of an n-stage feedback shift register used to generate the PN sequence. The binary sequences are shifted through the shift registers at the clock rate. The feedback logic consists of exclusive-OR gates generated by a unique algorithm or kernel. The output of the stages are logically combined and fed back as input, generating a PN sequence at the final output. The number of non-zero states that are possible for this linear PN sequence generator, called its maximal length (ML) will be

$$ML = 2^n - 1 \qquad (7.15)$$

Once the PN sequence is generated it is combined with the binary data sequence to produce the PN data stream sequence used in the CDMA process. Figure 7.9 shows the process used to generate the PN data stream. The PN clock, called the chip clock, generates the PN sequence shown in the center of the figure. The PN sequence, pPN(t), has a chip rate of rch, and a chip period of tch, as shown in the figure. The PN sequence is modulo-2 added to the data sequence m(t) to produce the data stream, e(t), i.e.,

$$e(t) = m(t) \bullet p_{PN}(t) \qquad (7.16)$$

where the operator indicates modulo-2 addition.

The PN data stream is at the chip rate, rch, which is significantly higher than the original data rate, rb, that is, rch >> rb. In the example shown on the figure

$$r_{ch} = 5\ r_b$$

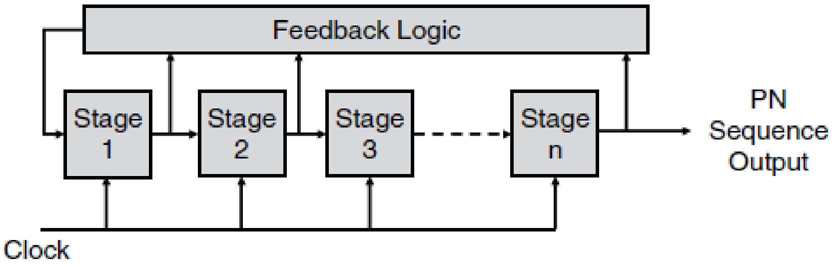


Figure 7.8 n-stage feedback shift register PN sequence generator

This condition that rch >> rb is essential for the successful implementation of CDMA, and is the reason that CDMA is often referred to as spread spectrum or spread spectrum multiple access, because the original data sequence is ‘spread’ out over a much greater frequency band in the transmission channel.

The chip rate is chosen to spread the signal over the total available channel bandwidth. Large spreading ratios are typical – for example, in mobile satellite voice networks, the original 16 kbps voice data stream may be spread at a chip rate to produce a PN data stream that operates over an 8MHz RF channel bandwidth. This is a spreading factor of 500, assuming 1 bit/Hz modulation such as BPSK.

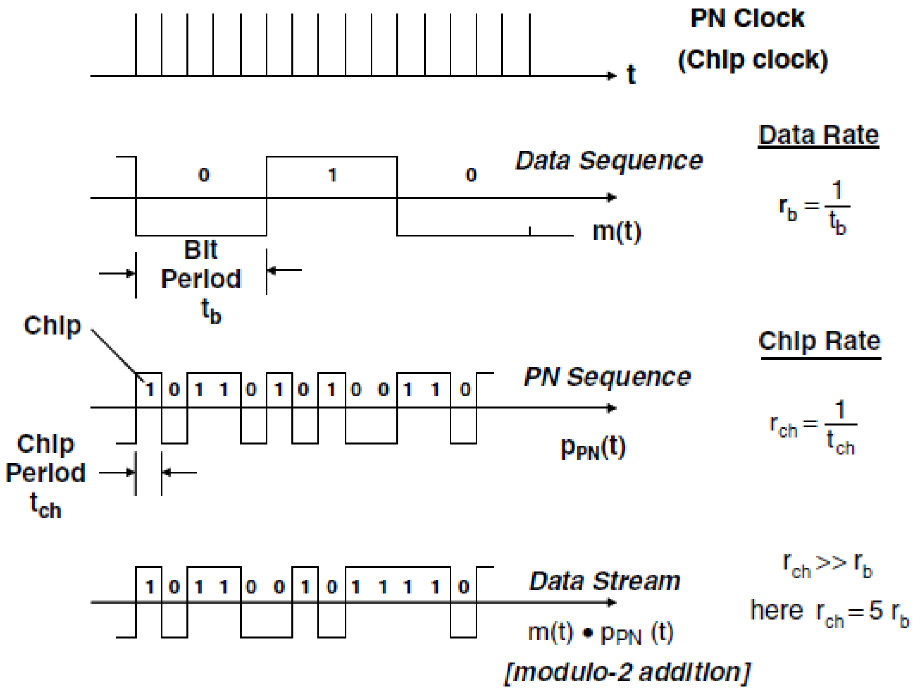


Figure 7.9 Generation of PN data stream

Two basic approaches are used in CDMA for spectrum spreading, based on the data elements being acted upon by the PN sequence:

- Direct Sequence Spread Spectrum (DS-SS): The baseband signal sequence is acted upon by the PN sequence (as discussed above).
- Frequency Hopping Spread Spectrum (FH-SS): The transmission (carrier) frequency is acted upon by the PN sequence, producing a sequence of modulated data bursts with time varying pseudorandom carrier frequencies.

Each of these techniques offers unique characteristics and performance.The approach DS-SS is discussed in the following subsection detailly.

7.3.1 Direct Sequence Spread Spectrum

A direct sequence spread spectrum (DS-SS) system spreads the baseband data bits with the PN sequence. In the most widely used satellite network implementation, a phase modulated baseband data stream is generated, then used to phase modulate an RF carrier with the PN spread signal.

Figure 7.10 shows the elements present in the DS-SS communications satellite system. The data bitstream is phase modulated onto a carrier, then directed to the PN Code Modulator which phase modulates the RF carrier to produce the spread signal. After passing through the satellite channel, the signal is 'despread' with a balanced modulator, then phase demodulated to produce the original data bitstream.

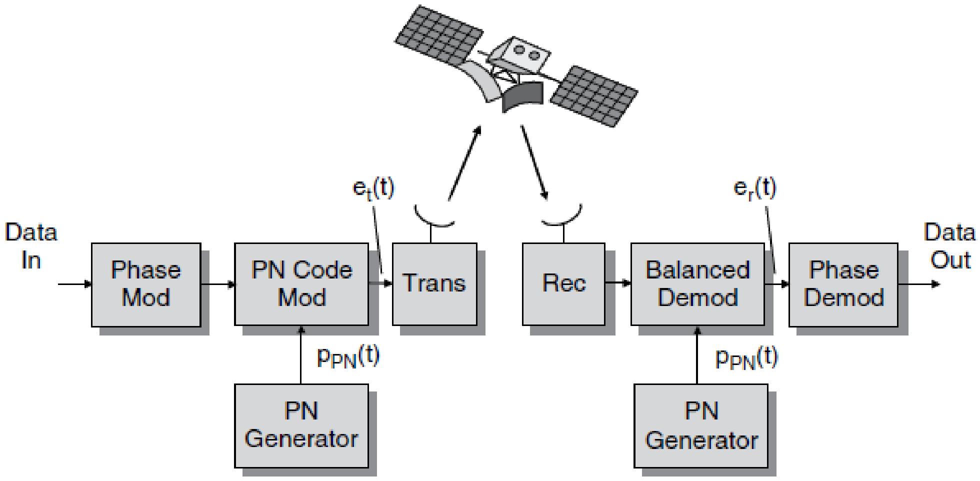


Figure 7.10 DS-SS satellite system elements

Let A cos[ωt+φ(t)] be the data modulated signal, where Φ(t) is the information bearing phase modulation. The PN Code Modulator phase modulates the data modulated signal with the PN sequence $p_{PN}(t)$. The output of the PN modulator is then

$$e_t(t) = p_{PN}(t) A \cos[\omega t + \varphi(t)] \quad (7.16)$$

$e_t(t)$, transmitted through the satellite transmission channel, is spread in frequency by the PN sequence to a bandwidth of B_{rf} .

At the receiver, the received spread signal is multiplied by a stored replica of $p_{PN}(t)$. The

output of the balanced demodulator is then

$$e_r(t) = p_{PN}^2(t) A \cos[\omega t + \varphi(t)] \quad (7.17)$$

Since $p_{PN}(t)$ is a binary signal

$$p_{PN}^2(t) = 1$$

Therefore,

$$e_r(t) = A \cos[\omega t + \varphi(t)] \quad (7.18)$$

and the information $\Phi(t)$ can be recovered through the final phase demodulator.

If the transmitter and receiver PN code sequences do not match, random phase modulation occurs and the spread signal looks like noise to the demodulator. If binary phase shift keying (BPSK) is used for the carrier modulation in the DS-SS satellite network, a simplification of the implementation is possible. Figure 7.11 shows a functional representation of DS-SS BPSK waveform generation process for the system. The binary data stream is used to BPSK modulate a carrier, then that signal used as input to the BPSK Code modulator.

Consider a constant envelope data modulated signal. The output of the BPSK data modulator, $s_x(t)$, will be of the form

$$s_x(t) = A\cos[\omega_\theta t + \theta_x(t)] \tag{7.19}$$

and the output of the BPSK code modulator, s(t), will be

$$s(t) = A\cos[\omega_0 t + \theta_x(t) + \theta_p(t)] \tag{7.20}$$

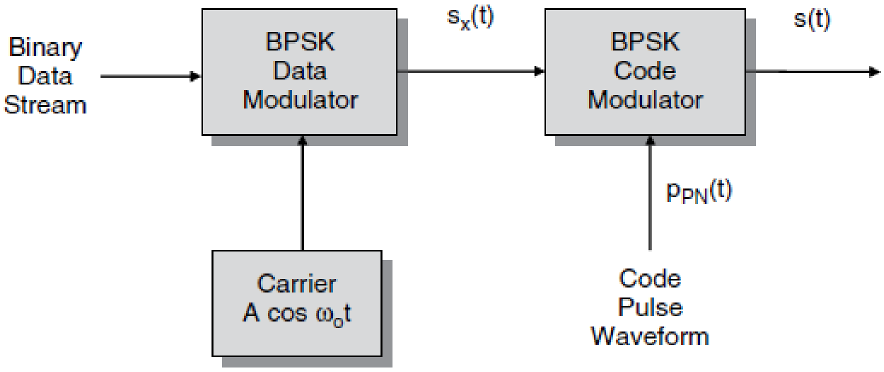


Figure 7.11 Functional representation of DS-SS BPSK waveform generation

where

$\theta_x(t)$ = the component of carrier phase due to the data stream

$\theta_p(t)$ = the component of carrier phase due to the spreading sequence

Since the data stream is binary, $s_x(t)$ will have the following values, depending on the data bit

Data Bit

0 $s_x(t) = A \cos[\omega_0 t + 0] = A \cos \omega_0 t$

1 $s_x(t) = A \cos[\omega_0 t + \pi] = -A \cos \omega_0 t$

This is equivalent to

$$s_x(t) = Ax(t) \cos \omega_0 t \qquad (7.21)$$

where

Data Bit

0 $x(t) = +1$

1 $x(t) = -1$

That is, $s_x(t)$ is an anti-podal pulse stream with the above assigned values.

The output of the BPSK code modulator Equation (7.20) is similarly of the form

$$s(t) = Ax(t)\, p_{PN}(t) \cos\omega_0 t \qquad (7.22)$$

where $p_{PN}(t)$ is again an anti-podal pulse stream with the values +1 and −1.

A modulator implementation producing the desired components for s(t) as described by Equation (7.22) is shown in Figure 7.12. This implementation reduces the number of phase modulators required to one, and simplifies the hardware elements significantly over the functional waveform generation process of Figure 7.11.

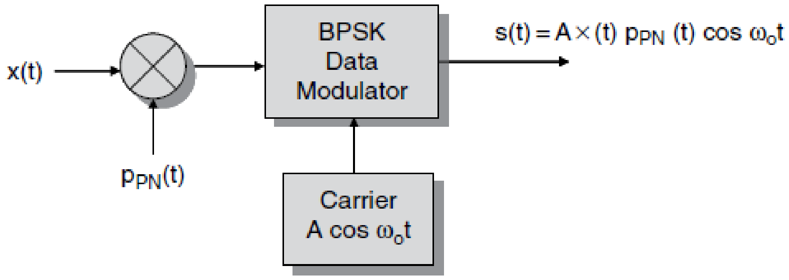


Figure 7.12 DS-SS BPSK modulator implementation

The output of the BPSK modulator, s(t), is transmitted through the satellite transmission channel, and will be subject to a propagation delay and possible degradations.The received downlink signal at the demodulator input will be of the form

$$r(t) = A‘x(t - t_d)\ p_{PN}(t - t_d)\ \cos[\omega_0(t - t_d) + \varphi] \qquad (7.23)$$

where

A‘ = the amplitude A modified by the transmission channel

t_d = the transmission channel propagation delay, in s

Φ = a random phase angle, introduced by the channel

Figure 7.13 shows a typical DS-SS/BPSK demodulator implementation. The demodulation process starts with a despreading correlator, shown by the dashed box in the figure. The despreading correlator mixes the received signal r(t) with the stored replica of the PN sequence, $p_{PN}(t)$, delayed by $t\hat{}_d$, which is an estimate of the propagation delay experienced in the transmission channel. $p_{PN}(t - t\hat{}_d)$ is a synchronized replica of the spreading PN code sequence.

The output of the correlator is then

$$r‘(t) = A‘\ x(t - t_d)\ p_{PN}(t - t_d)\ p_{PN}(t - \hat{}t_d)\ \cos[\omega_0(t - t_d) + \varphi] \qquad (7.24)$$

Since $p_{PN}(t)$ has the values +1 or −1, $p_{PN}(t - t_d)p_{PN}(t - \hat{}t_d) = 1$
when and only when$\hat{}t_d$ = td that is, when the receive code sequence is synchronized with the transmit receive code.

Then, at $\hat{}t_d = t_d$,

$$r'(t) = A’\ x(t - t_d)\ \cos[\omega_0(t - t_d)\Phi] \qquad (7.25)$$

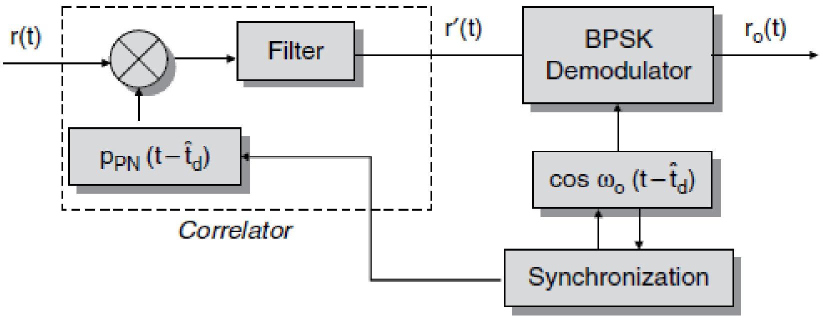


Figure 7.13 DS-SS BPSK Demodulator Implementation

which we found is identical to

$$r'(t) = A' \cos[\omega_0(t - t_d) + \theta_x(t - t_d)] \qquad (7.26)$$

This signal is then sent to a conventional BPSK demodulator, as shown in Figure 10.13, for recovery of the data, represented by the carrier phase component $\theta_x\ (t - t_d)$. The output of the BPSK demodulator is then

$$r_0(t) = x(t - t_d) \qquad (7.27)$$

The output of the BPSK demodulator is the desired data bit stream, delayed by the propagation delay of the channel, t_d.

References

[1] Roddy, D.: Satellite Communications, Third Edition, McGraw-Hill TELCOM Engineering, New York, 1989.

[2] Pratt, T.- Bostian, C.W. - Allnutt, J.E.: Satellite Communications, Second Edition, John Wiley & Sons, Inc., New York, 2003.

[3] Ippolito, L.J.: Satellite Communications Systems Engineering. Atmospheric Effects, Satellite Link Design and System Performance, Wiley, Singapore,2008

8 Earth Station

Three essential elements of any satellite communication network or system include the Earth segment, the space segment and the up/down link between the space segment and the Earth segment.

This chapter comprehensively covers different subsystems that make up a typical satellite Earth station and the key factors governing its design. Beginning with a brief introduction to the Earth station in terms of its role and significance in the overall satellite communication network, the chapter goes on to discuss different types of Earth stations along with their architecture and different subsystems constituting an Earth station. Key performance parameters and other factors governing the design of an Earth station are also discussed. After reading the chapter you will learn the following:

- The role of an Earth station in overall satellite communication set-up
- Types of Earth station with reference to size and complexity and type of service
- Earth station architecture
- Design considerations for an Earth station
- Earth station subsystems and function of each subsystem
- Earth station figure-of-merit
- Satellite tracking methodologies

8.1 Earth Station

An Earth station is a terrestrial terminal station mainly located on the Earth's surface. It could even be airborne or maritime. Those located on the Earth's surface could either be fixed or mobile. The Earth station is intended for communication with one or more manned or unmanned space stations as shown in Figure 8.1 or with one or more terrestrial stations of the same type via one or more reflecting satellites or other objects in space as depicted in Figure 8.2. In most of the applications related to communication satellites, Earth stations transmit to and receive from satellites. In some special applications, the Earth stations only transmit to or receive from satellites. Receive-only Earth station terminals are mainly of relevance in the case of broadcast transmissions. Transmit-only Earth station terminals are relevant to data gathering applications.

Major subsystems comprising an Earth station include (a) transmitter system whose complexity depends upon the number of different carrier frequencies and satellites simultaneously handled by the Earth station; (b) receiver system whose complexity again depends upon the number of frequencies and satellites handled by the Earth station; (c) antenna system that is usually a single antenna used for both transmission and reception with a multiplex arrangement to allow simultaneous connection to multiple transmit and receive chains; (d) tracking system to ensure that the antenna points to the satellite; (e) terrestrial interface equipment; (f) primary power to run the Earth station and (g) test equipment required for routine maintenance of the Earth station and terrestrial interface.

Earth station design is mainly governed by the type of service to be provided such as fixed satellite service (FSS), broadcast satellite service (BSS), mobile satellite service (MSS) etc.; quality of service to be provided mainly dictated by Earth station G/T ; type of Communication requirements such as telephony, data, television etc.; international regulations; cost considerations and site constraints. The Earth station is characterized by frequency band (6/4 GHz, 14/12 GHz etc.), polarization (linear, circular etc.), antenna diameter, effective isotropic radiated power (EIRP), G/T , receive antenna gain, modulation type, access method (FDMA, DMA etc.) and so on.

8.2 Types of Earth Station

Earth stations are generally categorized on the basis of type of services or functions provided by them though they may sometimes be classified according to the size of the dish antenna. Based on the type of service provided by the Earth station, they are classified into the following three broad categories. Communication requirements such as telephony, data, television etc.; international regulations; cost considerations and site constraints. The Earth station is characterized by frequency band (6/4 GHz, 14/12 GHz etc.), polarization (linear, circular etc.), antenna diameter, effective isotropic radiated power (EIRP), G/T , receive antenna gain, modulation type, access method (FDMA, DMA etc.) and so on.

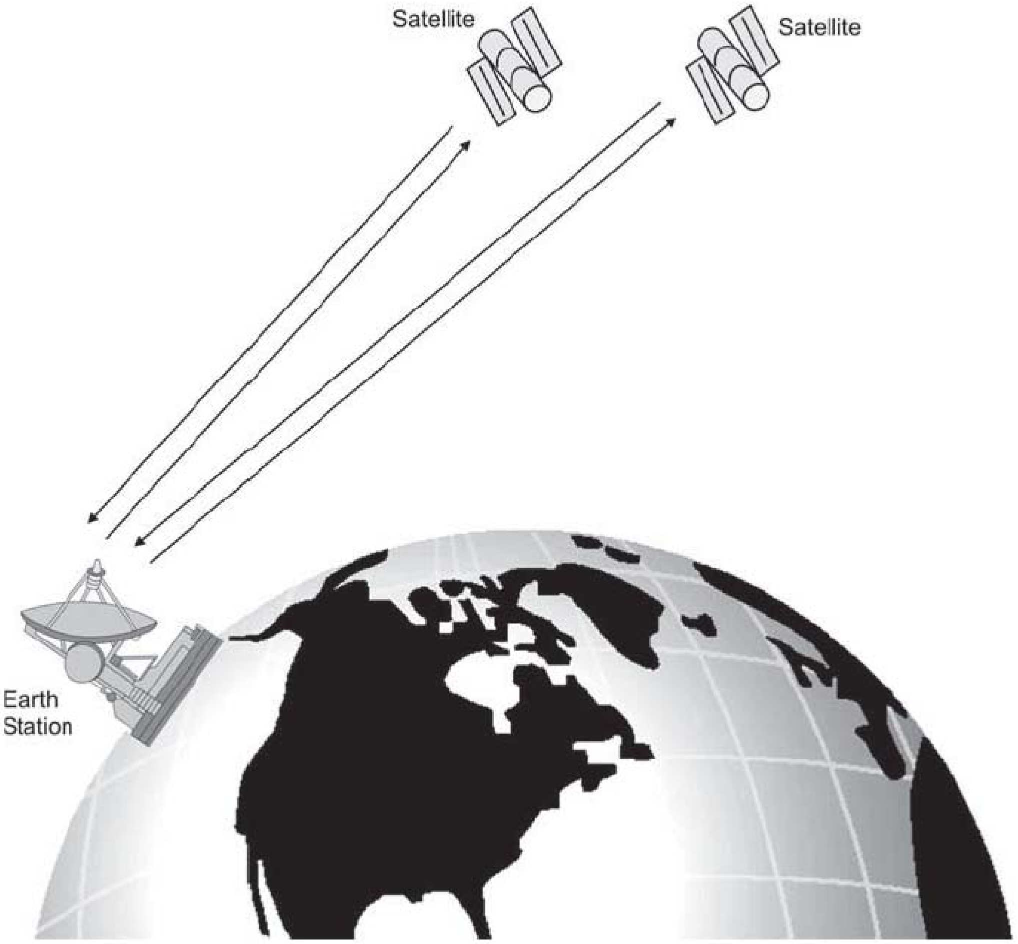


Figure 8.1 Earth station communicating with satellites

8.2 Types of Earth Station

Earth stations are generally categorized on the basis of type of services or functions provided by them though they may sometimes be classified according to the size of the dish antenna. Based on the type of service provided by the Earth station, they are classified into the following three broad categories.

1. Fixed Satellite Service (FSS) Earth Stations
2. Broadcast Satellite Service (BSS) Earth Stations
3. Mobile Satellite Service (MSS) Earth Stations

Earth stations are also sometimes conveniently categorized into three major functional groups depending upon their usage. These categories are the following.

1. Single function stations
2. Gateway stations
3. Teleports

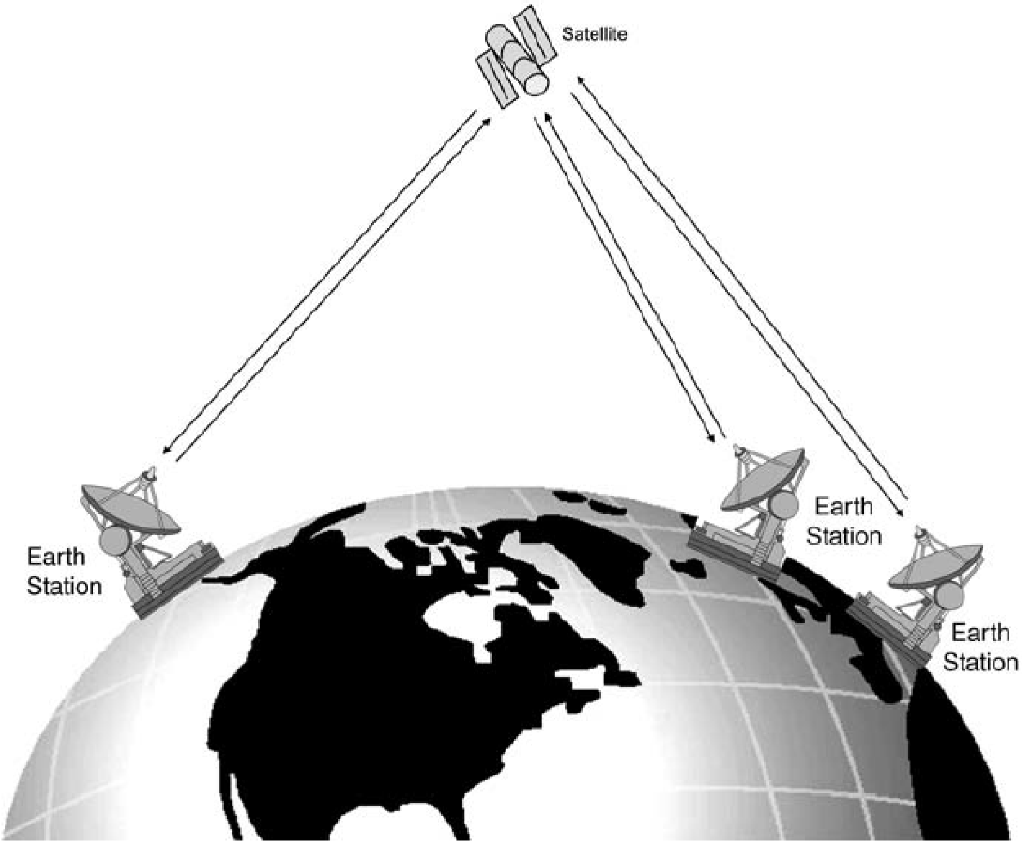


Figure 8.2 Earth station communicating with another Earth station

Each of the above mentioned types is briefly described in the following paragraphs.

8.2.1 Fixed Satellite Service (FSS) Earth Station

Under the group of FSS Earth stations, we have the large Earth stations (G/T≅40 dB/K) (Figure 8.3), medium Earth stations (G/T≅30 dB/K), small Earth stations (G/T≅ 25 dB/K), very small terminals with transmit/receive functions (G/T≅20 dB/K) (Figure 8.4) and verysmall terminals with receive only functions (G/T≅12 dB/K) (Figure 8.5). Fixed satellite service (FSS) is a term that is mainly used in North America. The service involves the use of geostationary communication satellites for telephony, data communications and radio and television broadcast feeds. FSS satellites operate in either the C band (3.7 GHz to 4.2 GHz) or the Ku band (11.45 GHz to 11.7 GHz and 12.5 GHz to 12.75 GHz in Europe, and 11.7 GHz to 12.2 GHz in the USA). FSS satellites operate at relatively lower power levels as compared to Broadcast Satellite.

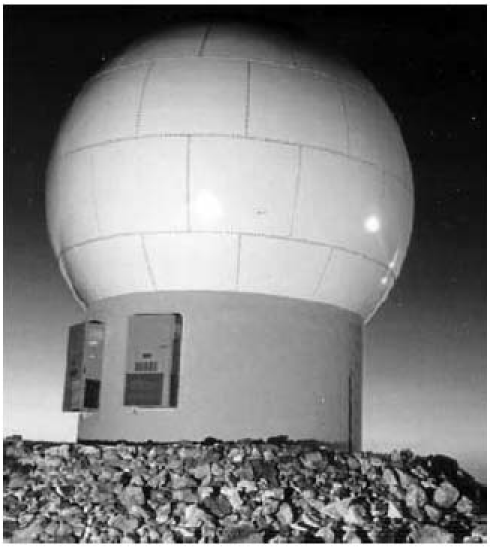


Figure 8.3 Large Earth station

Service (BSS) satellites and therefore consequently require a much larger dish. Also, FSS satellite transponders use linear polarization as compared to circular polarization employed by BSS satellite transponders. Fixed satellite service (FSS) is a term that is mainly used in North America. The service involves the use of geostationary communication satellites for telephony, data communications and radio and television broadcast feeds. FSS satellites operate in either the C band (3.7 GHz to 4.2 GHz) or the Ku band (11.45 GHz to 11.7 GHz and 12.5 GHz to 12.75 GHz in Europe, and 11.7 GHz to 12.2 GHz in the USA).

FSS satellites operate at relatively lower power levels as compared to Broadcast Satellite Service (BSS) satellites and therefore consequently require a much larger dish. Also, FSS satellite transponders use linear polarization as compared to circular polarization employed by BSS satellite transponders.

8.2.2 Broadcast Satellite Service (BSS) Earth Stations

Under the group of BSS Earth stations, we have large Earth stations (G/T≅15 dB/K) used for community reception and small Earth stations (G/T≅8 dB/K) used for individual reception. Technically, broadcast satellite service or BSS as it is known by the International Telecommunications Union (ITU) refers only to the services offered by satellites in specific frequency bands. These frequency bands for different ITU regions include 10.7 GHz to 12.75 GHz in ITU region-1 (Europe, Russia, Africa), 12.2 GHz to 12.7 GHz in ITU region-2 (North and South America) and 11.7 GHz to 12.2

Figure 8.4 Very Small terminal (Transmit/Receive)

GHz in ITU region-3 (Asia, Australia). ITU adopted an international BSS plan in the year 1977. Under this plan, each country was allotted specific frequencies for use at specific orbital locations for domestic services. It is also known by the name of Direct Broadcast Service or DBS or more commonly as Direct-to-Home or DTH. The term DBS is often used interchangeably with DTH to cover both analog and digital video and audio services received by relatively small dishes.

8.2.3 Mobile Satellite Service (MSS) Earth Stations

Under the group of MSS Earth stations, we have the large Earth stations (G/T≅4 dB/K), medium Earth stations (G/T≅12 dB/K) and small Earth stations (G/T≅24 dB/K). While both large and medium Earth stations require tracking, small MSS Earth stations are without tracking equipment.

Satellite phone is the most commonly used mobile satellite service. It is a type of mobile that connects to satellites instead of terrestrial cellular sites. Mobile satellite services are provided both by the geostationary as well as low Earth orbit satellites. In the case of the former, three or four satellites can maintain near continuous global coverage. These satellites are very heavy and therefore very expensive to build and launch. Geostationary satellite based mobile services also suffer from noticeable delay while making a telephone call or using data services. Yet another disadvantage of geostationary satellite system is frequent absence of line-of-sight between the satellite and the phone due to obstacles present in between the two. The disadvantages of the geostationary satellite system are overcome in Low Earth Orbit (LEO) satellite systems. In the case of LEO satellite systems, an obstacle would block the satellite access only for a short time until another satellite passes overhead. The major advantage of LEO satellites based communication system is worldwide wireless coverage with with no gaps. However, a constellation of LEO satellites would be required to maintain uninterrupted coverage. Iridium (Figure 8.6) and Globalstar are the two major LEO satellite systems offering mobile satellite services. Globalstar uses 44 satellites with the orbital inclination of the satellites being 52°. It may be mentioned here that the polar regions are not covered by the Globalstar constellation. Iridium operates 66 satellites orbiting in polar orbits.

Figure 8.5 Very small terminal (Receive only)

Radio links are used between the satellites in order to relay data to the nearest satellite connected to the Earth station.

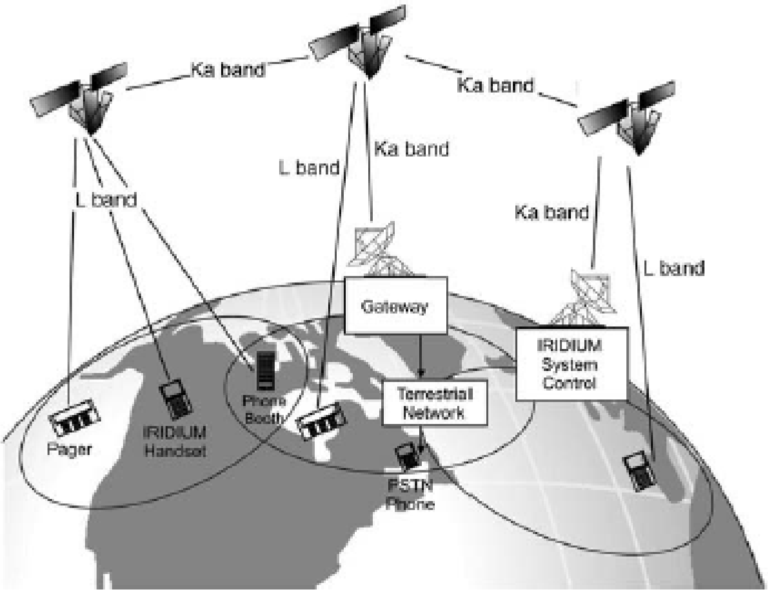


Figure 8.6 Iridium system

8.2.4 Single Function Stations

Single function stations are characterized by a single type of link to a satellite or a satelite constellation. These stations may be transmit-only, receive-only or both. Some common examples of single function stations include television receive-only (TVRO) terminals used for TV reception by an individual (Figure 8.7), satellite radio terminals, receive-only terminals used at a television broadcast station to pick up contribution feeds, two-way VSAT terminals used at a retail stores for point-of-sale communications with the corporate hub, handheld satellite telephone terminals designed to work with a single satellite constellation and many more.

8.2.5 Gateway Stations

Gateway stations serve as an interface between the satellites and the terrestrial networks and also serve as transit points between satellites. These stations are connected to terrestrial networks by various transmission technologies, both wired such as coaxial cables, optical fibres etc. and wireless such as microwave towers. Unlike single function Earth stations where it is just up-linking and down-linking operations that comprise the core activity; in the case of gateway stations, signal processing is the major activity.

Figure 8.7 TVRO terminal

A gateway station receives a large variety of terrestrial signals at any given time. These include telephone signals, television signals, and data streams and so on. These signals come in different formats; use various levels of multiplexing and telecommunication standards. A lot of signal manipulation activities therefore need to be carried out on these signals before they are routed to the intended satellite. There are both independent as well as satellite system owner's gateway stations. Antennas used at gateway stations working with a specific satelite system need to be designed and manufactured in accordance with the standards promulgated by the satellite fleet owner. Type approved Earth station equipment that is a particular satelite system specific is available from many manufacturers.

8.2.6 Teleports

Teleport is a type of gateway station operated by firms that are usually not a part of a specific satellite system. Teleports are useful for those companies whose not-too-high requirement of satellite connectivity does not justify having their own dishes. They are also useful for business houses located in crowded places inhibiting line-of-sight to the satellite of interest due to the close proximity of another tall building or some other obstacle. Teleports are usually located on the outskirts of the city and the connectivity from the subscriber company to the teleport station is usually provided through a hub. All subscribers are linked to the hub and the hub in turn is connected to the teleport through a fibre-optic or a microwave link.

Modern teleport stations are versatile and often have a wide range of dishes conforming to the standards of many satellite operators so as to be able to offer a wide range of services to the subscribers. The services offered by teleport stations typically include format conversion, encryption, production and post production, turn-around services and even leasing transportable uplinks for temporary events.

8.3 Earth Station Architecture

The major components of an Earth station include the RF section, the baseband equipment and the terrestrial interface. In addition, every Earth station has certain support facilities such as power supply unit with adequate back-up, monitoring and control equipment and thermal and environment conditioning unit (heating, air-conditioning etc.). Though the actual architecture of an Earth station depends on the application; the block schematic arrangement of Figure 8.8 is representative of a generalized Earth station.8.3

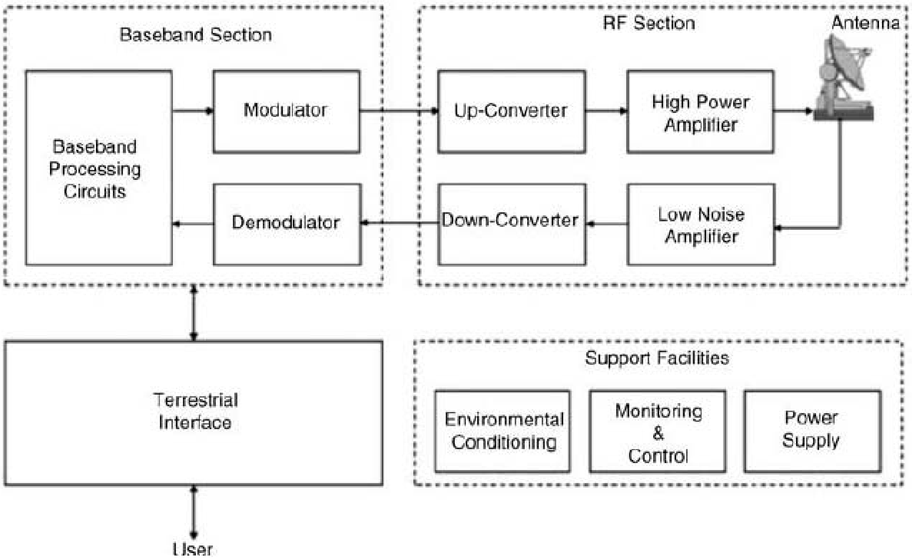


Figure 8.8 Block schematic arrangement of a generalized Earth station

The RF section as shown in the block schematic arrangement of Figure 8.8 mainly comprises of antenna subsystem, the up-converter and the high power amplifier (HPA) in the up-link channel and the antenna subsystem, low noise amplifier (LNA) and the down-converter in the down-link channel. In the case of an Earth station being a major hub of a network or if service reliability were a major concern; equipment redundancy is used in the RF section. RF section interfaces with the modem subsystem of the baseband section. The job of up-converter in the up-link channel is to up-convert the baseband signal to the desired frequency. The upconverted signal is then amplified to the desired level before it is fed to the feed system for subsequent transmission to the intended satellite. Similarly, a low noise amplifier amplifies the weak signals received by the antenna. The amplified signal is then down converted to the intermediate frequency level before it is fed to the modem in the baseband section. The antenna feed system provides the necessary aperture illumination, introduces the desired polarization and also provides isolation between the transmitted and the received signals by connecting HPA output and LNA input to the cross-polarized ports of the feed. The baseband section performs the modulation/demodulation function with the specific equipment required depending upon the modulation technique and the multiple access method employed. For example, in the case of a two-way digital communication link, the baseband section would comprise of a digital modem and a time division multiplexer.

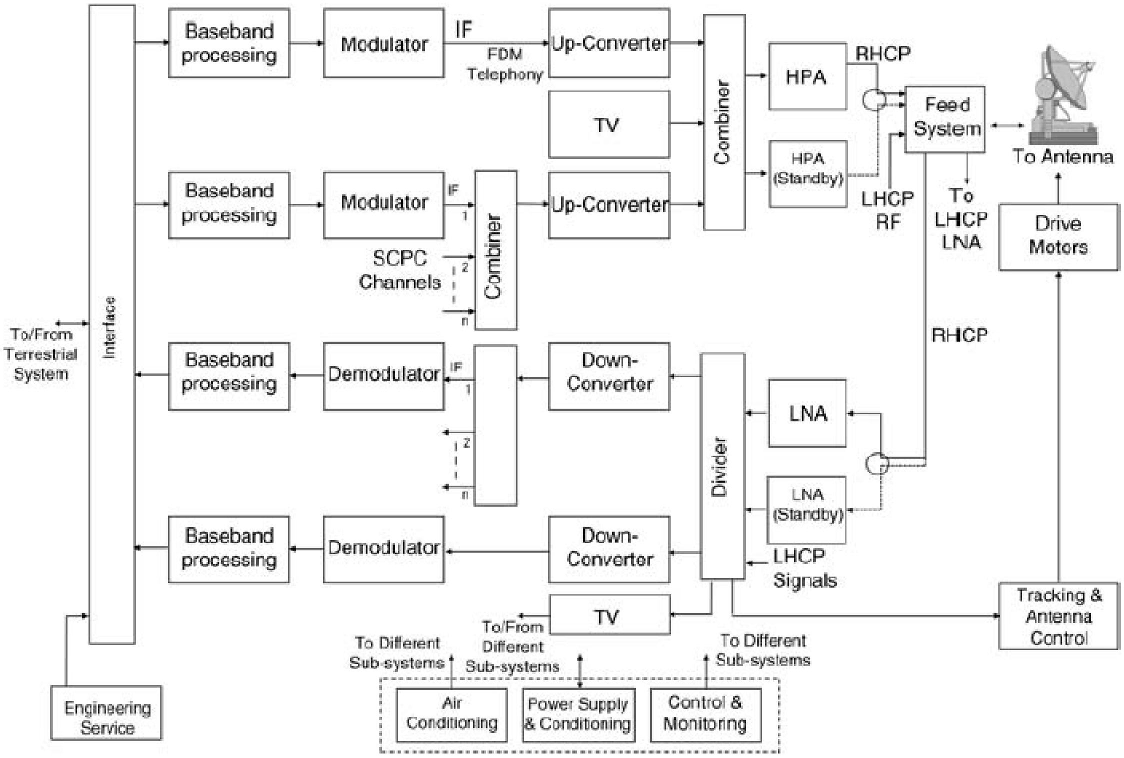


Figure 8.9 Block schematic of a typical large FSS Earth station

The baseband section input/output is connected to the terrestrial network through a suitable interface known as terrestrial interface. It may be connected directly to the user in some applications. The terrestrial network could be a fibre optic cable link or a microwave link or even a combination of the two. In addition to the three abovementioned components of an Earth station, every Earth station has support facilities such as tracking, control and monitoring equipment, power supply with back-up and environmental conditioning unit.

The complexity of Earth station architecture depends upon the application. For example, a TVRO Earth station would be far less complex than a FSS Earth station interconnecting large traffic nodes. Figure 8.9 shows the detailed block schematic of a typical large FSS Earth station. Redundancy of equipment as outlined earlier is evident in the RF and the baseband sections. The diagram shown is typical of the Earth station used in the INTELSAT network.

Figure 8.10 shows the block schematic of a typical VSAT remote terminal showing both the outdoor and the indoor units along with the dish antenna. The outdoor unit is typically of the size of a shoe box or even smaller and contains different subsystems of the RF section. The dish antenna is typically 0.55 to 2.4 metre in diameter. The indoor unit, typically of the size of a domestic video recorder, contains different subsystems of the baseband section. These include modulator and demodulator, multiplexer and demultiplexer and user interfaces.

8.4 Earth Station Design Considerations

Design of an Earth station is generally a two-step process. The first step involves identification of Earth station requirement specifications, which in turn govern the choice of system parameters. The second step is about identifying the most cost effective architecture that achieves the desired specifications.

Requirement specifications affecting the design of an Earth station include type of service offered (Fixed satellite service, Broadcast satellite service or Mobile satellite service), communication requirements (telephony, data, television etc.), required base band quality at the destination, system capacity and reliability. Major system parameters relevant to Earth station design include transmitter EIRP (Effective Isotropic Radiated Power), receiver figure-of-merit (G/T), system noise and interference and allowable tracking error.

When it comes to designing a satellite communication system, it is always advisable to minimize the overall system costs including both development as well as recurring costs of the Earth and space segments. A trade-off is always possible between the two where the cost of one segment can be reduced at the cost of the other. That is, cost incurred on the Earth station could be reduced by having a more expensive space segment. According to the most fundamental economic rule of satellite telecommunications, every dollar spent on the space segment gets divided by the number of potential users on the ground whereas every dollar spent on the user terminal gets multiplied by the same number. This leads to the practice of designing less expensive user terminals and more expensive satellites, a trend that started with advent of geostationary satellites way back in 1960s and continued for more than three decades. Several trade-offs are possible in Earth station design optimization, which are discussed in detail in Section 8.4.2. However, as we shall see, these trade-offs are subjected to some technical and regulatory constraints, which are also briefly outlined during the discussion.

8.4.1 Key Performance Parametres

Key performance parameters governing Earth station design include the EIRP (Effective or Equivalent Isotropic Radiated Power) and the figure-of-merit (G/T). While the former is a transmitter parameter, the latter is indicative of receiver performance in terms of sensitivity and the quality of the received signal.

Effective (or Equivalent) Isotropic Radiated Power (EIRP). EIRP gives the combined performance of the high power amplifier (HPA) and the transmitting antenna. It is given by the product of the power output of HPA at the antenna and the gain of the transmitting antenna. Expressed in decibels, EIRP is the sum of the power output of HPA in dB and the gain of transmitting antenna in dB. If a particular HPA-transmitting antenna combine had an EIRP of 60 dBW, it would imply that the RF power radiated by the antenna is the same as that radiated by an isotropic radiator in that direction when fed with million times more power at its input.

EIRP is defined for both Earth station transmitting antenna as well as satelite transmitting antenna. It is important to note that EIRP is always measured at the antenna. When we see a footprint map with EIRP numbers for a given transponder on a satellite, these numbers are indicative of the amount of power sent down to the Earth station and measured as it left the satellite's down-link dish. Some satellite operators have the practice of taking space loss in the satellite footprint maps for their users.

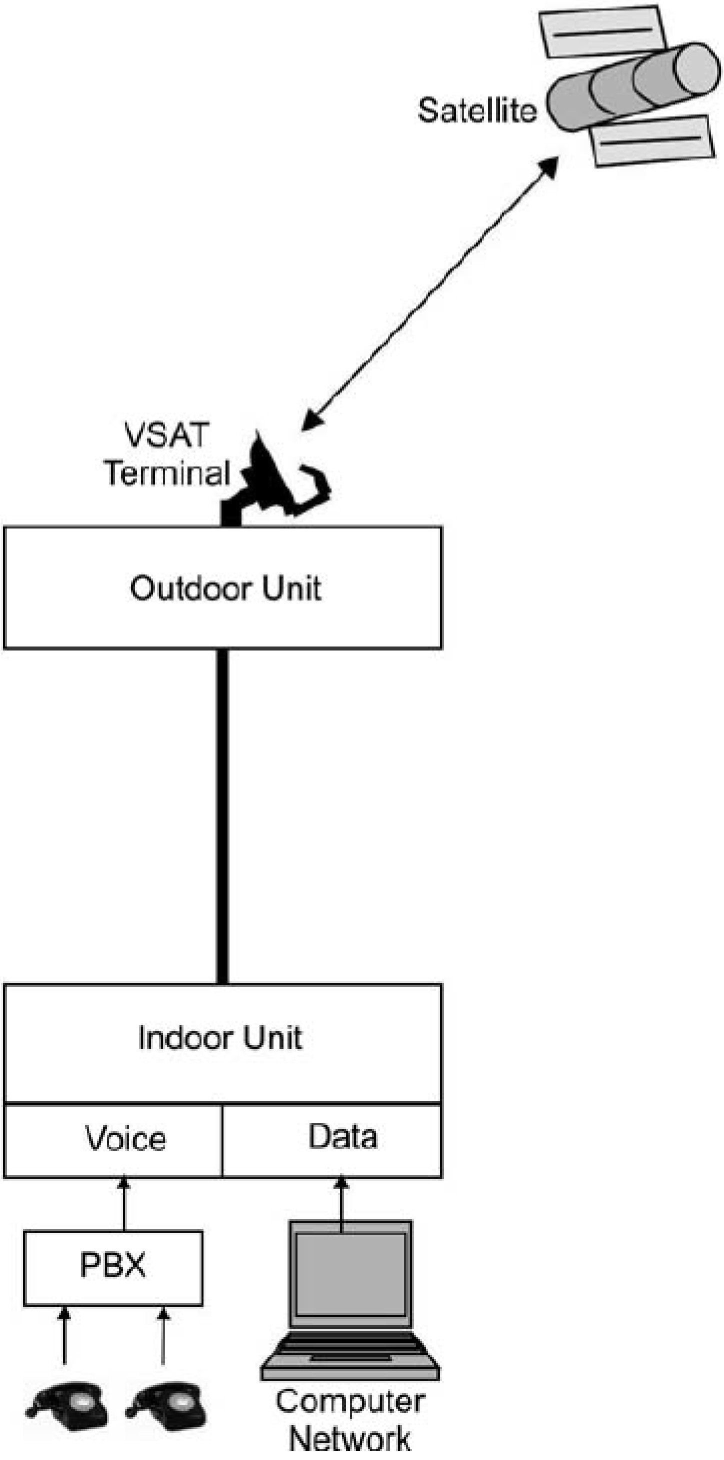


Figure 8.10 Block schematic of VSAT remote terminal

They prefer to give the signal strength as is received on ground thus correcting for the space loss at the frequency of operation. This number is known as Illumination Level and is given by (EIRP − Space loss). Some operators prefer to specify received power per unit bandwidth. The unit of bandwidth is typically taken as 4 kHz, which is the bandwidth of a typical analog telephony channel. In that case, the new value called Power Flux Density (PFD) is given by (EIRP − Space loss − Bandwidth). It may be mentioned here that the PFD is specified in the decibels scale.

Receiver Figure-of-merit (G/T). Receiver figure-of-merit is indicative of how the receiving antenna performs together with the receiving electronics to produce a useful signal. While the EIRP gives the performance of the transmitting antenna and HPA combination; receiver figure-of-merit, tells us about the sensitivity of the receiving antenna and the Low Noise Amplifier (LNA) combine to weak received signals. As it is effectively a measurement of the sensitivity of the receiving antenna to weak signals, the larger the value of receiver figure-of-merit, the better it is. The response of the receiving system to weak signals is largely governed by the receiving system gain and the overall system noise. The figure-of-merit is therefore defined by a parameter called G/T ratio, which is the ratio of receiving antenna gain to system noise temperature. G/T is expressed in dB/K. G/T of the Earth station may be enhanced by increasing the receiving antenna gain or lowering the noise temperature or both. For any practical communication link, EIRP of the satellite transmitting antenna and the G/T of the Earth station receiving antenna and the EIRP of the Earth station transmitting antenna and G/T of the satellite receiving antenna have to work together to get the desired results. A poorer G/T necessitates a higher EIRP and vice versa. Both EIRP and G/T were discussed at length along with illustrative examples in Chapter 3.

8.4.2 Earth Station Design Optimization

As outlined earlier, the transmitter EIRP and receiver G/T together dictate the performance of the communication system and therefore one can be traded off against the other during the design optimization process. In the early days of development of satellite technology, available EIRP from satellites was pretty low, which made complex and expensive Earth stations a necessity. In those days, Earth station antennas were several tens of metres in diameter and cost a few million US Dollars a piece. Current trend is to minimize Earth station complexity at the cost of a complex space segment. It is more so for applications that involve a large user population such as direct broadcast, business use, mobile communication and so on.

Possible trade-offs can be best understood by resorting to expression for Earth station G/T. The generalized expression for G/T is given by equation 8.1.

$$G/T = C/N_0 - EIRP + (L_p + L_m) + k \qquad (8.1)$$

Where C/N_0, EIRP, L_p, L_m and k are carrier-to-total noise power spectral density, satellite's effective isotropic radiated power, path loss, link margin and Boltzmann constant (in dBs) respectively. For a minimal cost Earth station, G/T should be minimized. This can be possible by either using relatively higher EIRP in the satellite or being able to afford a lower carrier-tonoise ratio or both. For desired base band quality at the receiver, this can be achieved by using modulation schemes that are more immune to noise. In the case of digital base band, coding allows a further reduction in G/T . Other factors governing Earth station complexity and hence its cost include the Earth station EIRP, antenna tracking requirements, traffic handling capacity and terrestrial interface requirements. In addition, there are international regulatory issues and technical constraints that drive the optimization process.

In the early days, International Telecommunications Union (ITU) had put certain limitations on the transmitted EIRP of the FSS satellites sharing their frequency bands with terrestrial systems in order to allow them to co-exist. For applications such as direct broadcast, mobile communications etc. where a small size terminal is a requirement, limiting the satellite EIRP would put a lower limit on the diameter of the dish antenna. This implies that G/T can not be reduced below a certain value. Even if G/T were reduced by using a smaller antenna, reduction in size would increase antenna side lobes to undesired levels, which would further lead to more interference to and from adjacent satellite systems. This has been overcome by having exclusive frequency allocations for these services, thus permitting relatively much higher EIRP for the satellites.

The satellite EIRP is also limited by the DC power available on the satellite, maximum power that can be generated by the high power amplifiers on board the satellite and the practical constraints imposed on the satellite antenna diameter limiting the antenna gain. Also, for a given antenna size, gain reduces with decrease in operational frequency. That is why, satelite EIRP limitation is more acute in L-band used for mobile communications.

Having decided on the EIRP and G/T values, the next obvious step is to choose an optimum configuration of the antenna, high power amplifier and the low noise amplifier to achieve the desired values. Specified EIRP and G/T may be obtained by any of the possible options. A small size antenna, which would be low cost, and a relatively low noise LNA, which would be expensive, is one option. Another would be the use of large size antenna and LNA with a higher noise figure. Antenna size also affects the EIRP as a small size antenna may require a prohibitively large HPA.

8.4.3 Environmental and Site Considerations

It is important to consider a number of environmental and locational factors while making a decision on the site of an Earth station. Environmental parameters of interest include external temperature and humidity, rainfall and snow, wind conditions, likelihood of Earth quakes, corrosive conditions of the atmosphere and so on. Careful site selection can take care of the ill effects of some but not all of these factors.

Minimizing radio frequency interference (RFI) and electromagnetic interference (EMI) is another requirement. RFI and EMI produced by the Earth station can cause interference to other RF installations. Also, RFI and EMI from external sources can adversely affect the Earth station performance. It is usually necessary to carry out a radio frequency survey at various possible sites before a final choice is made on the Earth station location.

An essential requirement is to have a clear line-of-sight to the satellites of interest. Availability of sufficient space for the Earth station equipment, easy transportation to the Earth station and reliable electrical power are the other requirements.

Though all efforts are made to take into account the abovementioned factors while choosing a suitable site for the Earth station; it is important that the satellite operators specify all possible environmental factors and site constraints to potential manufacturers of the Earth station equipment. Also, the manufacturers should build into the design of Earth station equipment the ability to operate reliably under specified environmental and interference conditions.

8.5 Earth Station Hardware

Most Earth station hardware can be categorized into one of the three groups namely RF equipment, IF and baseband equipment and terrestrial interface equipment. Basic functions performed by each one of these equipment classes were briefly outlined in

Section 8.3 on Earth station architecture. In the present section, these are described in more detail with focus on individual building blocks constituting these three groups.

8.5.1 RF Equipment

The RF equipment comprises of up-converters, high power amplifiers (HPA) and the transmit antenna in the transmit channel, and the receive antenna, low noise amplifiers (LNA) and down-converters in the receive channel. While the output of HPA feeds the transmit antenna; the receive antenna is connected to the input of theLNA.Transmit and receive antenna functions are almost invariably performed by the same antenna. Figure 8.11 shows the block schematic arrangement of the RF portion of the Earth station equipment.

From the viewpoint of EIRP and also Earth station G/T , it is always desirable to have minimal losses in the waveguide/cable connecting the antenna and the HPA output or LNA input. To achieve this, one option is to house the RF section in a separate shelter or cabinet adjacent to the antenna. Another option is to package the uplink and down link equipment separately. The uplink equipment mainly comprises of the modulator and the up-converter and the down-link equipment has down-converter and the demodulator. Yet another practice prevalent in the case of VSAT and TVRO terminals is to combine the LNA and first stage of the down-converter into a single block known as low noise block (LNB). This has the distinct advantage of offering low noise amplification and down conversion to L-band, which allow them the use of inexpensive coaxial cable to further carry the signal.

8.5.1.1 Antenna

Different types of antenna and their performance parametres of relevance to satellite communications have been discussed earlier in Chapter-6 on satellite hardware. A brief description of antennas of relevance to Earth stations is given in this section.

Different variants of reflector antenna are commonly used as Earth station antenna. These mainly include the prime focus fed parabolic reflector antenna, offset fed sectioned parabolic reflector antenna and cassegrain fed reflector antenna. The prime focus fed parabolic reflector antenna as shown in Figure 8.12 is used for an antenna diameter of less than 4.5 m, more so for receive only Earth stations. An offset fed sectioned parabolic reflector antenna (Figure 8.13) is used for antenna diameters of less than 2 m. Offset feed configuration eliminates the blockage of the main beam due to feed and its mechanical support system and thus improves antenna efficiency and reduces side lobe levels.

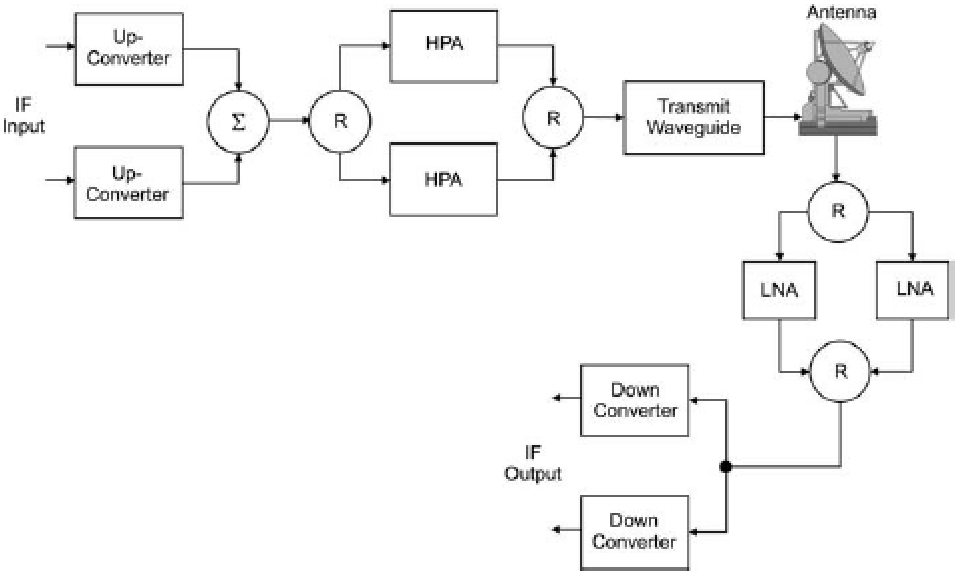


Figure 8.11 Block schematic of the RF portion of the Earth station

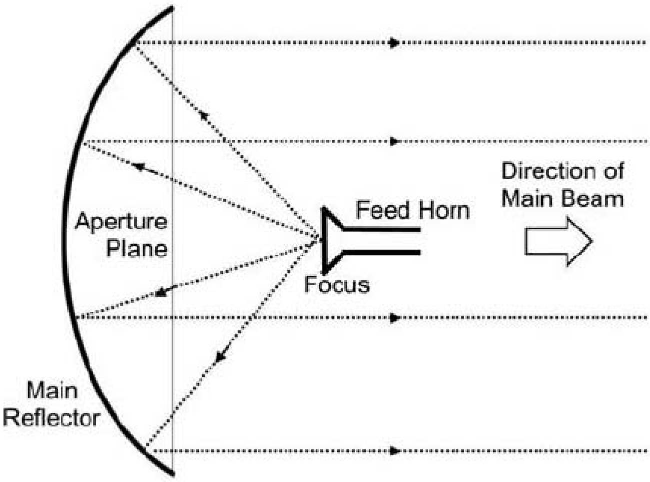


Figure 8.12 Prime focus fed parabolic reflector antenna

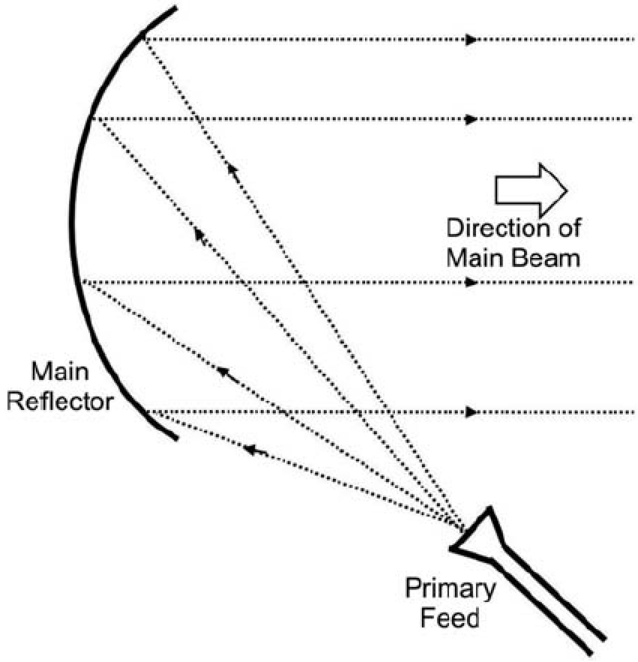


Figure 8.13 Offset fed sectioned parabolic reflector antenna

In a variation of the prime focus fed parabolic reflector antenna, a piece of hook shaped waveguide extending from the vertex of the parabolic reflector is connected to the feed horn. In this case, the low noise block (LNB) is connected to the waveguide behind the parabolic reflector. This allows placement of electronics without causing any obstruction to the main beam in addition to allowing an easy access to it.

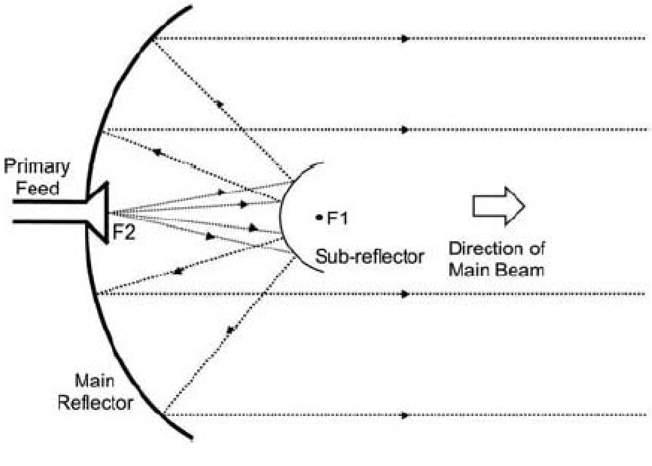


Figure 8.14 Cassegrain antenna

Cassegrain antennas overcome most of the shortcomings of the prime focus fed parabolic reflector antennas. The cassegrain antenna uses a hyperbolic reflector placed in front of the main reflector, closer to the dish than the focus as shown in Figure 8.14.

This hyperbolic reflector receives the waves from the feed placed at the centre of the main reflector and bounces them back towards the main reflector. In the case of cassegrain antenna, the front end electronics instead of being located at the prime focus is positioned on or even behind the dish. Offset feed configuration is also possible in case of Cassegrain antenna (Figure 8.15).

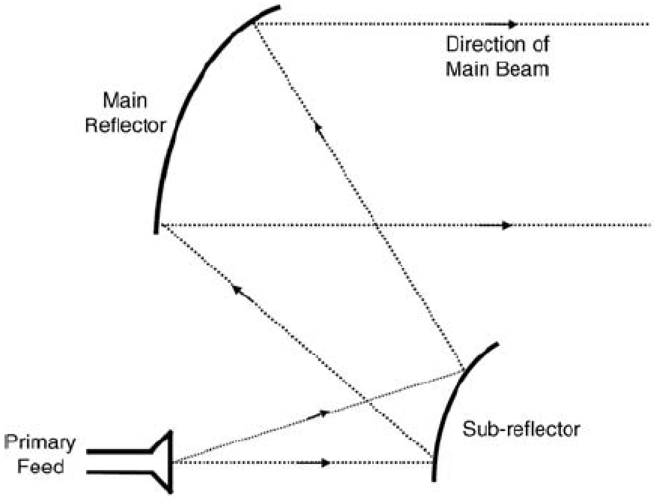


Figure 8.15 Offset fed Cassegrain antenna

Yet another common reflector antenna configuration is the Gregorian antenna [Figure 8.16(a)]. This configuration uses a concave secondary reflector just behind the prime focus. The purpose of this reflector is also to bounce the waves back towards the dish. The front end in this case is located between the secondary reflector and the main reflector. Offset feed configuration is also possible in case of Gregorian antenna [Figure 8.16(b)].

8.5.1.2 High Power Amplifier

EIRP, which is the product of the power output of the high power amplifier (HPA) minus the waveguide losses and gain of transmit antenna, is an important parameter in deciding the uplink performance of the Earth station. To achieve the specified EIRP of the Earth station, one could have a combination of moderate output power HPA and a high gain antenna. The other option is to have a relatively higher power output HPA feeding a moderate sized antenna. One could always draw a family of curves for different frequency bands (C, Ku, Ka) showing a variation of HPA power output against antenna diameter for desired value of EIRP. Figure 8.17 shows one such family of curves drawn for an EIRP of 80 dB. As is evident from the curves, one would need to have a 800 watt HPA for a C-band transponder, if the antenna diameter were to be around 6 m or so. A 10 m antenna on the other hand would need only a 300 watt HPA.

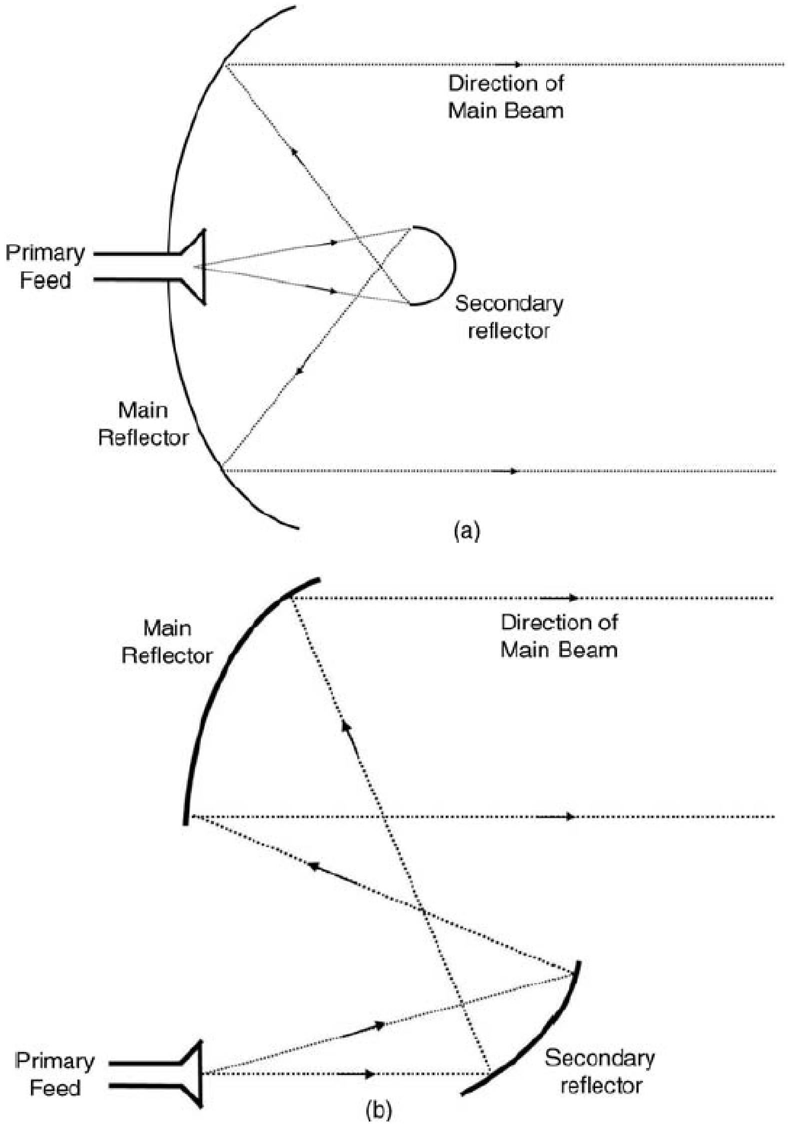


Figure 8.16 (a) Gregorian antenna (b) Offset fed Gregorian antenna

Different types of power amplifiers used in Earth stations include (a) Traveling wave tube (TWT) amplifiers (b) Klystron amplifiers and (c) Solid state power amplifiers (SSPA).

(TWT) amplifiers (b) Klystron amplifiers and (c) Solid state power amplifiers (SSPA). SSP are used for relatively lower power applications while tube based amplifiers are used when the required power levels are high. Klystrons are narrowband devices providing a bandwidth of the order of 40 to 80 MHz that is tunable over a range of 500 MHz or more. Power levels offered are from several hundred watts to few kilowatts. On the other hand, TWTA is a wideband amplifier offering a bandwidth as large as 500 MHz or more and a power level from a few watts to a few kilowatts.

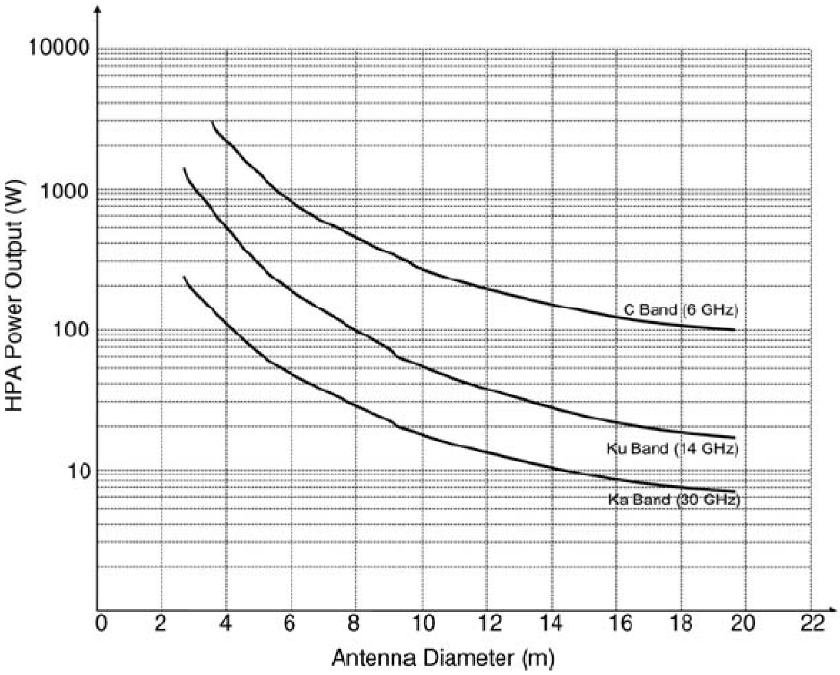


Figure 8.17 HPA power output versus antenna diametre for a given *EIRP*

However, klystrons are less expensive, simple to operate and easy to maintain. Solid state power amplifiers are comparatively cheaper and more reliable though the power level offered by them is limited as compared to klystrons and TWTAs. Apart from frequency, power level, linearity and bandwidth, other important characteristics of high power amplifiers include gain, variation of group delay with frequency, noise performance and AM/PM conversion. While variation of group delay with frequency is also a cause of intermodulation components; AM/PM conversion produces intelligible crosstalk and intermodulation noise.

Commonly used amplifier configurations for multi-carrier operation include the single amplifier and multiple amplifier configurations. In the case of single amplifier configuration (Figure 8.18), different carriers are combined before the amplifier and the composite signal is fed to the input of the amplifier. The amplifier is operated in the linear region of its operating characteristics to minimize the inter-modulation noise. In the figure shown, redundant HPA is used to improve the system reliability. It is terminated in a matched load. In the case of multiple amplifier configuration, each HPA amplifies either a single or a group of carriers as shown in Figure 8.19. Amplified signals are then combined at the output of HPAs. This configuration allows the HPA to be operated near its full power rating, which increases the overall efficiency of the Earth station. However this comes at the cost of additional HPAs.

8.5.1.3 Up-converters/Down-converters

Up-converters and down-converters are frequency translators that convert the IF used in the modems and baseband equipment to the operating RF frequency bands (C,Ku and Ka) and vice versa. The up-converter translates the IF signal at 70 MHz (or 140 MHz) from the modulator to the operating RF frequency in C or Ku or Ka band as the case may be. The down-converter translates the received RF signal in C or Ku or Ka band into IF signal, which is subsequently fed to the demodulator. Either single or double frequency conversion topologies are used for up-converters and down-converters.

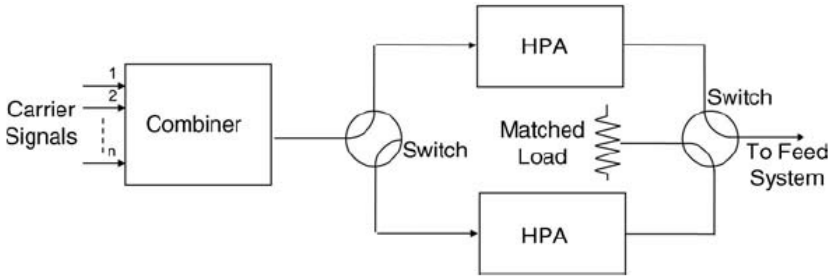


Figure 8.18 Single amplifier HPA configuration

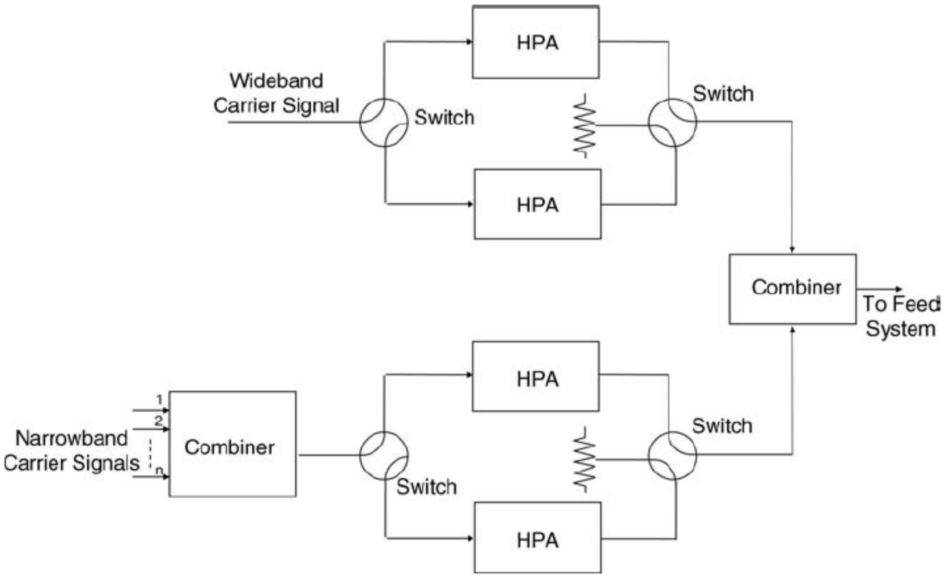


Figure 8.19 Multiple amplifier HPA configuration

Figures 8.20(a) and (b) respectively show the schematic diagrams of up-converters and down-converters employing single frequency conversion topology. A typical up-converter uses a stage of amplification before the mixer stage. Mixer along with local oscillator (LO) provides frequency conversion. A frequency synthesizer is used for LO so as to be able to generate any frequency within the satellite up-link band. The signal is further amplified after frequency conversion before it is fed to the high power amplifier. A band pass filter at the output of the mixer eliminates LO frequency and its harmonics from reaching the up-link path. Insertion loss in the filter causes a reduction of the effective isotropic radiated power (EIRP). The operation of down-converter can be explained on similar lines. Amplification stage provides gain and reduces the noise contribution of mixer and the IF equipment. The frequency synthesizer provides frequency agility in the receive frequency operation.

Double frequency conversion topology employs a two mixer conversion stage. In the case of an up-converter using double conversion, the IF frequency is first up-converted to another intermediate frequency usually in the L-band. The signals is then amplified and fed to the second mixer stage where it is up-converted to the final operational RF frequency band. As outlined in the case of single stage converters, an amplifier precedes the mixer and a band pass filter follows the same. Figure 8.21(a) shows block schematic of a C-band up-converter employing double frequency conversion topology. Figure 8.21(b) shows the arrangement of double frequency conversion topology based down-converter for C-band operation. The diagrams are self explanatory.

8.5.1.4 Low Noise Amplifier (LNA)

While the high power amplifier (HPA) is an important element of the up-link path that together with the transmit antenna gain decides the EIRP of the Earth station; the low noise amplifier (LNA) is one of the key components deciding the system noise temperature and hence the figure-of-merit G/T of the Earth station. The design of LNA and the active devices around which the design of a LNA is configured have undergone many changes since the advent of satellite communication. Design of LNA in the early days used to be configured around masers and subsequently parametric amplifiers. Requirements on LNA in those days used to be far more stringent than they are today.

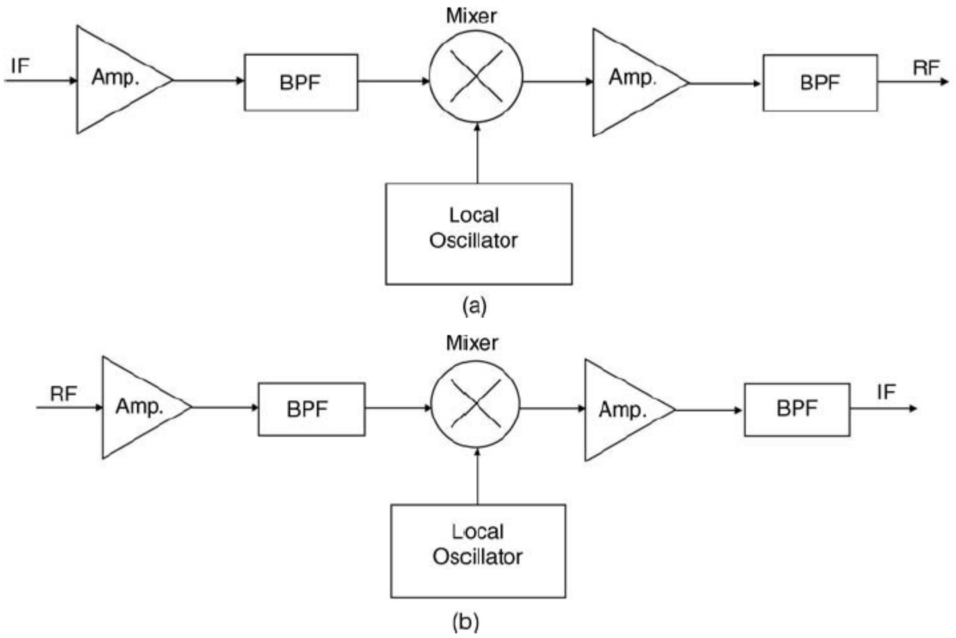


Figure 8.20 Simplified block diagram of single frequency conversion frequency converters (a) up-converter (b) down-converter

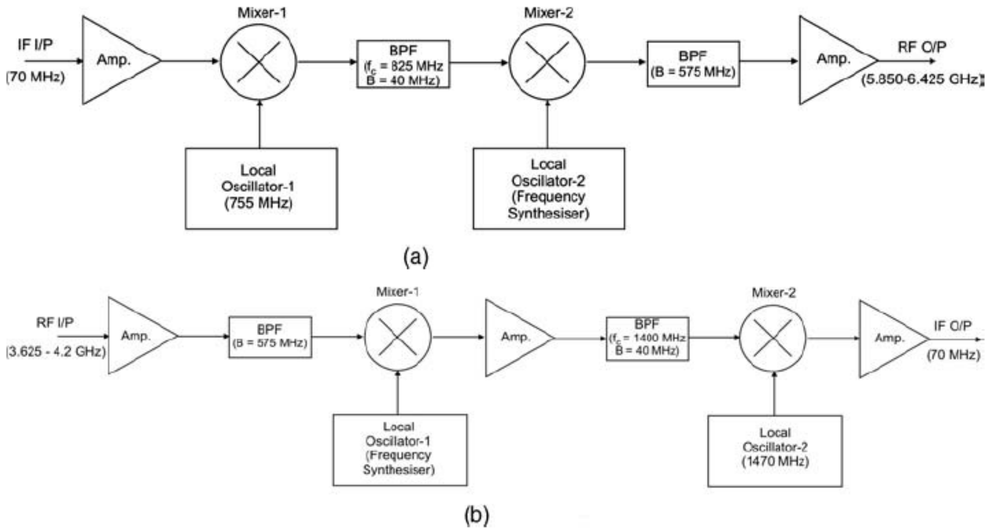


Figure 8.21 Simplified block diagram of double frequency conversion frequency converters (a) up-converter (b) down-converter

This has been made possible due to improvements in antenna efficiency and feed techniques and also increase in the transmit power capability of satellites.

Present day LNAs are configured around either Gallium Arsenide FET (GaAs FET) or High Electron Mobility Transistors (HEMT). These designs are far more compact and reliable than their parametric amplifier counterparts. The uncooled GaAs FET or HEMT based LNAs offer a noise temperature of about 75–170 K and compared with cryogenically cooled parametric amplifiers of early days giving noise temperature of 30–90 K. Table 8.1 gives a performance comparison of different LNA technologies.

There are two variations of low noise amplifier (LNA). In one of the variants, particularly where small size antennas are used such as those for TVRO or small business applications, the low noise amplifier section feeds a single stage down-converter in a single block called low noise block (LNB). Note that the frequency converters integrated into LNAs are all down converters shifting the received frequency to some lower frequency. This allows use of coaxial cables to transport the signal from the antenna to inside the premises. The output of LNB is a standard IF signal of around 1 GHz frequency. LNB is usually placed on the antenna structure itself and is connected to the feed directly. Figure 8.22 shows a photograph of a DTH dish and a co-located LNB.

Table 8.1 Performance comparison of LNA technologies

Type of Amplifier	Frequency Range (GHz)	Typical Noise Temperature (K)
Parametric Amplifier (Cooled)	3.7–4.2	30
	11–12	90
Parametric Amplifier (Uncooled)	3.7–4.2	40
	11–12	100
GaAs FET (Cooled)	3.7–4.2	50
	11–12	125
GaAs FET (Uncooled)	3.7–4.2	75
	11–12	170

Another variation of LNA is the LNC (‘C’ stands for converter). In LNC, the amplifier can typically be tuned to amplify over the entire bandwidth of a single transponder, whatever that bandwidth may be, before it down converts. The basic difference between LNB and LNC lies in the conversion bandwidth. LNB uses a block converter and is capable of handling block of frequencies from different transponders on the satellite. LNC uses the signal from a single transponder.

8.5.2 IF and Baseband Equipment

The nature and complexity of baseband equipment in an Earth station is mainly governed by the range of services offered by it and the requirement specifications that would be needed to provide those services. In the case of large Earth stations such as gateways, this portion of Earth station hardware also involves the largest investment. Important building blocks of IF and baseband equipment of the Earth station hardware include baseband processing circuits, modulator/demodulator (MODEM), multiplexer/

demultiplexer etc. The architecture of the IF and baseband section depends upon parametres like the modulation/demodulation scheme, multiple access method and so on. For example, in the case of an FDMA station, there must be one modem for each frequency resulting in use of a large number of such units. On the other hand, a TDMA Earth station needs to have only one modem for obvious reasons. However, the bandwidth requirement of the modem in the case of a TDMA station would be much larger than what it would be in the case of an FDMA station.

Figure 8.22 DTH dish and co-located LNB

Figure 8.23 shows the block schematic arrangement of FDMA Earth station capable of providing full duplex digital transmission for multiple carriers. In the arrangement shown, each carrier has its own dedicated modem tuned to a separate frequency in the transponder. The modems interface with the terrestrial network through a TDM multiplexer. Individual channels are combined into a single higher bandwidth channel.

Though the arrangement does not depict redundancy; full or partial redundancy is almost invariably provided to maintain high reliability.

Figure 8.24 shows the simplified block schematic arrangement of a typical TDM/ TDMA interactive VSAT terminal showing both the hub site and the remote locations. One can see the use of a single modem. In the case of TDMA, the frequency band occupied by the carrier is shared by several Earth stations on time basis. This implies that there needs to be only a single modem per Earth station. The modem receives bursts of data from different Earth stations in a manner that they do not overlap in time. In the case of CDMA however, different stations transmit on the same frequency simultaneously. Different multiple access techniques are discussed at length in Chapter 7.

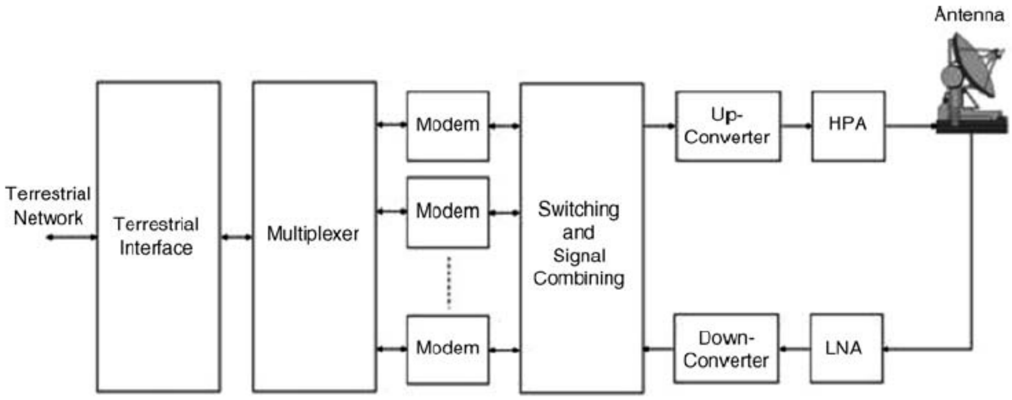


Figure 8.23 Block schematic of a full duplex FDMA digital communication Earth station

8.5.3 Terrestrial Interface

Terrestrial interface is that part of the Earth station that connects the Earth station to the users. Its importance lies in the fact that an improperly designed interface can significantly degrade the quality of service. The nature and complexity of the terrestrial interface depends upon the range of services or functions provided by the Earth station. The interface requirement varies from practically no interface in the case of portable user terminals such as satellite phones to a simple interface in the case of VSAT or TVRO terminals where the Earth station provides the services by directly feeding the consumer equipment, which could be a TV set or a personal computer. In the case of large Earth stations, depending upon service provided by it, terrestrial interface may even look like a telephone exchange or a broadcast studio.

Two major components of terrestrial interface include the terrestrial tail and the interface. Terrestrial tail links are needed to connect the main Earth station to one or more remote user locations with line-of-sight microwave and fibre optic cable being the two principle options. Commoninterfaces needed in satellite links and terrestrial networks include telephone interface (voice), data transmission interface (data) and television interface (video).

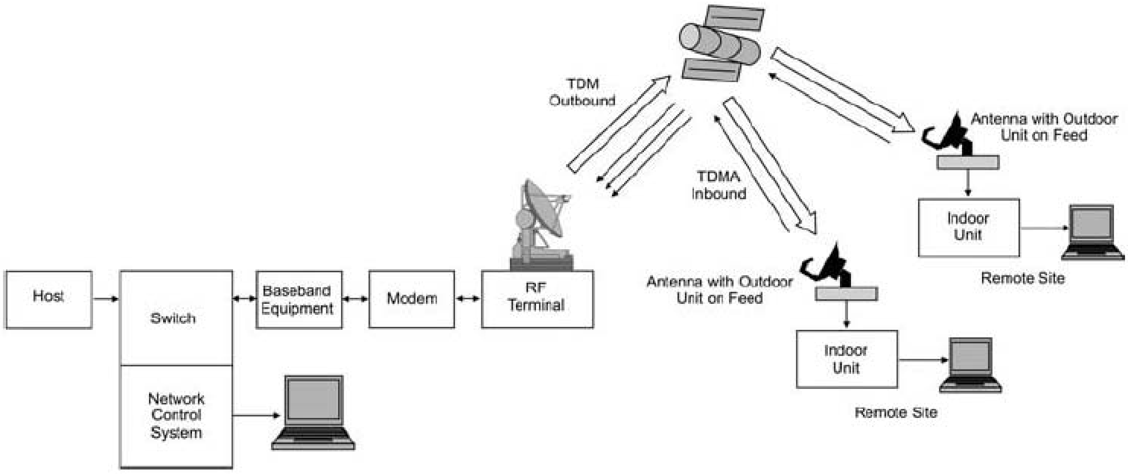


Figure 8.24 Block schematic arrangement of a typical TDM/ TDMA interactive VSAT terminal

8.5.3.1 Terrestrial Tail Options

The length of terrestrial tail may vary from fewtens of metres to hundreds of kilometres. C-band satellite systems suffer from problems of radio frequency interference (RFI). This necessitates that the Earth station be located far away from the city leading to use of an elaborate tail. On the other hand, Ku and Ka band systems do not have to worry much about interference related issues and therefore have relatively shorter tails. In the case of Ka band, site diversity may be used to maintain reliable service in the event of adverse weather conditions. In such situations, one would also need to have a tail link between diverse sites. This tail link would also need to handle large bandwidth and thus involve large investment. In addition, there may also be some short links connecting various facilities within the Earth station complex.

Figure 8.25 shows a typical set up depicting tail links connecting various centres. The diagram is self explanatory. It shows a fibre optic cable link connecting the RF terminal and the Earth station's main building, a microwave tail connecting the Earth station and a switching office, which in turn connects to user locations through public or private loops and an alternative fibre optic link between the Earth station and the customer location. Both fibre optic cable and microwave links are effective and reliable technologies. Fibre optic cable may be the preferred choice in the case of short tails such as those connecting the Earth station with other facilities or a VSAT terminal to customer. It is low noise and is immune to electromagnetic interference (EMI). Single hop microwave is also a good alternative for short tails.

For long and elaborate tails, microwave link is a better choice. Fibre optic cable in that case turns to be relatively more expensive option, more so in a metropolitan area. Fibre optic cable may retain the edge in terms of cost for tail lengths shorter than 20 km, beyond which the microwave link certainly entails relatively lower cost.

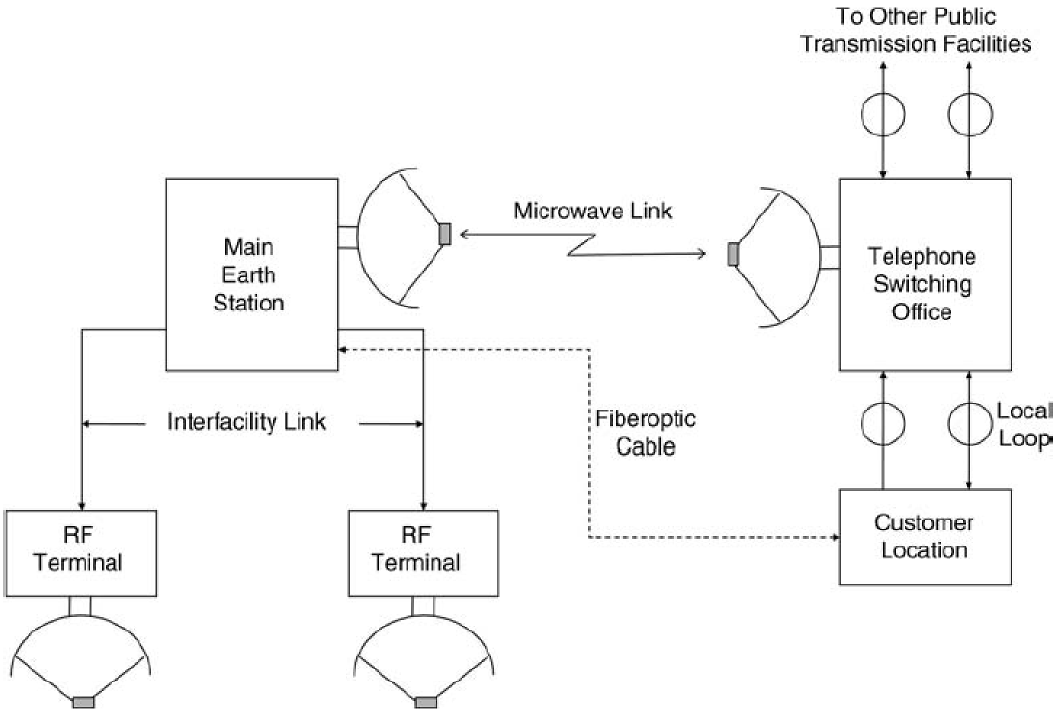


Figure 8.25 Typical Earth station set-up with terrestrial tail links

8.5.3.2 Interface

As outlined earlier, terrestrial interface equipment need could vary from practically no requirement as is the case with receive-only or satellite phone terminals to very elaborate interface equipment in the case of a large commercial satellite Earth station. Such stations are required to handle massive traffic comprising of hundreds of telephone channels together with data and video reaching the station through microwave and fibre optic systems using time division or frequency division terrestrial multiplex methods. Signals received from the terrestrial network therefore need to be de-

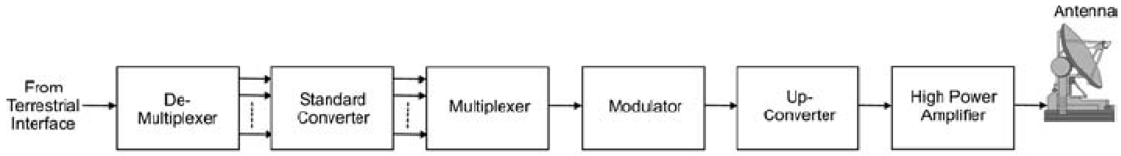


Figure 8.26 Terrestrial interface – up-link

multiplexed and then changed from the existing terrestrial formats to formats suitable for satellite transmission. After this format/standards conversion, the signals are processed further in the up-link chain of the Earth station as shown in Figure 8.31. On the down-link side, the signals received from satellite/s are processed in the down-link chain before they are sent to standard converter. After reformatting, the signals are multiplexed and put on the terrestrial network as shown in Figure 8.27.

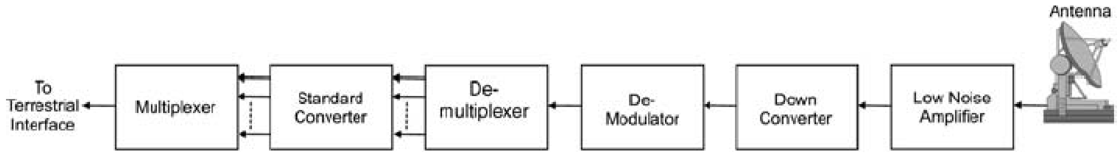


Figure 8.27 Terrestrial interface – down-link

Another interface related issue that is of particular relevance to handling digital signals arises out of the variation in the data rate at the receiving station over the period of the sidereal day caused by path length variation due to slight inclination and eccentricity of the orbit. As an example, at a nominal data rate of 9.6 kbps, a delay of 1.1 ms in the path length produces a peak-to-peak data rate variation of about 10.56 bits/s. It increases to 1.6984 kbps for a nominal data rate of 1.544 Mbps. This causes problems while interfacing with terrestrial networks that use synchronous transmission. Since these terrestrial networks cannot accommodate data rate variations of this magnitude, an elastic buffer that can absorb the expected peak-to-peak data rate variations is used between the satellite facilities and the terrestrial network as shown in Figure 8.28. An elastic buffer is nothing but a FIFO (First-in First-out) random access memory. The chosen elastic buffer should be large enough to absorb peak-to-peak data rate variations. For example, for peak-to-peak data rate variation of 10.56 bits/s, 16-bit buffer may be used. As the buffer is also going to add to the delay to the satellite link, smallest memory meeting the requirement should be used.

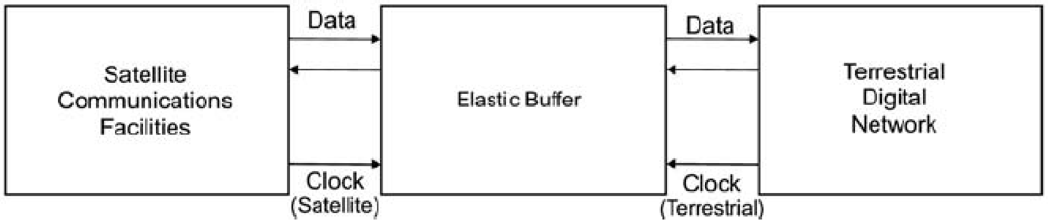


Figure 8.28 Use of elastic buffer to absorb data rate variations

8.6 Satellite Tracking

The Earth station antenna needs to track the satellite when the beam width of the antenna is only marginally wider than the satellite drift seen by it. Given the fact that satellite drift is typically in the range of 0.5–3° per day, antennas with large beamwidths such as DBS receivers do not require to track the satellite. On the other hand, large Earth stations do need some form of tracking with tracking accuracy depending upon the intended application. The tasks performed by the Earth station's satellite tracking system include some or all of the following.

1. Satellite acquisition
2. Manual tracking
3. Automatic tracking
4. Programme tracking

The acquisition system acquires the desired satellite by either moving the antenna manually around the expected position of the satellite or by programming the antenna to perform a scan around the anticipated position of the satellite. Automatic tracking is initiated only after the received signal strength due to the beacon signal transmitted by the satellite is above a certain threshold value, which allows the tracking receiver to lock to the beacon. Manual track option is used in the event of total failure of auto track system. Automatic tracking ensures continuous tracking of the satellite. Commonly used tracking techniques are described in the latter part of this section. In the case of programme tracking, the antenna is driven to the anticipated position of the satellite usually predicted by the satellite operator. Unlike automatic tracking, which is a closed loop system; programme track is an open loop system and therefore its accuracy is relatively much lower than that of auto track mode of operation.

8.6.1 Satellite Tracking System – Block Diagram

Figure 8.29 shows the generalized block schematic arrangement of the satellite tracking system. The Earth station antenna makes use of the beacon signal to track itself to the desired positions in both azimuth and elevation. The auto track receiver derives the tracking correction data or in some cases the estimated position of the satellite. The estimated position is compared with the measured position in the control subsystem whose output feeds the servomechanism. In the case of manual and programme track modes, the desired positions of the satellite in the two orthogonal axes are respectively set by the operator and the computer. The difference in actual and desired antenna positions constitutes the error signal that is used to drive the antenna.

8.6.2 Tracking Techniques

Tracking techniques are classified on the basis of the methodology used to generate angular

errors. Commonly used tracking techniques include the following.

1. Lobe switching
2. Sequential lobing
3. Conical scan
4. Monopulse track
5. Step track
6. Intelligent tracking

Of all the abovementioned techniques, the last four are more common in the case of satelite tracking. Sequential lobing with the rapid switching of a single beam has also been tried in

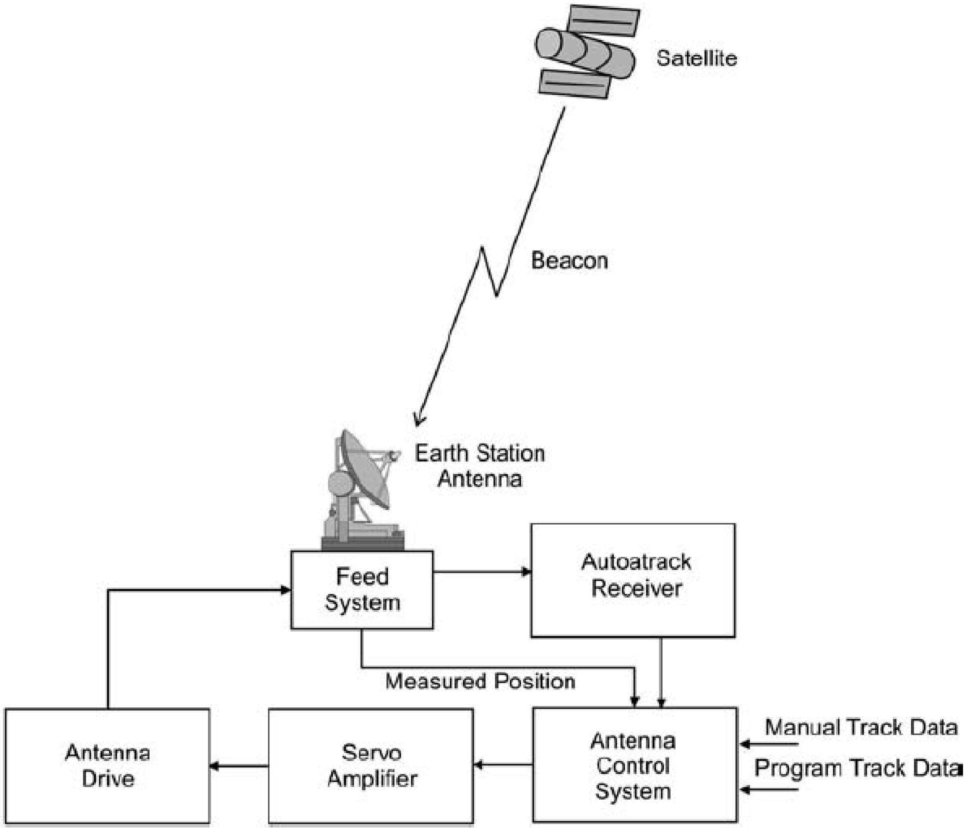


Figure 8.29 Block schematic arrangement of satellite tracking system

some cases. Each of the abovementioned concepts with relative merits and demerits is briefly described in the following paragraphs.

8.6.2.1 Lobe Switching

In the case of lobe switching tracking methodology, the antenna beam is rapidly switched between two positions around the antenna axis in a single plane as shown in Figure 8.30. The amplitudes of the echo from the object to be tracked are compared for the two lobe positions. The difference between the two amplitudes is indicative of the location of the target with respect to the antenna axis. When the object to be tracked is on the axis, the echo amplitudes for the two positions of the beam are equal and the difference between the two is zero. When the object is on one side of the antenna axis, the amplitude and sense of the difference signal tells how much and what side of the antenna axis the object is located. The difference signal can then be used to generate correction signal, which with the help of servo control loop can be used to drive the antenna to bring the object on to the antenna axis. The lobe switching technique is prone to inaccuracies if the object cross-section as seen by the antenna changes between different returns in one scan.

8.6.2.2 Sequential Lobing

In sequential lobing, the beam axis is slightly shifted off the antenna axis. This squinted beam is sequentially placed in discrete angular positions, usually four, around the antenna axis (Figure 8.31). The angular information about the object to be tracked is determined by processing several echo signals. The track error information is contained in the echo signal The track error information is contained in the echo signal amplitude variations. The squinting and beam switching is done with the help of electronically controlled feed and therefore can be done very rapidly practically simulating simultaneous lobing.

8.6.2.3 Conical Scan

This is similar to sequential lobing except that in the case of conical scan, the squinted beam is scanned rapidly and continuously in a circular path around the axis as shown in Figure 8.32. If the object to be tracked is off the antenna axis, the amplitude of the echo signal varies with antenna’s scan position. The tracking system senses the amplitude variations and the phase delay as function of scan position to determine the angular co-ordinates. The amplitude variation provides information on the amplitude of the angular error and the phase delay indicates direction. The angular error information is then used

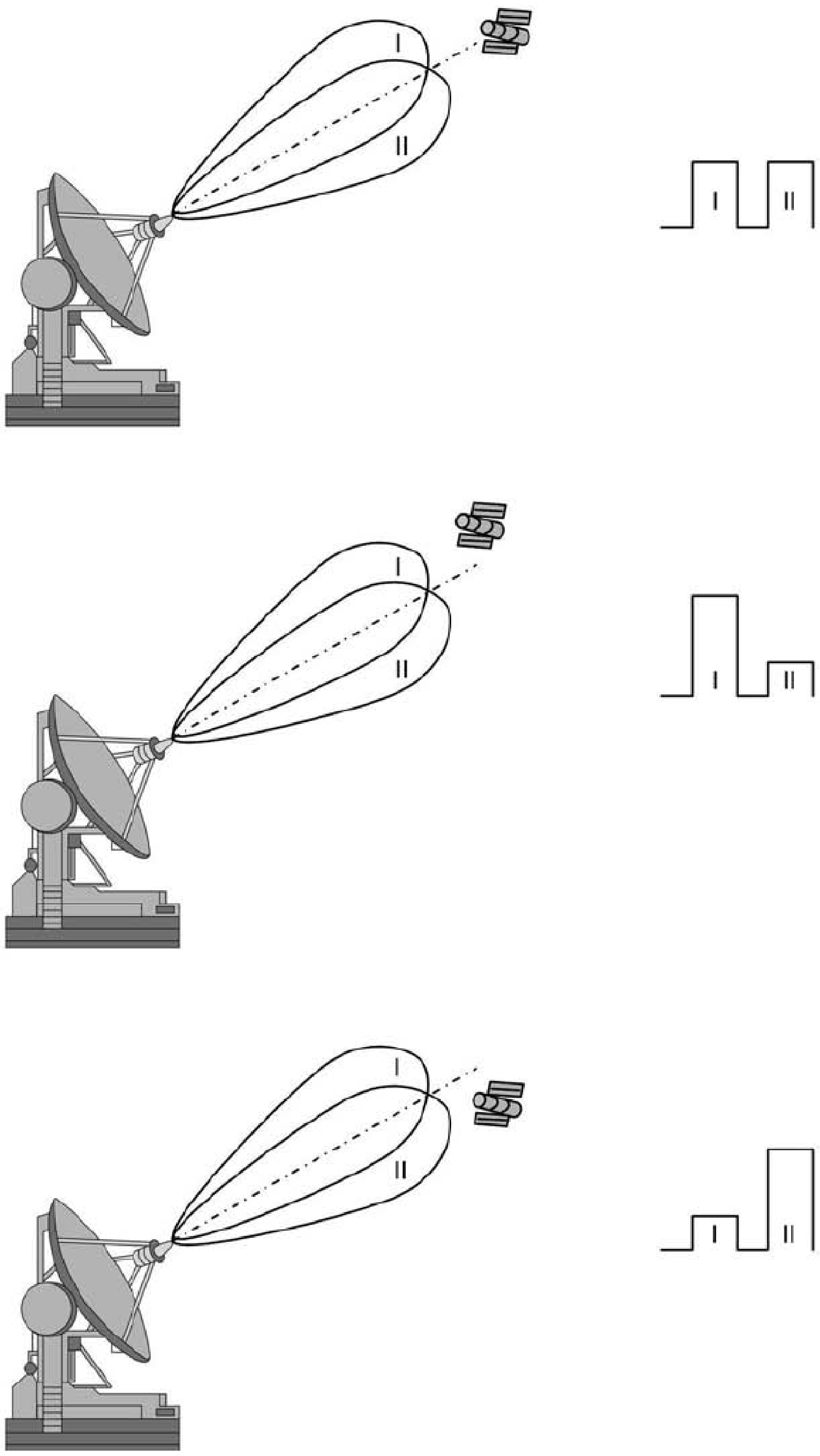


Figure 8.30 Principle of lobe switching technique

to steer the antenna axis to make it to coincide with the object location. The technique offers good tracking accuracy and an average response time. It is however not in common use now.

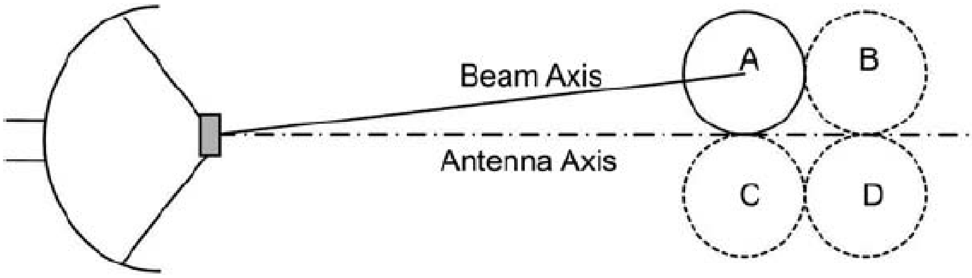


Figure 8.31 Principle of sequential lobing

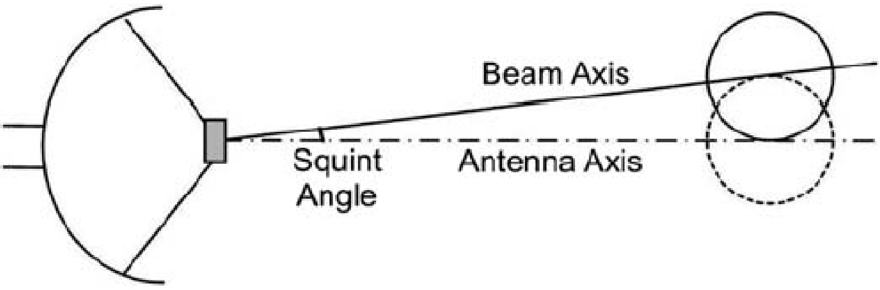


Figure 8.32 Principle of conical scan

8.6.2.4 Monopulse Tracking

One of the major disadvantages of sequential techniques including lobe switching, sequential lobing and conical scan is that the tracking accuracy is severely affected if the cross-section of the object to be tracked changes during the time the beam was being switched or scanned to get the desired number of samples. Monopulse tracking overcomes these shortcomings by generating the required information on the angular error by simultaneous lobing of the received beacon. There are two techniques of monopulse tracking namely amplitude comparison monopulse tracking and phase comparison monopulse tracking. In the case of amplitude comparison monopulse tracking, the antenna uses four feeds placed symmetrically around the focal point. The wavefronts of the received signal in the case of to-be-tracked satellite being on antenna axis and off antenna axis are shown in Figures 8.33(a) and (b) respectively. In the on-axis case, the wavefront will be focused onto a spot on the Figure 8.33 antenna axis as shown in Figure 8.33(a). For off-axis location of satellite, the focus spot will also be off the antenna axis. As a consequence, in the case of satellite being on-axis, the amount of energy falling on the four feeds representing four quadrants (A, B, C and D in Figure 8.34) will be the same.

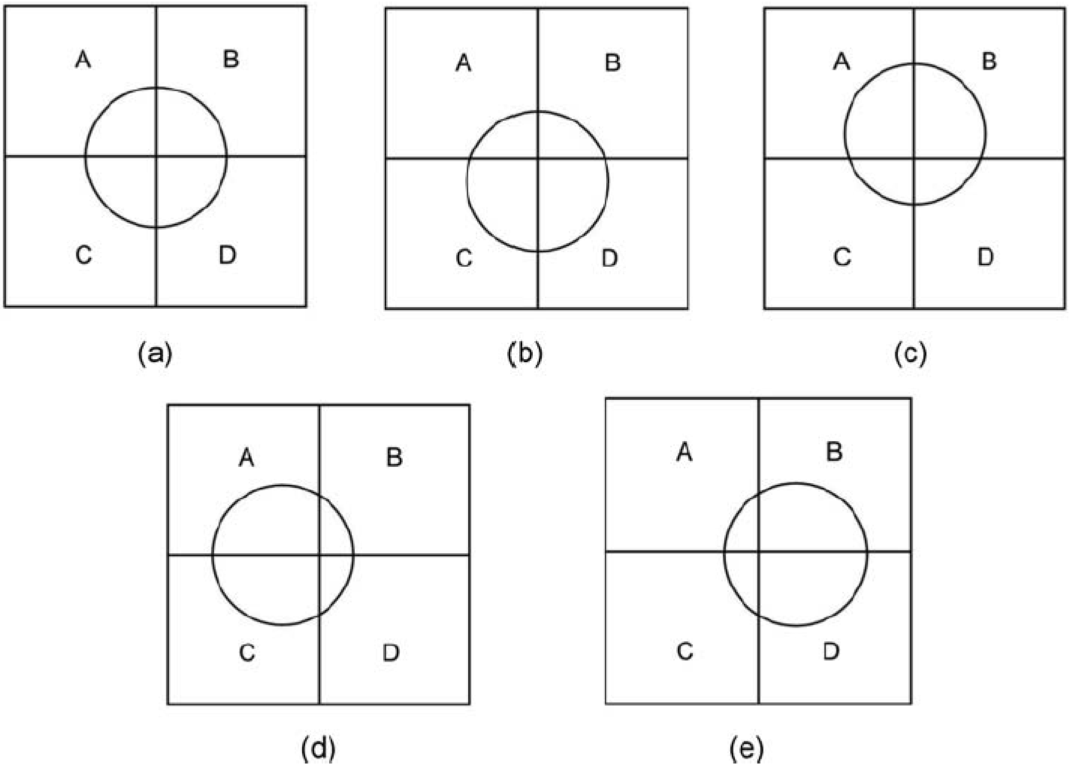
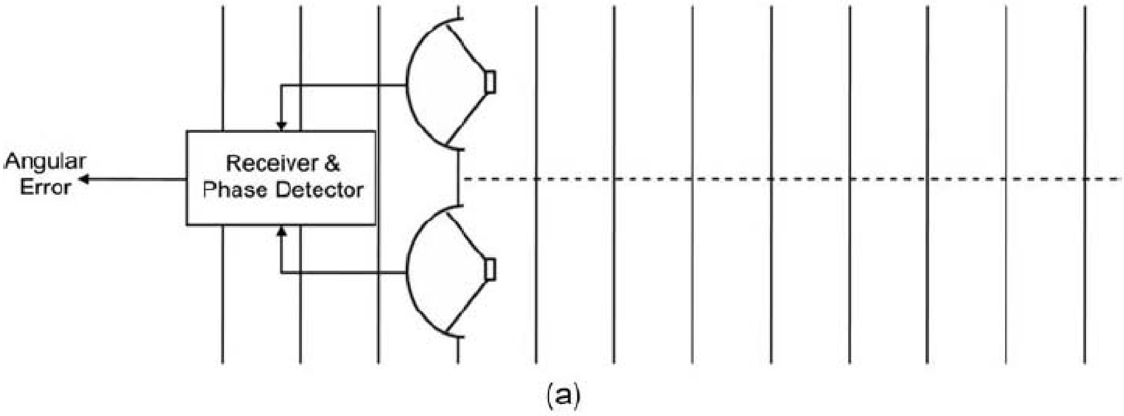


Figure 8.34 Amplitude comparison monopulse tracking – Spot for different angular positions

When the satellite is located off-axis, the amount of energy falling on the four feeds will be different depending upon which quadrant around the antenna axis, the satellite is located. Figures 8.34(a) to (e) show five different cases with satellite onaxis [Figure 8.3(a)], satellite located above antenna axis with same azimuth location [Figure 8.34(b)], satellite located below antenna axis with same azimuth loca-tion [Figure 8.34(c)], satellite located towards right of antenna axis with same elevation location [Figure 8.34(d)] and satellite located towards left of antenna axis with same elevation location [Figure 8.34(e)]. Figure 8.34 Amplitude comparison monopulse tracking – Spot for different angular positions The amplitudes of the received pulse at the output of the four feeds are appropriately processed to determine azimuth and elevation errors required for tracking. In the amplitude comparison monopulse tracking technique, it is important that the signals arriving at different feeds are in phase. This is not a problem when using reflector antennas with feeds that are physically small, usually a few wavelengths across. In the case of arrays, where antenna surface is very large, signals arriving from different off-axis angles present different phases to the different segments into which the array has been divided. These phases need to be equalized before the error signals are derived.

In the case of phase comparison monopulse tracking, it is the phase difference between the received signals in different antenna elements that contains information on angular errors. Atleast two antenna elements are required for both azimuth and elevation error detection. The magnitude and sense of the phase difference determines the magnitude and direction of the off-axis angle. When the satellite is on axis as shown in Figure 8.35(a), the magnitude of phase difference is zero. Figure 8.35(b) depicts the case when the satellite is off-axis.



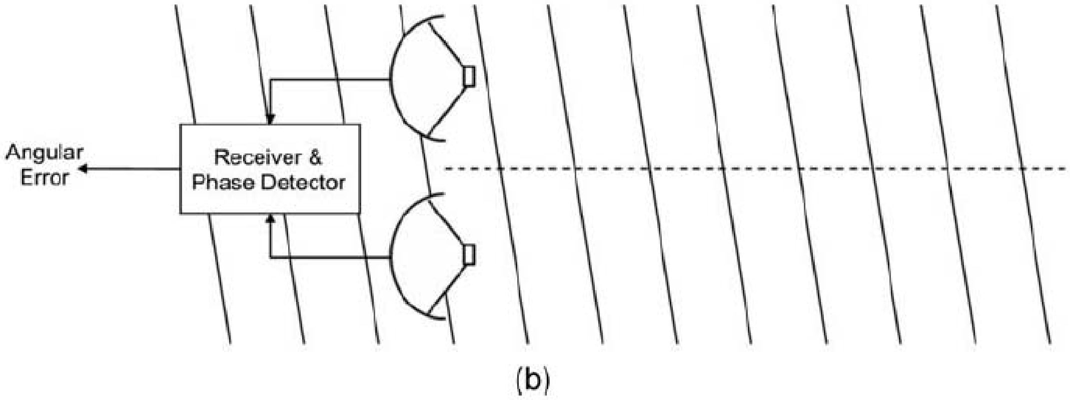


Figure 8.35 Phase comparison monopulse tracking technique

The sensitivity of this technique, i.e. the phase difference produced per unit angular error increases with increase in spacing between different the antenna elements. However, if they were too far apart, an off-axis signal may produce identical phases at the antenna elements. This gives rise to ambiguity. A practical system (Figure 8.36) could have two pairs of elements each for azimuth and elevation with outer pair giving the desired sensitivity and the inner pair resolving ambiguity.

Monopulse tracking technique offers very high tracking accuracy and fast response time. Due to absence of any mechanical parts, the feed system requires very little maintenance. The disadvantages include high cost, large and complex feed system and need to have at least two-channel coherent receivers and good RF phase stability. It is commonly employed in large Earth stations and also in those Earth stations that require accurate tracking of nongeostationary satellites.

8.6.2.5 Step Track

In the case of step track, antenna axes are moved in small incremental steps in an effort to maximize the received signal strength. Amplitude sensing is the basis of this tracking methodology. It is simple and low cost and RF phase stability is not important. It is best suited to small and medium Earth stations. As expected, the technique is susceptible to amplitude perturbations caused by scintillation, signal fading and so on. Tracking accuracy is primarily determined by the step size and signal to noise ratio. For a high signal-to-noise ratio, tracking error approaches the step size. Accuracy is sensitive to amplitude interference.

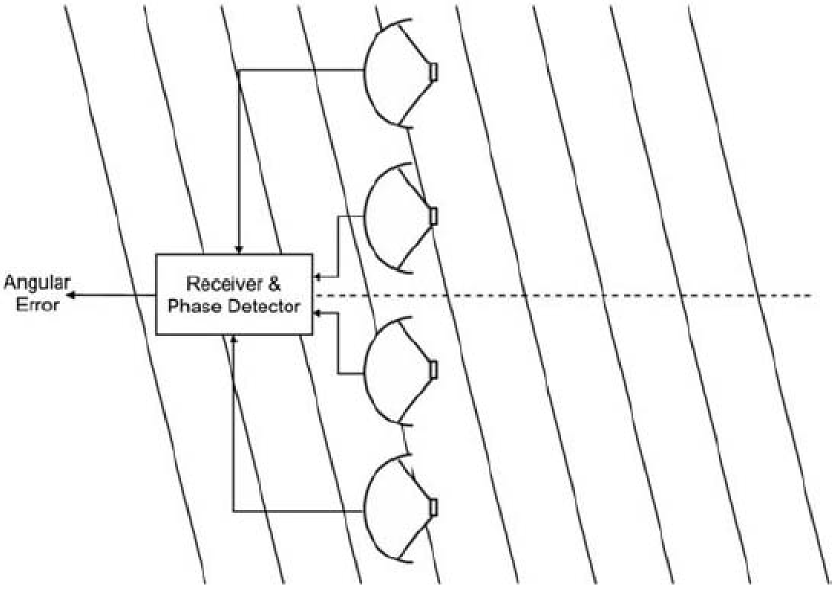


Figure 8.36 System to resolve ambiguity in phase comparison monopulse tracking technique

8.6.2.6 Intelligent Tracking

In the case of intelligent tracking, the satellite position is obtained by optimally combiningantenna position estimate data obtained from a gradient tracking algorithm with the prediction data on satellite position obtained from a satellite position model. In the case of signal amplitude fluctuations, the antenna position may be updated by using prediction data from satellite position model. Intelligent tracking offers all advantages of step track. It is however susceptible to amplitude fluctuations during initial acquisition. Full accuracy is achieved several hours after acquisition. Intelligent track may be used in small, medium and large Earth stations, particularly those susceptible to scintillation and signal fades.

References

[1]Maini, A.K.- Agrawal, V.: Satellite Technology.Principles and Applications.Wiley 2011

[2] Roddy, D.: Satellite Communications, Third Edition, McGraw-Hill TELCOM Engineering, New York, 1989.

[3] Pratt, T.- Bostian, C.W. - Allnutt, J.E.: Satellite Communications, Second Edition, John Wiley & Sons, Inc., New York, 2003.

[4] Ippolito, L.J.: Satellite Communications Systems Engineering. Atmospheric Effects, Satellite Link Design and System Performance, Wiley, Singapore,2008