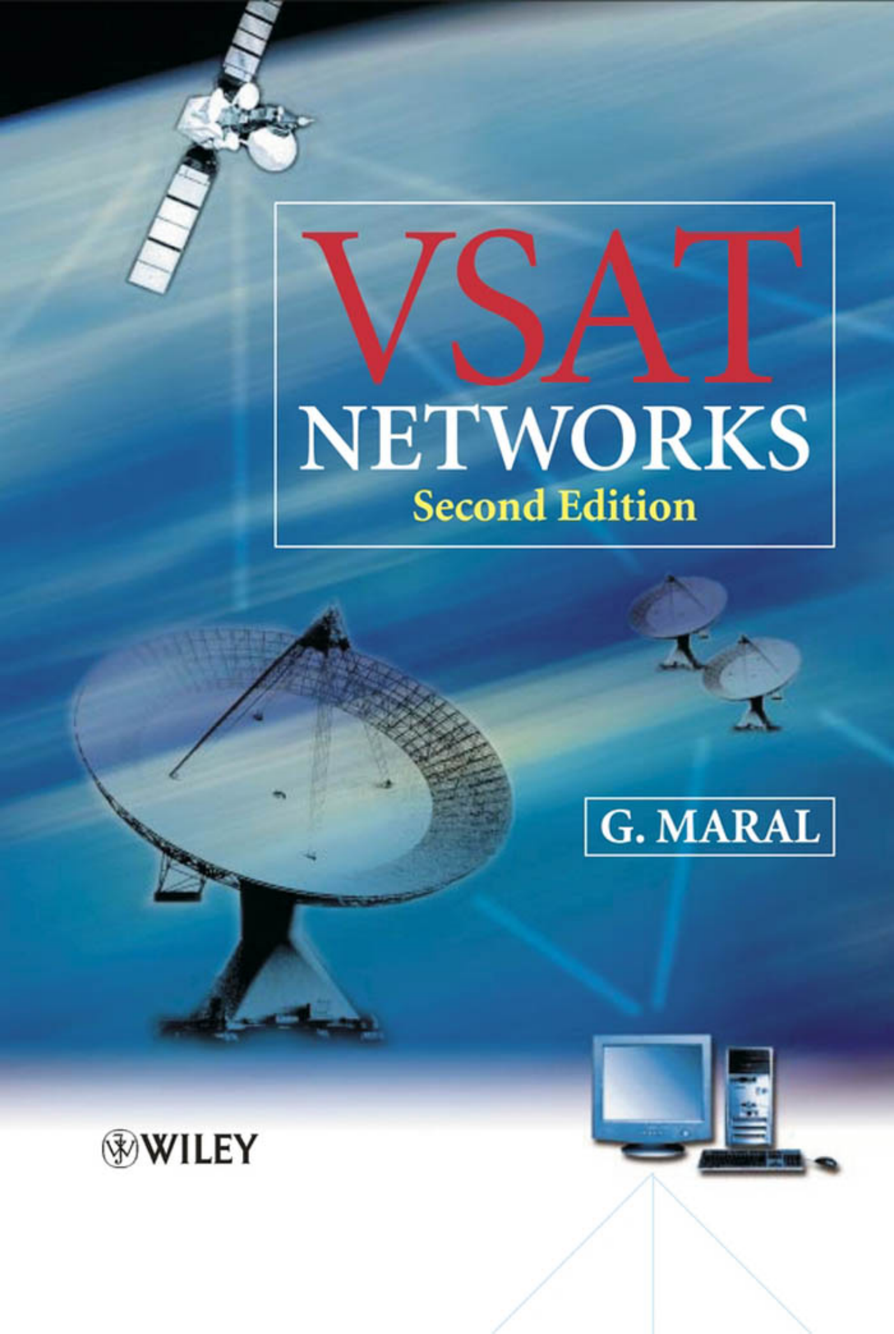


VSAT NETWORKS

Second Edition

G. MARAL

WILEY

VSAT Networks

Second Edition

Gérard Maral

Ecole Nationale Supérieure des Télécommunications,
Site de Toulouse France

John Wiley & Sons, Ltd

VSAT Networks

VSAT Networks

Second Edition

Gérard Maral

Ecole Nationale Supérieure des Télécommunications,
Site de Toulouse France

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Maral, Gérard.
VSAT networks / Gérard Maral. – 2nd ed.
p. cm.
ISBN 0-470-86684-5 (Cloth : alk. paper)
1. VSATs (Telecommunication) I. Title.
TK5104.2.V74 M37 2003
384.5′1 – dc22
2003022021

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-470-86684-5

Typeset in 11/13pt Palatino by Laserwords Private Limited, Chennai, India
Printed and bound in Great Britain by TJ International, Padstow, Cornwall
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

Preface

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 rn, 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 book on VSAT networks aims at introducing the reader to the important issues of services, economics and regulatory aspects. It is also intended to give detailed technical insight on networking and radiofrequency link aspects, therefore addressing the specific features of VSAT networks at the three lower layers of the OSI Reference Layer Model for data communications.

From my experience in teaching, I felt I should proceed from the general to the particular. Therefore, Chapter 1 can be considered as an introduction to the subject, with rather descriptive contents on VSAT network configurations, services, operational and regulatory aspects. The more intrigued reader can then explore the subsequent chapters.

Chapter 2 deals with those aspects of satellite orbit and technology which influence the operation and performance of VSAT networks.

Chapter 3 details the operational aspects which are important to the customer. Installation problems are presented, and a list of potential concerns to the customer is explored. Hopefully, this chapter will not be perceived as discouraging, but on the contrary as a friendly guide for avoiding misfortunes, and getting the best from a VSAT network.

The next two chapters are for technique oriented readers. Actually, I thought this would be a piece of cake for my students, and a reference text for network design engineers.

Chapter 4 deals with networking. It introduces traffic characterisation, and discusses network and link layers protocols of the OSI Reference Layer Model, as used in VSAT networks. It also presents simple analysis tools for the dimensioning of VSAT networks from traffic demand and user specifications in terms of blocking probability and response time.

Chapter 5 covers the physical layer, providing the basic radio frequency link analysis, and presenting the parameters that condition link quality and availability. An important aspect discussed here is interference, as a result of the small size of the VSAT antenna, and its related large beamwidth.

Appendices are provided for the benefit of those readers who may lack some background and have no time or opportunity to refer to other sources.

The second edition of this book takes into account my experience while using the first edition as a support for my lectures. It incorporates some theoretical developments that were missing in the first edition, which constitute useful tools for the dimensioning and the performance evaluation of VSAT networks. In particular, Chapter 4 provides a more detailed treatment on how to evaluate blocking probability and expands on the information transfer delay analysis of the first edition. This second edition also underplays the regulatory aspects, as during the seven year interval between this second edition and the first, many administrations have simplified and harmonised their regulatory framework. I felt this topic was not perhaps as important as it used to be.

I would like to take this opportunity to thank all the students I have taught, at the Ecole Nationale Supérieure des Télécommunications, the University of Surrey, CEI-Europe and other places, who, by raising questions, asking for details and bringing in their comments, have helped me to organise the material presented here.

Gérard Maral, Professor.

Acronyms and Abbreviations

ABCS	Advanced Business Communications via Satellite
ACI	Adjacent Channel Interference
ACK	ACKnowledgement
AMP	Amplifier
ARQ	Automatic repeat ReQuest
ARQ-GB(N)	Automatic repeat ReQuest-Go Back *N*
ARQ-SR	Automatic repeat ReQuest-Selective Repeat
ARQ-SW	Automatic repeat ReQuest-Stop and Wait
ASYNC	ASYNChronous data transfer
BEP	Bit Error Probability
BER	Bit Error Rate
BITE	Built-In Test Equipment
BPF	Band Pass Filter
BPSK	Binary Phase Shift Keying
BSC	Binary Synchronous Communications (bisync)
BSS	Broadcasting Satellite Service
CCI	Co-Channel Interference
CCIR	Comité Consultatif International des Radiocommunications (International Radio Consultative Committee)
CCITT	Comité Consultatif International du Télégraphe et du Téléphone (The International Telegraph and Telephone Consultative Committee)
CCU	Cluster Control Unit
CDMA	Code Division Multiple Access

CFDMA	Combined Free/Demand Assignment Multiple Access
CFRA	Combined Fixed/Reservation Assignment
COST	European COoperation in the field of Scientific and Technical research
DA	Demand Assignment
DAMA	Demand Assignment Multiple Access
dB	deciBel
D/C	Down-Converter
DCE	Data Circuit Terminating Equipment
DEMOD	DEMODulator
DTE	Data Terminal Equipment
DVB-S	Digital Video Broadcasting by Satellite
EIA	Electronic Industries Association
EIRP	Effective Isotropic Radiated Power
$EIRP_{ES}$	Effective Isotropic Radiated Power of earth station (ES)
$EIRP_{SL}$	Effective Isotropic Radiated Power of satellite (SL)
ES	Earth Station
ETR	ETSI Technical Report
ETS	European Telecommunications Standard, created within ETSI
ETSI	European Telecommunications Standards Institute
EUTELSAT	European Telecommunications Satellite Organisation
FA	Fixed Assignment
FCC	Federal Communications Commission, in the USA
FDM	Frequency Division Multiplex
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FET	Field Effect Transistor
FIFO	First In First Out
FODA	FIFO Ordered Demand Assignment
FSK	Frequency Shift Keying
FSS	Fixed Satellite Service
GBN	Go Back *N*
GVF	Global VSAT Forum
HDLC	High level Data Link Control
HEMT	High Electron Mobility Transistor
HPA	High Power Amplifier
IAT	InterArrival Time

IBO	Input Back-Off
IDU	InDoor Unit
IF	Intermediate Frequency
IM	InterModulation
IMUX	Input Multiplexer
IP	Internet Protocol
IPE	Initial Pointing Error
ISDN	Integrated Services Digital Network
ISO	International Organisation for Standardisation
ITU	International Telecommunication Union
LAN	Local Area Network
LAP	Link Access Protocol
LNA	Low Noise Amplifier
LO	Local Oscillator
MAC	Medium Access Control
MCPC	Multiple Channels Per Carrier
MIFR	Master International Frequency Register
MOD	MODulator
MTBF	Mean Time Between Failures
MUX	MUltipleXer
MX	MiXer
NACK	Negative ACKnowledgement
NMS	Network Management System
OBO	Output Back-Off
ODU	OutDoor Unit
OMUX	Output MUltipleXer
OSI	Open System Interconnection
PABX	Private Automatic Branch eXchange
PAD	Packet Assembler/Disassembler
PBX	Private (automatic) Branch eXchange
PC	Personal Computer
PDF	Probability Density Function
PDU	Protocol Data Unit
POL	POLarisation
PSD	Power Spectral Density
PSK	Phase Shift Keying
QPSK	Quaternary Phase Shift Keying
RCVO	ReCeiVe-Only
Rec	Recommendation
Rep	Report
RF	Radio Frequency
RX	Receiver
S-ALOHA	Slotted ALOHA protocol
SCADA	Supervisory Control and Data Acquisition

SCPC	Single Channel Per Carrier
SDLC	Synchronous Data Link Control
SKW	Satellite-Keeping Window
SL	SateLlite
SNA	Systems Network Architecture (IBM)
SNG	Satellite News Gathering
SR	Selective Repeat
SSPA	Solid State Power Amplifier
SW	Stop and Wait
TCP	Transmission Control Protocol
TDM	Time Division Multiplex
TDMA	Time Division Multiple Access
TTC	Telemetry, Tracking and Command
TV	TeleVision
TWT	Travelling Wave Tube
TX	Transmitter
VSAT	Very Small Aperture Terminal
XPD	Cross Polarisation Discrimination
XPI	Cross Polarisation Isolation

Notation

A	attenuation (larger than one in absolute value, therefore positive value in dB), also length of acknowledgement frame (bits)
A_{RAIN}	attenuation due to rain
Az	azimuth angle (degree)
a	semi-major axis (m)
B	bandwidth (Hz)
B_{i}	interfering carrier bandwidth (Hz)
B_{inb}	inbound carrier bandwidth (Hz)
B_{N}	receiver equivalent noise bandwidth (Hz)
B_{outb}	outbound carrier bandwidth (Hz)
B_{Xpond}	transponder bandwidth (Hz)
BU	burstiness
c	speed of light: $c = 3 \times 10^8 \mathrm{m/s}$
C	carrier power (W)
C_{D}	carrier power at earth station receiver input (W)
C_{U}	carrier power at satellite transponder input (W)
C_x	received carrier power on X-polarisation (W)
C_y	received carrier power on Y-polarisation (W)
C/N	carrier to noise power ratio
$(C/N)_{\mathrm{D}}$	downlink carrier to noise power ratio
$(C/N)_{\mathrm{Dsat}}$	same as above, at saturation
$(C/N)_{\mathrm{IM}}$	carrier to intermodulation noise power ratio (Hz)
$(C/N)_{\mathrm{U}}$	uplink carrier power to noise power ratio
$(C/N)_{\mathrm{Usat}}$	same as above, at saturation
$(C/N)_{\mathrm{T}}$	overall link (from station to station) carrier to total noise power ratio
C/N_{i}	carrier to interference power ratio
$(C/N_{\mathrm{i}})_{\mathrm{D}}$	downlink carrier to interference power ratio
$(C/N_{\mathrm{i}})_{\mathrm{U}}$	uplink carrier to interference power ratio
$(C/N_{\mathrm{i}})_{\mathrm{T}}$	overall link (from station to station) carrier to interference power ratio
C/N_0	carrier power to noise power spectral density ratio (Hz)
$(C/N_0)_{\mathrm{D}}$	downlink carrier power to noise power spectral density ratio (Hz)
$(C/N_0)_{\mathrm{Dsat}}$	same as above, at saturation (Hz)
$(C/N_0)_{\mathrm{IM}}$	carrier power to intermodulation noise power spectral density ratio (Hz)
$(C/N_0)_{\mathrm{U}}$	uplink carrier power to noise power spectral density ratio (Hz)
$(C/N_0)_{\mathrm{Usat}}$	same as above, at saturation (Hz)

$(C/N_0)_T$	overall link (from station to station) carrier power to total noise power spectral density ratio (W/Hz)
C/N_{0i}	carrier power to interference noise power spectral density ratio (Hz)
$(C/N_{0i})_D$	downlink carrier power to interference noise power spectral density ratio (Hz)
$(C/N_{0i})_U$	uplink carrier power to interference noise power spectral density ratio (Hz)
$(C/N_{0i})_T$	overall link (from station to station) carrier power to total interference noise power spectral density ratio (W/Hz)
D	antenna diameter (m), also number of data bits per frame to be conveyed from source to destination
dBx	value in dB relative to x
E	elevation angle (degree), also energy per bit (J)
E_b	energy per information bit (J)
E_c	energy per channel bit (J)
e	eccentricity
EIRP	equivalent isotropic radiated power of transmitting equipment (W)
$\mathrm{EIRP}_{\mathrm{ES}}$	EIRP of earth station (W)
$\mathrm{EIRP}_{\mathrm{ESmax}}$	maximum value of $\mathrm{EIRP}_{\mathrm{ES}}$ (W)
$\mathrm{EIRP}_{\mathrm{ESsat}}$	value of $\mathrm{EIRP}_{\mathrm{ES}}$, at transponder saturation (W)
$\mathrm{EIRP}_{\mathrm{ESi}}$	EIRP of interfering earth station (W)
$\mathrm{EIRP}_{\mathrm{ESi,max}}$	maximum value of earth station EIRP allocated to interfering carrier (W)
$\mathrm{EIRP}_{\mathrm{ESw}}$	EIRP of wanted earth station (W)
$\mathrm{EIRP}_{\mathrm{SL}}$	EIRP of satellite transponder (W)
$\mathrm{EIRP}_{\mathrm{SLsat}}$	EIRP of satellite transponder at saturation (W)
$\mathrm{EIRP}_{\mathrm{SL1sat}}$	EIRP of satellite transponder in beam 1 at saturation (W)
$\mathrm{EIRP}_{\mathrm{SL2sat}}$	EIRP of satellite transponder in beam 2 at saturation (W)
$\mathrm{EIRP}_{\mathrm{SLi,max}}$	maximum value of interfering satellite EIRP allocated to interfering carrier (W)
$\mathrm{EIRP}_{\mathrm{SLw,max}}$	maximum value of wanted satellite EIRP for wanted carrier (W)
$\mathrm{EIRP}_{\mathrm{SLww}}$	wanted satellite EIRP for wanted carrier in direction of wanted station (W)
$\mathrm{EIRP}_{\mathrm{SLiw}}$	interfering satellite EIRP for interfering carrier in direction of wanted station (W)
$\mathrm{EIRP}_{\mathrm{SL1ww}}$	EIRP of satellite transponder in beam 1 for wanted carrier in direction of wanted station (W)
$\mathrm{EIRP}_{\mathrm{SL2iw}}$	EIRP of satellite transponder in beam 2 for interfering carrier in direction of wanted station (W)
$\mathrm{EIRP}_{\mathrm{SL1wsat}}$	EIRP of satellite transponder in beam 1 in direction of wanted station at saturation (W)
$\mathrm{EIRP}_{\mathrm{SL2wsat}}$	EIRP of satellite transponder in beam 2 in direction of wanted station at saturation (W)
f	frequency (Hz): $f = c/\lambda$
f_D	downlink frequency (Hz)
f_{IM}	frequency of an intermodulation product (Hz)
f_{LO}	local oscillator frequency (Hz)
f_U	uplink frequency (Hz)
G	power gain (larger than one in absolute value, therefore positive value in dB), also normalised offered traffic, also gravitational constant: $G = 6.672 \times 10^{-11}$ m^3/kg s^2

G_{cod} coding gain (dB)
G_D power gain from transponder output to earth station receiver input
G_{IF} intermediate frequency amplifier power gain
G_{LNA} low noise amplifier power gain
G_{max} maximum gain
G_{MX} mixer power gain
G_R antenna receive gain in direction of transmitting equipment
G_{Rmax} antenna receive gain at boresight
G_{RX} receiving equipment composite receive gain: $G_{RX} = G_{Rmax}/L_R L_{pol} L_{FRX}$
$G_{RX\,max}$ maximum value of G_{RX}
G_{RXi} receiving equipment composite receive gain for interfering carrier
G_{RXw} receiving equipment composite receive gain for wanted carrier
G_T antenna transmit gain in direction of receiving equipment
G_{Tmax} antenna transmit gain at boresight
$G_{Ti,max}$ antenna transmit gain at boresight for interfering carrier
G_{T1w} satellite beam 1 transmit antenna gain in direction of wanted station
G_{T2w} satellite beam 2 transmit antenna gain in direction of wanted station
G_{TE} power gain from satellite transponder input to earth station receiver input
G_{Xpond} transponder power gain
G_1 gain of an ideal antenna with area equal to $1\ m^2 : G_1 = 4\pi/\lambda^2$
G/T figure of merit of receiving equipment (K^{-1})
$(G/T)_{ES}$ figure of merit of earth station receiving equipment (K^{-1})
$(G/T)_{ESmax}$ maximum value of $(G/T)_{ES}$
$(G/T)_{SL}$ figure of merit of satellite receiving equipment (K^{-1})

H total number of bits in the frame header (and trailer if any)

i orbit inclination
I_x received cross polar interference on X-polarisation (W)
IBO input back-off
IBO_{inb} input back-off for inbound carrier
IBO_{outb} input back-off for outbound carrier
IBO_1 input back-off per carrier with multicarrier operation mode
IBO_t total input back-off with multicarrier operation mode

J_x cross polar interference on X-polarisation generated by receive antenna (W)

k Boltzmann constant: $k = 1.38 \times 10^{-23}$J/K; k(dBJ/K) $= 10 \log k = -228.6$ dBJ/K

l Earth station latitude with respect to the satellite latitude (degrees)
L loss (larger than one in absolute value, therefore positive value in dB), also earth station relative longitude with respect to a geostationary satellite (degrees), also length of a frame (bits), also length of a message (bits)
L_a Earth station relative longitude with respect to the adjacent satellite (degrees)
L_w Earth station relative longitude with respect to the wanted satellite (degrees)
L_D downlink path loss
L_{FRX} feeder loss from antenna to receiver input
L_{FTX} feeder loss from transmitter output to antenna

L_{pol}	antenna gain loss as a result of antenna polarisation mismatch
L_R	off-axis receive gain loss
$L_{R\,max}$	maximum value of L_R
L_U	uplink path loss
L_{Ui}	Uplink path loss for interfering carrier
L_{Uw}	Uplink path loss for wanted carrier
M_e	mass of the Earth: $M_e = 5.974 \times 10^{24}$kg
N	noise power (W)
N_i	interference power (W)
N_{IM}	intermodulation noise power (W)
N_{0D}	downlink thermal noise power spectral density (W/Hz)
N_{0U}	uplink thermal noise power spectral density (W/Hz)
N_{0iD}	downlink interference power spectral density (W/Hz)
N_{0IM}	intermodulation noise power spectral density (W/Hz)
N_{0iU}	uplink interference power spectral density (W/Hz)
N_{0T}	total noise power spectral density at the earth station receiver input (W/Hz)
OBO	output back-off
OBO_1	output back-off per carrier with multicarrier operation mode
OBO_i	output back-off for interfering carrier
OBO_t	total output back-off with multicarrier operation mode
OBO_w	output back-off for wanted carrier
P	power (W)
P_f	probability for a frame to be in error
P_R	received power at antenna output (W)
P_T	power fed to transmitting antenna (W)
P_{TX}	transmitter output power (W)
$P_{TX\,max}$	transmitter output power at saturation (W)
P_x	transmitted carrier power on X-polarisation (W)
P_y	transmitted carrier power on Y-polarisation (W)
PSD	power spectral density (W/Hz)
PSD_i	interfering carrier power spectral density (W/Hz)
PSD_w	wanted carrier power spectral density (W/Hz)
Q_x	cross polar interference on X-polarisation generated by transmit antenna (W)
r	distance from centre of earth to satellite
R	range, also bit rate
R_a	slant range from earth station to adjacent satellite
R_b	information bit rate (b/s)
R_{binb}	information bit rate on inbound carrier (b/s)
R_{boutb}	information bit rate on outbound carrier (b/s)
R_c	transmission bit rate (b/s)
R_{cinb}	transmission bit rate on inbound carrier (b/s)
R_{coutb}	transmission bit rate on outbound carrier (b/s)
R_e	earth radius: $R_e = 6378$ km
R_0	geostationary satellite altitude: $R_0 = 35786$ km
R_w	slant range from earth station to wanted satellite
S	normalised throughput
SKW	satellite station keeping window halfwidth (degrees)
T	interval of time (s), also period of orbit (s), also medium temperature (K) and noise temperature (K)
T_A	antenna noise temperature (K)
T_D	downlink system noise temperature (K)
$T_{D\,min}$	minimum value of T_D (K)
T_F	feeder temperature (K)
T_{GROUND}	ground noise temperature in vicinity of earth station (K)

T_{IF}	intermediate frequency amplifier effective input noise temperature (K)
T_{LNA}	low noise amplifier effective input noise temperature (K)
T_m	average medium temperature (K)
T_{MX}	mixer effective input noise temperature (K)
T_p	propagation time (s)
T_R	receiver effective input noise temperature (K)
T_{SKY}	clear sky noise temperature (K)
T_U	uplink system noise temperature (K)
THRU	throughput (b/s)
W	window size
X	order of an intermodulation product
XPD	cross polar discrimination
XPI_{RX}	receive antenna cross polarisation isolation
XPI_{TX}	transmit antenna cross polarisation isolation
α	angular separation between two satellites (degrees)
Γ	spectral efficiency (b/s Hz)
Δ	ratio of co-polar wanted carrier power to cross-polar interfering carrier power
η	efficiency
η_a	antenna efficiency (typically 0.6)
η_c	channel efficiency
η_{cGBN}	channel efficiency with go-back-N protocol
η_{cSR}	channel efficiency with selective-repeat protocol
η_{cSW}	channel efficiency with stop-and-wait protocol
θ	angle from boresigth of antenna (degrees)
θ_{3dB}	half power beamwidth of an antenna (degrees)
θ_R	antenna off-axis of angle for reception (degrees)
$\theta_{R\,max}$	maximum value of antenna off-axis angle for reception (degrees)
θ_T	antenna off-axis angle for transmission (degrees)
θ_{Tmax}	maximum value of antenna off-axis angle for transmission (degrees)
λ	wavelength (m) $= c/f$, also traffic generation rate (s^{-1})
μ	product of gravitational constant G and mass of the Earth $M_e : \mu = 3.986 \times 10^{14} m^3/s^2$
ρ	code rate
τ	packet duration (s)
Φ	power flux density (W/m^2)
Φ_{sat}	power flux density at saturation (W/m^2)
Φ_t	total flux density (W/m^2)

1

Introduction

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 regulatory aspects. It can also be considered by the reader as a guide to the following chapters which aim to provide more details on specific issues.

1.1 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 book, 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).

VSAT Networks, 2nd Edition. G. Maral
 ISBN: 0-470-86684-5

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 hand held terminals for radiolocation such as the Global Postioning System (GPS) receivers. Present day hand held satellite phones (IRIDIUM, GLOBALSTAR) are pocket size. Figure 1.1 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 1.2 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.

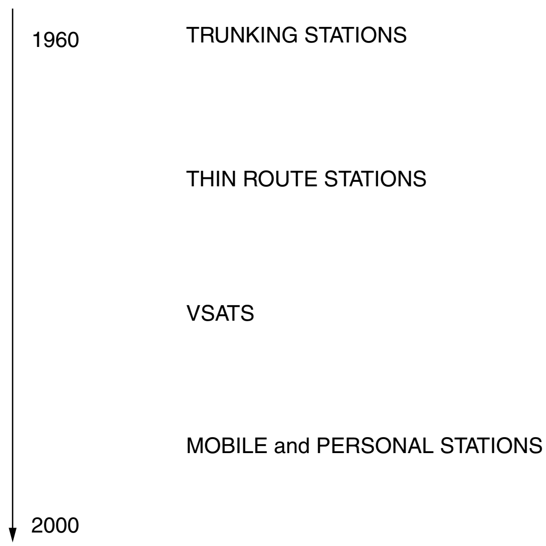


Figure 1.1 VSAT: a step towards earth station size reduction

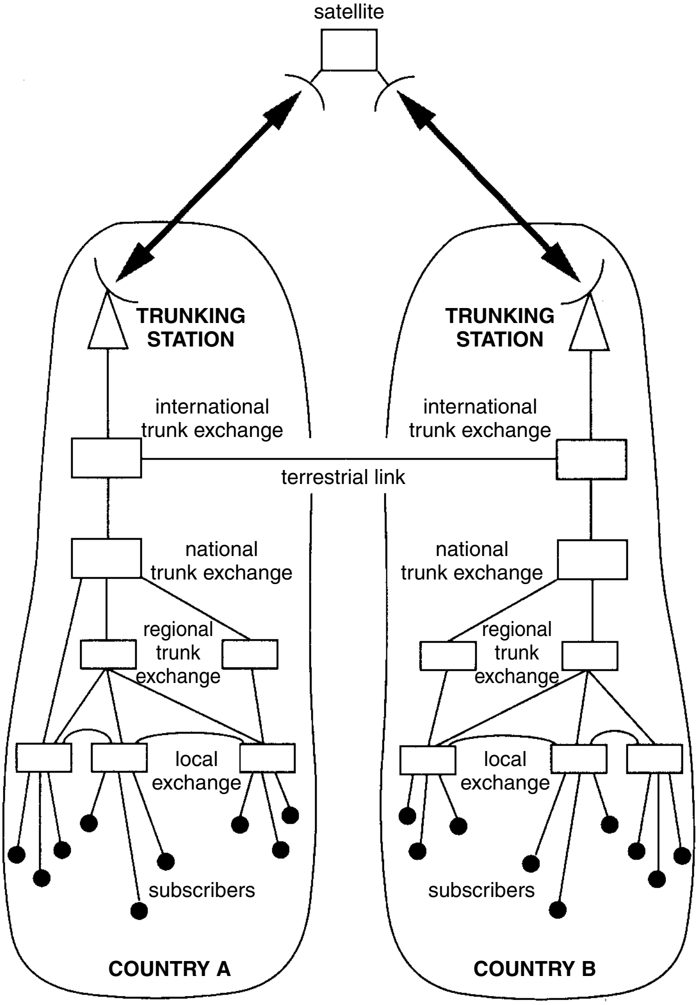


Figure 1.2 Trunking stations

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 of a few tens of kbs^{-1}, typically 56 or 64 kbs^{-1}.

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 1.3 illustrates this aspect by

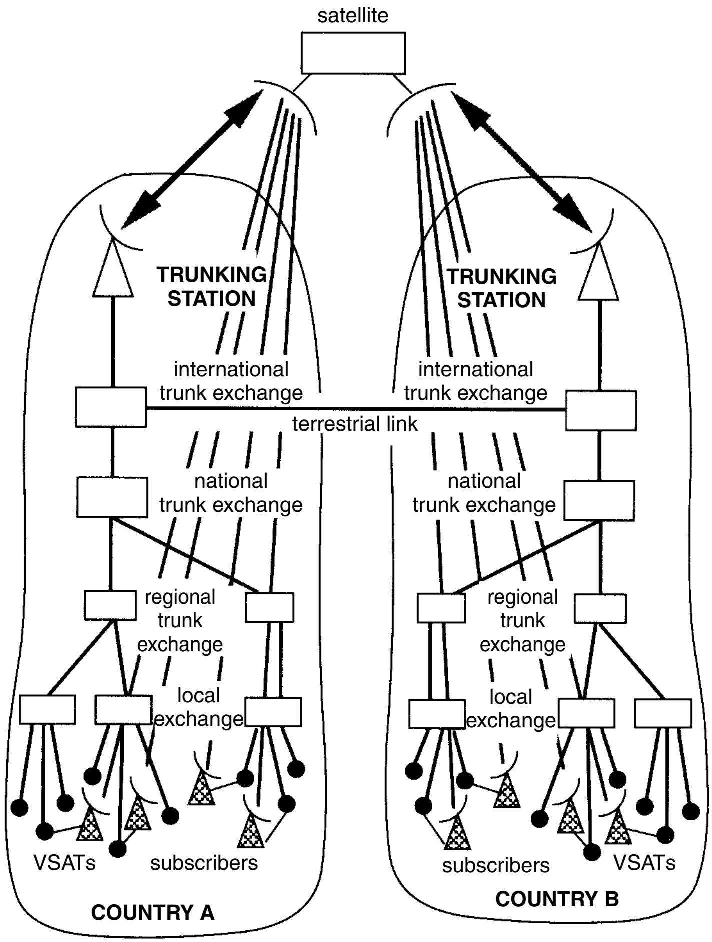


Figure 1.3 From trunking stations to VSATs

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.

1.2 VSAT NETWORK CONFIGURATIONS

As illustrated in Figure 1.3, 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 1.4). 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. A more detailed description of the satellite architecture is given in Chapter 2 (section 2.1).

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 will be shown in Chapter 2

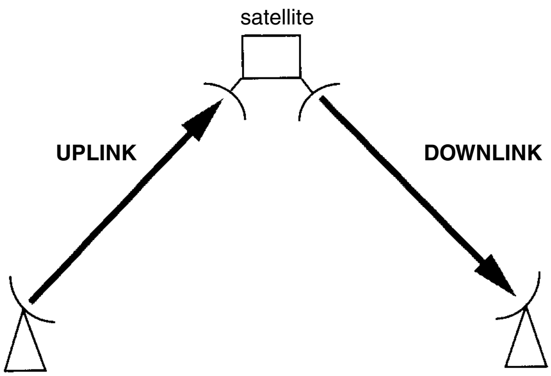


Figure 1.4 Definition of uplink and downlink

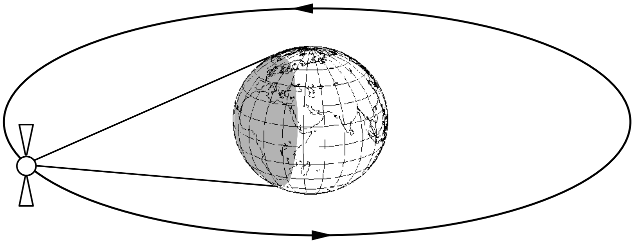


Figure 1.5 Geostationary satellite

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 1.5 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 (see Chapter 2, section 2.3).

As a result of its apparent fixed position in the sky, the satellite can be used 24 hours a day as a permanent relay for the uplinked radio frequency carriers. Those carriers are downlinked to all earth stations visible from the satellite (shaded area on the earth in Figure 1.5). Thanks to its apparent fixed position in the sky, there is no need for tracking the satellite. This simplifies VSAT equipment and installation.

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 1.6.

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 geostationary 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.

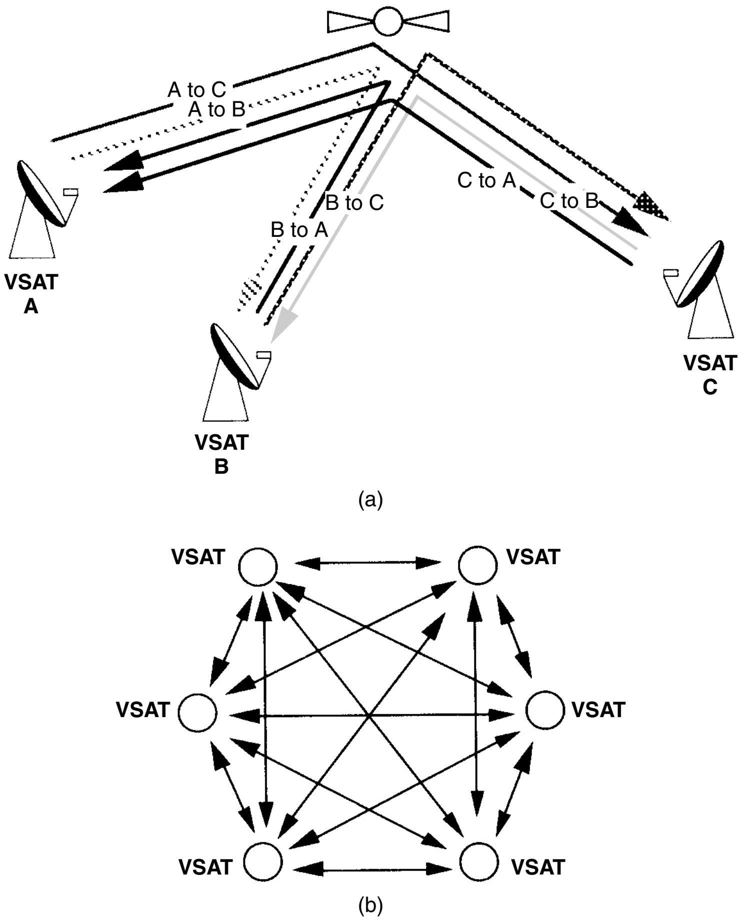


Figure 1.6 Meshed VSAT network. (a) Example with three VSATs (arrows represent information flow as conveyed by the carriers relayed by the satellite); (b) simplified representation for a larger number of VSATs (arrows represent bidirectional links made of two carriers travelling in opposite directions)

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 1.7 and 1.8. 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

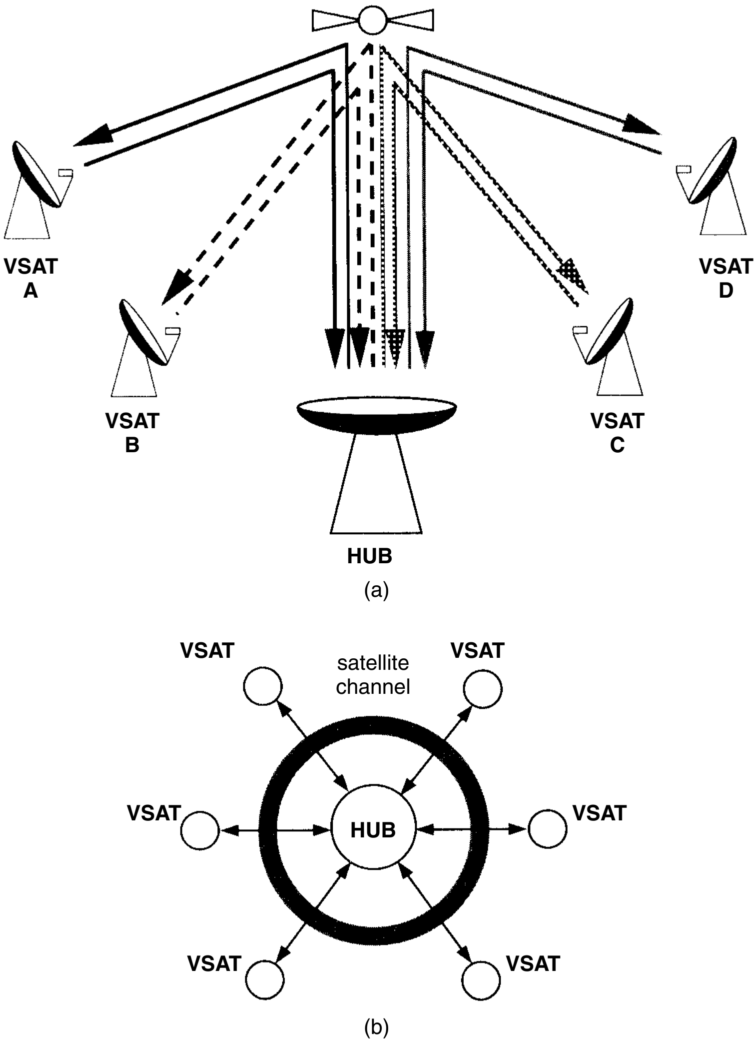


Figure 1.7 Two-way star-shaped VSAT network. (a) Example with four VSATs (arrows represent information flow as conveyed by the carriers relayed by the satellite); (b) simplified representation for a larger number of VSATs (arrows represent bidirectional links made of two carriers travelling in opposite directions)

consist of two links, uplink and downlink, to and from the satellite, as illustrated in Figure 1.4.

There are two types of star-shaped VSAT network:

- *two-way networks* (Figure 1.7), where VSATs can transmit and receive. Such networks support interactive traffic;
- *one-way networks* (Figure 1.8), where the hub transmits carriers to receive-only VSATs. This configuration supports broadcasting

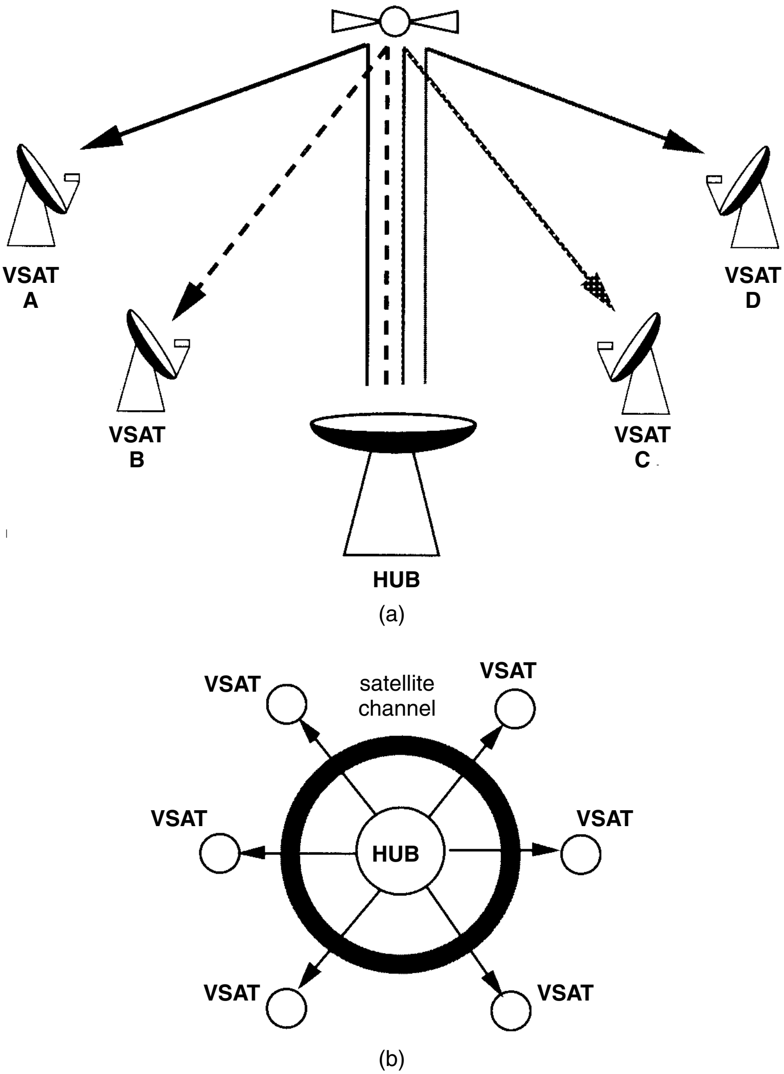


Figure 1.8 One-way star-shaped VSAT network. (a) Example with four VSATs (arrows represent information flow as conveyed by the outbound carriers relayed by the satellite); (b) simplified representation for a larger number of VSATs (arrows represent unidirectionnal links)

services from a central site where the hub is located to remote sites where the receive-only VSATs are installed.

1.3 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 1.6. The user terminal connectivity is illustrated in Figure 1.9;
– 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 1.10).

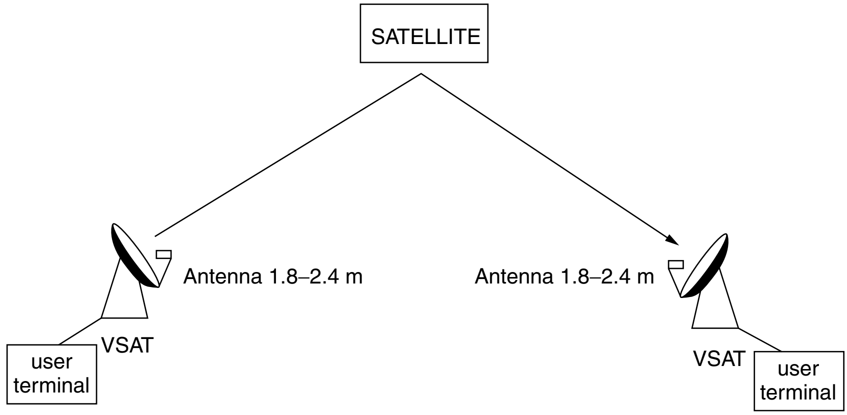


Figure 1.9 User terminal connectivity within meshed VSAT networks

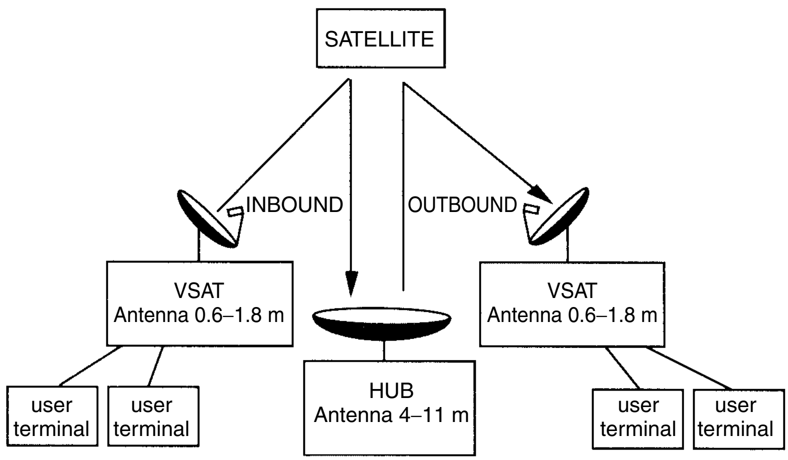


Figure 1.10 User terminal connectivity using the hub as a relay in star-shaped networks

Comparing Figure 1.9 and 1.10 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.

1.4 VSAT NETWORK APPLICATIONS AND TYPES OF TRAFFIC

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

1.4.1 Civilian VSAT networks

1.4.1.1 Types of service

As mentioned in the previous section, VSAT networks can be configured as one-way or two-way networks. Table 1.1 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 a 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 a location where news reporters can transmit video signals to a hub

Table 1.1 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

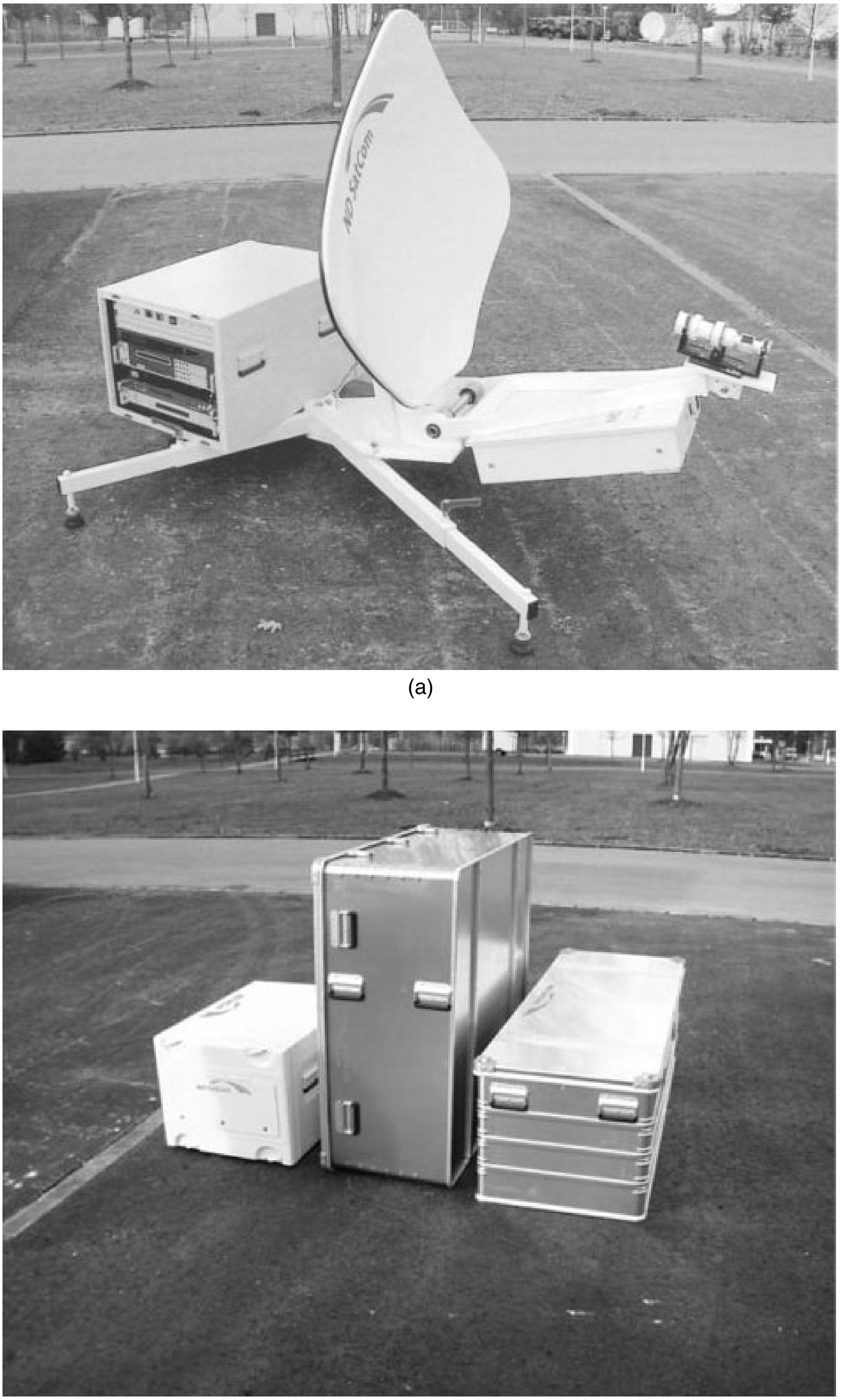
located near the company's studio. Of course the service could be considered as inbound only, if it were not for the need to check the uplink from the remote site, and to be in touch by telephone with the staff at the studio. As fly-away VSATs are constantly transported, assembled and disassembled, they must be robust, lightweight and easy to install. Today they weigh typically 100 kg and can be installed in less than 20 minutes. Figure 1.11 shows a picture of a fly-away VSAT station.

1.4.1.2 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.



(a)

(b)

Figure 1.11 'Fly-away' VSAT station. (a) In operation; (b) packed for transportation. (Reproduced by permission of ND Satcom)

Table 1.2 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

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. Examples of applications are control and monitoring of pipelines, offshore platforms, electric utilities and water resources.

Table 1.2 summarises the above discussion.

1.4.2 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 [EVA99, Chapter 23].

1.5 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 technology, as mentioned in section 1.1, cost savings and reliability, as will be discussed in section 3.3.

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.
– 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

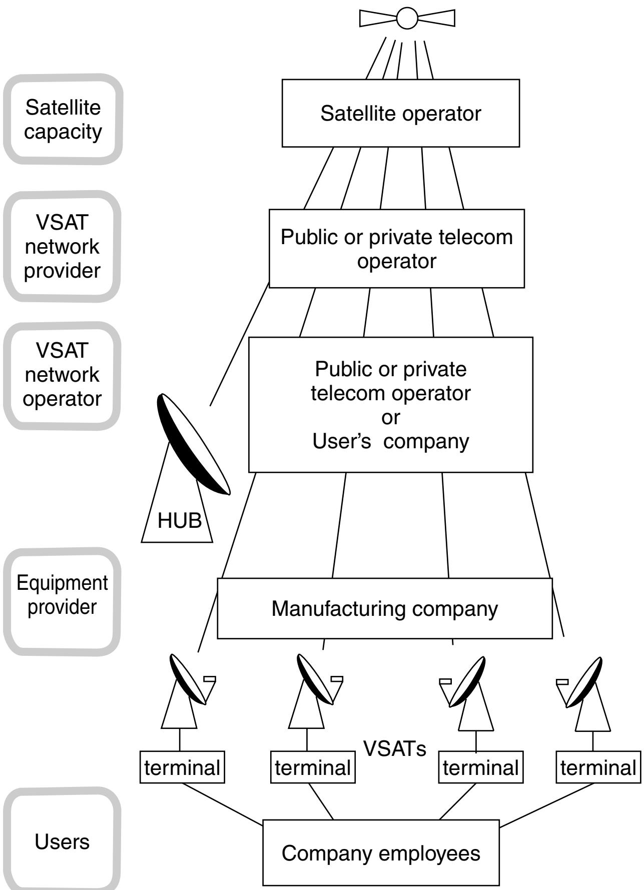


Figure 1.12 VSAT networks: involved parties

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.

The above parties are those involved in the contractual matters. Other parties are on the regulatory side and their involvement will be first presented in section 1.9.

Figure 1.12 summarises the above discussion. The terminology will be used throughout the book and therefore Figure 1.12 can serve as a convenient reference.

1.6 VSAT NETWORK OPTIONS

1.6.1 Star or mesh?

Section 1.2 introduced the two main architectures of a VSAT network: star and mesh. The question now is: 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.

1.6.1.1 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 1.3 summarises the above discussion.

Regulatory aspects are also to be taken into account (see section 1.9).

1.6.1.2 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 1.4), a double hop link would comprise two single hop links, one being inbound and the other outbound (Figure 1.10).

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 1.13). 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

Table 1.3 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)

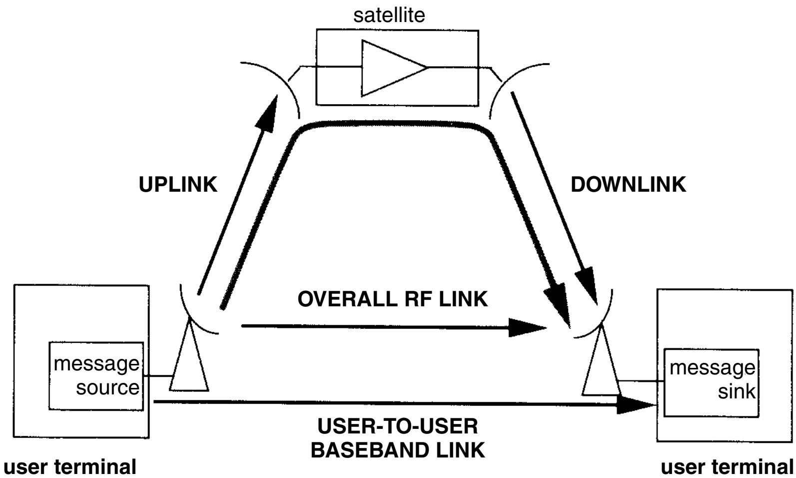


Figure 1.13 Overall radio frequency (RF) link and user-to-user baseband link

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).

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 [MAR02 Chapter 5].

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 5, section 5.7, 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*

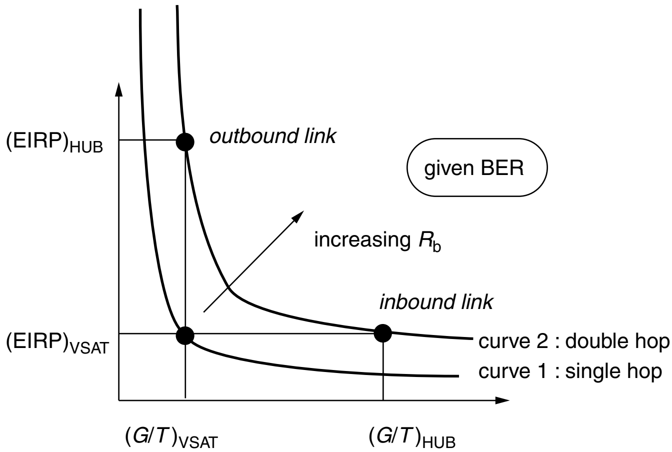


Figure 1.14 EIRP versus G/T in a VSAT network. Curve 1: single hop from VSAT to VSAT in a meshed network; Curve 2: double hop from VSAT to VSAT via the hub. Increased R_b means increased link capacity

rate R_b (bs^{-1}):

$$\frac{E_b}{N_0} = \frac{(C/N_0)_T}{R_b} \tag{1.1}$$

Figure 1.14 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 (see Chapter 5 for definition of the EIRP and of the figure of merit).

As can be seen from Figure 1.14, 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.

1.6.1.3 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. However it is not a severe problem for video or data transmission.

Table 1.4 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

Table 1.4 Characteristics of star and mesh network configurations

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

a lower capacity, by implementing either a star-shaped network, or a meshed one.

1.6.2 Data/voice/video

Depending on his needs, the customer may want to transmit either one kind of signal or a mix of different signals. Data and voice are transmitted in a digital format, while video may be analogue or digital. When digital, the video signal may benefit from bandwidth efficient compression techniques.

1.6.2.1 Data communications

VSATs have emerged from the need to transmit data. Standard VSAT products offer data transmission facilities. Rates offered to the user range typically from 50 bs^{-1} to 64 kbs^{-1} with interface ports such as RS-232, for bit rates lower than 20 kbs^{-1}, RS-422, V35 and X21 for higher bit rates. A local area network (LAN) interface is most often provided (using an RJ-45 connector, for instance). Appendix 3 gives some details on the functions of such ports.

Data distribution can be implemented in combination with video transmission, using for instance the DVB-S standard.

1.6.2.2 Voice communications

Voice communications are of interest on two-way networks only. They can be performed at low rate using voice encoding (vocoders). Typical information rates then range from 4.8 kbs^{-1} to 9.6 kbs^{-1}. They can also be combined with data communications (for instance up to 4 voice channels may be multiplexed with data or facsimile channels on a single 64 kbs^{-1} channel).

On VSAT networks voice communications suffer from delay associated with vocoder processing (about 50 ms) and propagation on satellite links (about 500 ms for a double hop). Therefore the user may prefer to connect to terrestrial networks which offer a reduced delay. Voice communications can be a niche market for VSATs as a service to locations where land lines are not available, or for transportable terminal applications.

1.6.2.3 Video communications

On the outbound link (from hub to VSAT)

Video communications make use of the usual TV standards (NTSC, PAL or SECAM) in combination with FM modulation, or can be implemented using the Digital Video Broadcasting by Satellite (DVB-S) standard, possibly in combination with distribution of data.

On the inbound link

As a result of the limited power of the VSAT on the uplink, video transmission is feasible at a low rate, possibly in the form of slow motion image transmission using video coding and compression.

1.6.3 Fixed/demand assignment

The earth stations of a VSAT network communicate via the satellite by means of modulated carriers. Any such carrier is assigned a portion of the resource offered by the satellite in terms of powered bandwidth. This assignment can be defined once for all, and this is called '*fixed assignment*' (FA), or in accordance with requests from the VSATs depending on the traffic they have to transmit, and this is called '*demand assignment*' (DA).

1.6.3.1 Fixed assignment (FA)

Figure 1.15 illustrates the principle of *fixed assignment*. A star-shaped network configuration is considered in the figure but the principle applies to a meshed network configuration as well. The satellite resource is shared in a fixed manner by all stations whatever the traffic demand. It may be that at a given instant the VSAT traffic load is larger than that which can be accommodated by capacity allocated to that VSAT as determined by its share of the satellite resource. The VSAT must store or reject the traffic demand, and this either increases the delay or introduces blocking of calls, in spite

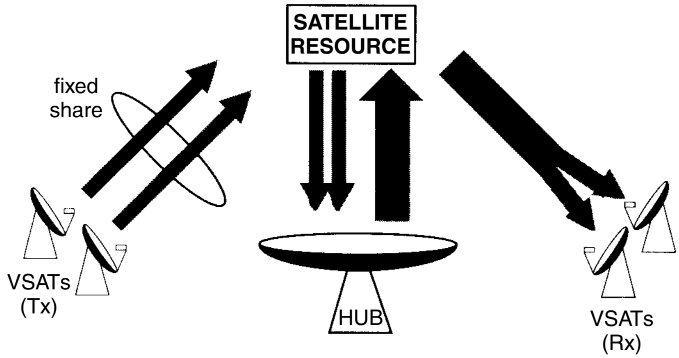


Figure 1.15 Principle of fixed assignment

of the fact that other VSATs may have excess capacity available. Because of this, the network is not optimally exploited.

1.6.3.2 Demand assignment (DA)

With *demand assignment*, VSATs share a variable portion of the overall satellite resource as illustrated in Figure 1.16. VSATs use only the capacity which is required for their own transmission, and leave the capacity in excess for use by other VSATs. Of course this variable share can be exercised only within the limits of the total satellite capacity allocated to the network.

Demand assignment is performed by means of requests for capacity transmitted by individual VSATs. Those requests are transmitted to the hub station, or to a traffic control station, should the management of the demand assignment technique be centralised, or to all other VSATs, if the demand assignment is distributed.

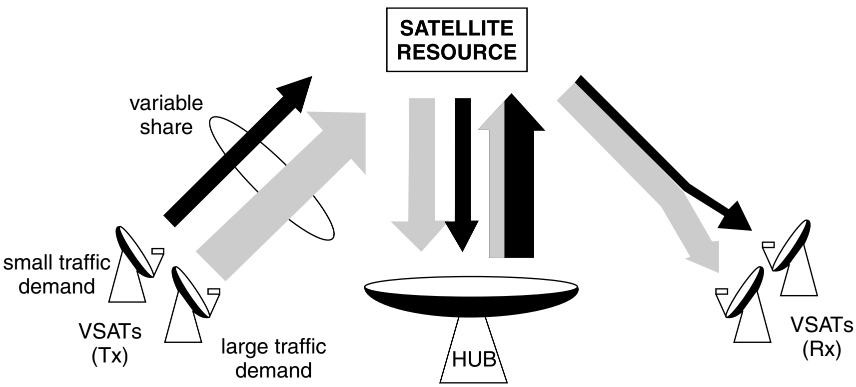


Figure 1.16 Principle of demand assignment

Those requests are transmitted on a specific signalling channel, or piggy-backed on the traffic messages. With centralised management, the hub station or the traffic control station replies by allocating to the VSAT the appropriate resource, either a frequency band or a time slot. With distributed management, all VSATs keep a record of occupied and available resource. This is discussed in more detail in Chapter 4, section 4.6.

From the above, it can be recognised that demand assignment offers a better use of the satellite resource but at the expense of a higher system cost and a delay in connection set-up. However, a larger number of stations can share the satellite resource. Hence the higher investment cost is compensated for by a larger return on investment.

The centralised/distributed management option depends on the network architecture: a centralised control is easier to perform with a star-shaped network, as all traffic flows through the hub, which then is the natural candidate for demand assignment control. With a mesh-shaped network, both centralised and distributed control can be envisaged. Delay for link set-up is shorter with distributed control, as a single hop (about 0.25 s) is sufficient to inform all VSATs in the network of the request and the corresponding resource occupancy, while a double hop (about 0.5 s) is necessary for the request to proceed to the central station, and for that station to allocate the corresponding resource. Finally, as demand assignment implies charging the remote sites according to the resource occupancy, billing and accounting is more easily handled by a centralised control.

1.6.4 Frequency bands

VSAT networks are supposed to operate within the so-called 'fixed satellite service' (FSS) defined within the International Telecommunication 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 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,

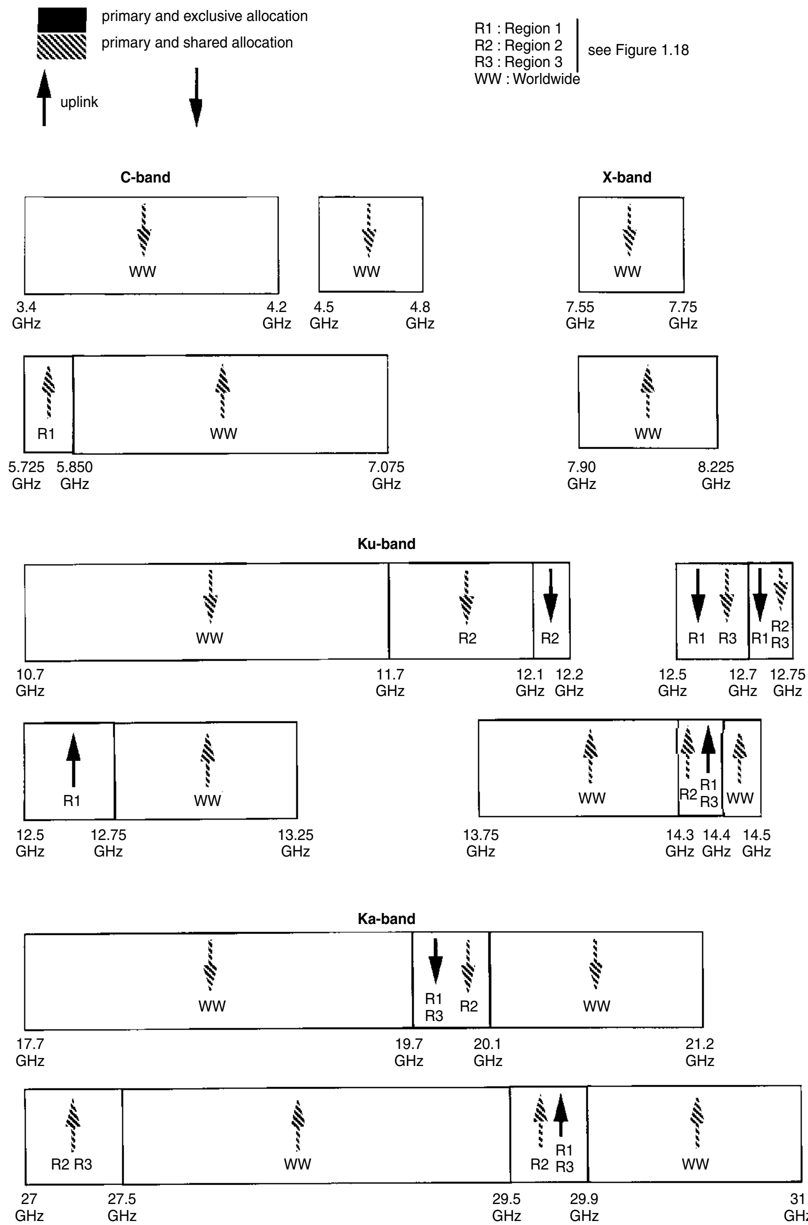


Figure 1.17 Frequency bands allocated to the fixed satellite service (FSS) and usable for VSAT networks (ITU00)

most are experimental. Figure 1.17 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

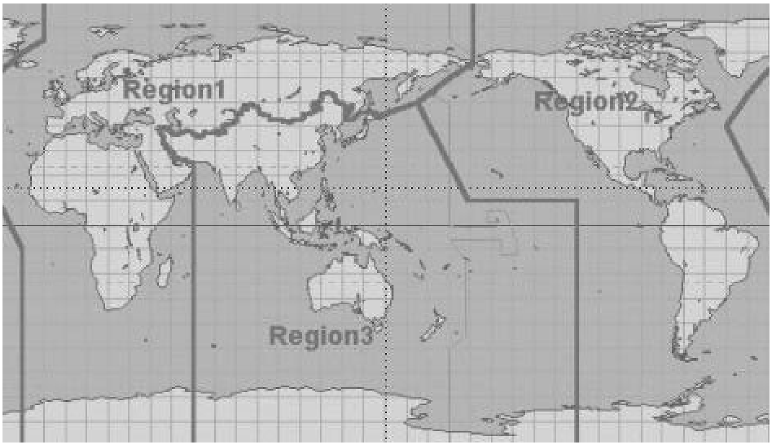


Figure 1.18 Regions 1, 2 and 3 in the world

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 1.18 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 1.17 which deals with the fixed satellite service only.

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 (1.2) the half power beamwidth θ_{3dB} of an antenna relates to the product of its diameter by frequency (see Appendix 4), as follows:

$$\theta_{3dB} = \frac{70\,c}{Df} \qquad \text{(degrees)} \tag{1.2}$$

where D (m) is the diameter of the antenna, f (Hz) is the frequency, and $c = 3 \times 10^8\ ms^{-1}$ is the velocity of light.

Therefore, the smaller the antenna diameter, the larger the beamwidth, and the off-axis interfering carriers are more likely to be emitted or received with high antenna gain. How important this perturbation can be is discussed in Chapter 5, section 5.5.

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 (1.2) 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 1.17), which is not the case for C-band. This simplifies the positioning of the VSAT and hub station as no coordination is implied.

Figure 1.19 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 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

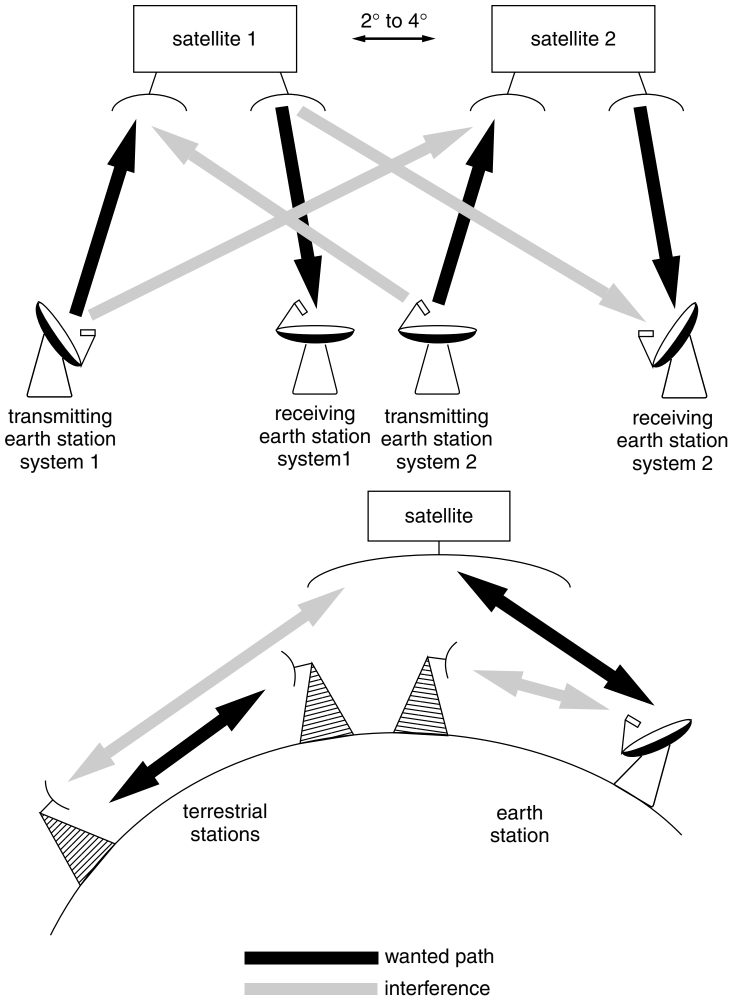


Figure 1.19 Interference paths

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 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.

Table 1.5 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

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 1.5 summarises the advantages and drawbacks of the most commonly available frequency bands.

1.6.5 Hub options

1.6.5.1 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 organisation 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.

1.6.5.2 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.

1.6.5.3 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.

1.7 VSAT NETWORK EARTH STATIONS

1.7.1 VSAT station

Figure 1.20 illustrates the architecture of a VSAT station. As shown in the figure, a VSAT station is made of two separate sets of equipment:

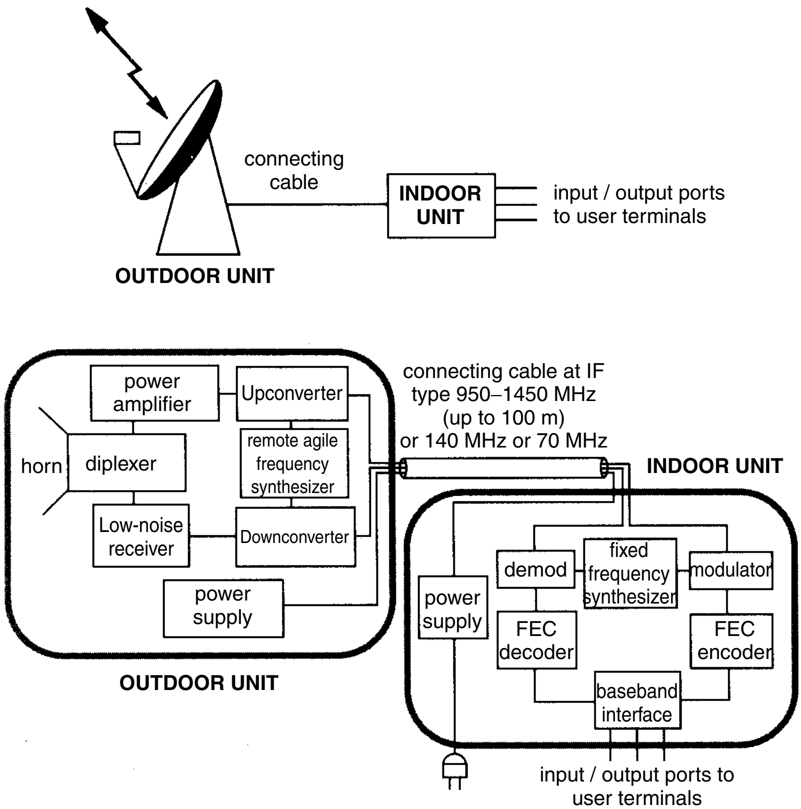


Figure 1.20 VSAT station equipment

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).

1.7.1.1 The outdoor unit (ODU)

Figure 1.21 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 1.22 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;

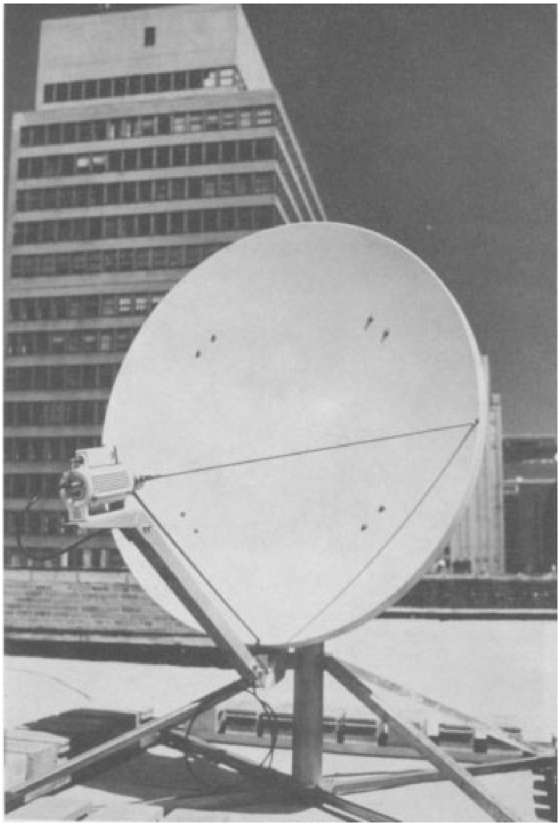


Figure 1.21 Photograph of the outdoor unit of a VSAT station. (Reproduced by permission of Gilat Satellite Networks, Ltd.)

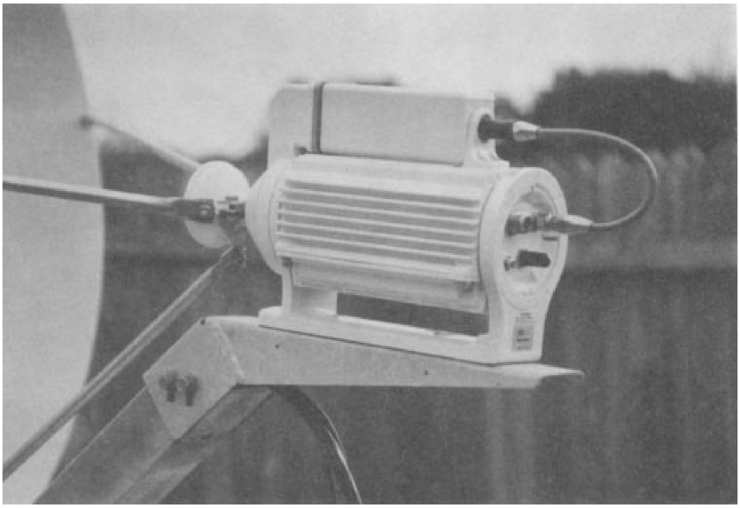


Figure 1.22 Photograph of the electronics container of the outdoor unit shown in Figure 1.21. (Reproduced by permission of Gilat Satellite Networks, Ltd.)

– 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 (see Chapter 5, section 5.2);
– 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 (see Chapter 5, section 5.3);
– 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 1.6 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.

1.7.1.2 The indoor unit (IDU)

The indoor unit installed at the user's facility is shown in Figure 1.23. 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 1.24.

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. This is often specified by reference to a standard, such as those mentioned in section 1.6.2 and in Appendix 3;
– 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.

Coherent modulation schemes such as biphase shift keying (BPSK) or quadrature phase shift keying (QPSK) are used. For acceptable

Table 1.6 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	$\pm 90^\circ$ 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	$-30^\circ C$ to $+55^\circ C$

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.

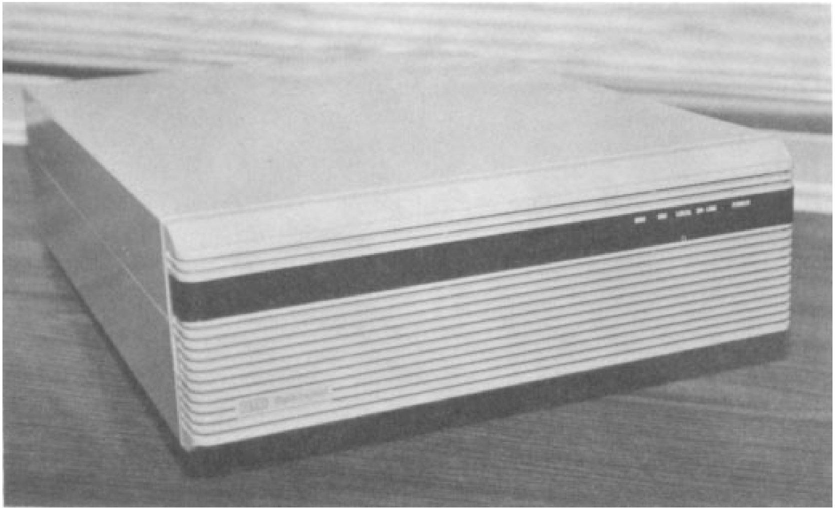


Figure 1.23 Front view of the indoor unit of a VSAT station. (Reproduced by permission of Gilat Satellite Networks, Ltd.)

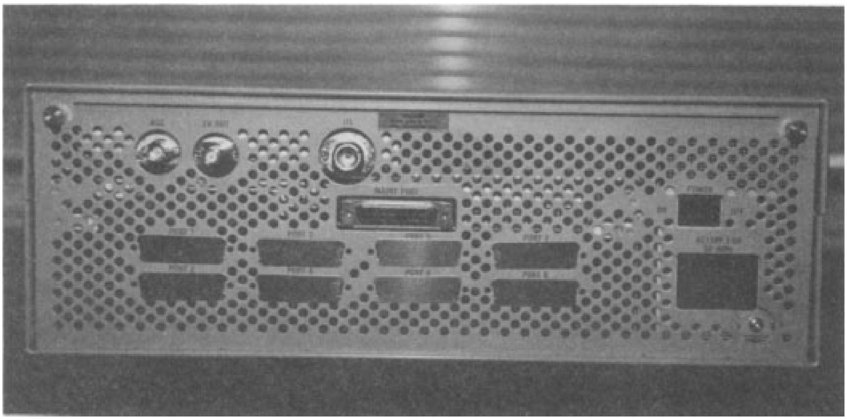


Figure 1.24 Rear view of the indoor unit shown in Figure 1.23. (Reproduced by permission of Gilat Satellite Networks, Ltd.)

1.7.2 Hub station

Figure 1.25 shows a photograph of a hub station and Figure 1.26 displays the architecture of the hub station with its equipment. Apart from the size and the number of subsystems, there is little functional difference between a hub and a VSAT, so that most of the content of the above section applies here. The major difference is that the indoor unit of a hub station interfaces to either a host computer or to a public switched network or private lines, depending on whether the hub is a dedicated or a shared one (see above section on

Figure 1.25 Photograph of the outdoor unit of a hub station. (Reproduced by permission of Hughes Network Systems, Inc.)

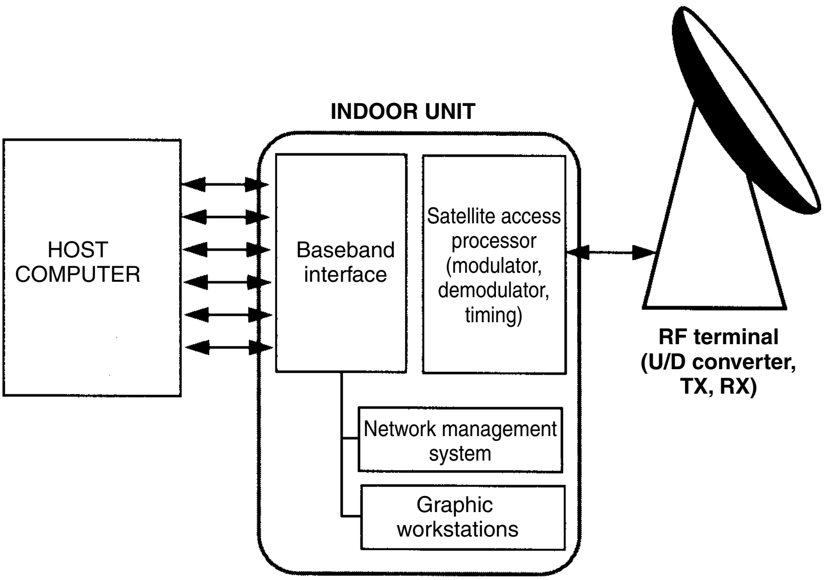


Figure 1.26 Hub subsystems

VSAT network options). Typical ODU hub station parameters are indicated in Table 1.7.

One can note in Figure 1.26 that the hub station is equipped with a network management system (NMS). The NMS is a mini-computer or a work station, equipped with its an dedicated software and displays, and used for operational and administrative functions.

This mini-computer is connected to each VSAT in the network by means of permanent virtual circuits. Management messages

Table 1.7 Typical values for the ODU parts of a hub 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	Axisymmetric dual reflector (Cassegrain)
Diameter	2 to 5 m (compact hub) 5 to 8 m (medium hub) 8 to 10 m (large hub)
TX/RX isolation	30 dB
Voltage Standing Wave Ratio (VSWR)	1.25:1
Polarisation	Linear orthogonal at Ku-band Circular orthogonal at C-band
Polarisation adjustment	$\pm 90^\circ$ for linear polarised antenna
Cross polarisation isolation	35 dB on axis
Sidelobe envelope	$29 - 25 \log \theta$
Azimuth travel	120 degrees
Elevation travel	3 to 90 degrees
Positioning	0.01°/s
Tracking	Steptrack at Ku-band if antenna larger than 4 m
Wind speed:	
operation	50 to 70 km/h
survival	180 km/h
Deicing	Electric
Power amplifier	
Output power	3–15 W SSPA at Ku-band 5–20 W SSPA at C-band 50–100 TWT at Ku-band 100–200 TWT at C-band
Power setting	0.5 dB steps
Frequency steps	100 kHz to 500 kHz
Low noise receiver	
Noise temperature	80–120 K at Ku-band 35–55 K at C-band
Operating temperature	-30°C to $+55^\circ$C

SSPA: Solid State Power Amplifier TWT: Travelling Wave Tube are constantly exchanged between the NMS and the VSATs and compete with the normal traffic for network resources.

1.7.2.1 Operational functions

Operational functions relate to the network management and provide the capability to reconfigure the network dynamically by adding, or deleting, VSAT stations, carriers and network interfaces. Operational functions also include monitoring and controlling the performance and status of the hub and each VSAT station, and all associated data ports of the network. This entails operational management tools which provide real-time assignment and connectivity of VSATs, and management and control of new installations and configurations.

The network control software allows automatic dynamic allocation of capacity to VSATs with bursty interactive traffic and to VSATs that will occasionally be used for stream traffic (see Chapter 4, section 4.3). No operator intervention is required to effect this temporary capacity reallocation.

The NMS notifies the operator in the case of capacity saturation, which prevents more VSAT users from entering the service. The NMS also handles all aspects related to alarm and failure diagnosis. In particular, in case of any power interruptions at the VSAT stations, the NMS downloads all the relevant software and system parameters for operation restart.

1.7.2.2 Administrative functions

Administrative functions deal with inventory of equipment, records of network usage, security and billing.

The NMS keeps an account of the VSAT stations installed and operated, the equipment configuration within the hub and each VSAT station, and the port configuration of each network interface. This information is available on request by the operator, along with statistical information on traffic, number of failures, average time of data transmission delay, etc. The information can be analysed and printed on a daily, weekly or monthly basis as well as being stored on magnetic tape for future reference. It forms the basis for traffic and trend performance analysis, cost distribution based upon usage, etc.

The above long and diverse list of functions to be performed by the NMS shows its important role for the network. Actually, the adequacy of the NMS's response to the user's needs makes the difference between popular network providers and those who fail to survive in the market.

1.8 ECONOMIC ASPECTS

VSAT business faces competition in regions of the world where terrestrial networks are available. For the same grade of service, terrestrial digital data service networks and packet switched networks are probable contenders. Provided that desired network reliability, response time and throughput are achieved, then the economic comparison should be made on the basis of cost per month per site. Such comparisons with alternative solutions must be made on a case-by-case basis as they depend on many local factors, and also may be time dependent. In general terms, a VSAT network is a cost effective solution if the unique property of the satellite to broadcast information is well used.

To illustrate the budget headings of a typical VSAT network and the impact of the number of VSATs, Tables 1.8 to 1.10 display the budget of three networks considering a typical interactive data application: Table 1.8 relates to a small network with 30 VSAT stations and a shared hub, Table 1.9 refers to a network of 300 VSAT stations and a dedicated mini-hub, and Table 1.10 relates to a network with 1000 VSAT stations and a dedicated large hub.

The cost per month per VSAT is calculated assuming a five year amortisation. The \$2 000 equipment cost of a VSAT is applicable to a station equipped with a 1.8 m antenna, a 2 W transmitter power, and 4 output user ports. This cost figure corresponds to an order for a small number (30 units). A discount of 20% is applicable should 300 VSATs be ordered, and 30% for 1000 VSATs. This is reflected in Tables 1.9 and 1.10. Installation cost is taken equal to one manday plus professional expenses, or typically \$700 per VSAT. Spare parts represent 10% of the equipment cost, and the annual maintenance cost is taken equal to \$500 per VSAT.

The VSAT transmits information at bit rate $R_b = 64\ \text{kbs}^{-1}$, with code rate $\rho = 1/2$. Therefore the transmitted bit rate is $R_c = 128\ \text{kbs}^{-1}$. Considering BPSK modulation, with spectral efficiency (ratio of bit rate upon used bandwidth) $\Gamma = 0.5\ \text{bs}^{-1}.\text{Hz}^{-1}$, including guard bands, the carrier bandwidth is $B_{inb} = R_c/\Gamma = 250$ kHz.

The access to the satellite is by time division multiple access (TDMA) and it is assumed that thirty VSAT stations share one inbound link. Therefore, the inbound *average* capacity per VSAT is $64\ \text{kbs}^{-1}$ divided by 30, i.e. $2133\ \text{bs}^{-1}$.

Table 1.8 Budget for a 30 VSAT network and a shared hub

	Cost per unit ($)	Units (over 5 years)	Total ($)	Cost/month ($)
VSAT				
Equipment	2 000	30	60 000	1 000
Installation	700	30	21 000	350
Spare parts	200	30	6 000	100
Maintenance per VSAT per year	500	$30 \times 5 = 150$	75 000	1 250
Hub				
Lease cost per year	40 000	5	200 000	3 333
Hub-to-host connection cost per year	20 000	5	100 000	1 667
Satellite				
Bandwidth lease per year (1.25 MHz)	62 500	5	312 500	5 208
Licence				
One time fee	8 000	1	8 000	133
Licence charge per VSAT per year	100	$5 \times 30 = 150$	15 000	250
Total cost			797 500	13 292
Cost/VSAT/month				$443

The outbound link is a time division multiplex (TDM) at information bit rate $R_b = 256$ kbs^{-1}, with code rate $= 1/2$. The transmitted bit rate is $R_c = 512$ kbs^{-1} and the utilised band is $B_{outb} = 1$ MHz. An outbound link is dedicated to each group of VSATs. Therefore, the outbound *average* capacity per VSAT is 256 kbs^{-1} divided by 30, i.e. 8533 bs^{-1}.

Table 1.8 refers to a network with 30 VSATs, with one inbound link and one outbound link, so that the utilised transponder bandwidth is $B_1 = B_{inb} + B_{outb} = 1.25$ MHz.

Table 1.9 refers to a network with 300 VSATs, with 10 inbound links and 10 outbound links (one per VSAT group), so that the utilised transponder bandwidth is $B_2 = 10(B_{inb} + B_{outb}) = 12.5$ MHz.

Table 1.10 refers to a network with 1000 VSATs, with 34 inbound links and 34 outbound links, so that the utilised transponder bandwidth is $B_3 = 34(B_{inb} + B_{outb}) = 42.5$ MHz.

A cost of $50 000 per year is considered for the non-pre-emptible lease of 1 MHz of transponder bandwidth.

To operate and maintain a dedicated large hub or a mini-hub, a staff of eight people working round the clock in eight-hour shifts is

Table 1.9 Budget for a 300 VSAT network and a dedicated mini-hub

	Cost per unit ($)	Units (over 5 years)	Total ($)	Cost per month ($)
VSAT				
Equipment	1 600	300	480 000	8 000
Installation	700	300	210 000	3 500
Spare parts	200	300	60 000	1 000
Maintenance per VSAT per year	500	300 × 5 = 1 500	750 000	12 500
Hub				
Equipment and installation	100 000	1	100 000	1 667
Operation and maintenance per year	320 000	5	1 600 000	26 667
Satellite				
Bandwidth lease per year (12.5 MHz)	625 000	5	3 125 000	52 083
Licence				
One time fee	8 000	1	8 000	133
Licence charge per VSAT per year	100	5 × 300 = 1 500	150 000	2 500
Total cost			6 483 000	108 050
Cost/VSAT/month				$360

considered with a cost per man-year of $40 000. Hence the annual staff cost in Tables 1.9 and 1.10 is $320 000. With a shared hub (Table 1.8) the staff is employed by the owner of the hub and the corresponding cost is charged to the client as part of the lease cost.

The licence structure is based on a one-time fee of $8000 and an annual fee per VSAT of $100. It can be noted that the one-time fee, although it may be considered an expensive one, has little impact on the cost per month per VSAT.

Comparing Tables 1.8 to 1.10 indicates a decreasing cost per site per month as the number of VSATs grows.

1.9 REGULATORY ASPECTS

Regulations in the field of VSAT networks entail several aspects:

- licensing;
- access to the space segment;
- permission for installation.

Table 1.10 Budget for a 1000 VSAT network and a dedicated large hub

	Cost per unit ($)	Units (over 5 years)	Total ($)	Cost per month ($)
VSAT				
Equipment	1 400	1 000	1 400 000	23 333
Installation	700	1 000	700 000	11 667
Spare parts	200	1 000	200 000	3 333
Maintenance per VSAT per year	500	$5 \times 1\,000 = 5\,000$	2 500 000	41 667
Hub				
Equipment and installation	1 000 000	1	1 000 000	16 667
Operation and maintenance per year	320 000	5	1 600 000	26 667
Satellite				
Bandwidth lease per year (42.5 MHz)	2 125 000	5	10 625 000	177 083
Licence				
One-time fee	8 000	1	8 000	133
Licence charge per VSAT per year	100	$5 \times 1000 = 5\,000$	500 000	8 333
Total cost			18 533 000	308 883
Cost/VSAT/month				$309

1.9.1 Licensing

A licence is to be delivered by the national telecommunications authority of a country where any earth station as a part of a network, be it the hub, a control station or a VSAT, is planned to be installed and operated.

The concern reflected here is to ensure compatibility between radio networks by avoiding harmful interference between different systems. By doing so, any licensed operator within a certain frequency band is recognised as not causing unacceptable interference to others, and is protected from interference caused by others.

In the past, national telecommunication authorities have required licensing of individual VSAT terminals in addition to requiring a network operator's licence. Then, the US Federal Communication Commission (FCC) implemented with success a *blanket licensing* approach for VSATs operated within the US. With blanket licensing, VSATs are configured based upon technical criteria (power level, frequency, etc.) to eliminate the risk of interference, so a single licence can be issued covering a large number of VSAT terminals. Blanket licensing has since gained interest among national telecommunications authorities all over the world, as a result of equipment manufacturers complying with the recommendations issued by international standardisation bodies, such as the International Telecommunication Union (ITU) and the European Telecommunications Standard Institute (ETSI). Relevant documentation from these bodies is available at http://www.itu.int/home/index.html and http://www.etsi.org/.

A licence usually entails the payment of a licence fee, which is most often in two parts: a one-time fee for the licensing work and an annual charge per station.

The licensing procedure is simpler when the network is national, as only one telecom authority is involved. For transborder networks, licences must be obtained from the different national authorities where the relevant earth stations are planned to be installed and operated, and rules often differ from one country to another. To facilitate the access to these rules, telecommunications authorities around the world have begun posting data related to their nations' VSAT regulatory conditions on the World Wide Web. Information on these websites can be obtained from the International VSAT Regulatory Portal of the Global VSAT Forum (GVF) (http://www.gvf.org/regulation/portal/index.cfm#).

1.9.2 Access to the space segment

This deals with the performance and operational procedures that satellite operators request from earth station operators for transmission of carriers to, and from, their satellite transponders. Most often such requirements are based on the ITU-R Recommendations, but some satellite operators may impose special constraints.

In any case, the applicant operator of a VSAT network is compelled to fulfil the requirements imposed by the satellite operator in terms of earth station maximum EIRP, G/T, frequency stability and control of transmission.

1.9.3 Local regulations

Installation of a VSAT encompasses problems relating to planning or zoning controls, building and person safety. The VSAT should comply to local regulations dealing with environmental protection such as antenna dish size, colour and shape. Finally, landlord permission to dig for cable ducts or install roof mounted antennas should be treated as contractual matters between the landlord and the tenant.

1.10 CONCLUSIONS

This conclusion summarises the perceived advantages and drawbacks of VSAT networks.

1.10.1 Advantages

1.10.1.1 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.

1.10.1.2 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 outbound 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.

1.10.1.3 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.

1.10.1.4 Private corporate networks

A VSAT network offers its operator end-to-end control over transmission 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.

1.10.1.5 Low bit error rate

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

1.10.1.6 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.

1.10.2 Drawbacks

1.10.2.1 Interference sensitivity

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

1.10.2.2 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.

1.10.2.3 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 transponder be located on another satellite, this may mean intervening on each site to repoint the antenna, and this may take some time.

1.10.2.4 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.

2

Use of satellites for VSAT networks

It is not so important for someone who is interested in VSAT networks to know a lot about satellites. However, a number of factors relative to satellite orbiting and satellite–earth geometry influence the operation and performance of VSAT networks. For instance, the relative position of the satellite with respect to the VSAT at a given instant determines the orientation of the VSAT antenna and also the carrier propagation delay value. The relative velocity of the satellite with respect to the earth station receiving equipment induces Doppler shifts on the carrier frequency that must be tracked and compensated for. This impacts on the specifications and the design of earth station receivers. For a geostationary satellite, which is supposed to be in a fixed position relative to the earth, one may believe that once the antenna has been properly pointed towards that position at the time of its installation, the adequate orientation is established once and for all. Actually, as a result of satellite orbital perturbations, there is no such thing as a geostationary satellite, and residual motions induce antenna depointing and hence antenna gain losses which affect the link performance.

Therefore it is worth mentioning these aspects, and this is the aim of this chapter. Orbit definition and parameters will be presented in the general case, with the ulterior motive to give the reader some conceptual tools that would be handy should VSAT networks be used someday in conjunction with non-geostationary satellite systems. However, as current VSAT networks use geostationary satellites, the bulk of the chapter will consider this specific scenario.

VSAT Networks, 2nd Edition. G. Maral
 ISBN: 0-470-86684-5

Many of the considerations developed in this chapter will be used in the following ones.

Before orbital aspects are dealt with, the role of the satellite and some related topics will first be introduced as an encouragement to the reader.

2.1 INTRODUCTION

2.1.1 The relay function

Satellites relay the carriers transmitted by earth stations on the ground to other earth stations, as illustrated in Figure 2.1. Therefore, satellites act similarly to microwave terrestrial relays installed on the tops of hills or mountains to facilitate long distance radio frequency links. Here the satellite, being at a much higher altitude than any terrestrial relay, is able to link distant earth stations, even from continent to continent.

Figure 2.1 indicates that the earth stations are part of what is called the *ground segment*, while the satellite is part of the *space segment*.

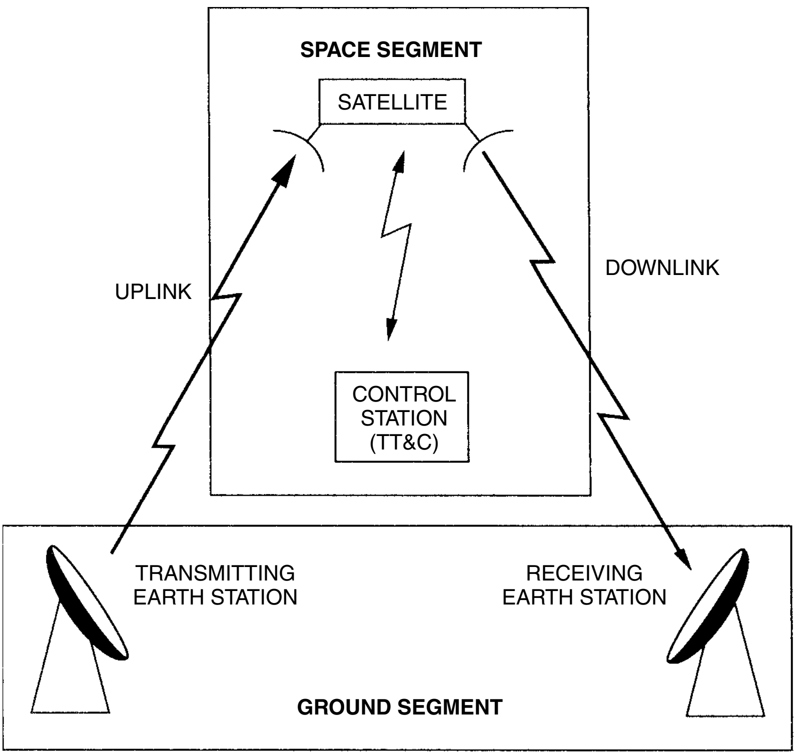


Figure 2.1 Architecture of a satellite system

The space segment also comprises all the means to operate the satellite, for instance the stations which monitor the satellite status by means of telemetry links and control it by means of command links. Such links are sometimes called TTC (telemetry, tracking and command) links.

The satellite roughly consists of a platform and a payload. The platform consists of all subsystems that allow the payload to function properly, namely:

- the mechanical structure which supports all equipment in the satellite;
- the electric power supply, consisting of the solar panels and the batteries used as supply during eclipses of the sun by the earth and the moon;
- the attitude and orbit control, with sensors and actuators;
- the propulsion subsystem;
- the onboard TTC equipment.

The payload comprises the satellite antennas and the electronic equipment for amplifying the uplink carriers. These carriers are also frequency converted to the frequency of the downlink. Frequency conversion avoids unacceptable interference between uplinks and downlinks.

Figure 2.2 shows the general architecture of the payload. The receiver (Rx) encompasses a wide band amplifier and a frequency downconverter. The input multiplexer (IMUX) splits the incoming carriers into groups within several sub-bands, each group being amplified to the power level required for transmission by a high

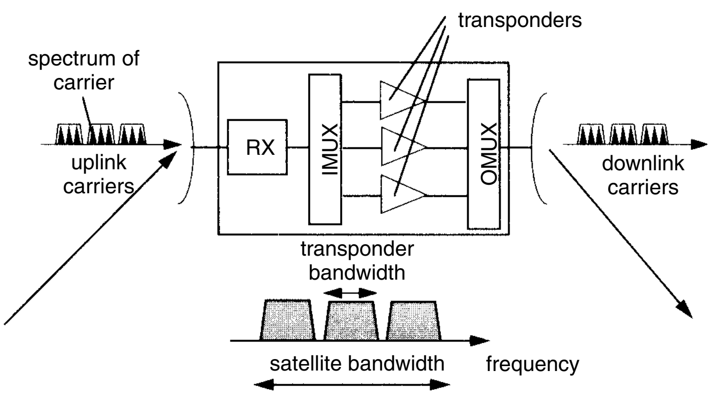


Figure 2.2 Payload architecture

power amplifier, generally a travelling wave tube (TWT). The different groups of carriers are then combined in the output multiplexer (OMUX) and forwarded to the transmitting antenna. The channels associated with the subbands of the payload from IMUX to OMUX are called *transponders*. The advantage of splitting the satellite band is three-fold:

- each transponder TWT amplifies a reduced set of carriers, hence each carrier benefits from a larger share of the limited amount of power available at the output of the TWT;
- the transponder TWT operates in a non-linear mode when driven near saturation. Saturation is desirable because the TWT then delivers more power to the amplified carriers than when operated in a backed-off mode, away from saturation. However, amplifying multiple carriers in a non-linear mode generates intermodulation, which acts as transmitted noise on the downlink. Less intermodulation noise power is transmitted with a reduced set of amplified carriers within each TWT;
- reliability is increased, as the failure of one TWT does not imply an overall satellite failure and each TWT can be backed up.

Typical values of bandwidth for a transponder are 36 MHz, 45 MHz and 72 MHz. However, there is no established standard. The TWT power is typically a few tens of watts. Some satellites are now equipped with solid state power amplifiers (SSPAs) instead of TWTs.

Figure 2.2 does not indicate any back-up equipment. To ensure the required reliability at the end of life of the satellite, some redundancy is built into the payload: for instance, the receiver is usually backed up with a redundant unit, which can be switched on in case of failure of the allocated receiver. The transponders are also backed up by a number of redundant units: a popular scheme is the ring redundancy, where each IMUX output can be connected to any of several transponders, with a similar arrangement between the transponder outputs and the OMUX inputs.

2.1.2 Transparent and regenerative payload

A satellite payload is transparent when the carrier is amplified and frequency downconverted without being demodulated. The frequency conversion is then performed by means of a mixer and a local oscillator as indicated in Figure 2.3. The carrier at a frequency equal to the uplink frequency f_U minus the local oscillator frequency

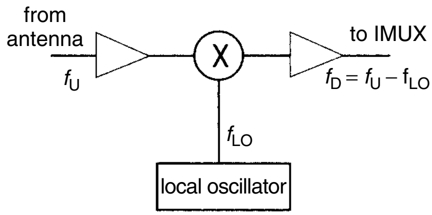


Figure 2.3 Receiver for a transparent satellite

f_{LO} is usually selected by filtering at the output of the mixer, and the local oscillator frequency is tuned so that the resulting frequency corresponds to the desired downlink frequency f_D. For instance, an uplink carrier at frequency $f_U = 14.25$ GHz mixed with a local oscillator frequency $f_{LO} = 1.55$ GHz results in a downlink carrier frequency $f_D = 12.7$ GHz.

A transparent payload makes no distinction between uplink carrier and uplink noise, and both signals are forwarded to the downlink. Therefore, at the earth station receiver, one gets the downlink noise together with the uplink retransmitted noise.

A regenerative payload entails on-board demodulation of the uplink carriers. On-board regeneration is most conveniently performed on digital carriers. The bit stream obtained from demodulation of a given uplink carrier is then used to modulate a new carrier at downlink frequency. This carrier is noise-free, hence a regenerative payload does not retransmit the uplink noise on the downlink. The overall link quality is therefore improved. Moreover, intermodulation noise can be avoided as the satellite channel amplifier is no longer requested to operate in a multicarrier mode. Indeed, several bit streams at the output of various demodulators can be combined into a time division multiplex (TDM) which modulates a single high rate downlink carrier. This carrier is amplified by the channel amplifier which can be operated at saturation without generating intermodulation noise as the carrier it amplifies is unique. This concept is illustrated in Figure 2.4.

It should be emphasised that today's commercial satellites which can be used for VSAT services are not equipped with regenerative payloads but only with transparent ones. Only a few experimental satellites such as NASA's Advanced Communications Technology Satellite (ACTS) and the italian ITALSAT satellite have incorporated a regenerative payload, but they are no longer in operation. Some satellites of the EUTELSAT fleet are equipped with a regenerative payload (Skyplex) but can be used only by earth stations operating according to the DVB-S standard.

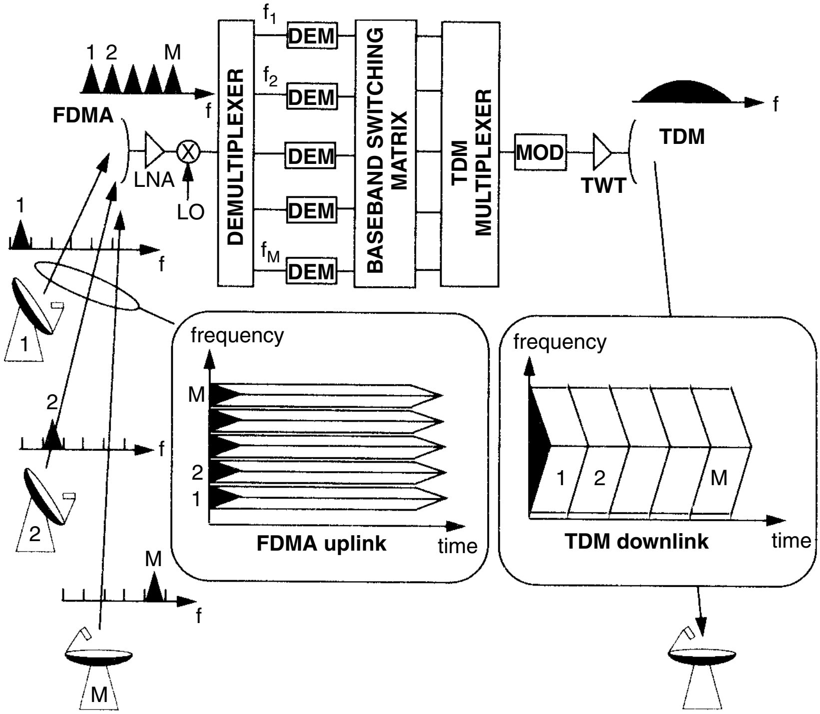


Figure 2.4 Regenerative satellite payload with multiplexed transmission on the downlink

2.1.3 Coverage

The coverage of a satellite payload is determined by the radiation pattern of its antennas. The receiving antenna and the transmitting antenna may have different patterns and hence there may be a different coverage for the uplink and the downlink. The coverage is usually defined by a specified minimum value of the antenna gain: for instance, the 3 dB coverage corresponds to the area defined by a contour of constant gain value 3 dB lower than the maximum gain value at antenna boresight. This contour defines the *edge of coverage*.

There are four types of coverage:

– *Global coverage*: the pattern of the antenna illuminates the largest possible portion of the earth surface as viewed from the satellite (Figure 2.5). A geostationary satellite sees the earth with an angle equal to 17.4°. Selecting the beamwidth of the antenna as 17.4° imposes a maximum gain at boresight of 20 dBi, thus the gain at the edge of the −3 dB coverage is 17 dBi.

– *Zone coverage*: an area smaller than the global coverage area is illuminated (Figure 2.6). The coverage area may have a simple

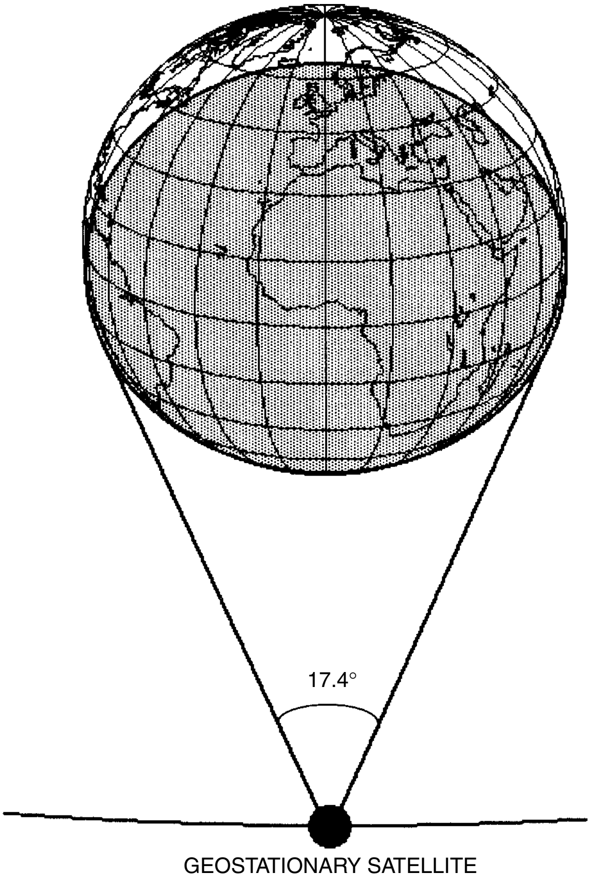


Figure 2.5 Global coverage

shape (circle or ellipse) or a more complex shape (contoured beam). For a typical zone coverage the antenna beamwidth is of the order of 5°. This imposes a maximum gain at boresight of 30 dBi, and a gain at the edge of the −3 dB coverage of 27 dBi.

– *Spot beam coverage*: an area much smaller than the global coverage area is illuminated. The antenna beamwidth is of the order of one to two degrees (Figure 2.7). Considering a 1.7° beamwidth imposes a maximum gain at boresight of 40 dBi and a gain at the edge of the −3 dB coverage of 37 dBi.

– *Multibeam coverage*: a spot beam coverage has the advantage of higher antenna gain than any other type of coverage previously discussed, but it can service only the limited zone within its coverage area. A service zone larger than the coverage area of a spot beam can still be serviced with high antenna gain thanks

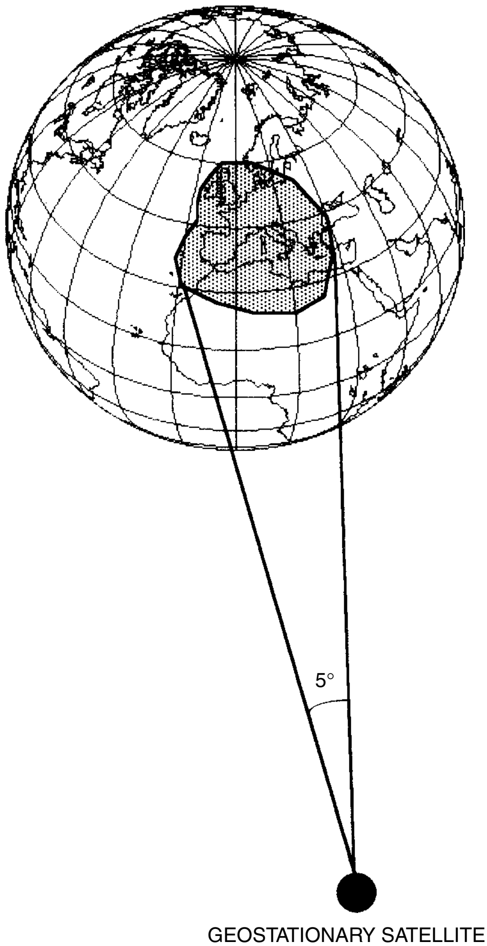


Figure 2.6 Zone coverage

to a multibeam coverage made of several individual spot beams. An example of such a coverage with adjacent spot beams is shown in Figure 2.8. This requires a multibeam satellite payload with more complex antenna farms. Maintaining interconnectivity between all stations of the service zone also implies a more complex payload architecture than that considered in Figure 2.2. Interconnectivity between stations implies that beams are interconnected; this can be achieved either by permanent connections from the uplink beams to the downlink ones, as illustrated in Figure 2.9, or by temporary connections established through an on-board switching matrix, as shown in Figure 2.10.

Permanent connections entail a larger number of transponders than on-board switching. On-board satellite switching requires that

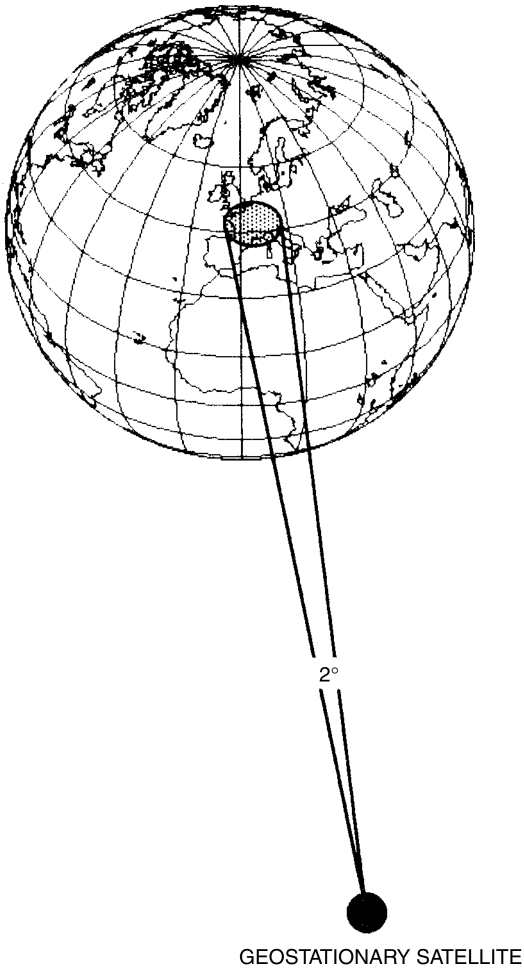


Figure 2.7 Spot beam coverage

earth stations transmit bursts of carriers, synchronous to the satellite switch state sequence, in such a way that they arrive at the satellite exactly when the proper uplink beam to downlink beam connection is established. More details on the operation of such multibeam satellite systems can be found in [MAR02, Chapter 7].

Usually the extension of a VSAT network is small enough for all VSATs and the hub station to be located within one beam.

2.1.4 Impact of coverage on satellite relay performance

The relay function of the satellite as described in section 2.1.1 entails adequate reception of uplink carriers and transmission of downlink

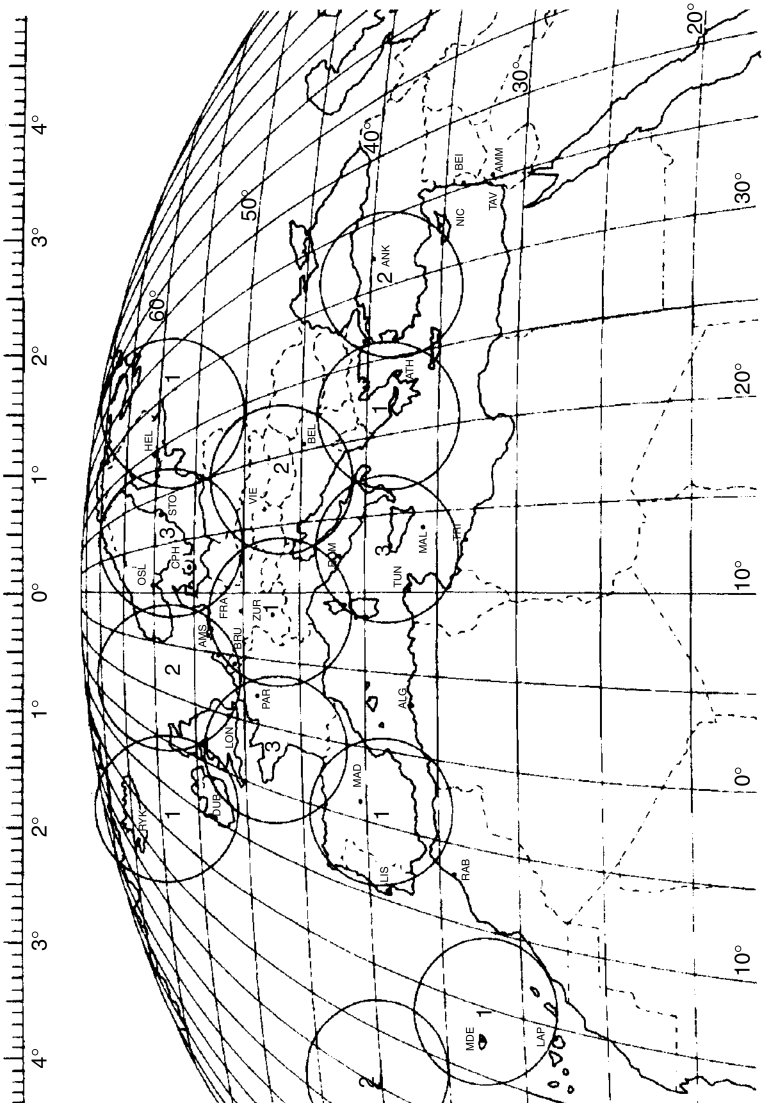


Figure 2.8 Coverage of a larger zone than covered by a single beam using a multibeam satellite. (Reproduced from (Mar02) by permission of John Wiley & Sons Ltd)

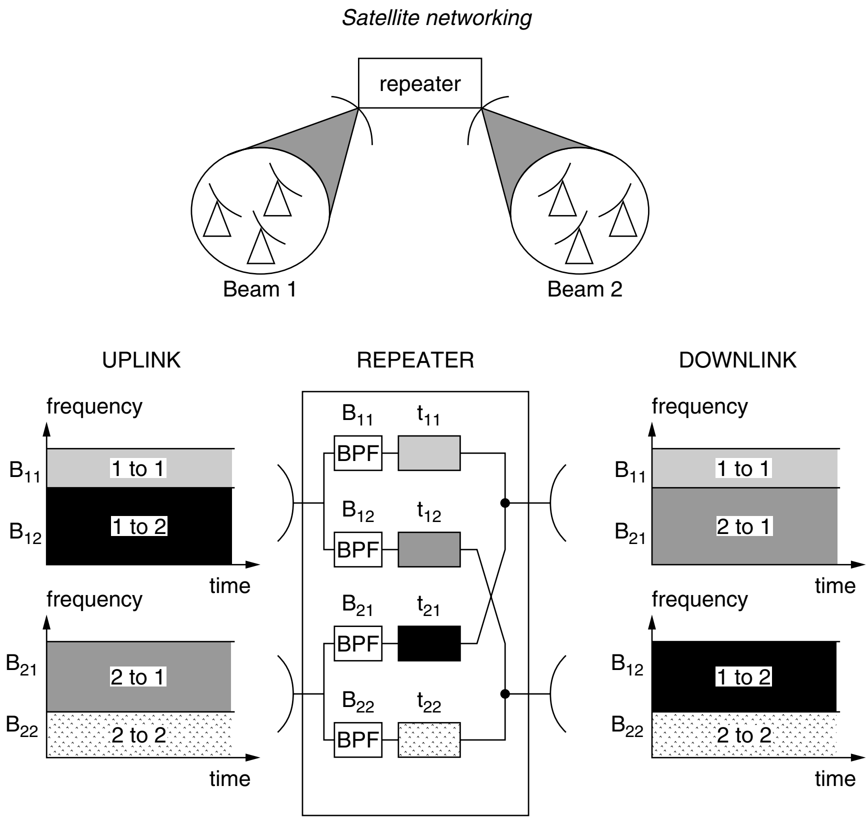


Figure 2.9 Interconnectivity of beams by permanent connections. (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd) BPF: band pass filter. t_{ij}: transponder connecting beam i to beam j and operating in band B_{ij}

carriers. As will be demonstrated in Chapter 5, the ability of the satellite payload to receive uplink carriers is measured by the figure of merit G/T of the satellite receiver, and its ability to transmit is measured by its effective isotropic radiated power (EIRP). These characteristics are defined in more detail in Chapter 5. Basically, G/T is the ratio of the receiving satellite antenna gain to the uplink system noise temperature, and the EIRP is the product of the transmitting satellite antenna gain G_T and the power P_T fed to the antenna by the transponder amplifier. Therefore, both parameters are proportional to the satellite antenna gain.

The specified values of G/T and EIRP are to be considered at the edge of coverage. Usually the edge of coverage is defined by the contour on the earth corresponding to a constant satellite antenna gain, say 3 dB below the gain G_{max} at boresight.

Now the maximum satellite antenna gain, G_{max}, as obtained at boresight, is inversely proportional to the square of its half power

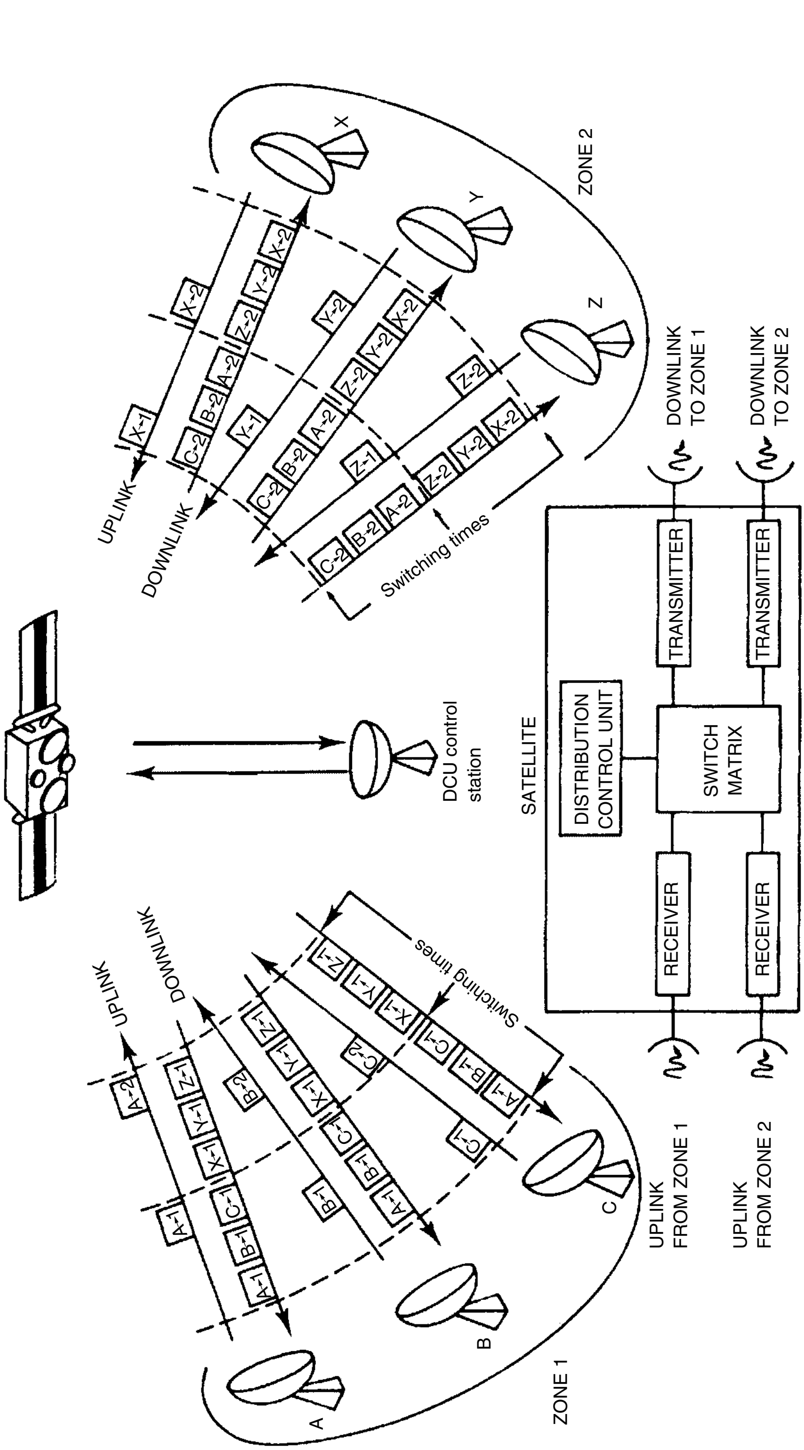


Figure 2.10 Interconnectivity of beams by temporary connections (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd)

beamwidth θ_{3dB}:

$$G_{max} = \frac{29000}{(\theta_{3dB})^2}$$

$$\text{or} \quad G_{max}(\text{dBi}) = 44.6 - 20\log\theta_{3dB} \tag{2.1}$$

Hence, one can consider that the specified values of G/T and EIRP are determined by the value of the satellite antenna gain at the edge of coverage G_{eoc}, given by:

$$G_{eoc} = \frac{G_{max}}{2}$$

$$\text{or} \quad G_{eoc}(\text{dBi}) = G_{max}(\text{dBi}) - 3\ \text{dB} \tag{2.2}$$

From (2.2) and (2.1), it can be seen that the specified values of G/T and EIRP at the edge of the coverage area determined by the satellite antenna beamwidth θ_{3dB}: the larger the beamwidth, the lower the G/T and EIRP.

So, the coverage of the satellite influences its relaying performance in terms of G/T and EIRP. A global coverage leads to smaller values of satellite G/T and EIRP, compared to a spot beam coverage. Should the VSAT network be included in a single satellite beam, then the larger its geographical dispersion, the poorer the satellite performance: this has to be compensated for by installing larger VSATs. For networks comprised of highly dispersed VSATs, say spread over several continents, the advantages of simple networking in terms of easy interconnectivity by placing all VSATs within a single beam have to be weighed against the cost of increasing the size of the VSATs, which might not be necessary if one agrees to service the network with a multibeam satellite, at the expense, however, of a more complex network operation.

2.1.5 Frequency reuse

Frequency reuse consists of using the same frequency band several times in such a way as to increase the total capacity of the network without increasing the allocated bandwidth. Frequency reuse can be achieved within a given beam by using polarisation diversity: two carriers at the same frequency but with orthogonal polarisations can be discriminated by the receiving antenna according to their respective polarisation. With multibeam satellites the isolation resulting from antenna directivity can be exploited to reuse the same frequency band in different beams.

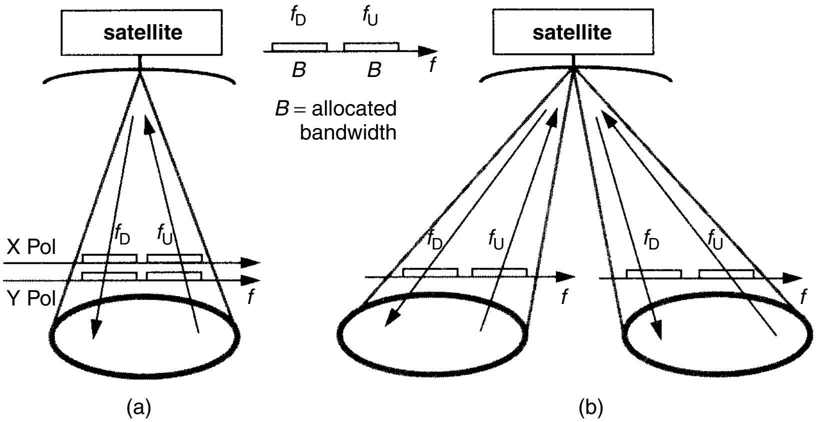


Figure 2.11 Frequency reuse; (a) by orthogonal polarisation; (b) by angular separation of the beams in a multibeam satellite system

Figure 2.11 compares the principle of frequency reuse by orthogonal polarisation (Figure 2.11(a)) and the principle of frequency reuse by angular beam separation (Figure 2.11(b)). In both cases the bandwidth allocated to the system is B. The system uses this bandwidth B centred on frequency f_U for the uplink and on the frequency f_D for the downlink. In the case of frequency reuse by orthogonal polarisation, the bandwidth B can only be reused twice. In the case of reuse by angular separation, the bandwidth B can be reused for as many beams as the permissible interference level allows. Both types of frequency reuse can be combined.

2.2 ORBITS

2.2.1 Newton's universal law of attraction

Satellites orbit the earth in accordance with Newton's universal law of gravitation: two bodies of mass m and M attract each other with a force which is proportional to their masses and inversely proportional to the square of the distance, r, between them:

$$F = \frac{GMm}{r^2} \quad \text{(N)} \tag{2.3}$$

where G (gravitational constant) $= 6.672 \times 10^{-11}\ \text{m}^3\ \text{kg}^{-1}\ \text{s}^{-2}$

As the mass of the earth is $M_e = 5.974 \times 10^{-23}$ kg, the product GM_e for an earth orbiting body has a value:

$$\mu = GM_e = 3.986 \times 10^{14}\ \text{m}^3\ \text{s}^{-2}$$

From Newton's law, the following results can be derived, which actually were formulated prior to Newton's works by Kepler from his observation of the movement of the planets around the sun:

- the trajectory of the satellite in space, called the orbit, lies in a plane containing the centre of the earth. For communication satellites, the orbit is selected to be an ellipse and one focus is the centre of the earth. Should the orbit be circular, then the orbit centre coincides with the earth centre;
- the vector from the centre of the earth to the satellite sweeps equal areas in equal times;
- the period T of revolution of the satellite around the earth is given by:

$$T = 2\pi(a^3/\mu)^{1/2} \quad \text{(seconds)} \qquad (2.4)$$

2.2.2 Orbital parameters

Six parameters are required to determine the position of the satellite in space (Figure 2.12) [MAR02, Figure 2.4, p 32]:

- two parameters for the determination of the plane of the orbit: the inclination of the plane (i) and the right ascension of the ascending node (Ω);

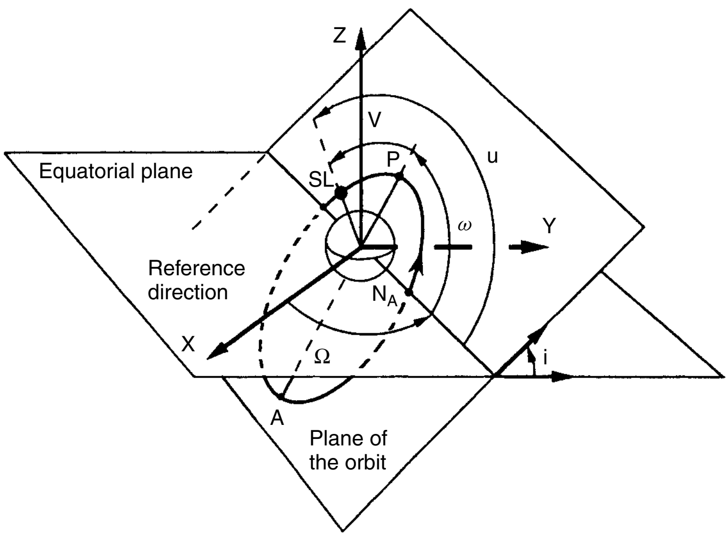


Figure 2.12 Positioning of satellite in space. (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd)

– one parameter for positioning the orbit in its plane: the argument of the perigee (ω);
– two parameters for the shape of the orbit: the semi-major axis (a) of the ellipse, and its eccentricity (e);
– one parameter for the positioning of the satellite on the elliptic curve: the true anomaly (v).

2.2.2.1 Plane of the orbit (Figure 2.13)

The plane of the orbit is obtained by rotating the earth's equatorial plane about the *line of nodes* of the orbit. The nodes are the intersections of the orbit with the equatorial plane of the earth. There is one ascending node where the satellite crosses the equatorial plane from south to north, and one descending node where the satellite crosses the equatorial plane from north to south. The rotation angle about the line of nodes is i, defined as the *inclination of the orbital plane*. This angle is counted positively in the forward direction between 0° and 180° between the normal n_1 (directed towards the east) to the line of nodes in the equatorial plane, and the normal n_2 (in the direction of the satellite velocity) to the line of nodes in the orbital plane.

The line of nodes must be referenced to some fixed direction in the equatorial plane. The commonly used reference direction is the line of intersection of the earth's equatorial plane with the plane of the ecliptic, which is the orbital plane of the earth around the sun (Figure 2.14). This line maintains a fixed direction in space with time, called the *direction of the vernal point* γ. Actually, as a result of some irregularities in the rotation of the earth, with the earth axis experiencing nutation, the direction of the vernal point is not perfectly fixed with time. Therefore the reference direction is taken as the direction of the vernal point at some instant, usually noon

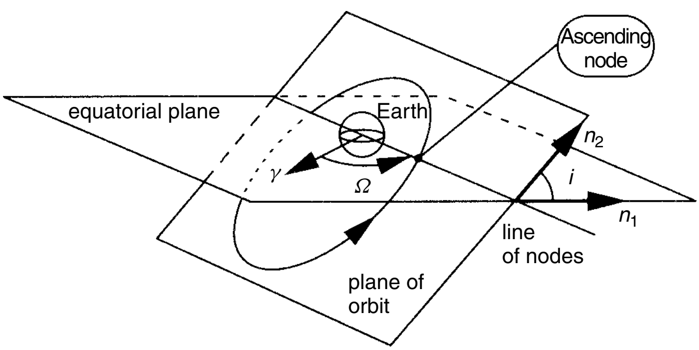


Figure 2.13 Orbit plane positioning: Ω, i

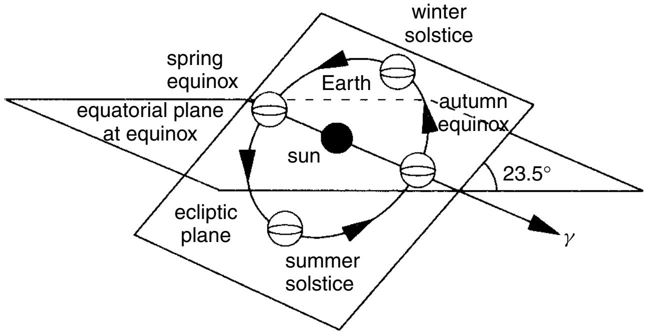


Figure 2.14 The direction of the vernal point γ is used as the reference direction in space

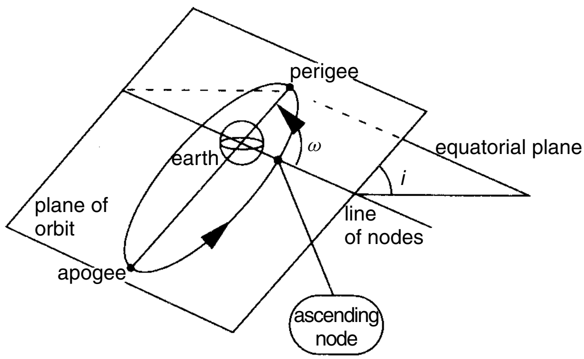


Figure 2.15 Positioning the orbit in its plane: the argument of the perigee (ω)

on January 1,2000, designated γ_{2000}. The angle which defines the direction of the line of nodes is the *right ascension of the ascending node* Ω. It is counted positively from 0° to 360° in the forward direction in the equatorial plane about the earth axis.

2.2.2.2 Positioning the orbit in its plane (Figure 2.15)

The centre of the earth is one of the focuses of the elliptical orbit. Therefore, the major axis of the ellipse passes through the centre of the earth. The direction of the perigee in the plane of the orbit is determined by the *argument of the perigee* ω, which is the angle, with vertex at the centre of the earth, taken positively from 0° to 360° in the direction of the motion of the satellite between the direction of the ascending node and the direction of the perigee. The *perigee* is the point of the orbit that is nearest to the centre of the earth. At the opposite of the major axis is the *apogee*, which is the point of the orbit that is farthest from the centre of the earth.

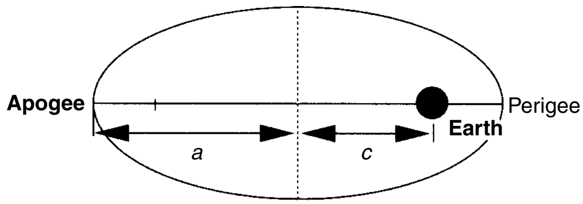


Figure 2.16 Defining the shape of the orbit: a, $e = c/a$

2.2.2.3 Shape of the orbit (Figure 2.16)

The shape of the orbit is determined by its *eccentricity*, e, and the length, a, of its *semi-major axis*. The eccentricity is given by:

$$e = \frac{c}{a} \quad (2.5)$$

where c is the distance from the centre of the ellipse to the centre of the earth. For a circular orbit the eccentricity is zero, and the centre of the earth is the centre of the circular orbit.

The distance from the centre of the earth to the apogee is $a(1 + e)$, and the distance from the centre of the earth to the perigee is $a(1 - e)$.

2.2.2.4 Positioning the satellite in its orbit (Figure 2.17)

The position of the satellite in its orbit is conveniently defined by the *true anomaly* v, which is the angle with vertex at the centre of the earth counted positively in the direction of movement of the satellite from 0° to 360°, between the direction of the perigee and the direction of the satellite.

The distance from the centre of the earth to the satellite is given by:

$$r = \frac{a(1 - e^2)}{(1 + e\cos v)} \quad \text{(m)} \quad (2.6)$$

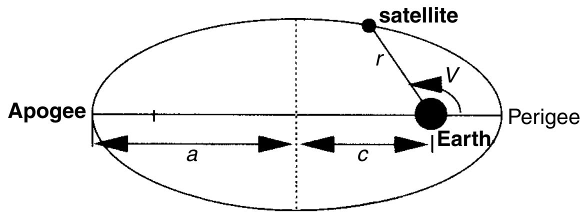


Figure 2.17 Positioning the satellite in its orbit

The satellite velocity is given by:

$$V = \mu^{1/2}(2/r - 1/a)^{1/2} \quad \text{(m/s)} \tag{2.7}$$

2.3 THE GEOSTATIONARY SATELLITE

2.3.1 Orbit parameters

A geostationary satellite proceeds in a circular orbit ($e = 0$) in the equatorial plane ($i = 0°$). The angular velocity of the satellite is the same as that of the earth, and in the same direction (direct orbit), as illustrated in Figure 1.5. To a terrestrial observer, the satellite seems to be fixed in the sky.

The above conditions make the period of the circular orbit, T, equal to the duration of a sidereal day, that is the time it takes for the earth to rotate 360°. Hence $T = 23$ h 56 min 4 s = 86 164 s. From expression (2.4) one can calculate the semi-major axis, a, of the orbit which identifies the radius of the orbit. One obtains $a = 42\,164$ km. Subtracting from this value the earth radius $R_e = 6378$ km, one obtains the satellite altitude $R_o = a - R_e = 35\,786$ km. The satellite velocity V_s can be calculated from expression (2.7) selecting $r = a$. It results in $V_s = 3075\ \text{ms}^{-1}$.

Table 2.1 summarises the characteristics of a geostationary satellite orbit.

2.3.2 Launching the satellite

The principle of launching a satellite into orbit is to provide it with the appropriate velocity at a specific point of its trajectory in the plane of the orbit, starting from the launching base on the earth surface. This usually requires a launch vehicle for the take-off, and an on-board specific propulsion system.

With a geostationary satellite, the desired orbit is circular, in the equatorial plane, and it is attained by an intermediate orbit called the *transfer orbit*. This is an elliptical orbit with perigee at an altitude

Table 2.1 Characteristics of a geostationary satellite orbit

Eccentricity (e)	0
Inclination of orbit plane (i)	0°
Period (T)	23 h 56 min 4 s = 86 164 s
Semi-major axis (a)	42 164 km
Satellite altitude (R_o)	35 786 km
Satellite velocity (V_s)	3075 ms^{-1}

of about 200 km, and apogee at the altitude of the geostationary orbit (35 786 km). Most conventional launch vehicles (Ariane, Delta, Atlas Centaur) inject the satellite into the transfer orbit at its perigee, as shown in Figure 2.18.

At this point, the launch vehicle must communicate a velocity $V_p = 10\,234\ \text{m/s}^{-1}$ to the satellite (for a perigee at 200 km). Then the satellite is left to itself and proceeds forward in the transfer orbit. When arriving at the apogee of the transfer orbit, the satellite propulsion system is activated and a velocity impulse is given to the satellite. This increases its velocity to the required velocity for a geostationary orbit, that is $V_s = 3075\ \text{ms}^{-1}$. The satellite orbit is now circular, and the satellite has the proper altitude.

Note the advantage of a launch towards the east as the launch vehicle benefits from the velocity introduced into the trajectory by the rotation of the earth.

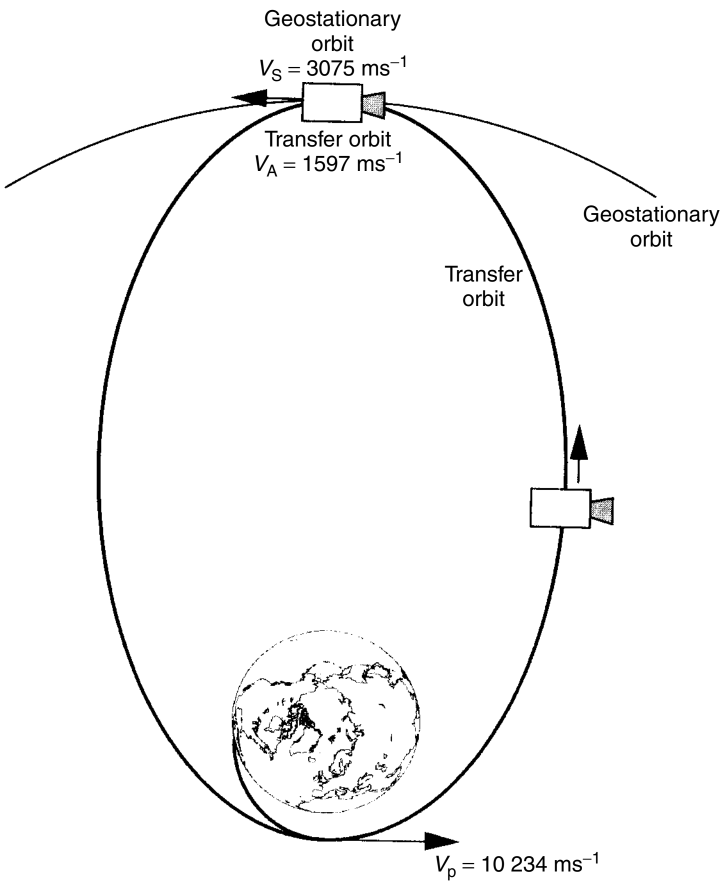


Figure 2.18 Transfer orbit and injection phases

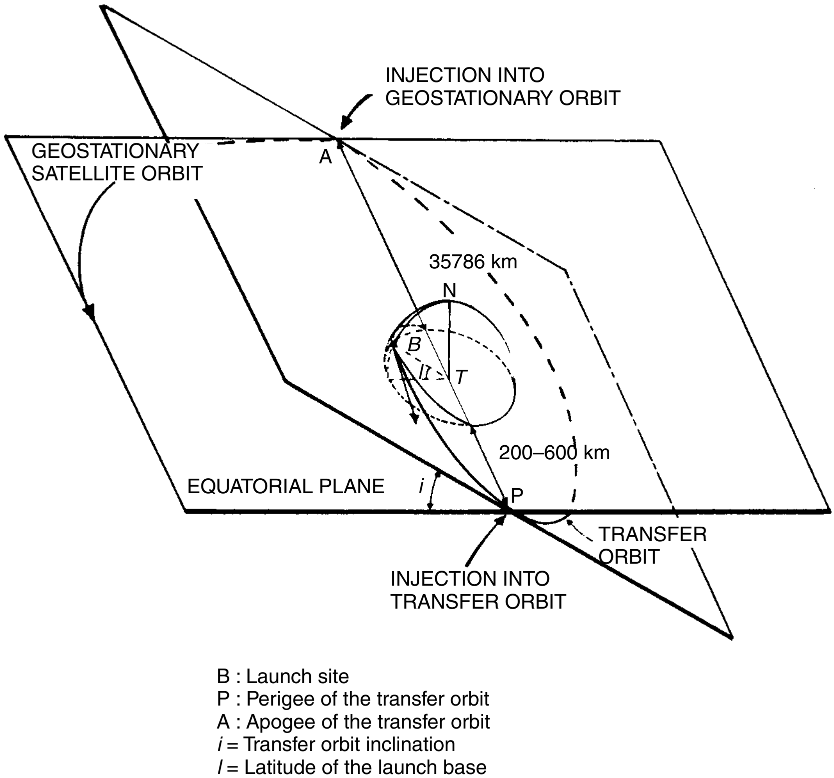


Figure 2.19 Sequence for launch and injection into transfer and geostationary orbit when the launch base is not in the equatorial plane. (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd)

In practice, there are some slight deviations to the above procedure:

1. The launch base may not be in the equatorial plane. The launch vehicle follows a trajectory in a plane which contains the centre of the earth and the launch base (Figure 2.19). The inclination of the orbit is thus greater than or equal to the latitude of the launching base, unless the trajectory is made non-planar, but this would induce mechanical constraints and an additional expense of energy. So the normal procedure is to have it planar. Should the launch base not be on the Equator, then the transfer orbit and the final geostationary satellite orbit are not in the same plane, and an *inclination correction* has to be performed. This correction requires a velocity increment to be applied as the satellite passes through one of the nodes of the orbit such that the resultant velocity vector, V_S, is in the plane of the Equator, as indicated in Figure 2.20. For a given inclination correction, the

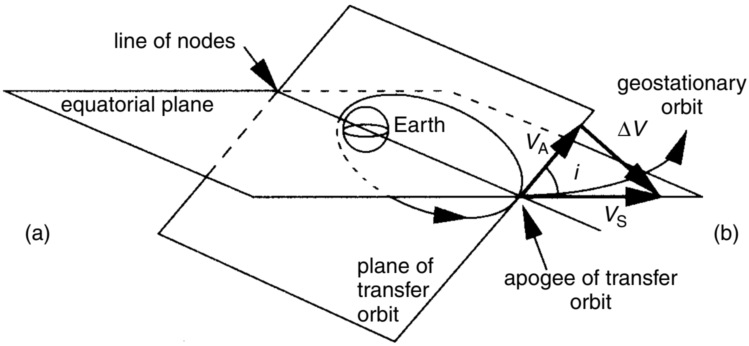


Figure 2.20 Inclination correction: (a) transfer orbit plane and equatorial plane; (b) required velocity increment (value and orientation) in a plane perpendicular to the line of nodes

velocity impulse ΔV to be applied increases with the velocity V_A of the satellite. The correction is thus performed at the apogee of the transfer orbit where V_A is minimum, at the same time as circularisation.

2. A precise determination of the transfer orbit parameters requires *trajectory tracking* during several orbits. Hence, the satellite propulsion system is activated only after several transfer orbit periods.
3. The injection into geostationary orbit does not necessarily take place in the meridian plane of the earth where the geostationary satellite is to be positioned for operation. To reach this position, a relative non-zero small angular velocity between the satellite and the earth must be kept so that the satellite undergoes a longitudinal drift. This leads to injecting the satellite from transfer orbit into a circular orbit, called *drift orbit*, with a slightly different altitude than that of geostationary satellite orbit. Once the satellite has reached the intended station longitude, a correction is initiated by activating the thrusters of the satellite orbit control system.

2.3.3 Distance to the satellite

The distance from an earth station to the satellite impacts on the propagation time of the radio frequency carrier and hence on the delay for information delivery (see Chapter 4, section 4.6.6). It also determines the path loss which intervenes in the link budget calculation (see Chapter 5).

Figure 2.21 displays the geometry of the position of the earth station with respect to the satellite.

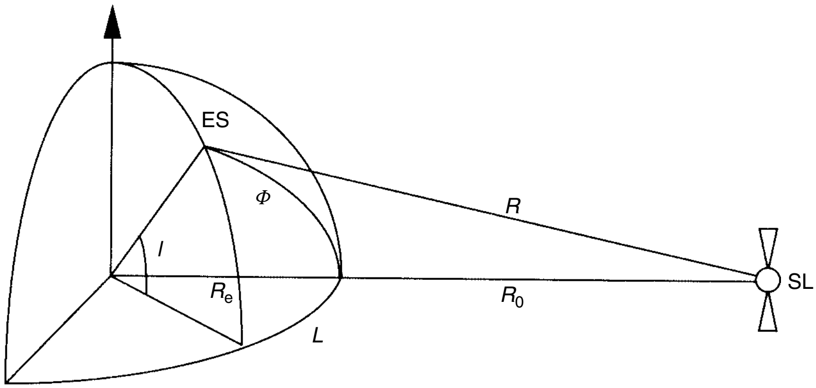


Figure 2.21 Relative position of the earth station (ES) with respect to the satellite (SL)

If we denote by l the geographical latitude of the earth station, and L the difference in longitude between that of the earth station and that of the satellite meridian, the distance R from the satellite to the earth station is then given by:

$$R = [(R_e^2 + (R_o + R_e)^2 - 2R_e(R_o + R_e)\cos\Phi]^{1/2} \quad \text{(m)} \qquad (2.8)$$

where:

$$R_e = \text{earth radius} = 6378 \text{ km}$$

$$R_o = \text{satellite altitude} = 35\,786 \text{ km}$$

$$\cos\Phi = \cos l \cos L$$

With the above numerical values, equation (2.8) can be written as:

$$R = R_o[1 + 0.42(1 - \cos\Phi)]^{1/2} \quad \text{(m)} \qquad (2.9)$$

2.3.4 Propagation delay

The single hop propagation delay (from earth station to earth station) is given by:

$$T_p = \frac{2R}{c} = \frac{2R_o}{c}[1 - 0.42(1 - \cos\Phi)]^{1/2} \quad \text{(s)} \qquad (2.10)$$

where c is the velocity of light $= 3 \times 10^8$ ms^{-1}.
Figure 2.22 displays T_p as a function of l and L.

2.3.5 Conjunction of the sun and the satellite

Conjunction of the satellite and the sun at the site of the earth station means that the sun is viewed from the earth station in the same

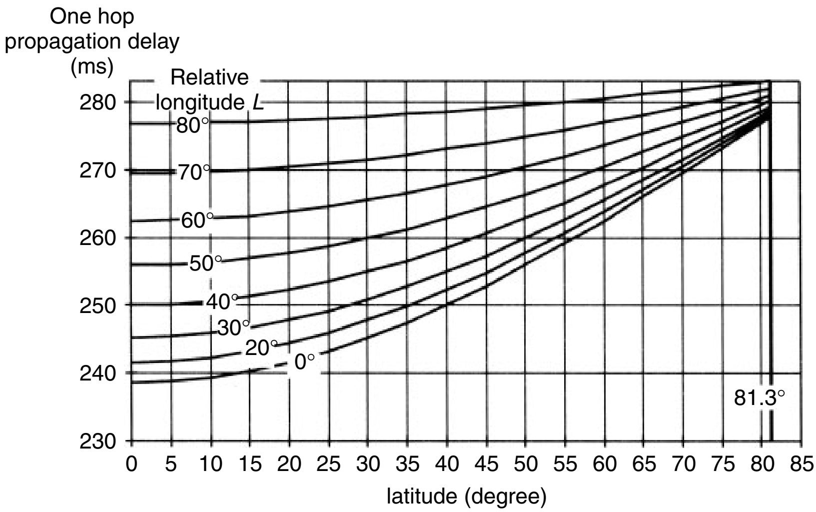


Figure 2.22 Single hop propagation delay as a function of the earth station latitude, *l*, and its relative longitude, *L*, with respect to the geostationary satellite meridian

direction as the satellite. As the earth station antenna is pointed towards the satellite, it now becomes also pointed towards the sun. The antenna captures the radio frequency power radiated by the sun and this increases the noise at the antenna. The antenna noise increase is discussed in section 3.3.10.

As the satellite rotates along with the earth, conjunction of the satellite and the sun is a momentary event. It is predictable and actually happens twice per year for several minutes over a period of five or six days [MAR02, Chapter 2]:

- before the spring equinox and after the autumn equinox for a station in the northern hemisphere;
- after the spring equinox and before the autumn equinox for a station in the southern hemisphere.

2.3.6 Orbit perturbations

Actually, a geostationary satellite does not exist: indeed, Newton's law considers an attracting force exerted on the satellite by a point mass, and oriented towards that point mass. Actually, the earth is not a point mass, there are other attracting bodies apart from the earth, and other forces than attraction forces are exerted on the satellite. These effects result in orbit perturbations.

For a geostationary satellite, the major perturbations originate in:

– the earth neither being a point mass nor being rotationally symmetric: this produces an asymmetry of the gravitational potential;
– the presence of the sun and the moon as other attracting bodies;
– the radiation pressure from the sun, which produces forces on the surfaces of the satellite body facing the sun.

These effects are discussed in detail in [MAR02, Chapter 2]. The practical consequences are summarised below:

– the asymmetry of the gravitational potential generates a *longitudinal drift* of the satellite depending on its station longitude. Actually, there are four equilibrium positions around the earth where this drift is zero, two of which are stable (at 102° longitude west and 76° longitude east) and two unstable (at 11° longitude west and 164° longitude east). Left to itself, a geostationary satellite would undergo an oscillatory longitudinal drift about the stable positions with a period depending on its initial longitude relative to the nearest point of stable equilibrium. The evolution of the longitude drift with respect to a point of stable equilibrium is shown in Figure 2.23.

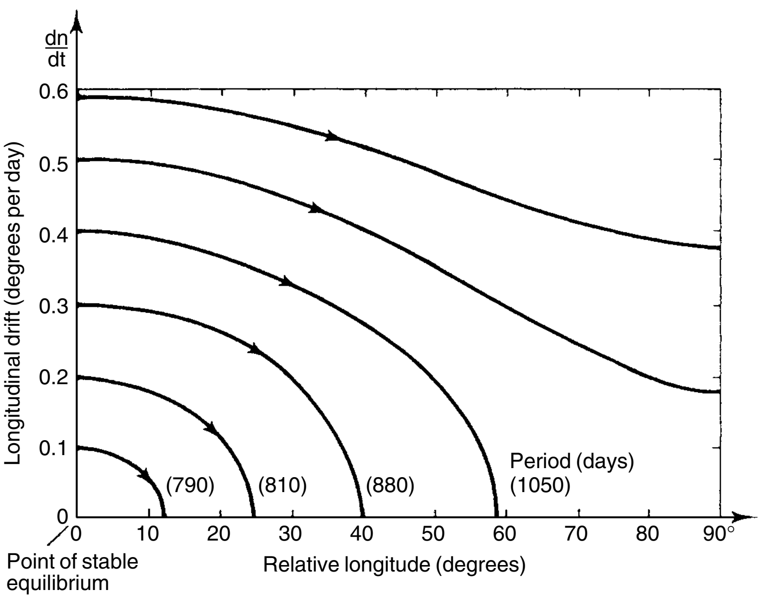


Figure 2.23 Evolution of the longitude drift of a geostationary satellite as a function of the longitude with respect to a point of stable equilibrium. (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd)

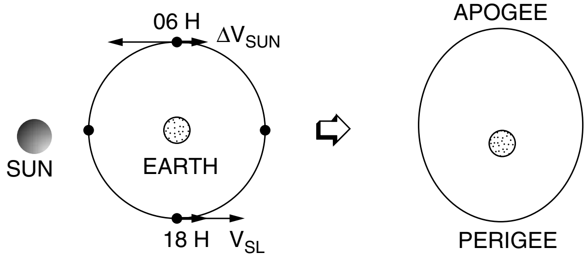


Figure 2.24 Effect of sun radiation pressure on the eccentricity of the orbit. (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd)

- the attraction of the moon and the sun modifies the *inclination of the orbit* at a rate of about 0.8° per year;
- the radiation pressure from the sun creates a force which acts in the direction of the velocity of the satellite on one half of the orbit and in the opposite direction on the other half. In this way the circular orbit of a geostationary satellite tends to become *elliptical*, as illustrated in Figure 2.24.

The ellipticity of the orbit does not increase constantly: with the movement of the earth about the sun, since the apsidial line of the satellite orbit remains perpendicular to the direction of the sun, the ellipse deforms continuously and the eccentricity remains within limits.

2.3.7 Apparent satellite movement

2.3.7.1 Effect of non-zero inclination

The track on the earth of a geostationary satellite with non-zero inclination displays a 'figure of eight' with a 24-hour period, as indicated in Figure 2.25.

This can be understood by considering that the satellite is at its nominal position in the equatorial plane when it passes through the nodes of the orbit, then proceeds on a trajectory that is above the equatorial plane from the ascending node to the descending node and below the equatorial plane from the descending node to the ascending node. This explains the north–south component of the 'figure of eight'. The longitudinal component can be understood by observing that the projection B of the satellite on the equatorial plane, as illustrated in Figure 2.26 by the dotted curve, does not have the constant angular velocity of the point A which is subjected

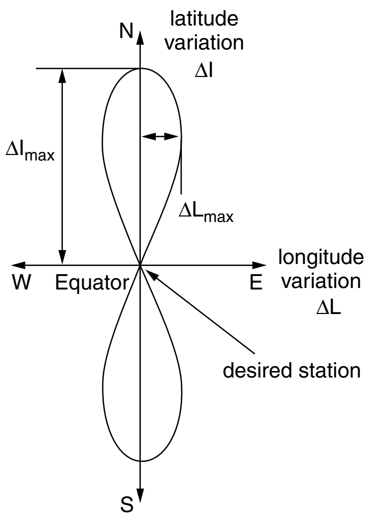


Figure 2.25 Figure of eight as a result of non-zero inclination (24-hour period). (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd)

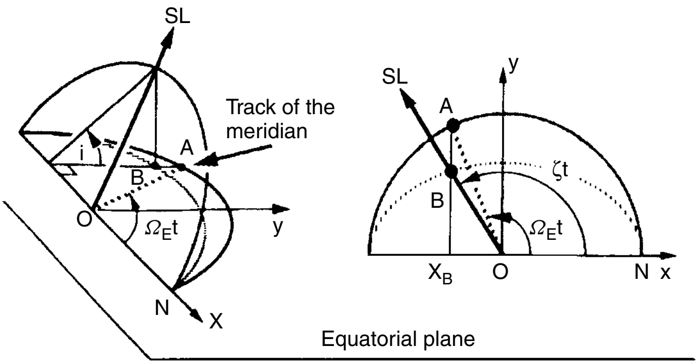


Figure 2.26 Projection of the movement of an earth synchronous satellite in the equatorial plane. (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd)

to the constant angular velocity of the satellite in its orbit. There is consequently an apparent east–west movement of the satellite with respect to the reference meridian on the surface of the earth (that of the satellite on passing through the nodes).

The latitudinal variation corresponding to the amplitude of the figure of eight above or below the Equator is equal to the inclination angle i, and the maximum longitudinal variation with respect to the reference meridian is equal to $4.36 \times 10^{-3} i^2$ (all values in degrees). Therefore for small inclination values, say 0.1°, the figure of eight can be considered to reduce to a north–south oriented segment, with latitudinal extent from $+i$ degrees to $-i$ degrees.

The transformation of the track of the satellite from a dot on the Equator for a perfect geostationary satellite to a north–south oriented segment about this dot for a slightly inclined orbit translates into an apparent movement of the satellite in the sky as viewed from the earth station. This apparent movement leads to a 24-hour period variation of the elevation angle for a station located in the meridian of the nominal satellite meridian, and of the azimuth angle for a station located on the Equator east or west of the satellite. For any other station the apparent motion leads to a combined variation of both elevation and azimuth angles.

2.3.7.2 Effect of non-zero eccentricity

Figure 2.27 illustrates the effect of non-zero eccentricity: the successive positions of two satellites are represented. One is in a circular orbit, and the other is in an elliptical orbit of the same period.

The subsatellite point remains in the equatorial plane and displays a 24-hour period oscillation about its nominal position when the satellite is at perigee or apogee. It can be shown that for small eccentricity values, as considered here, typically less than 0.001, the maximum longitudinal amplitude of the oscillation is $114e$ degrees, where e is the value of the eccentricity [MAR02, Chapter 2].

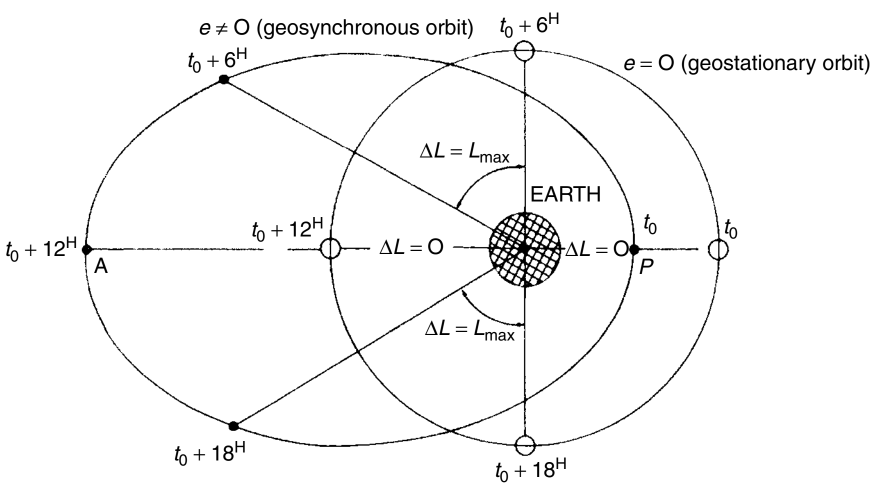


Figure 2.27 Effect of non-zero eccentricity (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd)

2.3.7.3 Effect of east–west drift

The east–west drift of the satellite induces a similar drift of the subsatellite point. This drift is oscillatory and periodic about the nearest point of stable equilibrium. The period and the amplitude of the longitudinal oscillation can be fairly large as indicated by Figure 2.23. Therefore, the apparent movement of the satellite is to be considered as a long-term movement compared to the previously discussed movements.

2.3.7.4 Combined effects

As a consequence of perturbations, the satellite displays an apparent movement in the sky relative to its nominal position. This apparent movement is the resultant of the combined effects of oscillations of period 24 h due to the non-zero inclination (north–south oriented oscillations) and non-zero eccentricity (east–west oriented oscillations) and the long-term drift of the mean longitude (east–west oriented drift). For an earth station, the apparent movement translates into variations of the elevation and azimuth angles with a 24-hour periodic component superimposed on a long-term drift. Figure 2.28 gives an example of such variations.

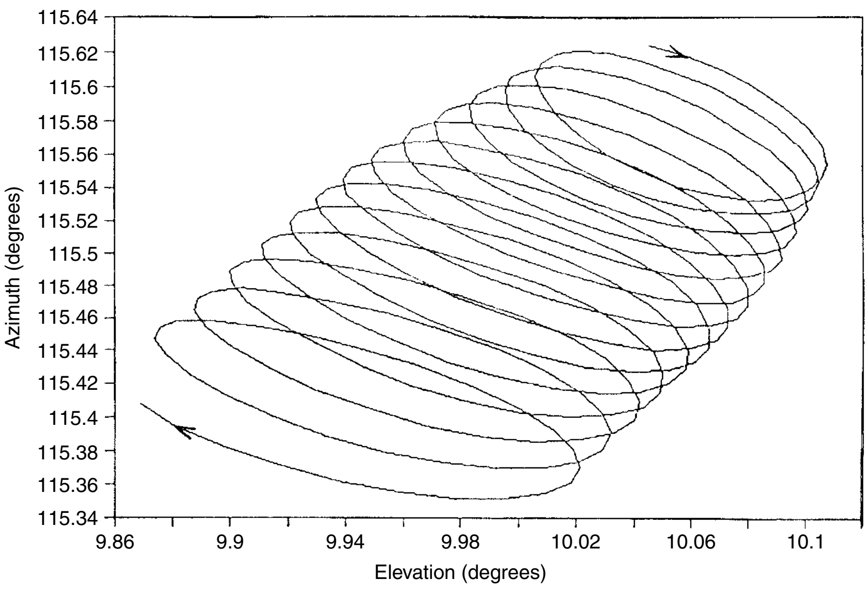


Figure 2.28 Variations of azimuth and elevation angles with time as a result of orbit perturbations. (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd)

2.3.8 Orbit corrections

2.3.8.1 Station keeping box

It is mandatory to take corrective actions to prevent the satellite from straying away from its nominal position in the Equator in a given earth meridian. These corrective actions are part of the so-called 'station keeping' procedures. The objective is to maintain the satellite within a station keeping box, which is defined as the volume in space represented in Figure 2.29. This side of the box viewed from the earth centre is called the station keeping window.

The station keeping window is defined by two angles at the earth centre which limit the maximum excursions of the satellite in longitude and latitude: one within the equatorial plane, the other in the satellite meridian. The maximum value of the residual eccentricity determines the overshoot of the radial distance. The indicated dimensions of the box in Figure 2.29 correspond to a typical window specification of ±0.05° in longitude and latitude, and a residual eccentricity of 0.0004.

2.3.8.2 Correction procedures

To maintain the satellite within the box, orbit corrections are achieved by applying velocity impulses to the satellite at a point in the orbit. These impulses are generated by activating the thrusters that are mounted on the satellite as part of the propulsion subsystem. The satellite can be kept at station as long as there is enough propellant left for the thrusts to be produced. When no propellant is left, the satellite drifts in space out of control. This is the end of its operational life. To avoid a possible collision with other geostationary satellites, satellite operators usually keep some amount of

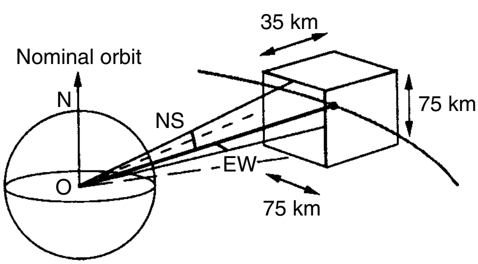


Figure 2.29 Station keeping 'box'. (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd)

propellant for generating a final impulse that positions the satellite into an orbit sufficiently far away from the trajectories followed by its geostationary neighbours.

There is a tendency today, among satellite operators, to relax the north–south specification of the station keeping window. This allows substantial propellent savings, and therefore satellite lifetime extension. However, this results in a higher depointing loss for fixed mounted antennas on the ground (see Chapter 5, section 5.2).

2.3.9 Doppler effect

The Doppler effect is a change in the receive frequency with respect to the transmit frequency as a result of a non-zero velocity of the transmitter relative to the receiver. If the transmitter transmits at a frequency f, the frequency received by the receiver is $f \pm \Delta f$. The frequency shift Δf is given by:

$$\Delta f = \frac{fV_r}{c} \quad \text{(Hz)} \tag{2.11}$$

where V_r is the absolute value of the relative velocity of the receiver with respect to the transmitter and c is the speed of light ($c = 3 \times 10^8\ \mathrm{ms}^{-1}$).

Communications with geostationary satellites experience a small Doppler effect as a result of the movement of the satellite within its station keeping window. With non-regenerative satellites, the Doppler effect acts twice: once on the uplink, with value Δf_U, and a second time on the downlink with value Δf_D, so the maximum overall frequency shift $\Delta f_{T,max}$ at the receiving earth station is given by:

$$\Delta f_{T,max} = \frac{(f_U + f_D)V_r}{c} \quad \text{(Hz)} \tag{2.12}$$

A typical maximum value for the satellite relative velocity $V_{r,max}$ is 10 km/h^{-1}, i.e. 3 ms^{-1}, so this generates a maximum frequency shift $\Delta f_{T,max}$ with respect to the transmitted carrier typically equal to 100 Hz at C-band and 260 Hz at Ku-band. This must be accounted for in the design of demodulators, especially at low data rate (a few kbs^{-1}), by implementing carrier recovery devices with the ability to track the carrier over the expected frequency span.

2.4 SATELLITES FOR VSAT SERVICES

The selection of satellites for VSAT services entails technical, administrative and commercial aspects.

Table 2.2 Typical values of EIRP and G/T for geostationary satellites

	Type of coverage	EIRP	G/T
C-band	Global beam	24 to 30 dBW	−13 to −8 dBK^{-1}
	Zone beam	30 to 36 dBW	−8 to −3 dBK^{-1}
	Spot beam	36 to 42 dBW	−3 to +3 dBK^{-1}
Ku-band	Zone beam	36 to 42 dBW	−7 to −1 dBK^{-1}
	Spot beam	42 to 52 dBW	−1 to +5 dBK^{-1}

First the satellite must be positioned at a longitude where it is visible from any earth station in the network. The longitude of existing satellites can be obtained from the ITU, but the documents identify the registered satellites, many of which may not yet be or may never become operational. Apart from the ITU, relevant information can be obtained from the websites of satellite operators. Then the elevation angle should be computed for extreme stations and checks made that it is large enough (10° is a minimum value in open areas).

Second, the satellite coverage should be matched to the network geographical extension. So further investigation should include examination of maps indicating contours at constant EIRP (effective isotropic radiated power) and G/T (receiving figure of merit). As such contours correspond to constant satellite antenna gain, the achieved values are dependent on the type of coverage: a global coverage offers a worse link performance than zone or spot beam coverage. Candidate satellites should be selected upon the minimum required values of the EIRP and G/T which allow the requested link performance.

Third, a check of all station pointing angles (azimuth and elevation) should be performed for all planned sites in such a way that no obstacles prevent the earth station from accessing the satellite, once installed.

Finally, the selection procedure includes negotiations related to regulatory and financial matters.

Table 2.2 indicates typical values of EIRP and G/T for geostationary satellites, depending on the type of coverage and frequency bands.

3

Operational aspects

This chapter aims to provide a survey of the main items to be considered when installing and operating a VSAT network. Installation is considered first. Then the user's most obvious concerns are discussed.

3.1 INSTALLATION

3.1.1 Hub

As the hub is relatively large, the installation of it is relatively complex and expensive. Civil works may be necessary. Typically, it takes between one and four weeks to install a hub station depending on its size and the selected site. This does not include on-site testing of the equipment.

3.1.2 VSAT

The major problem with VSAT installation is that it involves potentially hundreds of remote VSAT locations with a very wide variety of users, landlords, site conditions, and local zoning requirements.

VSATs can be roof mounted (Figure 3.1), wall mounted (Figure 3.2) or ground mounted (Figure 3.3). A non-penetrating roof mount makes use of an angle iron frame covered in concrete slabs. The antenna support tube is held vertical by several angled braces. A penetrating roof mount is fixed to a horizontal surface using expansion bolts or chemical fixings (bolts that are permanently

VSAT Networks, 2nd Edition. G. Maral
 ISBN: 0-470-86684-5

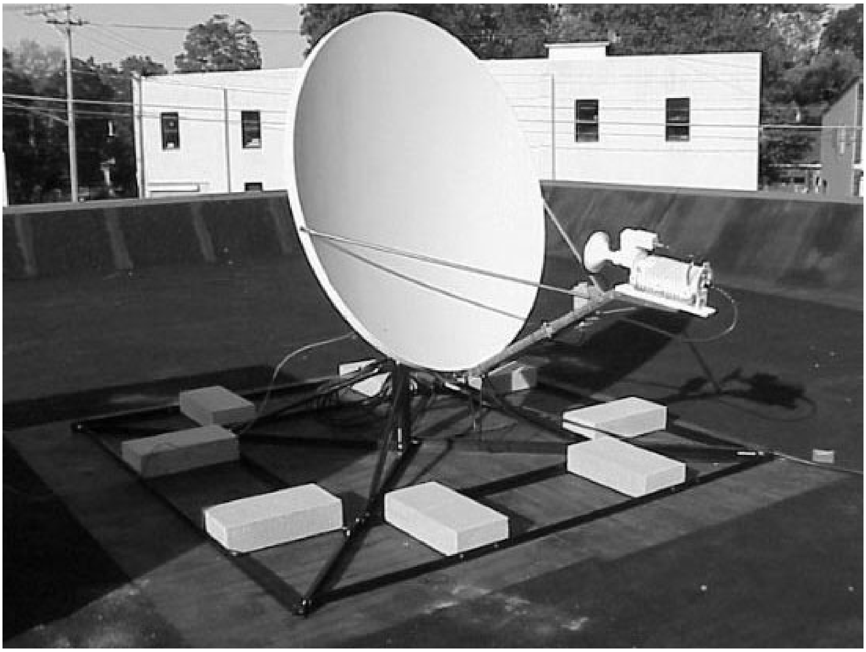


Figure 3.1 Non-penetrating roof mount for VSAT. http://www.freedom-commconsulting.com

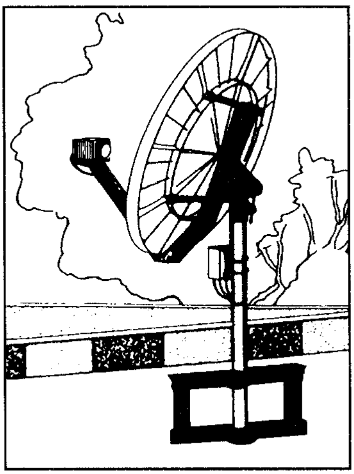


Figure 3.2 Wall mount for a VSAT (JON88). Reproduced by permission of Nelson Publishing Inc.

'glued' into holes made in the surface using epoxy resin). Wall mounting makes use of plates or assemblies of struts designed to be fixed on the face of a wall, or on the inside corner of two walls. Ground mounting involves a tube lowered into a hole which is then filled in with concrete. Alternatively, the tube may have a base plate

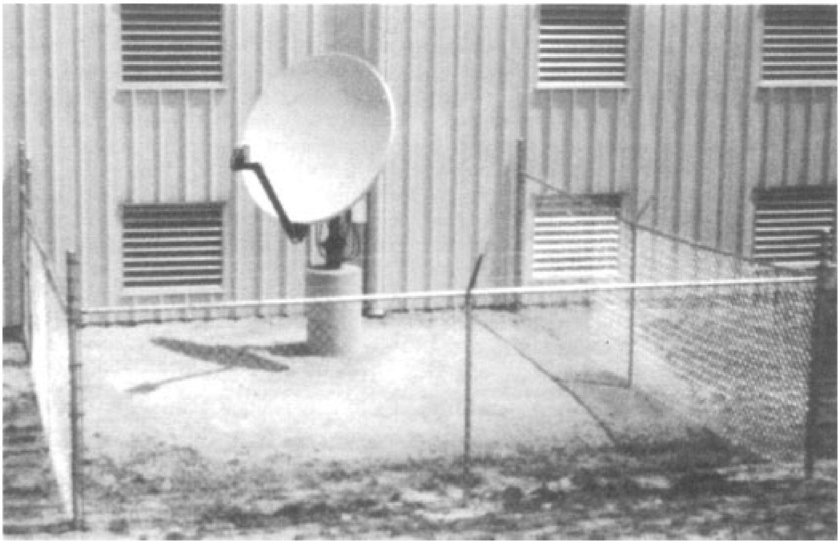


Figure 3.3 VSAT surrounded by fences on the user's premises (SAL88). Reproduced by permission of Nelson Publishing Inc.

attached so that it may be screwed to a plain concrete base using expanding bolts or similar. When ground mounted, the VSAT is best secured with fences (Figure 3.3) to prevent people or animals from getting hurt or damaging the outdoor unit. However, fences are not a strong protection against vandalism.

A typical VSAT installation generally requires three visits to each site: a site survey, basic site preparation, equipment installation and test. About 20% of the sites will require an installation revisit.

The following figures may be used for installation planning: the average VSAT cable run is 60 metres, 30% of VSATs are ground mounted, 70% roof mounted. Possibly 15% of roof mounts require special engineering.

3.1.3 Antenna pointing

Accurate antenna pointing is paramount for transmitting and receiving maximum power to and from the satellite the earth station antenna aims at. At least two angles need to be considered in the pointing procedure:

- the azimuth angle, Az;
- the elevation angle, E.

If the transmission is based on linearly polarised carriers, a third angle should be considered: the polarisation angle, ψ.

Figure 3.4 introduces the azimuth angle and the elevation angle:

– the azimuth angle, Az, is the rotation angle about a vertical axis through the earth station counted clockwise from the geographical north, which brings the antenna boresight into the vertical plane that contains the satellite. This plane contains the centre of the earth, the earth station and the satellite. The value of Az is obtained by means of an intermediate parameter, a, determined from the family of curves of Figure 3.5 and used to calculate Az according to the table shown in the figure [SMI72]. The curves are obtained from the following expression which can be used for greater accuracy:

$$a = \text{arc tan}\left(\frac{\tan L}{\sin l}\right) \quad \text{(degrees)} \tag{3.1}$$

where l is the geographical latitude of the earth station and L is the absolute difference between the longitude of the satellite and the longitude of the earth station.

– the elevation angle E is the rotation angle about a horizontal axis perpendicular to the above-mentioned vertical plane counted from 0° to 90° from the horizontal, which brings the antenna boresight in the direction of the satellite. The elevation angle is obtained from the corresponding family of curves of Figure 3.5 which correspond to the following expression:

$$E = \text{arc tan}[(\cos \Phi - R_e/(R_e + R_o))/(1 - \cos^2 \Phi)^{1/2}] \quad \text{(degrees)} \tag{3.2}$$

where:

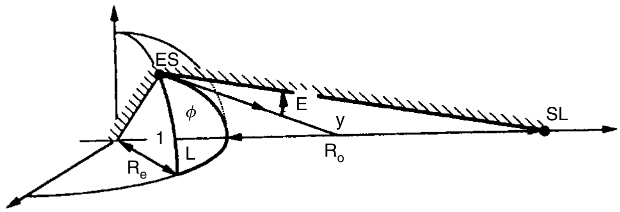
$$\cos \Phi = \cos l \cos L$$

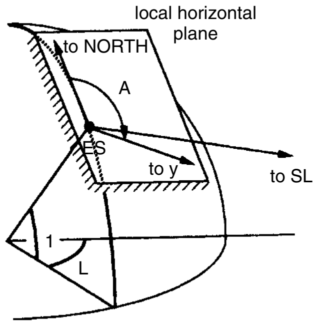
$$R_e = \text{radius of the earth} = 6378 \text{ km}$$

$$R_o = \text{altitude of the satellite} = 35\,786 \text{ km}$$

Expressions (3.1) and (3.2), or Figures 3.4 and 3.5 can be used for a coarse orientation of the antenna. The azimuth angle is defined from the geographic north while the magnetic north is given by a compass used on the site. The difference is the magnetic declination, the value of which depends on the site location and the year. The elevation angle must be measured from the horizon, which is defined by the local horizontal plane and easily determined from a spirit level.

At Ku-band the polarisation of the wave received from the satellite is most often linear, and the earth station antenna feed must have its polarisation aligned with the polarisation plane of the received



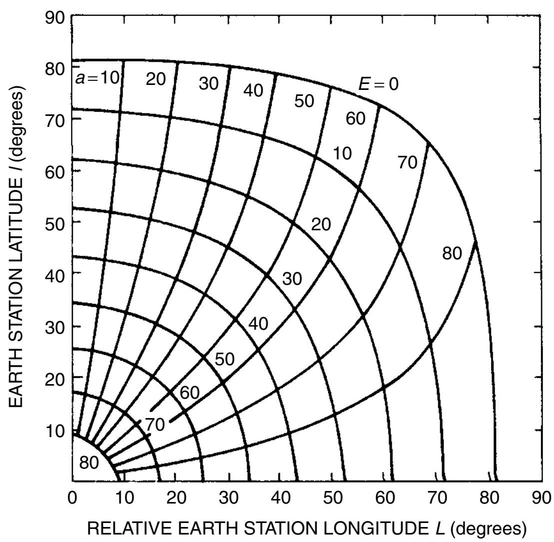


	SL East of ES	SL West of ES
ES in NH*	A = 180 − a	A = 180 + a
ES in SH*	A = a	A = 360 − a

* NH = North hemisphere
SH = South hemisphere

with:
a = Arctan (tan L/sin 1)

Figure 3.4 Definition of azimuth and elevation angles. (ES: earth station, SL: satellite). (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd)



	SL EAST OF ES	SL WEST OF ES
NORTH HEMISPHERE	$A = 180 - a$	$A = 180 + a$
SOUTH HEMISPHERE	$A = a$	$A = 360 - a$

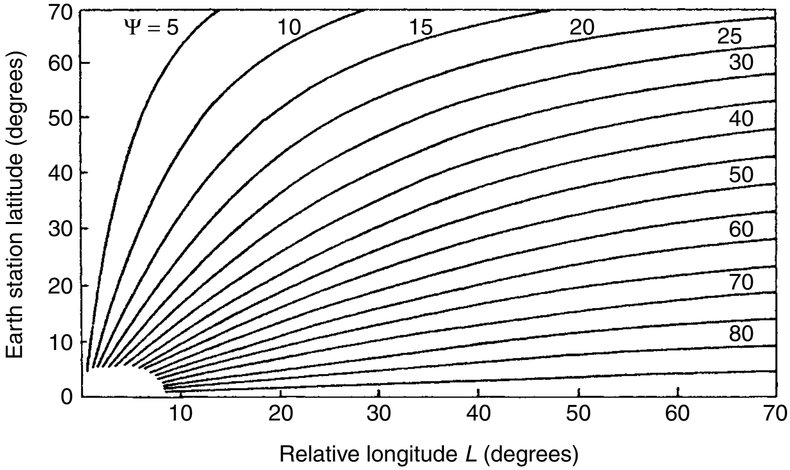
Figure 3.5 Azimuth and elevation angles as a function of the earth station latitude *l* and satellite relative longitude *L*. (Reproduced from (MAR02) by permission of John Wiley & Sons Ltd)

wave. This plane contains the electric field of the wave. The polarisation plane at the satellite is defined by the satellite antenna boresight and a reference direction. This reference direction is, for instance, the perpendicular to the equatorial plane for vertical polarisation (VP) or parallel to the plane of the Equator for horizontal polarisation (HP). The *polarisation angle* at the earth station is the angle ψ between the plane defined by the local vertical at the earth station and the antenna boresight, and the polarisation plane. $\psi = 0$ corresponds to the reception or emission at the earth station of a linearly polarised wave with its polarisation plane containing the local vertical. The polarisation angle at the earth station for a reference direction at the satellite in a plane perpendicular to the equatorial plane is given by [MAR02 p391]:

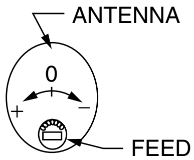
$$\cos\psi = \frac{\sin l}{\sqrt{(1 - \cos^2 l \cos^2 L)}} \tag{3.3}$$

where l and L are defined as in (3.1). Figure 3.6(a) displays absolute values of the polarisation angle ψ according to equation (3.3). The feed of the antenna should be rotated from the vertical by an angle ψ either clockwise or counterclockwise when facing the antenna dish, depending on the location of the earth station (northern or southern hemisphere) and its relative position with respect to the satellite (eastward or westward). The sense of rotation is indicated in Figure 3.6b.

Once coarse orientation has been achieved, a more precise orientation is performed according to maximisation of the power received from a satellite beacon or a downlink carrier. For a large hub equipped with a tracking antenna, the tracking equipment can be activated and the orientation of the antenna will remain in the direction of the satellite within the precision of the tracking equipment, whatever the subsequent motion of the satellite within its station keeping window. The tracking error is of the order of 0.2 θ_{3dB}, where θ_{3dB} is the half power beamwidth of the earth station antenna (see Appendix 4 for definition). Small hub stations and VSATs are not equipped with a tracking antenna and the orientation of the antenna will remain at its initial pointing, assuming that no severe hazardous constraint acts upon the antenna equipment (e.g. a blow or strong wind). Any subsequent motion of the satellite translates into a depointing angle, and the corresponding loss of gain has to be accounted for in the link margin. The maximum gain loss then depends on the initial pointing error and the limits of the satellite motion. This is discussed in more detail in Chapter 5, section 5.2.



(a)

	VSAT in northern hemisphere	VSAT in southern hemisphere
VSAT west of satellite	−	+
VSAT east of satellite	+	−

(b)

Figure 3.6 (a) Polarisation angle ψ as a function of the VSAT latitude l and satellite relative longitude L considering a reference polarisation plane of the satellite perpendicular to the equatorial plane, (b) sense of feed rotation with respect to vertical when facing the antenna, depending on the position of the VSAT

3.2 THE CUSTOMER'S CONCERNS

A VSAT network most often replaces an existing leased line data network. The reasons for using VSAT services are, in order: cost savings, flexibility, reliability, data rates supported and no other services meet needs. This section attempts to list some aspects to be looked at when considering VSAT technology.

3.2.1 Interfaces to end equipment

The indoor unit (IDU) is the part of the network most visible to the user, as it is most often installed in his own office. The IDU is the terminating equipment of the VSAT network, to which the user connects his own terminals. The IDU incorporates a number of input/output ports with specific connectors which must be compatible with that of the user's terminal.

With data networks, the customer wants to be able to use satellite channels and VSATs in a manner which is transparent to existing and future applications. Often the customer is interested in replacing an existing network but he is usually not willing to replace current equipment such as cluster controllers, front end processors, or other data concentration devices, nor to change the interfaces to that equipment. A customer may be reluctant even to reconfigure the equipment by changing device addresses or the duration of timers [EVE92, p 156–157]. Therefore, it is important that all physical interfaces be software defined and downline loadable from the network management system (NMS) located at the hub station. Modifications to individual operational interfaces within a VSAT should not affect other operational interfaces at the same location.

3.2.2 Independence from vendor

The general functions of a VSAT network, as discussed in Chapter 1, are the same across all vendor products. However, each VSAT has a proprietary design and proprietary protocols. Therefore, VSAT equipment from different vendors cannot share the same satellite channels nor the same network hub equipment in the case of a star network [EVE92, p 161].

3.2.3 Set-up time

This topic encompasses two aspects:

1. The time required to set up the network in a given initial configuration: typically, it takes 90 days to implement a hundred node network.
2. The time to expand the network by addition of new sites: a VSAT can be added within a few days. This compares favourably with several weeks waiting time for the installation of a terrestrial leased line.

With satellite news gathering (SNG), the VSAT can be installed and in operation typically within twenty minutes.

3.2.4 Access to the service

Many VSAT networks are initially one way networks used, for instance, for broadcasting of video. Most often, the customer then wishes to upgrade the service into a two-way network for data transmission. Alternatively, broadcast video is a cheap option once the network is installed for data transmission.

It may be worthwhile for the network operator to ask the network provider to perform installation tests prior to full deployment of the network. This is an opportunity for equipment testing, and also checking that the requested service is actually offered over the subnetwork under test. At this point, the network provider can proceed to traffic measurements and check that the actual traffic is in conformity with design assumptions. Moreover, should the client not be satisfied, the network can be turned down at little expense compared to the cost of turning down the full network.

3.2.5 Flexibility

One of the main advantages of VSAT networks is that network expansion, addition of new terminals and provision of new services can be accommodated without reconfiguring or impacting the operation of the rest of the network.

However, the performance of the network and hence the quality of the service delivered to the user are sensitive to the amount of traffic, which increases as more and more terminals of VSATs are added to the network. It is therefore important to allow spare capacity in the space segment and the hub, typically 20% more traffic and 20% more VSATs than initially expected. Growth beyond initial capacity must be orderly and modular. Considering that frequent acquisitions and corporate restructuring are part of today's business world, it is important that the customer not be constrained on its potential growth in telecommunication needs.

3.2.6 Failure and disaster recovery

As telecommunications are a sensitive part of a company's ability to support its business, the customer is concerned with the general feeling that satellite communications are risky by nature, as the network operation relies on a single satellite far away in space without the possibility of repair. Most company managers have very little natural confidence in this telecommunications technology which is unknown to them. It is therefore important to establish adequate failure monitoring and diagnostic facilities, restoration procedures, and consistent disaster recovery scenarios. The disaster scenario should be adapted to the customer's particular requirements.

Actions should include the following:

– hub restoration;
– VSAT restoration;
– satellite back-up;
– back-up terrestrial connections.

A hub failure may only affect some of the hub functions and still allow a reduced capability in networking. Should the hub fail, or be destroyed, to a point where the network suffers a complete breakdown, one may consider a properly equipped fixed or transportable earth station to resume immediate operations with no changes to either the satellite or the VSATs. The shared hub option, presented in Chapter 1, section 1.6.5, with its landline connections to the host site, may be more prone to disaster. Further, with many users clamouring for service, priority for restoration could be a problem at the shared hub. The shared hub operator must have a sound, tested plan for such an occurrence.

The network management system (NMS) should perform a centralised failure identification and diagnostic functions at each VSAT. Failure to the card level should be detectable at the NMS. The failure of a VSAT station implies an event which cannot be rectified by commands and the downloading of parameters from the NMS. The successful handling of failures requires accurate and timely detection of the fault. The inclusion of built-in test equipment (BITE) in the VSAT station is essential to support this monitoring facility [EVE92, p 202].

In case the failure threatens the network integrity, for example if the impaired VSAT transmission would generate interference to other links, there should be immediate termination of transmission from that terminal. A solution is to implement a continuous hub signal which is monitored by the VSAT. The VSAT automatically stops its transmission when not receiving the hub signal.

Satellite failures are rare, but over the typical fifteen year lifetime of a satellite, one must be prepared to face some kind of failure. Satellite depointing is the most probable event and results in a dramatic network breakdown. However, it normally takes no more than a few hours to bring back the satellite to normal operation, and the networking interruption is accounted for in the network availability.

Transponder failure requires shifting the network to another transponder on the same satellite. This possibility is highly dependent on the contractual conditions between the satellite operator and the network operator: the satellite capacity may be leased either as non preemptible or preemptible. Non-preemptible lease means that the satellite operator warrants the use of the transponder bandwidth and commits himself to do his best to offer the same bandwidth on another transponder in case of failure of the leased one. Preemptible means that the leased capacity cannot be guaranteed over time, and the network operator may be asked to give back the used bandwidth on request from the satellite operator.

Migrating to another transponder on the same satellite implies changing the operating frequencies or polarisation of the entire network. This must be planned in advance so that in case of signal loss for a predetermined time, the VSAT could automatically tune to a new frequency and a new polarisation plane and search for signals from the hub. It should be possible to download the back-up assignment of frequencies and polarisation from the network management system (NMS) to take into account possible updating of the back-up transponder scenario.

Finally, there exists a possibility for total satellite failure and subsequent necessity to migrate to another satellite. This implies repointing all remote VSAT antennas, which can be done manually but takes time, especially with large networks, or automatically, but at a higher cost per VSAT.

In any case, partial or complete breakdown of the network can be avoided if back-up terrestrial connections are available. Should a link be disrupted, the traffic on that link can be routed automatically by automatic dial-up modems to a public, either circuit or packet switched, terrestrial network. Figure 3.7 shows a possible implementation for remote-to-host back-up interconnections.

The sensing of link failure and the automatic recovery via the terrestrial network increases the service availability. Vendors usually offer such features.

3.2.7 Blocking probability

Blocking probability is considered in relation to demand assignment operation, when the total number of VSATs registered in the network

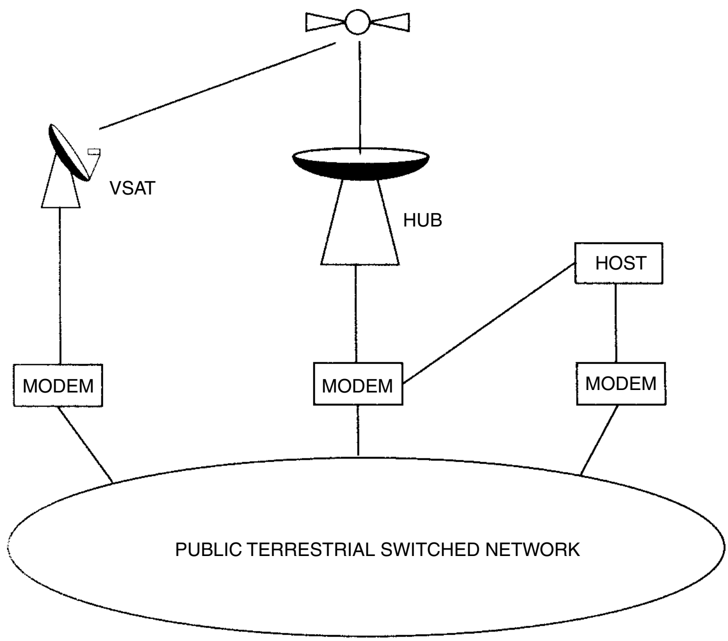


Figure 3.7 Implementation of remote-to-host back-up interconnections

possibly generates a traffic demand that exceeds the capacity of the network. When a station needs to establish a connection with another or with the hub, it initiates a request to the network management system (NMS), and this request is satisfied only if capacity is available. If not, the call is blocked. Chapter 4, section 4.3 gives means for determining the blocking probability. For VSAT networks it ranges from 0.1% to 1%.

3.2.8 Response time

Response time is defined as the time elapsed between emission of speech and reception of the other talker's response in the case of voice telephony communications, or time elapsed between transmission of an enquiry message initiated by pressing the return key of the computer keyboard and the appearance of the first character of the response message on the computer screen.

Response time for data transfer builds up from several components:

– queuing time at the transmitting side as a result of possible delay for capacity reservation before transmission occurs;

– time for transmission of the emitted message which depends on the length of the message and the transmission bit rate;
– propagation time which depends on the network architecture and the number of satellite hops: for a single hop, the propagation time is 0.25 s, and 0.5 s for a double hop. This propagation time occurs on the ongoing link from transmitter to receiver and on the return link;
– processing time of the enquiry message at the receiver, and time necessary for generating and transmitting the response;
– protocol induced delay, as a result of error recovery and flow control between the transmitting site and the receiving one.

The VSAT network is only responsible for the routeing delay, which includes propagation delay and processing delay, as a result of protocol handshake between VSATs and hub front end processor, but excludes the processing delay of the data terminal equipment.

A more detailed analysis of the origin of network delays is given in Chapter 4, section 4.6.

Contrary to a well established belief, a VSAT network will very likely result in a much better response time than the typical private line network. The only physical limitation is the 0.5 second round trip satellite transit delay. The issue then becomes one of cost: how short does the response time really need to be?

3.2.9 Link quality

VSAT networks typically offer a bit error rate (BER) of 10^{-7}. This guarantees an acceptable quality for digital voice and video. For data communications, the bit error rate is not a significant parameter, as the transmission can be made error-free thanks to the retransmission protocols that are usually implemented between end-to-end terminals. However, the bit error rate influences the number of required retransmissions, and hence influences the delay.

As a result of the symmetry of all links, VSAT networks provide the same service quality to each user. This may not be the case for terrestrial networks.

3.2.10 Availability

In general terms, availability is defined as the ratio of the time a unit is properly functioning to the total time of usage:

$$A(\%) = \frac{100\ (\text{total usage time} - \text{down-time})}{\text{total usage time}} \qquad (3.1)$$

Network link availability is the percentage of time the service is delivered at a given site with the requested quality (bit error rate less than specified value, for instance one part in 10^7, response time within specified limits, for instance less than five seconds). Network availability builds upon equipment reliability, propagation impairments, and sun outage.

More precisely, network availability can be expressed as:

$$A_{\text{net}} = A_{\text{Tx}} A_{\text{sat}} A_{\text{link}} A_{\text{Rx}} \tag{3.2}$$

where A_{Tx} is the transmitting earth station availability, A_{sat} the space segment availability, A_{link} the link availability, and A_{Rx} the receiving station availability.

Table 3.1 gives some typical figures.

A 99.7% availability corresponds to a cumulated down-time of 26 hours per year. However, it is likely that the user will not accept a service interruption lasting typically more than four hours in a row. Should the service interruption be caused by equipment failure, an appropriate maintenance procedure should be implemented to restore the service within the requested time. Should propagation impairments be responsible for the service interruption, then site diversity can be considered. Finally, back-up terrestrial connections may be a means to achieve service continuity.

3.2.10.1 Earth station availability (transmit or receive)

There are two aspects to this topic: (a) equipment failure; (b) antenna depointing.

Equipment failure

A typical mean time between failures (MTBF) for an earth station is 50 000 hours (6 years). The availability of a remote VSAT station depends on the total repair time. This depends on how easy it is to access the equipment. Spare parts are usually easy to get.

Table 3.1 Typical figures for availability

Equipment	Availability (%)
Remote VSAT	99.9
Space segment	99.95
Link	99.9
Hub	99.999
Network	99.7

Typically, the repair time is from a few hours to a few days. Hence, availability per remote VSAT is typically 99.9% (9 hours per year down-time). For a hub station where there is built-in redundancy, the equipment availability is higher, typically 99.999% (five minutes per year down-time).

Antenna depointing

This may happen as a result of severe mechanical constraints on the antenna reflector, resulting from meteorological events such as strong winds (storms, hurricanes) or heavy snowfall, where wet snow or ice accumulates in the dish. Deicing devices reduce the risk.

3.2.10.2 Space segment availability

There are two aspects to this topic:

– availability of capacity for coping with traffic growth or unexpected demand for a variety of services;
– availability of other transponders should a transponder or the entire satellite fail.

Availability of capacity in case of traffic growth

The network operator should be aware of any available satellite capacity in case he needs to expand. It may be good practice to have spare capacity on the same satellite for occasional video once a data network is operating.

Availability of capacity depends on the specific region of the world, and varies with time. Truly, adapting the space segment capacity to the demand is a severe challenge for satellite operators, as estimating the demand and scheduling satellite launches to meet the demand entails predictions over a period of ten to twenty years, with the uncertainties associated with the spacecraft launching schedule and success rate.

Transponder failure

Should the failed transponder be a non preemptible one, the network can be transferred within hours to another transponder. The deal is risk versus cost: the network operator may want to lease a non-preemptible transponder with the warranty of being given capacity on another transponder in case of failure. But this costs more than a preemptible transponder with no warranty at all. He then faces the risk of being asked to disembark from his working transponder to leave room for a customer on a failed non preemptible one.

Satellite failure

Should the entire satellite become useless, then one must transfer the network to another satellite. This implies that one is available in the same band, and with spare capacity. Transferring to this satellite then imposes an intervention on each site to repoint each antenna. This could take days to weeks, depending on the number of sites. Alternatively, VSATs can be equipped with microprocessor controlled, locally activated automatic repositioning mechanisms. Pointing of the antenna then should be controlled both locally and from the hub station.

3.2.10.3 Link availability

Availability requires that the link performance in terms of carrier power to noise power spectral density, C/N_0, be larger than a given value for the considered percentage of time. C/N_0 varies according to propagation effects (mainly rain fade) and sun transit (increase in noise). These effects tend to decrease C/N_0 below its required value and cause link outage as illustrated in Figure 3.8.

Given a specific margin, C/N_0 decreases with rain and sun transit, and may become lower than the required value. C/N_0 resumes its value with margin when rain ceases or sun transit is over. The down-time is smaller as the margin increases; hence, a larger margin value leads to a higher link availability.

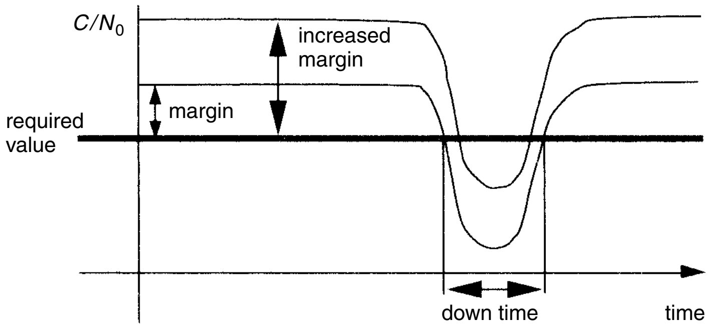


Figure 3.8 Link outage as produced by variation of C/N_0 with rain and sun transit

Rain fade

Rain is an issue at Ku-band and even more at Ka-band. It has little effect at C-band. A quantitative discussion of rain effects is presented in Chapter 5.

For voice and video services, reduction of C/N_0 translates into reduction of baseband signal quality (increase in bit error rate, or reduction of baseband signal power to noise power ratio). In data networks, as a result of the end-to-end protocol for error correction, rain fade results in a gradually increasing response time and not a precipitous failure.

An effective way to combat rain attenuation and reduce link outage probability is site diversity [MAR02, Chapter 5]: this means routeing information by either of two stations connected by terrestrial lines, depending on which station is less affected by rain attenuation. The two stations must be sufficiently geographically separated in order not to endure rain simultaneously. As large rain attenuation is caused mostly by storms with limited extension, typically ten kilometres, site diversity is quite feasible with VSAT networks, selecting an appropriate nearby VSAT station as a back-up to the failed one. Some systems provide as an option automatic dial-up in case of short term outage for rerouteing of the ongoing connection to the diverse VSAT, via a terrestrial network which will route data from the failed VSAT to the hub station. Service is automatically restored when the failed VSAT is returned to service.

Sun transit

Sun transit occurs when the conjunction of the satellite and the sun is effective at the site of the earth station. Then the sun passes directly in the line of sight path between the earth station antenna and the satellite, hence the name 'sun transit'.

Conjunction of the satellite and the sun has been introduced in Chapter 2, section 2.3.6. This happens twice per year for several minutes over a period of five or six days:

- before the spring equinox and after the autumn equinox for a station in the northern hemisphere;
- after the spring equinox and before the autumn equinox for a station in the southern hemisphere.

The sun radiation enters the earth station receiving antenna and increases its noise temperature. This results in a reduced C/N_0. The

Table 3.2 Typical peak antenna noise temperatures at sun transit for various antenna sizes

Antenna diameter (m)	Effective rise in sky noise (K)
1.2	700
1.8	1500
2.4	2700
3.7	5800
9	6000

reduction can be calculated using the equations given in Chapter 5, section 5.3.4.

Table 3.2 gives typical values of peak antenna noise temperatures at sun transit for various antenna sizes at Ku-band. When no sun transit effect is present the typical antenna noise temperature at Ku-band is in the range 40 to 60 K (see Chapter 5, section 5.3).

The ability of the link to operate through peak sun transit times depends on the built-in link margin. The margin usually implemented for reducing the down time due to rain effects to the typical desired level of link availability of 99.9% is usually large enough to protect the link from sun transit effects. Therefore, the sun transit has little impact on the link availability.

At C-band, the sun transit effect is less pronounced than at Ku-band.

3.2.11 Maintenance

The maintenance concerns the equipment on the ground: the hub station and the VSATs. The customer may wish to be responsible for all of the maintenance, or have the vendor handle some or all of the work.

Maintenance at a shared hub is normally the responsibility of the hub service provider. For a dedicated hub, the network operator may wish to contract out or to perform the maintenance on his own. In terms of maintenance staff, two different categories are required: radio frequency and data communications.

A VSAT station should require as little maintenance as possible as the operational cost of maintenance over a large number of sites scattered over a large service zone would hamper the operational cost of the network. Therefore, it is highly desirable that the maintenance of the VSAT be performed by local people in charge of other duties. For instance, the local technician who maintains the existing PC network on the site can also perform the normal maintenance of the VSAT station. Self diagnostics and box level repair make his task much simpler. In the case of extensive trouble shooting, he can call the equipment vendor for ad hoc assistance.

The network provider should warrant the availability of the provided hardware and software for a given period, typically two years. He should present a plan as to how all hardware and software will be supported for a minimum of ten years.

3.2.12 Hazards

VSATs are usually located in urban areas or near areas where people and animals may be around. They are usually unattended. Problems are:

3.2.12.1 Protection of people and animals from radiation

The electromagnetic radiation should be kept to a harmless low level: typically not more than 10 $mWcm^{-2}$ per 6 hours. ETSI's ETS 300 159 specifies that a warning notice should be posted indicating regions where the radiation may exceed 10 Wm^{-2} (= 1 $mWcm^{-2}$).

3.2.12.2 Protection of hardware against ill-intentioned people

Fences are a solution (see Figure 3.3), but it is safer if the outdoor equipment is not easily accessible, although this renders maintenance more difficult.

3.2.13 Cost

The cost of a VSAT per month per site has been shown to be dependent on the number of VSATs in the network (see Chapter 1, section 1.8), and the cost of the space segment is a sensitive issue. Unfortunately, in most regions of the world the network operator has little freedom of selecting the satellite operator to contract with. The requested availability also has a strong impact on the network cost, therefore the user should not ask for tight specifications unless strictly needed.

One advantage often advocated by VSAT network operators is the control of communications cost. To the initial investment cost is added the maintenance costs, these can both be under the control of the network operator.

Therefore, cost containment is a fact. As mentioned above, a company most often turns to VSAT technology in replacement of existing leased lines. As the cost of a link in a VSAT network is not distance sensitive, both immediate and long term cost savings compared to terrestrial alternatives will result if the company encompasses a large number of dispersed sites to be connected.

4

Networking aspects

4.1 NETWORK FUNCTIONS

As mentioned in Chapter 1, VSAT networks usually offer a communications service between user terminals. These terminals generate baseband signals that are analogue or digital, predominantly digital.

For signals generated by a source terminal and to be delivered to a destination terminal, the VSAT network must provide the following functions:

– establish a connection between the calling terminal and the called one;
– route the signals from the calling terminal to the called one, although the physical resource offered for the considered connection may be shared by other signals on other connections;
– deliver the information in a reliable manner.

Reliable delivery of *data* means that data is accepted at one end of a connection in the same order as it was transmitted at the other end, without loss and without duplication. This implies four constraints:

– no loss (at least one copy of each part of the information content is delivered);
– no duplication (no more than one copy is delivered);
– first in first out (FIFO) delivery (the different parts of the information content are delivered in the original order);
– the information content is delivered within a reasonable time delay.

VSAT Networks, 2nd Edition. G. Maral
 ISBN: 0-470-86684-5

It has been indicated in Chapter 1 section 1.4 that VSAT networks could be envisaged to support many different types of traffic. However, the network cannot convey all such different types of traffic in a cost effective way. Therefore, VSAT networks are optimised for a given set of traffic types, which reflect the dominant service demand from the user, and may offer as an option other types of service, but not as efficiently. Most VSAT networks are optimised for *interactive* exchange of data.

This chapter aims to present the characteristics of traffic the network may have to convey for interactive data services, and the relevant techniques used for conveying such traffic.

4.2 SOME DEFINITIONS

4.2.1 Links and connections

A link serves as a physical support in a network for a connection between a sending terminal and a receiving one. The network consists of several links and nodes. Every link has two end nodes: one that sends and one that receives.

In a VSAT network, one finds:

- radio frequency links (uplinks and downlinks);
- cable links between the outdoor and the indoor units, or between the indoor unit and the user terminal;
- possibly terrestrial lines (microwaves or leased terrestrial lines, or lines as part of a public switching network) between the hub and the customer's central facility.

Some connections are one-way, thus requiring that information only travels in one direction: for such connections *simplex* links can be used. An example of a simplex link is a radio frequency wave.

Other connections require interactivity, and hence two-way flow of information. It may be that the information flow is not simultaneous in both ways, but alternate. The supporting links for such connections are named *half duplex* links. An example is when a given radio frequency bandwidth is used alternately by two receiving and transmitting units on a 'push-to-talk' mode: one unit transmits on the bandwidth for some time, while the other unit operates in the receive mode. Once this is done, the transmitter turns to the receive mode, and the receiver to the transmit mode, and information flows the other way round.

When the information must travel both ways simultaneously the supporting links are called *full duplex* links. An example is the line from a telephone handset to the indoor unit (IDU).

Radio frequency links of VSAT networks are inherently simplex links, but a connection requiring full duplex links can be implemented using two radio frequency links: one for each direction of information flow. In a star-shaped network, a duplex link between a given VSAT and the hub is constituted for one part by the inbound link and for the other by the outbound link.

A link can support one connection at a time, in a so-called *Single Channel Per Carrier* (SCPC) mode, or be shared by several connections, in a so-called *Multiple Channels Per Carrier* (MCPC) mode. Figure 4.1 illustrates these concepts.

4.2.2 Bit rate

Basically, the bit rate is the number of bits transferred per time unit (second) on a given link. A distinction should be made between:

- the information bit rate R_b, which is the rate at which information bits conveying data messages of interest to the end users are delivered on the link by the data source;
- the channel bit rate R_c, which corresponds to the actual bit rate on a given link while the connection is active. Along with information bits, other bits for error correction and signalling purposes may also be transmitted, so that the channel bit rate on the link is higher than the information bit rate. The channel bit rate imposes bandwidth requirements on the physical support, depending on the format used at baseband to represent a bit or a group of bits, also called symbol, and, at radio frequency, on the type of coded modulation used;
- the average bit rate $\langle R \rangle$: links may not be active at all times as connections may be used intermittently, and actually are frequently inactive in case of bursty traffic, made of short data bursts at random intervals. Therefore the average transmitted bit rate is lower than the observed bit rate at times when the link is active. Averaging may apply to either the information bit rate or to the channel bit rate.

Consider, for instance, a user terminal acting as a signal source and delivering messages at an average rate of one message per second to a VSAT for transfer to the hub station over a satellite link (Figure 4.1(a)). Every message contains 1000 information bits.

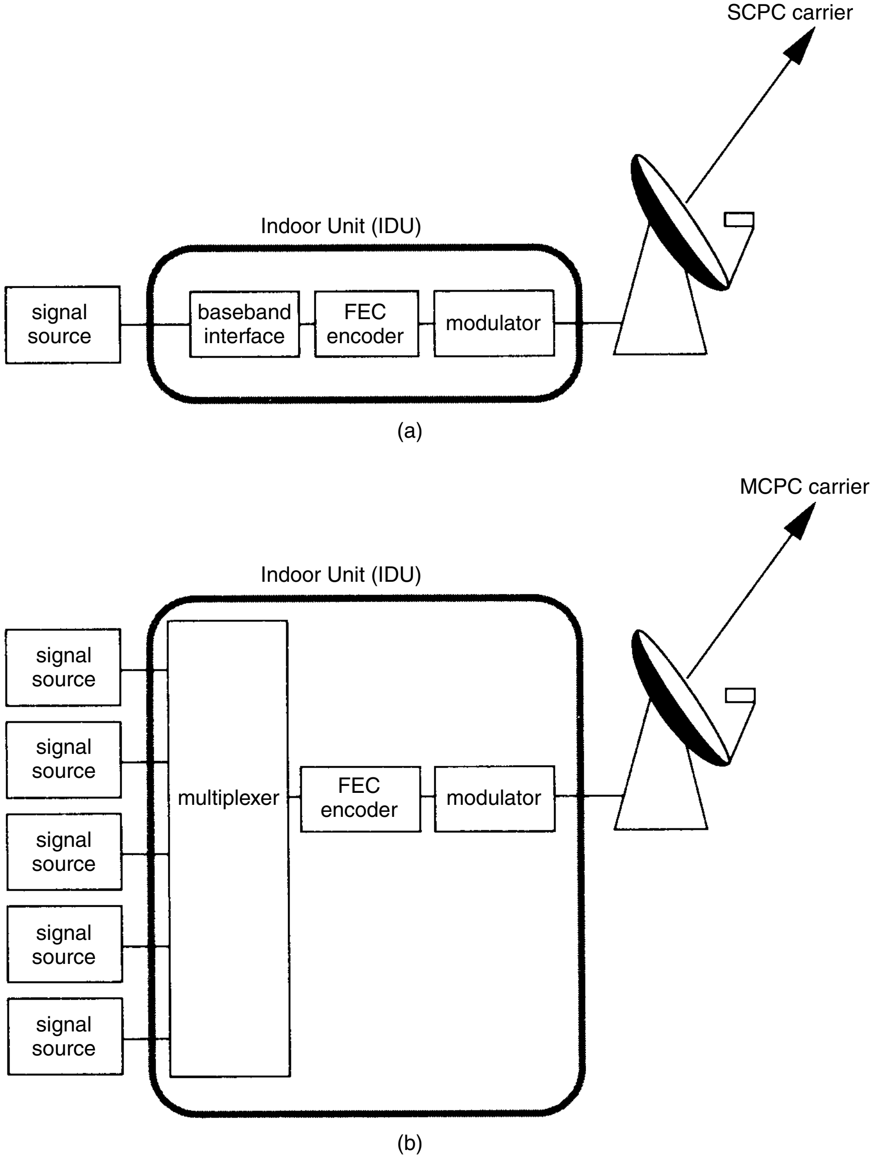


Figure 4.1 (a) Single Channel Per Carrier (SCPC); (b) Multiple Channels Per Carrier (MCPC)

The baseband interface of the indoor unit (IDU) of the VSAT adds some overhead $H = 48$ bits to the message and sends a data unit consisting of the data field $D = 1000$ bits preceded by the overhead $H = 48$ bits to the FEC encoder at a rate $R_b = 64$ kbs^{-1}. Therefore the data unit has a duration of 1048/64 000 seconds, which is equal to 16.375 ms. The FEC encoder adds one redundant bit to every received bit which means a code rate $\rho = 1/2$. The data unit now modulating the carrier consists of $(D + H)/\rho = 2 \times 1048 = 2096$ bits, and those bits are still occupying a time interval of 16.375 ms corresponding to the duration of the data unit. Thus, the channel bit rate is $R_c = ((D + H)/\rho) \times (1/16.375\ \text{ms}) = 128\ \text{kbs}^{-1}$. The average time interval between two messages being 1 second, the average information bit rate $\langle R_b \rangle$ is:

$$\langle R_b \rangle = 1000\ \text{bits}/1\ \text{s} = 1\ \text{kbs}^{-1}$$

The link being active at rate $R_c = 128\ \text{kbs}^{-1}$ only 16.375 ms out of every second, the average channel bit rate $\langle R_c \rangle$ is:

$$\langle R_c \rangle = R_c \times (16.375\ \text{ms}/1\ \text{s}) = 2.096\ \text{kbs}^{-1}$$

4.2.3 Protocols

A protocol is a procedure for establishing and controlling the interchange of information over a network.

For non-data type traffic, the protocols are usually simple and reduce to connection set-up between two end points of a link (TV, voice).

Data communications between the different parts of a network, or between different networks, entail a layered functional architecture which describes how data communications processes are handled. A data protocol is a set of rules for establishing and controlling the exchange of information between peer layers of the network functional architecture.

An example of such a layered architecture is that of the Open Systems Interconnection (OSI) developed by the International Standards Organisation (ISO). This reference model is illustrated in Figure 4.3 and will be discussed in more detail in section 4.5.

4.2.4 Delay

Transfer of information from one user connected to a network to another entails some delay. As mentioned in Chapter 3, section 3.2.8, delay originates from queueing time, transmission time, propagation time, processing time, and protocol induced delay.

Delay conditions the network *response time* perceived by the user from the instant he requests a service to the instant the service is performed. The network response time is highly dependent on the type of service considered. For instance:

- for a data transfer service, the response time would be measured as the time elapsed from the instant the first bit of the transmitted data message leaves the sender terminal to the instant the last bit of the message is received at the destination terminal;
- for an interactive data or an enquiry/response service, the response time of a VSAT network would be measured as the time elapsed between when the 'enter' key is pressed at the remote terminal and the first character of the response appears on the screen.

Delay is one aspect but delay jitter is also of importance for some applications, such as voice or video transmission. Delay jitter represents the amplitude variation of delay value about its average value, and can be characterised for instance by the value of delay standard deviation.

4.2.5 Throughput

The throughput THRU is the average rate at which a connection in the network delivers information bits to the receiver.

$$\mathrm{THRU} = \langle R_\mathrm{b} \rangle \quad (\mathrm{bs}^{-1}) \tag{4.1}$$

The throughput cannot exceed the rate R_b at which the source sends information bits into the network. It may even be lower than this rate because of overheads, message loss, or source blocking time due to flow control. It is bounded by the maximum throughput, which is a function of the network load. As the source increases its input rate, the actual throughput will grow up to a limit and then remain constant or even deteriorate [FER90].

4.2.6 Channel efficiency

The channel efficiency measures the efficiency of the connection by comparing its throughput to the rate R_b at which the source sends information bits:

$$\eta = \frac{\langle R_\mathrm{b} \rangle}{R_\mathrm{b}} \tag{4.2}$$

4.2.7 Channel utilisation

The channel utilisation is the ratio of the time the connection is used to the sum of the idle time plus the time the connection is used.

4.3 TRAFFIC CHARACTERISATION

Traffic characterisation entails different aspects depending on the involved parties and the considered time in the evolution of a network.

4.3.1 Traffic forecasts

This means estimating the type and volume of traffic at peak hours that will be conveyed by the network. Such forecasts should include: traffic breakdown among the different services, variability of the traffic volume and breakdown from site to site, degree of asymmetry of bidirectional services. This information represents valuable input to the network provider for his network design, prior to any operation, and for the dimensioning of links and interface equipment. Unfortunately, the user is most often incapable of stating a precise activity plan, so it is difficult to make any accurate traffic forecasts. It is less of a problem if measurements can be done on an existing terrestrial network to be replaced by the VSAT network.

4.3.2 Traffic measurements

Measuring the traffic deals with collecting actual values of the traffic flows in order to provide representative values of the parameters included in the traffic models. This implies a clear perception of which parameters are to be measured, and when and where they are to be measured. Measurements are available only once the network is operational or, prior to its installation, on the existing network it is supposed to replace. There is some risk in basing the dimensioning of a VSAT network on traffic measurements performed on an existing network to be replaced by the VSAT network, as the client's staff may change working and communicating habits once the VSAT network is in operation. Therefore, as mentioned in Chapter 3, section 3.2.4, it is prudent to proceed with such measurements during the installation tests prior to the full deployment of the VSAT network, and to make provision for spare capacity, in case of a higher traffic demand than anticipated. Experience shows that the statistical information provided by a network management system (NMS), indicating for example the number of calls and the volume of messages sent into the network by the user's terminal, may be adequate for network monitoring and billing procedures but is not accurate enough for a proper dimensioning of the network. Indeed, it does not take into account the actual volume of messages generated in the network as a result of information transfer according to end-to-end or local protocols. Such protocols are responsible for error recovery and flow control, and influence the actual traffic volume in the network and the network throughput.

4.3.3 Traffic source modelling

This involves developing adequate synthetic inputs to the network designer, sufficiently simple to allow mathematical treatment, or to limit the load of the simulation tool, and still sufficiently complex to represent the traffic generated by a source in a realistic manner. Traffic source models should, as far as possible, include parameters that can be interpreted physically. Examples of popular traffic source models are given in Appendix 1.

Traffic sources can be characterised statistically at call level and burst level.

A call is the means by which a terminal connected to a VSAT in the network indicates its intention to send messages to some other terminal. Some networks offer permanent connections between terminals in the form of leased terrestrial lines. In such circumstances, initiating a call is useless, as a physical path is always available along which the sender can send messages to the destination terminal.

VSAT networks may also offer permanent connections between any two terminals: for this, some bandwidth must be reserved for any carrier between the two VSATs to which the terminals are connected (meshed network) or between the two VSATs and the hub (star network). Most often, this solution is not cost effective, and the required bandwidth will be allocated for the time interval when messages are to be exchanged. Thus, demand assignment is a built-in feature of most VSAT networks. Therefore, before sending messages, a terminal must initiate a call which will be processed by the VSAT network management system (NMS).

Once a connection is established, as a result of call generation and acceptance, the sending terminal is able to transfer messages. Should the message transfer correspond to a continuous flow of data during the call, then the traffic on the connection is of 'stream' type. The characterisation of the traffic during the call (arrival time and duration) has the same parameters as that of the call. Should now the message transfer occur by sequences of small packets, also called bursts, then the traffic is said to be 'bursty', with characteristics of its own.

Figure 4.2 illustrates the two above situations.

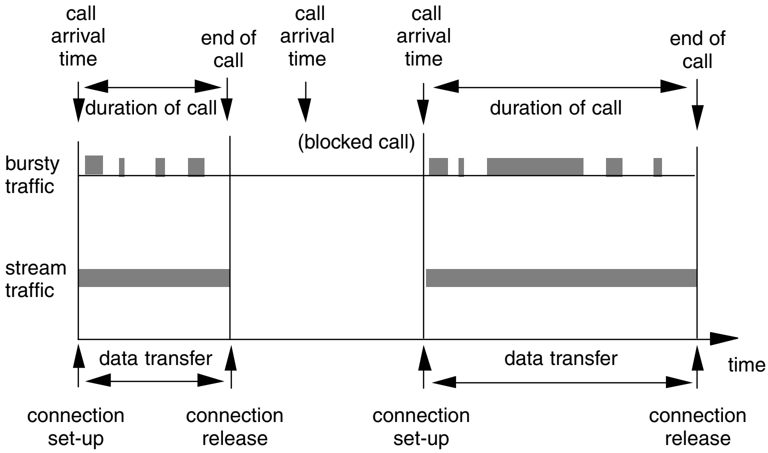


Figure 4.2 Call arrival, connection set-up and data transfer for bursty and stream traffic

4.3.3.1 Call characterisation

Parameters are:

- call generation rate: λ_c (s^{-1})
- mean duration of call: T (s)

When a call is generated, a network resource has to be allocated by the network management system (NMS), in the form of a connection over links with the required capacity. The probability of calls being blocked as a result of lack of network capacity can be estimated from the Erlang formula, which assumes that blocked calls are cleared (the NMS does not keep memory of blocked calls). The formula gives the probability that n connections out of C are occupied:

$$E_n(A) = \frac{A^n/n!}{\sum_{k=0}^{C}(A^k/k!)} \tag{4.3}$$

where A is the traffic intensity, defined as:

$$A = \lambda_c T \quad \text{(Erlang)} \tag{4.4}$$

and C is the network capacity.

Call blocking occurs when $n = C$, therefore the call blocking probability is given by:

$$E_C(A) = \frac{A^C/C!}{\sum_{k=0}^{k=C}(A^k/k!)} \tag{4.5}$$

Formula (4.5) can easily be implemented on a calculator, by using the following iteration:

$$E_n(A) = \frac{AE_{n-1}(A)}{n + AE_{n-1}(A)} \tag{4.6}$$

where $E_0(A) = 1$

Figure 4.3 displays the required number of connections, C, versus the traffic intensity, A, given the probability of blocking.

An approximation for the required number of connections given the traffic intensity A is:

$$C = A + \alpha A^{1/2} \tag{4.7}$$

Where α is the exponent in the $10^{-\alpha}$ blocking probability objective.

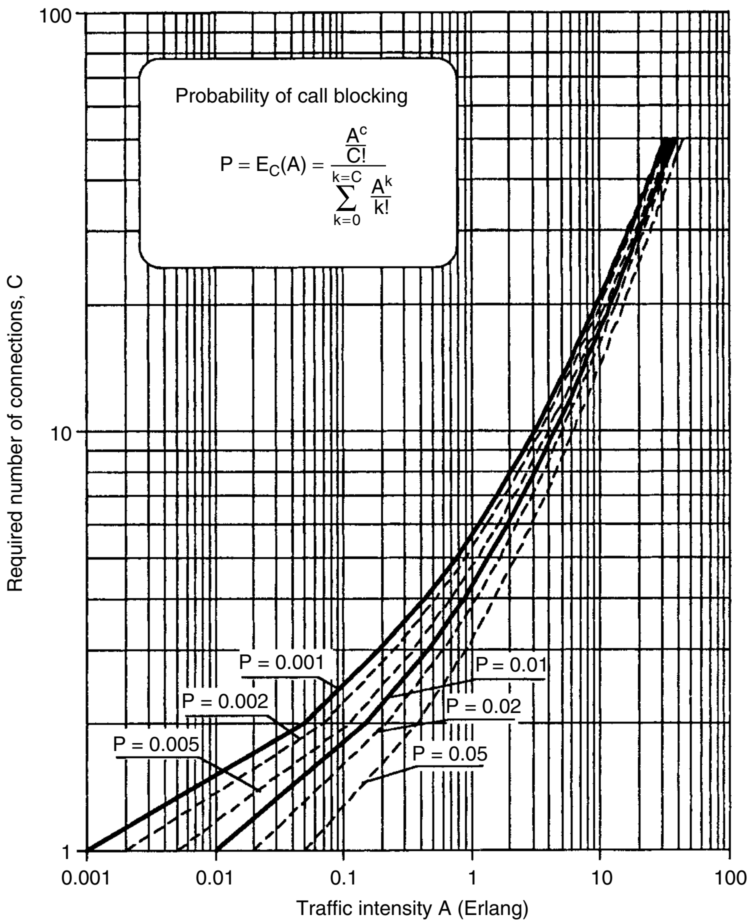


Figure 4.3 Required number of connections to ensure call set-up with a given call blocking probability as a function of the traffic intensity

4.3.3.2 Stream traffic

Stream traffic refers to the situation where a continuous transfer of information occurs once the connection between two terminals has been set up for the purpose of that transfer. Therefore, stream traffic can be characterised by the call connection set-up rate λ_c, as this parameter indicates how frequently the traffic is generated by the transmitting terminal. Once the connection is set up, the information transfer is constant and performed at peak bit rate.

An example of such traffic is transfer of video or audio signals. Telephony signals can be considered as stream traffic, although the interactivity between users implies a duplex connection, and transfer of information usually is not continuous on each of the two connections, as normally one end user would remain silent while the other talks. Therefore, telephony signals, although classified in the stream traffic category, entail some of the characteristics of bursty traffic.

4.3.3.3 Bursty traffic

Bursty traffic refers to intermittent transfer of information during a connection, in the form of individual messages. Messages are short data bursts at random intervals. Typically, this situation arises when a human operated PC is activated by its operator after some thinking time (activation being performed, for instance, by pressing the 'enter' key on the key pad), thus generating the transfer of some data to another terminal. It also results from the specific protocols that are used for data transfer, with information being segmented by the transmitting terminal and segments being acknowledged in the form of short messages by the receiving terminal prior to further transmission by the transmitting terminal.

Bursts introduce new temporal features, characterised as follows:

- the burst generation rate: λ (s^{-1})
- the average length of a burst: L (bits)

The interarrival time (IAT) is the time between two successive generations of burst (see Appendix 1). The average interarrival time $\langle \text{IAT} \rangle$ is equal to:

$$\langle \text{IAT} \rangle = \frac{1}{\lambda} \quad \text{(s)} \tag{4.8}$$

It is convenient to introduce a measure of how bursty the traffic is. A practical definition for burstiness, BU, is the ratio of the peak

Table 4.1 Typical parameter values for examples of stream and bursty traffic

Stream traffic

Service	Call generation rate	Average length of message/duration at 64 kbs^{-1}	Traffic intensity (Erlang)
Telephony	1 per hour	3 minutes	0.05
Television	1 per day	1 hour	0.042
File transfer (electronic	1 per minute	10^4 bits/0.16 s	0.0026
mail, batch)	1 per day	10^8 bits/1560 s	0.018

Bursty traffic

Service	Message generation rate	Average length of message	Burstiness (at 64 kb/s)
Packetised voice	$1\,\text{s}^{-1}$	2800 bytes (22 400 bits)	3
Interactive transactions	$0.02–0.2\,\text{s}^{-1}$	50–250 bytes (400–2000 bits)	160–8000
Enquiry/response	$0.02–0.2\,\text{s}^{-1}$	30–100 bytes (240–800 bits)	400–13 300
Supervisory control and data acquisition (SCADA)	$1\,\text{s}^{-1}$	100 bytes (800 bits)	80

bit rate, i.e. the rate at which bits are transmitted in burst, to the average bit rate:

$$\text{BU} = \frac{R}{\langle R \rangle} = \frac{R}{\lambda L} \tag{4.9}$$

4.3.3.4 Typical values

Table 4.1 indicates typical values of the above parameters for different types of service.

4.4 THE OSI REFERENCE MODEL FOR DATA COMMUNICATIONS

The Open Systems Interconnection (OSI) reference model was originally formulated to provide a basis for defining standards for the interconnection of computer systems. Such standards became a necessity when it was found that different hardware and software installed in different branches of the same organisation were incapable of exchanging information as a result of incompatibilities. In an attempt to overcome these incompatibilities and create a basis for vendor-independent capabilities of information systems, the International Standards Organization (ISO) has created a model which defines seven functional layers for protocols, as indicated in Figure 4.4.

The figure displays two stacks of layers, one for each of the two interconnected systems. The system on the left is the source machine, generating data to be transmitted in a reliable manner to the system on the right, which is the destination machine. Within one machine, a layer presents an interface consisting of one or more service access points and provides services to the next higher layer while utilising the services provided by the next lower layer. Layers in different stacks at the same level are called 'peer' layers. At every layer, there is a pair of cooperating processes, one in each machine, which exchange messages according to the corresponding layer protocol.

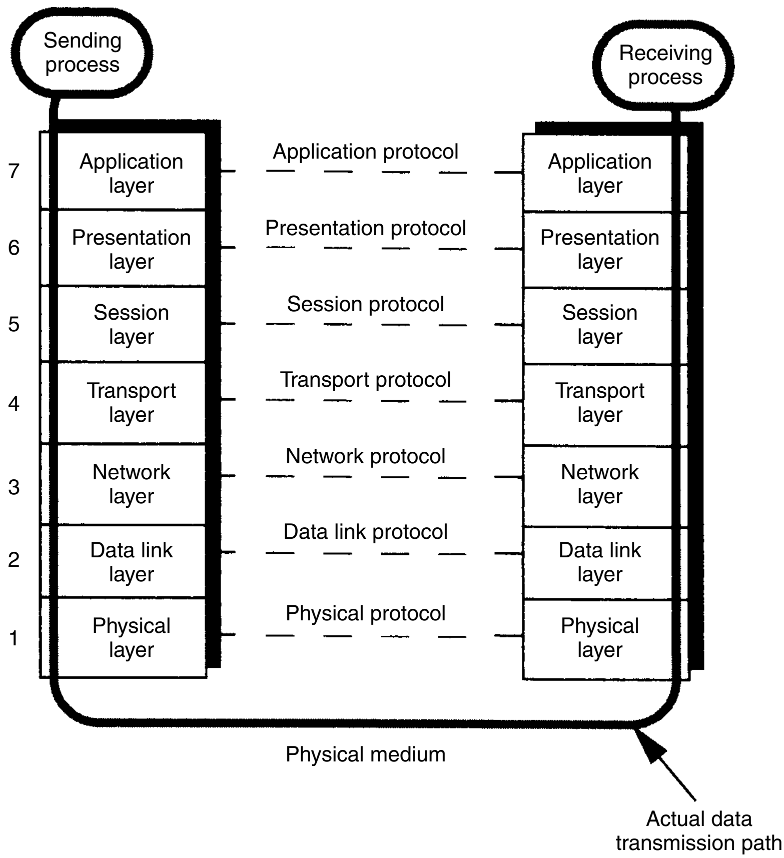


Figure 4.4 The Open Systems Interconnection (OSI) reference model

The message generated at a given layer is actually passed down to the next lower layer, which is physically implemented by hardware and software on the same machine. In this way, the actual data transmission path is down each stack, along the physical medium below the physical layer which connects the two systems, and up the stacks again.

Messages between layers are called protocol data units (PDUs). A PDU consists of data preceded by a header (H) and possibly followed by a trailer (T). When a given layer wants to transmit a PDU to its peer layer on the other system, it passes down that PDU to the next lower layer along with some parameters related to the service being requested. Every lower layer accepts the higher layer's PDU as its data, uses the parameters to determine what should be included in the header and appends its own header, and possibly a trailer, so that its peer layer on the other system will know what to do with the data. This procedure is called 'encapsulation', and is illustrated in Figure 4.5.

When a message is received in a machine, it passes through the layers. Every layer deciphers its header to derive information on how to handle the data and then strips the header before passing the data up to the next higher layer.

The lower three layers are responsible for the transmission and communications aspects, whereas the upper four layers take care of the end-to-end communication and information exchange. A computer system (hardware and software) which conforms to these rules and standards is termed an 'open system'. These systems can be interconnected into an 'open systems environment' with full interoperability.

4.4.1 The physical layer

The physical layer deals with actual transfer of information on the physical medium which constitutes a link, as described in section 4.2.1. Hence, it is concerned with all aspects of bit transmission: bit format, bit rate, bit error rate, forward error correction (FEC) encoding and decoding, modulation and demodulation, etc.

4.4.2 The data link layer

The data link layer ensures the reliable delivery of data across the physical link. It sends blocks of data called 'frames' and provides the necessary frame identification, error control, and flow control.

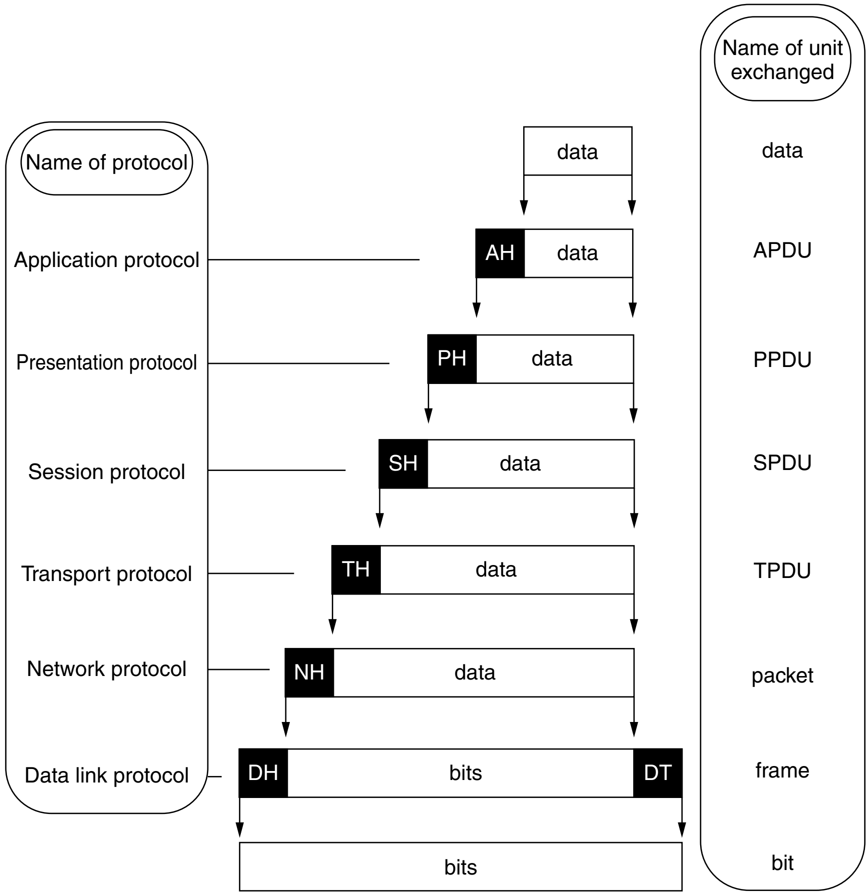


Figure 4.5 Encapsulation from layer to layer in the OSI reference model

4.4.2.1 Detection of damaged, lost or duplicated frames and error recovery

The sender organises data in frames of typically a few hundred bytes and transmits the frames sequentially. Frames are identified by means of special bit patterns at the beginning and the end of every frame. Precautions are taken to avoid these bit patterns occurring in the data field.

Upon reception of frames, the receiver sends acknowledgement frames. However, as noise on the link may introduce bit errors, the receiving device must be able to detect such an occurrence. This is performed thanks to checksum bits in the trailer of the frame.

Should the checksum be incorrect, the receiving device sends no acknowledgement frame to the sending device. Not receiving an acknowledgement frame within a given time limit, the sending device retransmits the frame. Hopefully this frame will be correctly received. Otherwise, no acknowledgement is delivered and retransmissions will occur until completion of error-free reception.

Multiple transmissions of a frame introduce the possibility of duplicate frames: this would happen, for instance, if the routeing delay exceeded the time limit for retransmission. One or several duplicate frames may be generated before the receiving device has had a chance to transmit its acknowledgement. To obviate this problem, a sequence number in the frame header indicates to the receiver if the received frame is a new frame or a duplicate. Duplicated frames can therefore be discarded.

4.4.2.2 Flow control

A fast sender must be kept from saturating a slow receiver in data. Some traffic regulation must be employed to inform at any instant the sender how much buffer space the receiver has available. This is done by means of sliding window techniques [TAN89, p224]: at any instant, the sender maintains a list of consecutive sequence numbers corresponding to frames it is permitted to send. These frames are said to fall within the sending window. Similarly, the receiver also maintains a receiving window corresponding to frames it is permitted to accept.

4.4.3 The network layer

The network layer is responsible for routeing packets from the source to the destination. Therefore, it is concerned by transfer of data over multiple links in the network. This implies identifying the destination (addressing function), identifying the path (routeing), and making sure that the resource is available (congestion control). It also has to identify the link user for purposes of billing (accounting function).

4.4.3.1 Addressing

The network layer is in charge of identifying the destination of data. The receiving device is known by its address. However, this address may be different from one network to the other. For instance, the end terminal may be part of a local area network (LAN) connected to a VSAT. The address of that specific terminal in the LAN may be different from the address of that same terminal in the VSAT network. Therefore, it is up to the network layer to perform the proper address mapping.

4.4.3.2 Routeing of information

The network layer is in charge of determining which links are used. The utilisation of links can be on a fixed assignment basis or on demand.

4.4.3.3 Congestion control

The network layer is also in charge of determining how links are used. For instance it is up to the network layer to regulate the traffic flow in order to avoid congestion on the link, should the traffic flow exceed the capacity of the link.

4.4.3.4 Accounting

The network layer supervises the amount of information delivered at any network input and output so as to produce billing information.

4.4.4 The transport layer

The transport layer is in charge of providing reliable data transport from the source machine to the destination machine. Hence, it is an end-to-end layer: it deals with functionalities required between end terminals, possibly communicating through several different networks.

4.4.4.1 End-to-end transfer of data

The transport layer accepts data from the session layer, splits it into smaller units if needed, passes these to the network layer and ensures that all pieces arrive correctly at the other end.

4.4.4.2 Multiplexing

The transport layer may organise the routeing of several transport connections onto a unique network connection. This is called multiplexing and should be transparent to the session layer.

4.4.4.3 Flow control

The transport layer is in charge of controlling the flow of information between the end terminals so that a fast terminal does not saturate a slow one. This flow control is distinct from that of the data link layer, although it can be done by similar means.

4.4.5 The upper layers (5 to 7)

These are: the session layer, the presentation layer, and the application layer. All are end-to-end layers. These layers are of no concern to VSAT networks. Hence they will not be discussed here. For further information the reader may refer to books on computer communications.

4.5 APPLICATION TO VSAT NETWORKS

4.5.1 Physical and protocol configurations of a VSAT network

A VSAT network essentially provides a connection between any remote user terminal and the host computer. Figure 4.6 illustrates two representations of the end terminals (host computer and user terminals) and the VSAT network in between. One is a physical configuration which indicates the kind of equipment that support the connection, the other is the protocol configuration which displays the peer layers between the above equipment. The physical configuration shown here displays the hub baseband interface which is part of the indoor unit of the hub, as shown in Figure 1.26, to which the host computer is connected, and the VSAT baseband interface which is part of the VSAT indoor unit, as shown in Figure 1.20, to which the user terminals are connected. The protocol configuration displays the respective stacks of layers from one to seven within the host computer and the user terminal and reduced stacks for the front end processor and the baseband interface of the VSAT indoor unit. Such a configuration allows for protocol conversion, also named emulation, which will now be presented.

4.5.2 Protocol conversion (emulation)

For the protocol configuration of Figure 4.6, it would be convenient to have a similar one to that of Figure 4.4, where peer-to-peer interactions between end terminals are end-to-end. The VSAT network

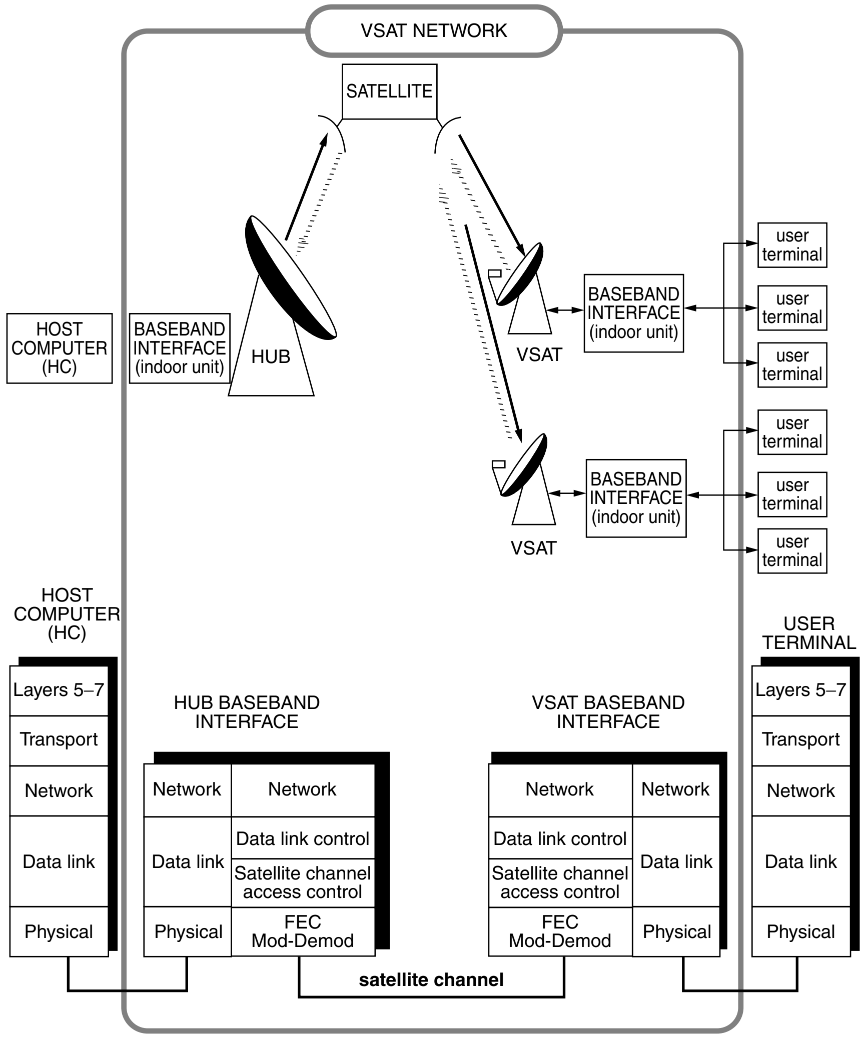


Figure 4.6 Physical and protocol configurations of a VSAT network

would act as a pure 'cable in the sky' at the physical layer level, and interconnection of the customer's machines (user terminals and host computer) would be most easy to perform. However, this is not feasible as a result of the characteristics of the satellite channel where information is conveyed at radio frequency with propagation delay and bit error rate. These characteristics differ from those of the terrestrial links for which the protocols used on the customer's machines have been designed. Terrestrial links usually display shorter delays and lower bit error rates than those encountered on satellite links. Consequently, terrestrial oriented protocols may become inefficient over satellite links. Therefore, different protocols must be considered for data transfer over satellite links. Such protocols, however, cannot be end-to-end protocols, as this would imply changing the protocols implemented on the customer's machine, which would be unacceptable to the customer. So, finally, the solution is to implement some form of protocol conversion both at the hub baseband interface and the VSAT baseband interface. The conversion of end terminal protocols into satellite link protocols is called *emulation*, or in a more colloquial manner, *spoofing*. Indeed, if the conversion is adequate, that is if it ensures end-to-end transparency, the end terminals will have the impression of being directly interconnected, although they are not.

In Figure 4.6, only the three lower layers (network, data, and physical) are emulated. This corresponds to a common situation. However, some services might require that emulation be carried up to the transport layer.

The network layer protocol emulation performs address mapping for the customer's machines. This enables the network addresses to be independent of the customer addresses.

The data link layer is split into two sublayers: the sublayer named 'data link control' provides data link control over the satellite links independently from the data link control between the VSAT network interfaces and the customer's machines. The sublayer called 'satellite channel access control' is responsible for the access to the satellite channel by multiple carriers transmitted by the VSATs or the hub station. An important aspect here, which is specific to VSAT networks, is that the powered bandwidth of the satellite required for the carrier which provides the connection at the physical level, if allocated on a permanent basis, is poorly used in the case of infrequent stream traffic or with bursty traffic. It is, therefore, desirable that this satellite resource be allocated to any VSAT earth station on a demand assignment (DA) basis, as presented in Chapter 1, section 1.6.3, according to the traffic demand and characteristics.

Finally, at the physical level, any earth station (hub or VSAT) has to provide a physical interface which actually supports the physical connection. On the customer's side, the physical interface should be compliant with the customer's hardware. On the satellite side, the physical level should provide protection of data against errors by means of forward error correction (FEC) encoding and decoding techniques, and modulate or demodulate carriers conveying the data.

4.5.3 Reasons for protocol conversion

This section aims to discuss in more detail some of the underlying reasons exposed above.

Satellite links differ from terrestrial links in two ways:

1. The large propagation delay, about 270 ms, from one site to another over satellite links in comparison to the much smaller delays encountered on terrestrial networks, typically a few milliseconds to tens of milliseconds.
2. The bit error rate: satellite links are corrupted by noise which affects the carrier demodulation process. The bit error rate can be reduced to levels typically of 10^{-7} thanks to the use of forward error correction (FEC) but this is still higher than the bit error rate level encountered on terrestrial links.

It will now be shown how these characteristics impact on protocols when used over satellite links.

4.5.3.1 Impact on error control

The following example deals with transmission of a data stream over a connection from the host computer to a user terminal, using automatic repeat request (ARQ) protocols for error control. The data link layer gets its protocol data unit (packet) from the network layer and encapsulates it in a frame by adding its data link header and trailer to it (see Figure 4.5). This frame is then transmitted over the network to the data link layer of the user terminal which verifies the integrity of data by means of the error detection information contained in the trailer of the frame. Should the received data be error free, the user terminal sends a positive acknowledgement (ACK) back to the host computer. If not, it sends a negative acknowledgement (NACK). If it receives no ACK and no NACK after a time out delay, then it retransmits the frame.

In the following, three ARQ protocols are considered (Figure 4.7):

– a stop-and-wait (SW) protocol: the host computer waits until it receives a positive acknowledgement, ACK, before sending the next frame. If a negative acknowledgement, NACK, is received, the host computer retransmits the same frame (Figure 4.7(a));
– a go-back-*N* protocol (GBN): the host computer transmits frames in sequence as long as it does not receive any negative acknowledgement, NACK. Receiving NACK for frame *N*, it retransmits frame *N* and all subsequent frames (Figure 4.7(b));
– a selective-repeat (SR) protocol: the host computer transmits frames in sequence as long as it does not receive a negative acknowledgement, NACK. Receiving a NACK for frame *N* while

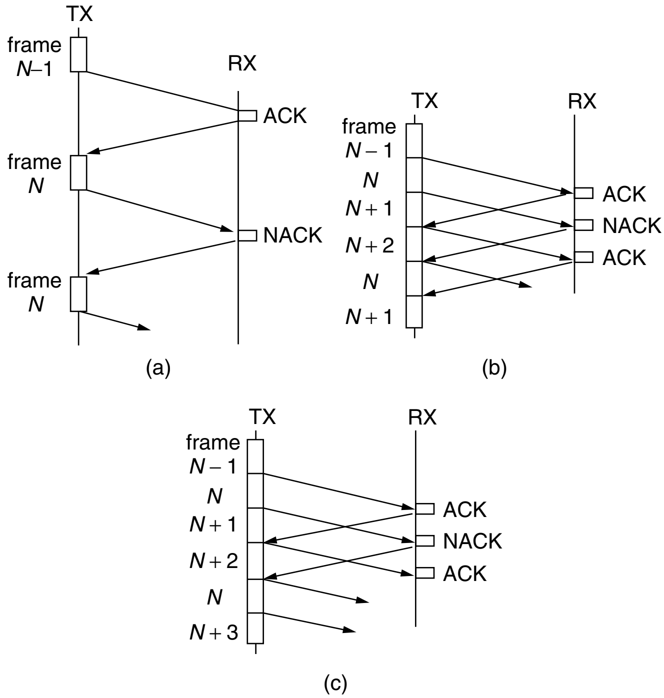


Figure 4.7 (a) Stop-and-wait protocol; (b) go-back-*N* protocol; (c) selective-repeat protocol

sending frame $N + n$, it retransmits frame N after frame $N + n$, then continues with frame $N + n + 1$ and the subsequent ones (Figure 4.7(c), here $n = 2$).

The three protocols will now be compared on the basis of the channel efficiency. Appendix 2 demonstrates that the channel efficiency η_c for every protocol is equal to:

stop-and-wait: $$\eta_{cSW} = \frac{D(1 - P_f)}{(R_b T_{RT})} \tag{4.10}$$

go-back-N: $$\eta_{cGBN} = \frac{D(1 - P_f)}{(L(1 - P_f) + R_b T_{RT} P_f)} \tag{4.11}$$

selective-repeat: $$\eta_{cSR} = \frac{D(1 - P_f)}{L} \tag{4.12}$$

where:

D = number of information bits per frame (bits)
L = length of frame (bits) = $D + H$ (information plus overhead)
P_f = frame error probability = $1 - (1 - \text{BER})^L$, where BER is the bit error rate

R_b = information bit rate over the connection (bs^{-1})
T_{RT} = round trip time (s)

The round trip time T_{RT} corresponds to the addition of service times and propagation delays. At time L/R_b, the last bit of the frame has been sent. At time L/R_b plus the propagation delay T_p from the sender to the receiver, the last bit has arrived at the receiver. From a host computer to a user terminal over a *terrestrial link*, T_p is about 5 ms. From a VSAT to the hub station over a *satellite link*, T_p is about 260 ms (see Figure 2.22). Neglecting the processing time, the receiver is now ready to send the acknowledgement message. Denote by A the acknowledgement frame length and R_{back} the information bit rate at which the acknowledgement is sent on the return link. At time $L/R_b + T_p + A/R_{back}$, the last bit of the acknowledgement frame has been sent. At time $L/R_b + T_p + A/R_{back} + T_p$ the sender has received the acknowledgement. So the round trip time is:

$$T_{RT} = \frac{L}{R_b} + 2T_p + \frac{A}{R_{back}} \quad \text{(s)} \tag{4.13}$$

One can neglect A/R_{back} relative to L/R_b as the acknowledgement frame length is much smaller than the frame length L (it often reduces to the header H), and for a star VSAT network generally R_{back} is the outbound link bit rate which is usually larger than R_b. Therefore the round trip time can be approximated by:

$$T_{RT} = \frac{L}{R_b} + 2T_p \quad \text{(s)} \tag{4.14}$$

Figure 4.8 compares channel efficiency η_{cSW} and η_{cGBN} as a function of the round trip time T_{RT} for different values of the bit error rate. The parameter values selected here are:

$$D = 1000 \text{ bits}$$
$$H = 48 \text{ bits}$$
$$L = 1048 \text{ bits}$$
$$R_b = 64 \text{ kb/s}$$

On a terrestrial link, taking $T_p = 5$ ms, T_{RT} would be about 26 ms. On a satellite link, taking $T_p = 260$ ms, T_{RT} would be about 536 ms.

With the selective-repeat protocol, the channel efficiency is independent of the round trip time. It can be seen that η_{cSR} is always greater than with the two other protocols. With the selected parameter values, one obtains the values indicated in Table 4.2.

Figure 4.8 indicates that the channel loses much of its efficiency when a stop-and-wait protocol is implemented on a satellite link, as

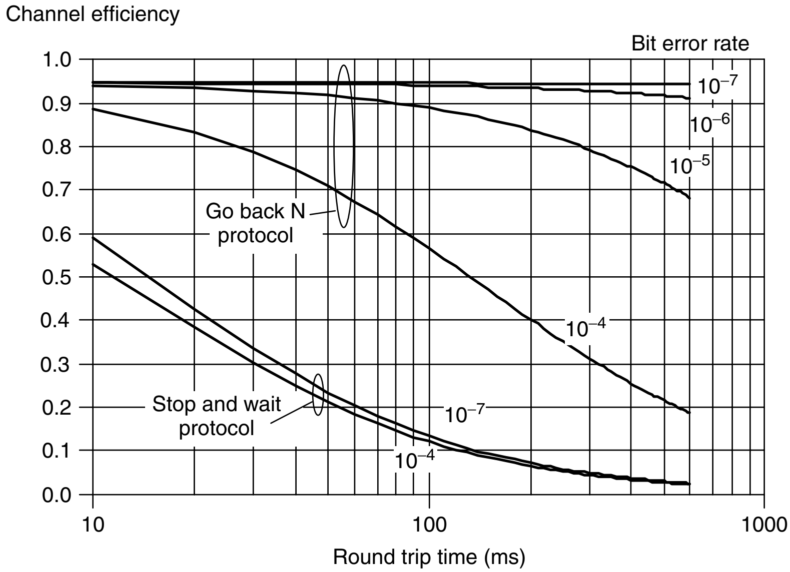


Figure 4.8 Channel efficiency for stop-and-wait, and go-back-*N* as a function of round trip time T_{RT} and bit error rate (BER)

Table 4.2 Values of channel efficiency η_{cSR} for selective-repeat protocol as a function of bit error rate (BER)

BER	10^{-4}	10^{-5}	10^{-6}	10^{-7}
η_{cSR}	0.86	0.95	0.95	0.95

a result of the increased round trip time compared to terrestrial links. The same is true of a go-back-*N* protocol should the satellite link be of poor quality (bit error rate in the range 10^{-4} to 10^{-5}). If the satellite link has a bit error rate in the order of 10^{-7}, then no degradation is observed. Finally, as seen from Table 4.2, the selective-repeat protocol, which offers a good performance for reasonably low bit error rates, is a good candidate for satellite links as it is not sensitive to a long round trip delay.

4.5.3.2 Impact on flow control

Protocols at the data link layer and transport layer levels often make use of sliding windows for flow control purposes. In such cases, only positive acknowledgements (ACK) are sent. Not receiving an acknowledgement before a given time-out interval, the sender retransmits the protocol data unit which has not been acknowledged. The sender can only send a limited number of protocol data units following the last acknowledged one. These are said to fall within the sending window. The window slides by one position at every received acknowledgement, therefore initiating the clearance for sending a subsequent protocol data unit. Similarly, the receiver accepts only a limited number of protocol data units before sending a positive acknowledgement. Any incoming protocol data unit that falls outside the window is discarded. The window slides by one position at every emitted acknowledgement, and the receiver can subsequently accept one more protocol data unit.

It can be shown that the channel efficiency for a sliding window protocol with error control based on the selective-repeat procedure is [TAN89, p 243]:

$$\eta_{cSR} = \frac{D(1 - P_f)W}{(L + R_b T_{RT})} \quad \text{if } W < 1 + \frac{R_b T_{RT}}{L} \tag{4.15}$$

$$\eta_{cSR} = \frac{D(1 - P_f)}{L} \quad \text{if } W \geqslant 1 + \frac{R_b T_{RT}}{L} \tag{4.16}$$

where W is the window size, and the other parameters are as in the previous section.

Figure 4.9 represents the channel efficiency as a function of the round trip time T_{RT} for the example discussed in the previous section. Different window sizes are considered, from 1 to 31. The case $W = 1$ corresponds to a stop-and-wait protocol, as presented in the previous section. One can see from the figure and from

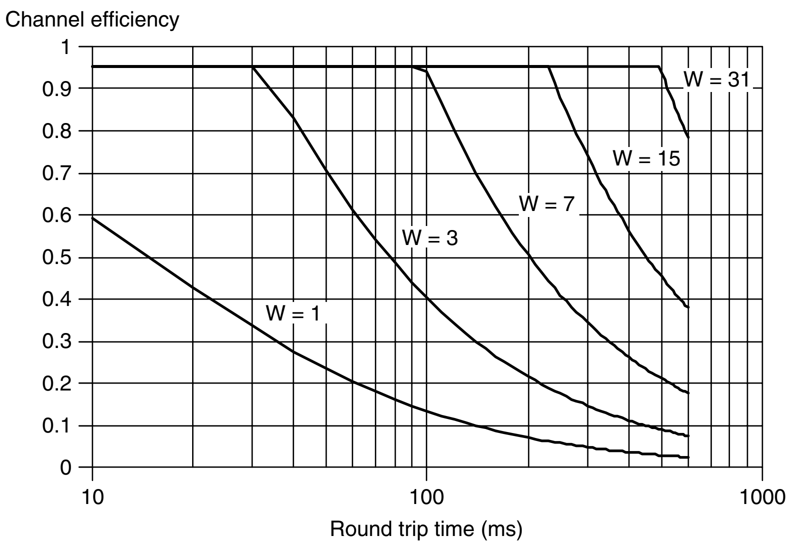


Figure 4.9 Channel efficiency when using a sliding window protocol as a function of round trip time for different window sizes W. The bit error rate on the link is 10^{-7}

equation (4.16) that satellite links require large window values to be efficient, while for terrestrial links, small values can be implemented.

This can be explained as follows: the quantity $R_b T_{RT}/2L$ represents the number of protocol data units that the link from the sender to the receiver can hold. The quantity $R_b T_{RT}/L$, conditions the selection of either equation (4.15) or (4.16) to express the channel efficiency. The horizontal parts of the curves in Figure 4.9 are obtained from equation (4.16), while the sharp decreases are expressed by (4.15). The quantity $R_b T_{RT}/L$ represents the total number of protocol data units filling both the direct and return links from sender to receiver. Should the window exceed that quantity, then transmission goes on continuously and the protocol leads to an efficient use of the channel (horizontal part of the curve). If the window is less than that quantity, the sender is blocked when the window is full and has to wait for an acknowledgement to come in before resuming transmission. As the channel is not being used during this waiting time (the longer the round trip time, the longer the waiting time), the use of the channel is reduced and the channel efficiency decreases.

One can conclude by stating that flow control over satellite links using sliding window protocols is feasible without loss of channel efficiency if the window is large enough.

4.5.3.3 Polling over satellite links

Figure 4.10 illustrates a typical synchronous data link control (SDLC) environment, where the host computer communicates with a series of remote user terminals by means of a multidrop line. The host computer manages the transfer of data between itself and the user terminals, on the shared capacity of the multidrop line by means of a technique named *polling*.

Polling means that the host sends a message to every user terminal it controls, inquiring whether or not the user terminal has anything to send. Every user terminal acknowledges its own poll and sends along data if it has data to send. The host then acknowledges reception of data. If the user terminal has no data to send, it sends a 'poll reject' message and the host polls the next user terminal in a round-robin fashion.

Alternatively, the host when having data to send to a given user terminal sends this data along with its poll. The user terminal sends an acknowledgement to the host for the accepted data, possibly along with its own data.

A network would possibly comprise a host and several multidrop lines such as the one illustrated in Figure 4.10, using permanent terrestrial leased lines or temporary connections via the public

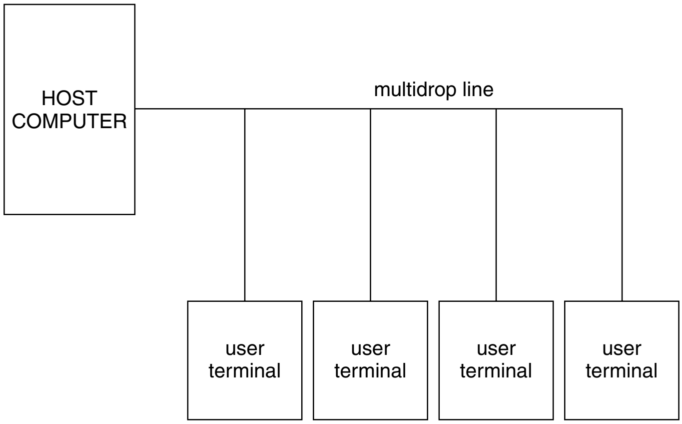


Figure 4.10 Multidrop line

switched network. It is assumed that the company which utilises the private network wishes to replace part or all of it by a VSAT network, as shown in Figure 4.11.

It is undesirable to pass every polling message and its acknowledgement across the satellite. Such a scheme would produce high transmission delays, as every handshake would take 0.5 s, and waste transponder capacity, as every poll is not necessarily followed by transfer of data.

To put this into perspective, consider a terrestrial network incorporating one host, $n = 10$ user terminals and a multidrop line. Assuming:

- maximum transmission delay over the multidrop line: $t_d = 5$ ms;
- maximum message processing time (poll and/or acknowledgement plus data if any): $t_m = 1$ ms

Each polling involves transmitting the message from the host to the user terminal ($t_d = 5$ ms), processing the message at the user terminal ($t_m = 1$ ms) and transmitting the reply from the user terminal to the host ($t_d = 5$ ms). The polling time of a user terminal then is:

$$t_p = t_d + t_m + t_d = 11 \text{ ms}$$

The maximum network response time is the cycle time (n user terminals in the cycle):

$$t_r = nt_p = 110 \text{ ms}$$

The mean network response time, assuming constant probability density for all delays, is one half the maximum delay which amounts to 55 ms.

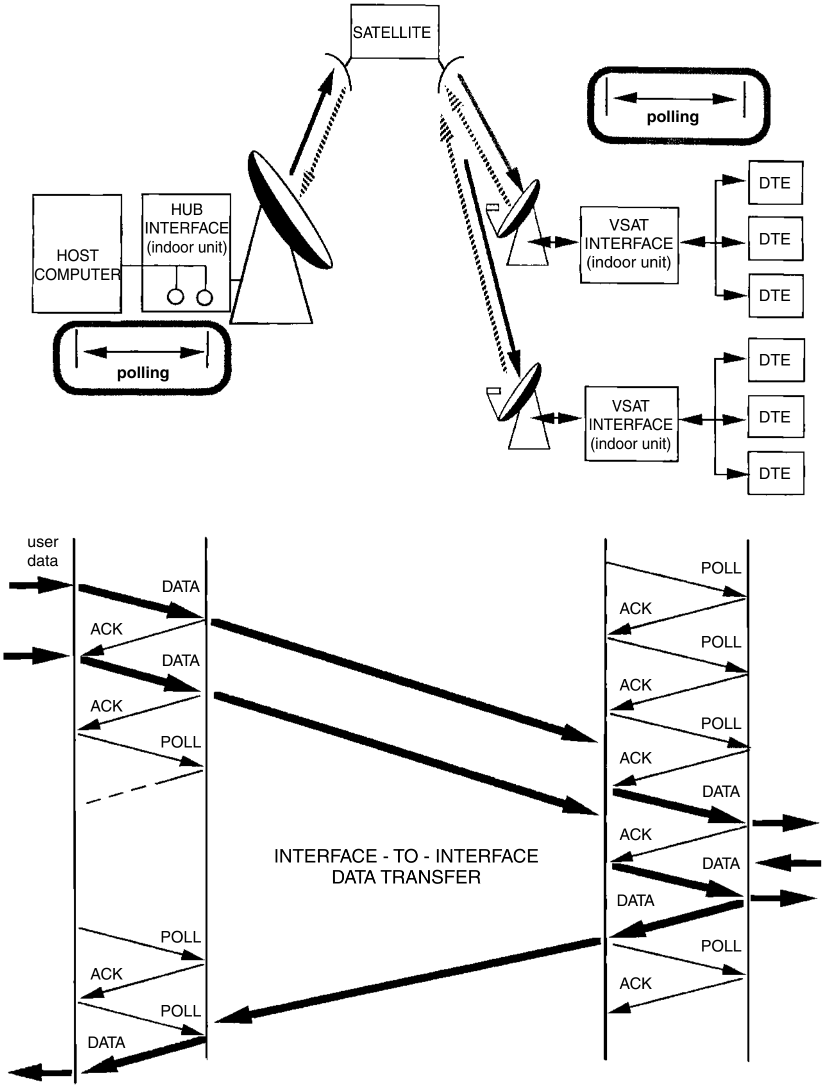


Figure 4.11 Polling over a VSAT network

Now consider that the multidrop line is replaced by satellite links. Then $t_p = 250$ ms, and the cycle time becomes:

$$t_r = n(t_d + t_m + t_d) = 10(250 + 1 + 250) = 5 \quad \text{seconds}$$

So the maximum network response time increases from 110 ms to 5 s with mean value increasing from 55 ms to 2.5 s. Notice that in the absence of traffic, the satellite links still convey polls and acknowledgements, thus using bandwidth for no data transfer.

To avoid these undesirable effects, polling emulation may be implemented at the remote sites and at the central facility, as illustrated in Figure 4.11: at the remote site, the VSAT interface polls the user terminals, thus acting as the host computer in Figure 4.10, while at the central facility, the ports of the hub interface, as many as the remote sites, are polled by the host, acting as 'virtual' user terminals.

Acknowledgements are provided by the hub interface ports as soon as they receive the polling message from the host, and the VSAT interface independently polls every user terminal. Data messages are transmitted on the satellite links only if a user terminal responds to polling by transmitting data, or if the host selects one of the hub interface ports to transmit data. VSAT and hub interfaces must provide buffering and flow control on the satellite links. This avoids satellite bandwidth being used in the absence of traffic.

One can evaluate the network response time with polling emulation over satellite links. It amounts to a maximum of one cycle time at the host/hub interface plus propagation time from host to VSAT, plus a maximum of one cycle time at the VSAT/user terminal interface, plus propagation time from VSAT to hub, plus a maximum of one cycle time at the hub/host interface, i.e.:

$$T_r = 10(5 + 1 + 5) + 250 + 10(5 + 1 + 5) + 250$$
$$+ 10(5 + 1 + 5) = 830\ \text{ms}.$$

The increase in response time from terrestrial network to VSAT network is now less than with direct polling over satellite links (from 110 ms to 830 ms maximum, instead of 5 seconds), in conjunction with using satellite bandwidth only when needed.

4.5.3.4 Conclusion

The above examples show that protocols that perform adequately on terrestrial links may work poorly when used as such on satellite links. Hence, there is a need for protocol tuning or protocol conversion at the interface between end terminals and the VSAT network, or between other networks and the VSAT network.

4.6 MULTIPLE ACCESS

The earth stations of a VSAT network communicate across the satellite by means of modulated carriers. Depending on the network configuration, different types and numbers of carriers must be routed simultaneously within the same satellite transponder. Figure 4.12 illustrates different possible situations:

– with one-way networks, where the hub broadcasts a time division multiplex of data to many receive-only VSATs, only one carrier is to be relayed by the satellite transponder. Accordingly, there is no other carrier competing for satellite transponder access, and there is no need for any multiple access protocol;
– with two-way star-shaped networks, carriers from VSATs and the hub station are competing to access a satellite transponder;
– with two-way meshed networks, there is no hub station and the only carriers competing to access a satellite transponder are those transmitted by the VSAT stations.

Multiple access is therefore to be considered in the two latter situations only. Multiple access schemes differ in the way the satellite

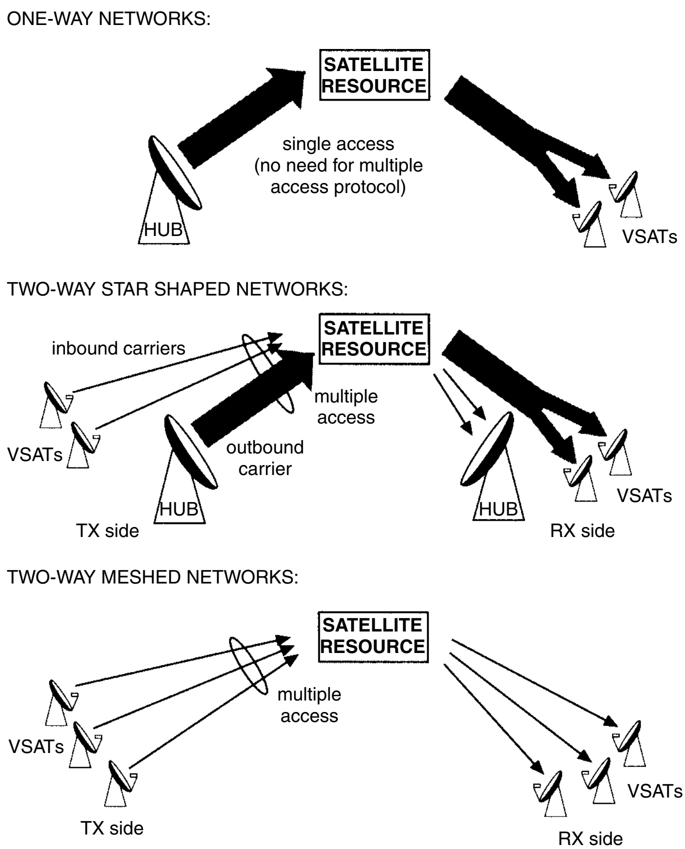


Figure 4.12 Multiple access for different network configurations

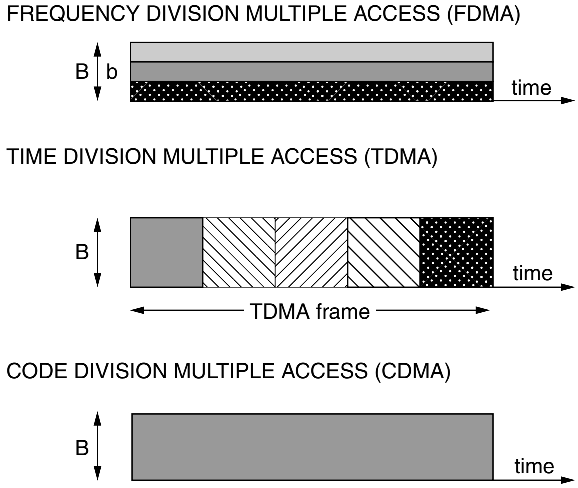


Figure 4.13 Basic multiple access protocols

transponder resource, which is powered bandwidth during the lifetime of the satellite, is shared among the contenders.

4.6.1 Basic multiple access protocols

Figure 4.13 illustrates the ways of partitioning the bandwidth of a satellite transponder between multiple carriers with time [MAR02, Chapter 6]:

- *Frequency division multiple access (FDMA)* means allocating a given subband of overall transponder bandwidth, B, to every carrier. The allocated subband, shown as b for a specific carrier in Figure 4.13, must be compatible with the carrier bandwidth which depends on the bit rate it conveys and the type of modulation and coding (see Chapter 5, section 5.8). The bit rate on the carriers may correspond to the traffic of one one-way connection: this is a Single Channel Per Carrier (SCPC) mode, or to several one-way connections which are time division multiplexed (TDM), and then this is a Multiple Channels Per Carrier (MCPC) mode.
- *Time division multiple access (TDMA)* means allocating the overall bandwidth of the transponder, B, to every carrier in sequence for a limited amount of time, called a *time slot*.

  The sequence may be random, every station transmitting a data packet on a carrier burst with duration equal to a time slot whenever it has data to transmit, without being coordinated with respect to other stations. This is named 'random TDMA' and is best represented by the so-called ALOHA type protocols. As a result of the random nature of transmissions, such multiple access schemes do not protect two or more carrier bursts transmitted by separate stations from possibly colliding within the transponder (that is overlapping in time). The interference which results then prevents the receiving stations from retrieving the data packets from the corrupted bursts. To provide error-free transmission, ALOHA protocols make use of ARQ strategies by sending acknowledgements for every packet correctly received: in case of collision, the transmitting stations not receiving any acknowledgement before the end of their time-out interval, will retransmit the unacknowledged packet at the end of a random time interval calculated independently at every station, so as to avoid another collision.

Alternatively, the sequence may be synchronised in such a way that bursts occupy assigned non-overlapping time slots. This implies that the time slots are organised within a periodic structure, called a TDMA frame, with as many time slots as active stations (note that the term 'frame', with TDMA, should not be confused with the term used in computer communications, where a 'frame' is a block of data sent or received by a computer at the data link layer of the OSI reference model of Figure 4.4).

With TDMA, carriers are transmitted in bursts and received in bursts. Every burst consists of a header made of two sequences of bits: one for carrier and bit timing acquisition by the receiving VSAT demodulator, another named 'unique word' indicating to the receiver the start of the data field. The header is followed by a data field containing the traffic associated with either one or several one-way connections. If only one, the burst is a Single Channel Per Carrier (SCPC) burst, if several, the burst is a Multiple Channel Per Carrier (MCPC) burst and is divided into subbursts, each subburst corresponding to one one-way connection. Synchronisation is necessary between earth stations, and the earth station must be equipped with rapid acquisition demodulators in order to limit burst preambles to a minimum.

– *Code division multiple access (CDMA)* is a multiple access technique which does not consider any frequency–time partition: carriers are allowed to be transmitted continuously while occupying the full transponder bandwidth, B. Therefore interference is inevitable, but is resolved by using spread spectrum transmission techniques based on the generation of high-rate chip sequences (or 'code'), one for every transmitted carrier. These sequences should be orthogonal so as to limit interference. Such techniques allow the receiver to reject the received interference and retrieve the wanted message.

The selection of a multiple access scheme should take into account the requirement for power and bandwidth, not only of the satellite transponder, but also of the earth stations (VSATs and hub station). Generally speaking, operating a satellite transponder in a multi-carrier mode (several carriers sharing the transponder bandwidth at a given time) as with FDMA and CDMA, entails the generation of intermodulation noise which adds to the thermal noise (see Chapter 5, section 5.4). Carriers conveying a high bit rate are more demanding for bandwidth and power than smaller carriers. This impacts on the EIRP requirement of the transmitter: it translates into a higher demand for power from VSAT transmitters on the inbound links, from the hub station transmitter on the outbound links, and from the satellite transponder on all links. It also translates into a higher demand for bandwidth on the satellite transponder.

We will now discuss the practical implementation of these multiple access schemes in VSAT networks. It will be assumed that a fraction of a satellite transponder bandwidth is allocated to the VSAT network, hence it may be that the rest of the transponder is occupied by carriers originating from earth stations other than those belonging to the considered VSAT network. Indeed, it seldom happens that the demand for capacity of a VSAT network requires full transponder usage. Therefore, the transponder is actually divided into subbands, every subband being used by different networks. In a way this represents 'network FDMA'. This means that the satellite resource available to a given network is only a fraction of a satellite transponder's overall resource, as not only the transponder bandwidth has to be shared but also the output power. Therefore, a considered VSAT network benefits neither from the entire transponder effective isotropic radiated power (EIRP), nor from its full bandwidth.

4.6.2 Meshed networks

The meshed network comprises *N* VSATs. Every VSAT should be able to establish a link to any other one across the satellite.

A first approach is to have every VSAT transmitting as many carriers as there are other VSATs: the information conveyed on every carrier represents the traffic on a one-way connection from one user terminal attached to one of the VSATs to another user terminal attached to another VSAT. Such carriers are *Single Channel*

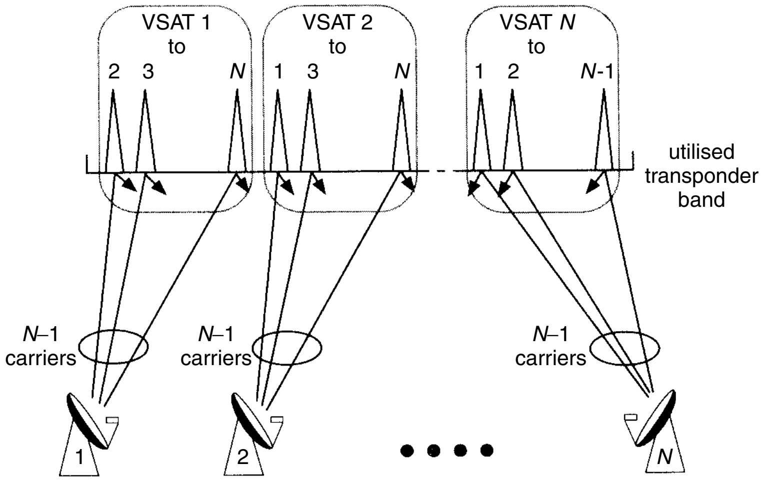


Figure 4.14 Meshed network with N VSATs transmitting as many SCPC carriers as there are other VSATs, using Frequency Division Multiple Access (FDMA)

Per Carrier (SCPC) carriers. A two-way connection between two user terminals entails the use of two SCPC carriers, each one being transmitted by each of the involved VSATs. For permanent full network connectivity, every VSAT should be able to receive at any time all SCPC carriers transmitted by the other VSATs in the network.

Figure 4.14 proposes an implementation based on FDMA. Such a configuration requires that every VSAT be equipped with $N - 1$ transmitters and $N - 1$ receivers. This is costly if N is large, and poses operational difficulties as more transmitters and receivers must be installed at every VSAT each time the network incorporates new VSATs. Moreover, the satellite transponder is occupied by $N(N - 1)$ carriers. Such carriers are narrow band ones as they convey low bit rates. This may require frequency stable modulators because guard bands between carriers must be kept to a minimum in order to save satellite bandwidth. As an example, consider a VSAT network with $N = 100$ VSATs. The number of transmitters and receivers per VSAT is $N - 1 = 99$. The number of carriers is $N(N - 1) = 9900$.

A variant of Figure 4.14 is to consider the broadcasting capability of the satellite: any carrier uplinked by a VSAT is actually received by all VSATs. Therefore, the overall traffic conveyed by the $N - 1$ carriers transmitted by a given VSAT in Figure 4.14 can be multiplexed onto a unique *Multiple Channels Per Carrier* (MCPC) carrier. Receiving that carrier, any VSAT can demodulate it and extract from the baseband multiplex the traffic dedicated to the user terminals attached to it. Now, every VSAT still needs $N - 1$ receivers, but only

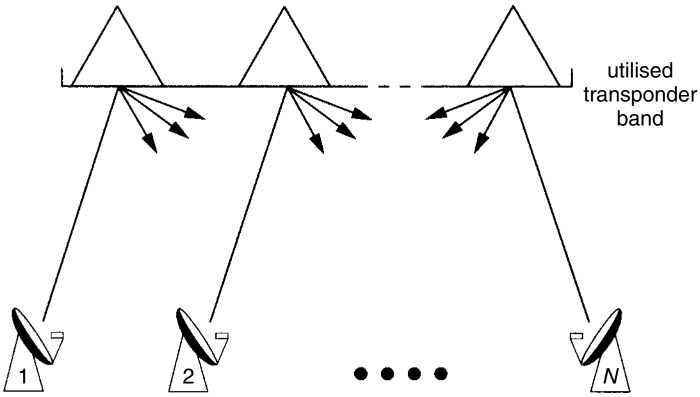


Figure 4.15 Variant of Figure 4.14 where the overall traffic from a VSAT to all other VSATs is multiplexed on a single carrier

one transmitter. However, the capacity of the transmitted carrier is higher, thus the VSAT transmitter must be more powerful. This scheme is represented in Figure 4.15.

The problem of having many receivers and transmitters comes from the requirement for permanent full connectivity. Actually, there is seldom need for such a requirement: indeed, apart from some broadcasting applications, the customer usually only requests that temporary two-way connections be set up between any two remote terminals attached to two different VSATs in the network. This works out most conveniently through demand assignment (see Chapter 1, section 1.5.3): should a terminal ask for such a connection, then the VSAT it is attached to sends a request on a signalling channel to a traffic control station, which replies by allocating some of the available satellite resource to both the calling and the called VSATs. With FDMA, this resource consists of two subbands on the satellite transponder, one for each carrier transmitted by the two VSATs. So any VSAT needs only to be equipped with one transmitter and one receiver, both tunable on request to any potential frequency band allocation within the transponder bandwidth.

Should now TDMA be used in Figure 4.15 instead of FDMA, then permanent full connectivity can be achieved with only one carrier being transmitted and received by every VSAT. This looks appealing but one must consider the higher cost of the TDMA equipment, and the fact that permanent full connectivity is not really needed.

With CDMA, the analysis follows the same lines as with FDMA. With demand assignment, temporary connections are set up by allocating to every transmitting VSAT a specific code. However, there does not seem to be any advantage to using CDMA apart from for small VSAT networks operating at C-band. CDMA then offers protection against interference generated by other systems.

Most of today's commercial meshed networks are based on demand assignment FDMA.

4.6.3 Star-shaped networks

The star-shaped network comprises N VSATs and a hub. Every VSAT can transmit up to K carriers, corresponding to connections between terminals attached to the VSAT and the corresponding applications at the host computer connected to the hub station.

4.6.3.1 FDMA–SCPC inbound/FDMA–SCPC outbound

Figure 4.16 illustrates the case where two-way connections between any remote user terminal and the host computer are supported by means of two Single Channel Per Carrier (SCPC) carriers: one from the VSAT to the hub station, and one from the hub to the VSAT. Every carrier requires its own modulator and demodulator. Hence, this configuration requires K modulators and demodulators at every VSAT and KN modulators and demodulators at the hub station. This is costly if the number of VSATs is large and K larger than 1. For instance, with $N = 100$ and $K = 3$, three hundred modulators and demodulators are to be installed at the hub.

With demand assignment, frequency agility is required for both transmitting and receiving VSATs.

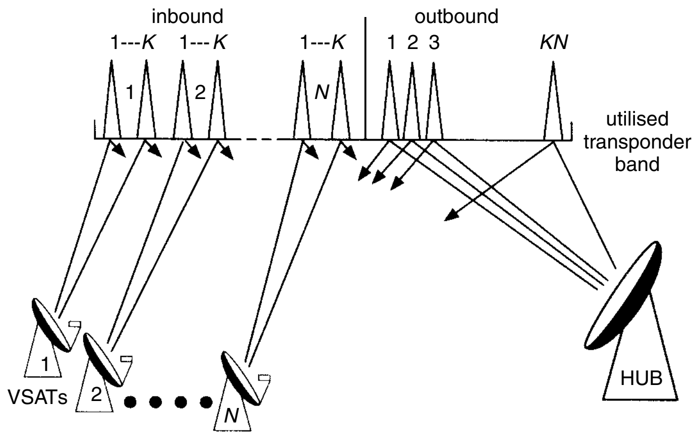


Figure 4.16 Star-shaped network with a two-way connection being conveyed on two SCPC carriers: one from the VSAT to the hub station, and one from the hub to the VSAT. Satellite transponder access is FDMA

4.6.3.2 FDMA–SCPC inbound/FDMA–MCPC outbound

Considering that any carrier transmitted by the hub is received by all VSATs, the number of modulators at the hub can be reduced, as indicated in Figure 4.17, by time division multiplexing the traffic from the hub to one VSAT on an outbound Multiple Channels Per Carrier (MCPC) carrier. The number of modulators at the hub is now equal to the number, N, of VSATs.

As the number of multiplexed connections on any outbound carrier may vary with time, the hub modulators and VSAT demodulators must be able to accommodate variable rates. The transmitted rate from the hub is higher with MCPC carriers. This translates into a higher demand for power from the hub transmitter.

With demand assignment, frequency agility is required for transmitting VSATs only.

4.6.3.3 FDMA–SCPC inbound/TDM–MCPC outbound

The number of modulators at the hub and demodulators at the VSATs can even be reduced to one, as illustrated in Figure 4.18, by time division multiplexing all connections from the hub to the VSATs on one MCPC outbound carrier. The hub modulator and the VSAT demodulator can operate at constant bit rate, equal to the maximum capacity of the network. But as a result of the higher bit rate, the demand for power from the hub transmitter is increased.

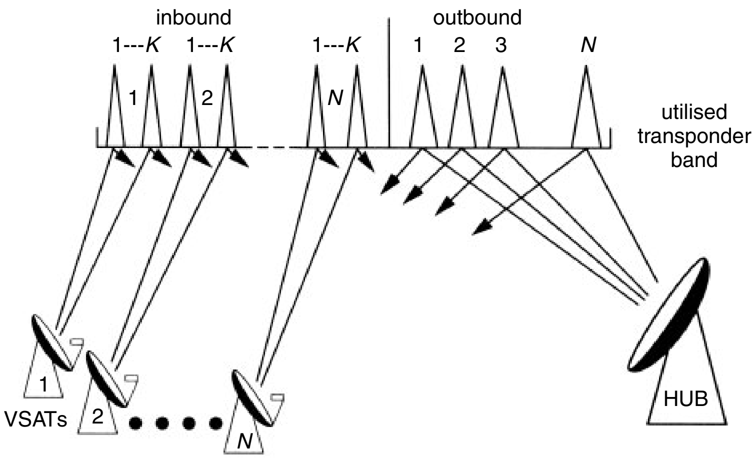


Figure 4.17 Star-shaped network with a two-way connection being conveyed on one SCPC carrier from the VSAT to the hub station, and multiplexed with others on one MCPC carrier from the hub to the VSAT. Satellite transponder access is FDMA

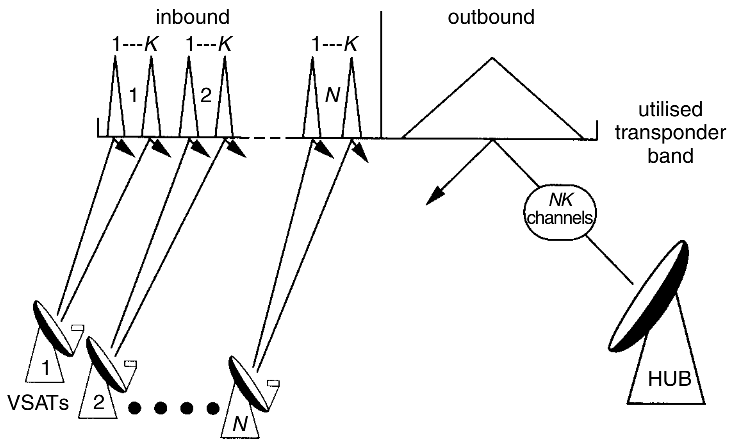


Figure 4.18 Star-shaped network with a two-way connection being conveyed on one SCPC carrier from the VSAT to the hub station, and time division multiplexed (TDM) with all others on the MCPC carrier transmitted by the hub. Satellite transponder access is FDMA

The large inbalance in input power between the low powered inbound carriers and the high powered outbound carrier results in a '*capture effect*' at the output of the satellite transponder, when used near saturation: the outbound carrier has a larger share of the output transponder power than its share at the input [MAR02, p. 452]. Therefore, less power is available to the inbound carriers.

With demand assignment, frequency agility is required for transmitting VSATs only.

4.6.3.4 FDMA–MCPC inbound/TDM–MCPC outbound

The number of modulators at the VSATs can be reduced to one, as illustrated in Figure 4.19, by time division multiplexing the traffic on the *K* inbound carriers from every VSAT to the hub station onto a single MCPC carrier. As the number of multiplexed connections on the inbound link may vary with time, the VSAT modulator must be at variable rate. Also, as the transmission rate is higher, the VSAT transmitter must be more powerful. The hub station needs to be equipped with *N* demodulators only.

With demand assignment, frequency agility is required for transmitting VSATs only.

4.6.3.5 TDMA inbound/TDM–MCPC outbound

The VSAT may now access the satellite transponder in a TDMA mode, every VSAT transmitting its carrier burst in sequence, at

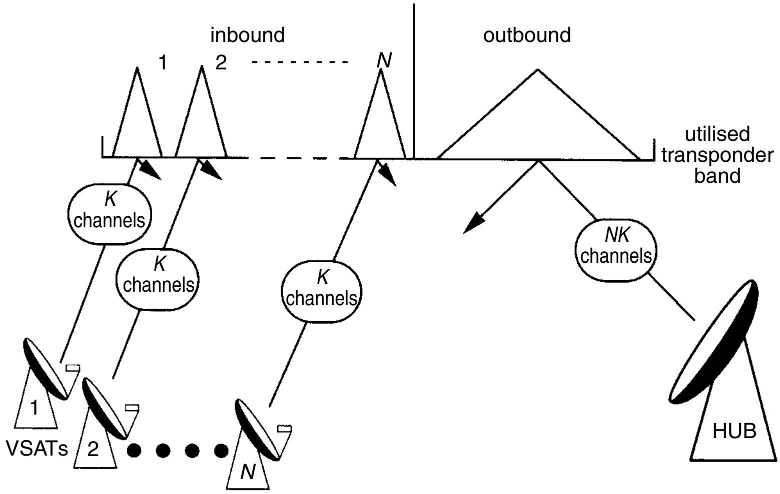


Figure 4.19 Star-shaped network with time division multiplexed (TDM) two-way connections conveyed on two MCPC carriers: one from the VSAT to the hub station, and one from the hub to the VSAT. Satellite transponder access is FDMA

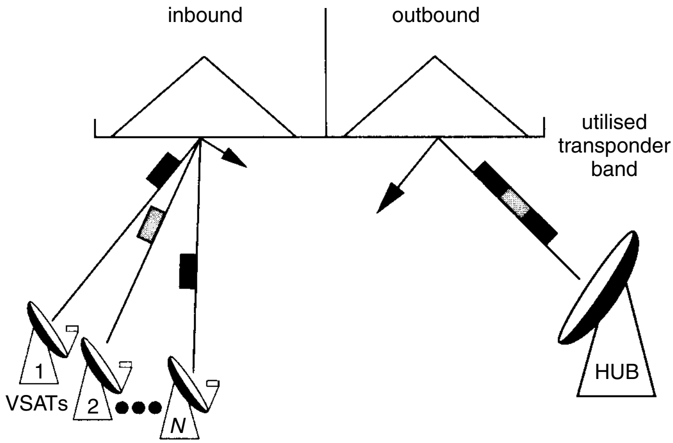


Figure 4.20 Star-shaped network with TDMA

the same bandwidth and the same frequency, as illustrated in Figure 4.20. Each burst may convey the traffic of either one one-way connection (SCPC) or several one-way connections (MCPC). In the latter case, the burst is divided into subbursts, each subburst being associated with one one-way connection. Denoting by T_B the duration of a carrier burst and by T_F the duration of the TDMA frame, any VSAT transmits with a duty cycle T_B/T_F.

The capacity of a radio frequency link from a VSAT is equal to the number of transmitted bits per unit of time. In a TDMA scheme, if

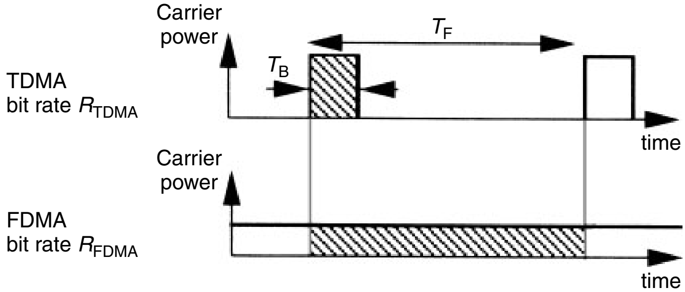


Figure 4.21 Comparison of bit rate and carrier power for FDMA and TDMA

a VSAT is to benefit from the same radio frequency link capacity as with FDMA, then it has to transmit at a higher bit rate. Indeed, with FDMA, the radio frequency link capacity is equal to the continuous transmitted bit rate. With TDMA, the radio frequency link capacity of the VSAT is given by the number of bits transmitted per TDMA frame duration.

As can be seen from Figure 4.21, where R_{TDMA} and R_{FDMA} are the transmitted bit rates for, respectively, TDMA and FDMA, T_{B} is the burst duration and T_{F} the TDMA frame duration, the number of bits transmitted per frame duration is equal to $R_{\mathrm{TDMA}}\ T_{\mathrm{B}}$ for TDMA, while it is equal to $R_{\mathrm{FDMA}}\ T_{\mathrm{F}}$ for FDMA. Equating these two expressions leads to:

$$R_{\mathrm{TDMA}} = R_{\mathrm{FDMA}} \left(\frac{T_{\mathrm{F}}}{T_{\mathrm{B}}} \right) \tag{4.17}$$

Clearly the transmission rate is higher by a factor equal to the inverse of the duty cycle. If one neglects guard time between bursts, the inverse of the duty cycle is equal to the number of VSATs in the network. Therefore, for a given capacity, a large number of VSATs entails a high bit rate transmission.

It will be shown in Chapter 5 that the power of the carrier is proportional to the bit rate. Therefore, TDMA demands more power than FDMA from the VSAT transmitters.

Consider, for instance, a VSAT network with $N = 50$ VSATs, each with a radio frequency link capacity of 64 kbs^{-1}. With FDMA, all VSATs transmit at $R_{\mathrm{FDMA}} = 64$ kbs^{-1}, and the satellite transponder bandwidth has to support 50×64 kbs^{-1} = 3.2 Mbs^{-1}. With TDMA, the same bandwidth would be used but now all VSATs would be requested to transmit at a rate of 3.2 Mbs^{-1}, thus increasing the demand for power by a factor of 50, or 17 dB, which is beyond the capability of cheap VSATs. Therefore, it would be necessary to reduce the capacity of every VSAT.

The following scheme, which is a hybrid combining FDMA and TDMA, brings some flexibility to a cost effective design.

4.6.3.6 FDMA–TDMA inbound/FDMA–MCPC outbound

In order to lower the requirement on the VSAT transmitter power by reducing the transmitted bit rate, an elegant solution is to organise VSATs into groups, with L VSATs per group, a group sharing the same frequency band and accessing the satellite transponder in a TDMA mode. The different groups use different frequency bands: this is a *combined FDMA–TDMA* scheme, as illustrated in Figure 4.22. With this approach, given the number N of VSATs in the network and capacity per VSAT, the transmitted bit rate on the carrier burst, and hence the required carrier power, is divided by G, the number of groups.

For instance, in the previous example, by splitting the 50 VSATs into 5 groups of 10 VSATs, the transmitted bit rate reduces from 3.2 Mbs^{-1} for pure TDMA to 640 kbs^{-1}, and the increase in power demand compared to pure FDMA is only 10 dB instead of 17 dB.

It may be convenient to consider time division multiplexing of all connections from hub to VSATs of the same group on one MCPC outbound carrier. The MCPC outbound carriers for the different

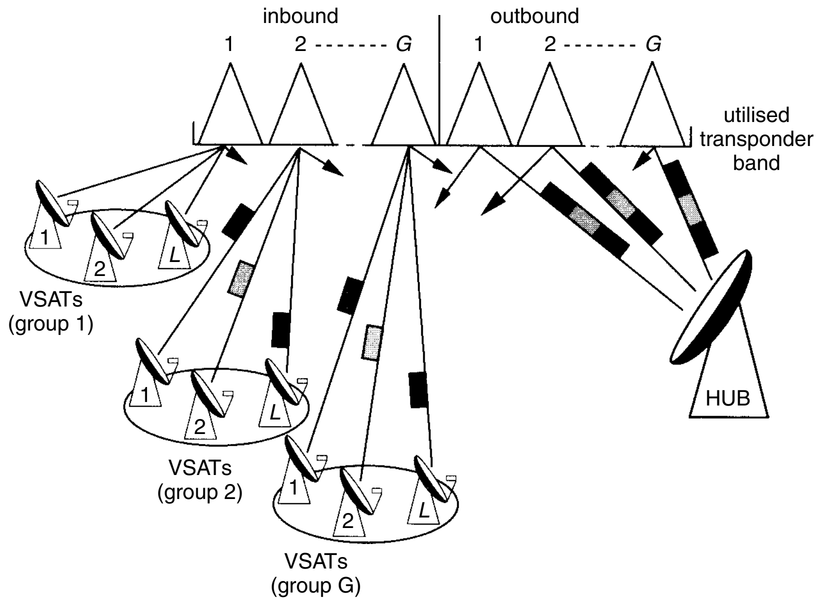


Figure 4.22 Star-shaped network using a combined FDMA–TDMA inbound scheme, and an FDMA–MCPC outbound scheme

groups then access the transponder in an FDMA mode. This reduces the bit rate transmitted by the hub, and hence its transmitter power, and offers the network manager the opportunity of implementing groups of VSATs as independent networks sharing a common hub.

4.6.3.7 CDMA

Figure 4.23 illustrates the variety of schemes that can be considered in connection with full CDMA access, or a combination of CDMA and FDMA for the inbound and the outbound links. CDMA access can also be combined with SCPC or MCPC by grouping inbound connections.

With CDMA, carriers are assigned spreading pseudo-random codes instead of frequencies because all carriers use the same centre frequency. Hence frequency agility is no longer needed for demand assignment, eliminating the problem caused by frequency instability and phase noise encountered by SCPC/FDMA carriers that require precise frequency assignments. The major drawback to CDMA is its low throughput [MAR02, section 6.6.5 p 315] which can be accepted

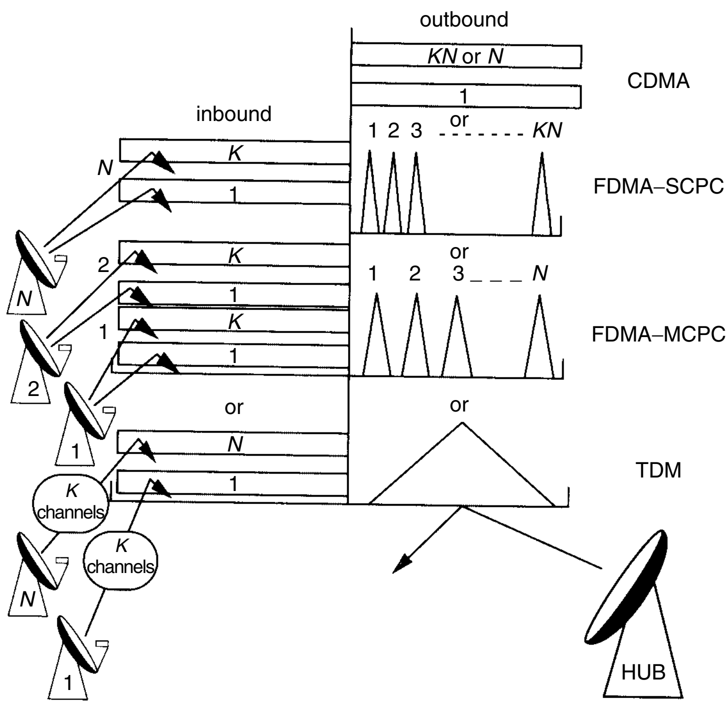


Figure 4.23 Star-shaped VSAT network using CDMA or a combination of CDMA and FDMA

only if it is balanced by the advantages gained from rejection of interference caused by other systems sharing the same frequency bands and polarisation.

4.6.4 Fixed assignment versus demand assignment

Demand assignment has been presented in Chapter 1, section 1.6.3 as an appealing option within VSAT networks. It has been shown in the context of meshed networks (see section 4.6.2) that demand assignment permits implementing the desired connectivity between VSATs by setting up temporary links, with reduced VSAT equipment compared to fixed assignment allowing permanent links.

Therefore, it is interesting to discuss the impact of demand assignment, compared to fixed assignment, in a general sense.

4.6.4.1 Fixed assignment with FDMA (FA–FDMA)

The network comprises L VSATs, each possibly transmitting up to K carriers at bit rate R_c. So we have at most $L = KN$ carriers, and every carrier is allocated a given subband of the satellite transponder bandwidth. This subband is used by a VSAT when active (carrier 'on'), and remains unused when the VSAT has no traffic to convey (carrier 'off'). Should this happen, the capacity corresponding to the subband allocated to the VSAT is lost to the network. Figure 4.24 illustrates how fixed assignment works for FDMA in the case where $K = 1$.

Fixed assignment has the advantage of simplicity, and provides no blocking nor waiting time for setting up a carrier. However, the required capacity for the VSAT network, equal to LR_c, is poorly utilised if the traffic demand is highly variable.

Blocking may occur at a user terminal attached to a given VSAT should several of these terminals wish to establish connections simultaneously with other terminals in the network, and the number of requested connections exceed the capacity of the VSAT.

For instance, consider the VSAT network of Figure 4.16, with $N_t = 8$ user terminals per VSAT, $N = 50$ VSATs and up to $K = 4$ SCPC carriers per VSAT, each transmitting at bit rate $R_c = 128\ \text{kbs}^{-1}$ and requiring bandwidth $b = 200$ kHz. The transponder bandwidth used by the inbound links is split into $L = KN = 4 \times 50 = 200$ subbands. Every subband is allocated to one $128\ \text{kbs}^{-1}$ carrier. Therefore, the required network capacity is $LR_c = 200 \times 128\ \text{kbs}^{-1} = 25.6\ \text{Mbs}^{-1}$, and the required bandwidth is $Lb = 200 \times 200\ \text{kHz} =$

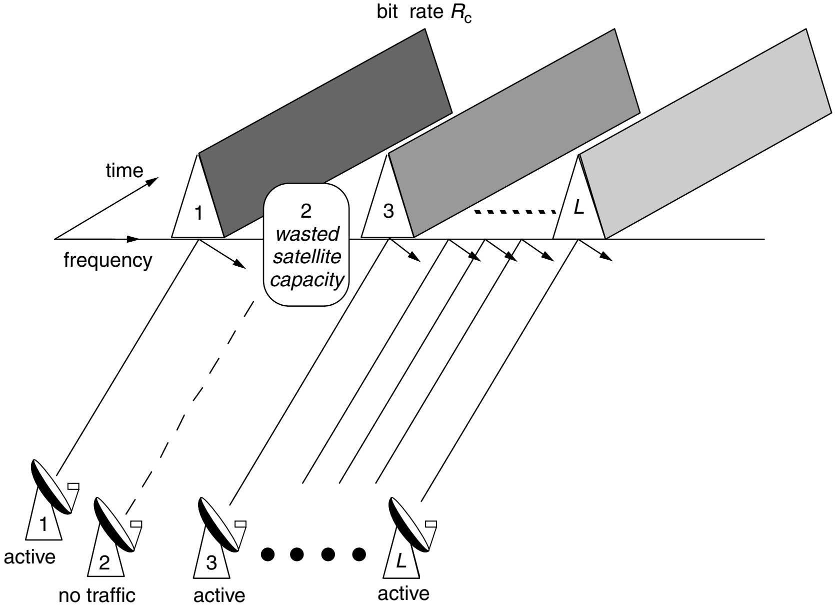


Figure 4.24 Fixed assignment FDMA (each VSAT transmits at most $K = 1$ carrier)

40 MHz. Assuming user terminals generate traffic with intensity $A_t = 0.1$ erlang, then the traffic intensity offered to the $K = 4$ VSAT channels' capacity is $A_{VSAT} = N_t A_t = 0.8$ erlang. The probability of blocking, as given by formula (4.5), is equal to:

$$E_4(0.8) = 8 \times 10^{-3} = 0.8\%$$

In order to avoid any blocking, each user terminal should be permanently allocated a channel, this would imply $K = 8$, and therefore a total of $KN = 8 \times 50 = 400$ SCPC carriers. The required network capacity would then be $KNR_c = 8 \times 50 \times 128\ \text{kbs}^{-1} = 51.2\ \text{Mbs}^{-1}$, and the required transponder bandwidth would be 80 MHz.

4.6.4.2 Demand assignment with FDMA (DA–FDMA)

The network again comprises N VSATs, each possibly transmitting K carriers, and sharing a pool of L frequency subbands, but now $L < KN$. These subbands are used by the active VSATs. Figure 4.25 illustrates how demand assignment works for FDMA in the case where $K = 1$. Should the number of carriers exceed the number that can be supported by the allocated satellite transponder bandwidth, then there is blocking at the VSAT level: at the time of the call, no new carrier can be established.

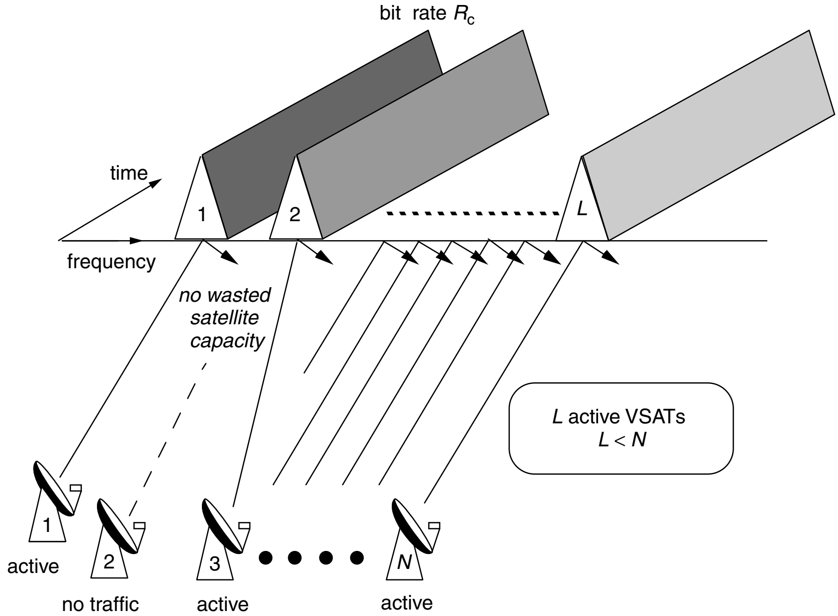


Figure 4.25 Demand assignment FDMA ($K = 1$)

For example, again consider the VSAT network of Figure 4.16, with $N_t = 8$ user terminals per VSAT, and up to $K = 4$ SCPC carriers per VSAT, each transmitting at bit rate $R_c = 128$ kbs^{-1} and requiring bandwidth $b = 200$ kHz. Assuming user terminals generate traffic with intensity $A_t = 0.1$ erlang, then the traffic intensity offered to the $K = 4$ VSAT channels' capacity is $A_{VSAT} = N_t A_t = 0.8$ erlang. Assuming again $N = 50$ VSATs, at a given time at most K terminals per VSAT generate traffic, so the traffic intensity offered to the L subbands is $A = NKA_t = 50 \times 4 \times 0.1 = 20$ erlang. The blocking probability for setting up a link can be maintained at a low level by having a sufficiently large pool of subbands. For instance, using (4.5), and taking $L = 35$ subbands, the carrier set-up blocking probability is $E_{35}(20) = 7 \times 10^{-4} = 0.07\%$.

A call is blocked either because a terminal cannot access any one of the K channels, or because the VSAT cannot access any one of the L subbands. The probability for a call to be blocked is given by:

$$\begin{aligned} P_{\text{blocked}} &= E_4(0.8) + E_{35}(20) \\ &= 8 \times 10^{-3} + 7 \times 10^{-4} \\ &= 0.87\% \end{aligned}$$

The required network capacity is now $LR_c = 35 \times 128$ kbs^{-1} = 4.48 Mbs^{-1}, and the required bandwidth is $Lb = 35 \times 200$ kHz =

7 MHz. Therefore, with a negligible increase in call blocking probability, demand assignment offers a potential saving of $100(1 - L/KN)\% = 100(1 - 35/200) = 82.5\%$ of the used satellite transponder bandwidth.

4.6.4.3 Fixed assignment with TDMA (FA–TDMA)

Figure 4.26 illustrates fixed assignment in connection with TDMA operation. Every VSAT transmits a carrier burst within a dedicated time slot. The number of time slots is equal to the number, L, of VSATs. Position and duration of bursts are fixed, therefore the capacity of every VSAT is constant whatever the traffic demand.

If R_c is the transmitted bit rate, then the total network capacity is R_c, and the capacity allocated to a VSAT is R_c/L. Should a VSAT have no traffic to transmit, then the slot remains unoccupied, and the corresponding capacity is lost for the network.

Fixed assignment has the advantage of simplicity, and provides no blocking nor waiting time for setting up a carrier. However, the total network capacity of the VSAT network (transponder bandwidth allocated to the network) is poorly utilised if the traffic demand is highly variable.

Blocking may occur at the user terminal. The blocking probability can be calculated following derivations similar to that of the example of fixed assignment with FDMA. One can assume that the carrier burst is split into K subbursts, each corresponding to one channel available to the attached user terminals.

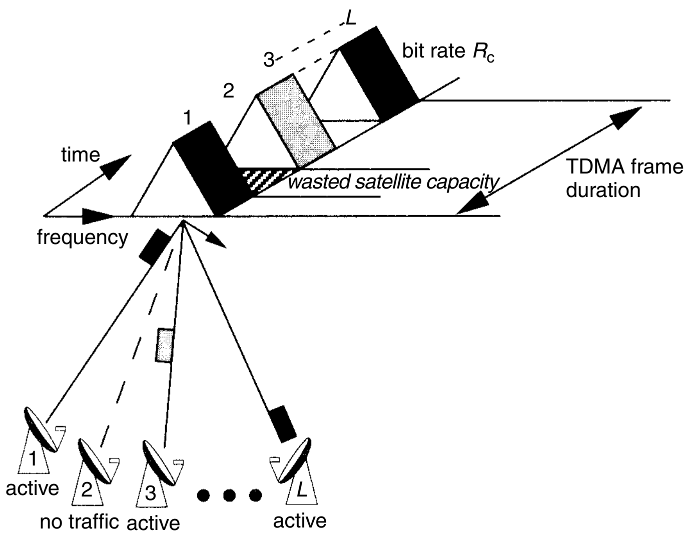


Figure 4.26 Fixed assignment TDMA

Considering, as previously, $L = 50$, $K = 4$, $N_t = 8$ and $A_t = 0.1$ erlang, the blocking probability for a terminal is $E_4(0.8) = 8 \times 10^{-3} = 0.8\%$, as previously.

It is worth recalling that in order to achieve comparable capacities per VSAT, the value of R_c must be increased by a factor L relative to the FDMA scheme (see equation (4.17)).

4.6.4.4 Demand assignment with TDMA (DA–TDMA)

Figure 4.27 shows how the L time slots of the frame are now shared by N VSATs, with $N > L$. Any VSAT wishing to set up a link (carrier 'on' from carrier 'off' state) can access any unoccupied time slot on the frame, or should it already be active, it can increase its capacity by increasing the duration of its burst, and then support a larger number of connections. This requires a change in the burst time plan, and is performed under the control of the network management system (NMS) at the hub station.

As the traffic demand from all VSATs may exceed the offered capacity R_c, blocking of link set-up may occur as a result of the TDMA frame being filled with carrier bursts.

For example, the network capacity $KL = 4 \times 50 = 200$ channels considered in the above FA–TDMA scheme is now available as a pool to all user terminals, whose total traffic intensity is $A = N \times 8 \times 0.1 = 0.8\ N$ erlang, where N is the number of VSATs in the network. For a comparison with FA–TDMA, N can be selected

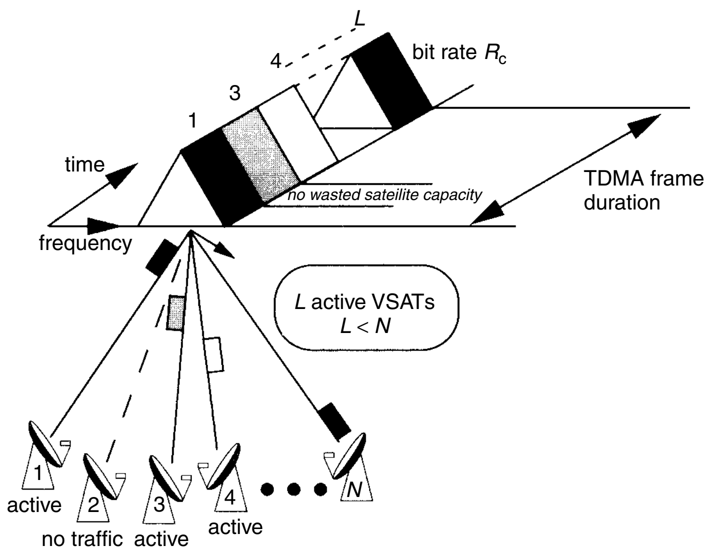


Figure 4.27 Demand assignment TDMA

so as to achieve a 0.8% blocking probability for a terminal. This means solving $E_{200}(A) = 0.8\%$, which corresponds to $A = 178$ erlang. Therefore $N = 178/0.8 = 222$, which indicates that the number of VSATs in the network can be increased by a factor $222/50 = 4.4$.

4.6.4.5 Demand assignment multiple access (DAMA) procedure

With demand assignment, a VSAT receives a call demand from one of the user terminals attached to it. This is indicated in Figure 4.28 as 'call arrival time'. This call may concern:

1. An application at the host computer attached to the hub, or
2. A user terminal attached to some other VSAT in the network.

As illustrated in Figure 4.28, the VSAT then sends a request to the hub by means of a specific inbound signalling channel, and the hub allocates the requested capacity, if available, to the corresponding VSAT by means of response messages transmitted on an outbound

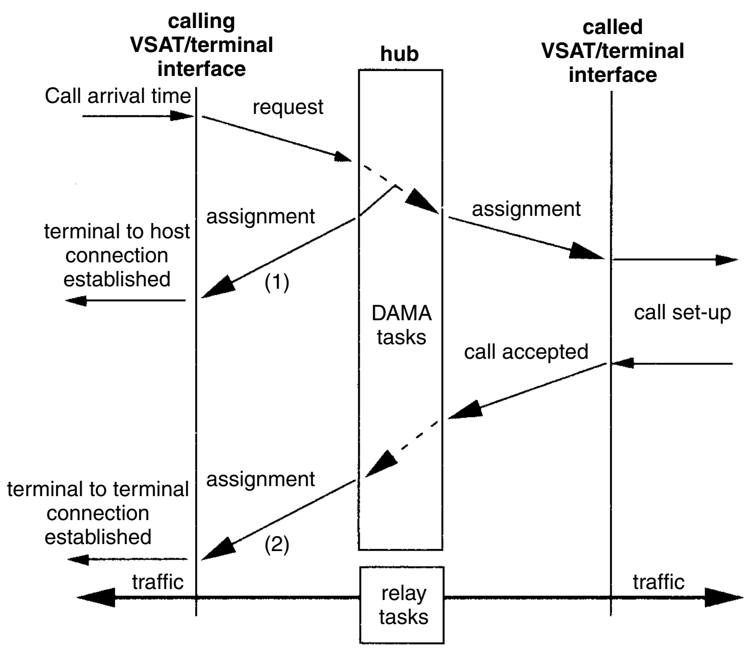


Figure 4.28 Demand assignment multiple access (DAMA) procedure: (1) terminal to host computer connection set-up, (2) terminal to terminal connection set-up

signalling channel, and conveying capacity assignments (carrier frequency, time slot or code).

If the connection is between a user terminal and the host computer, the delay of the incoming response is a two-hop delay, plus processing time. If the connection is between two user terminals, the hub transmits an assignment to the destination terminal and waits for a call acceptance message. This avoids assigning capacity to the incoming call before making sure the call can be accepted by the destination terminal. Then the hub assigns capacity to the calling VSAT. The connection set-up then takes a four-hop delay plus processing time. Once the connection is established from terminal to terminal, the hub acts as a relay.

Demand assignment requires that some network capacity be dedicated to request and response signalling. Moreover, some reservation delay is to be expected as a result of the time for signalling to be routed and processed.

When the access of VSATs to the signalling channel is organised according to a fixed assignment scheme, the need for limiting the capacity of the signalling channel to a reasonable fraction of the total network capacity will limit the number of VSATs in the network. In order to avoid such a limitation and offer the possibility of easy addition of new VSATs to the network, a random access scheme to the signalling channel is often preferred.

4.6.4.6 Demand assignment limitations

From the above, it can be seen that demand assignment allows more VSATs to share the satellite resource, or, for a given network size, allows a reduction in the utilised satellite bandwidth. Also to be considered is the penalty for signalling capacity. However, it can be kept small enough.

Of greater concern is the time delay: a message must wait at the VSAT before being transmitted while requests are forwarded and channel allocation is completed. As a result of propagation time and processing time, the delay may be as high as one to two seconds. This is not compatible with efficient transmission of short messages which constitute bursty traffic, should a connection be set up every time a message comes in.

To illustrate the problem, consider Figure 4.29 where the connection set-up delay is compared to the transmission time of a message (also called 'service time'). Table 4.1 indicates that the typical message length for bursty traffic is a few hundred bytes. Consider, for example, a message length of 200 bytes, the service time at a rate

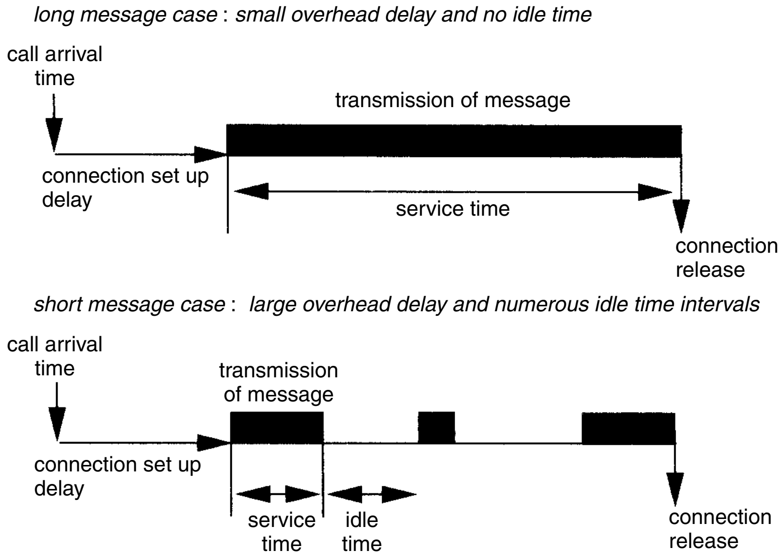


Figure 4.29 Overhead delay associated with connection set-up and idle time during connection for the case of a long message and for that of a short one (bursty traffic)

of 64 kbs^{-1} is 25 ms. If one considers a connection set-up delay of 1.5 s, then the overhead delay before transmission of the message is as large as 60 times the message transmission time.

Moreover, with bursty traffic the temporary connection established on demand for message transfer is poorly utilised as a result of idle times. The greater the burstiness, the worse the problem. Indeed, considering an information transmission rate R_b and a message length L, the service time is $\tau = L/R_b$. If we denote by $\langle \text{IAT} \rangle$ the average interarrival time, and recall the expression for burstiness from (4.9):

$$\text{BU} = \frac{R_b \langle \text{IAT} \rangle}{L} \tag{4.18}$$

then, from the definition of channel utilisation (section 4.2.7):

$$\begin{aligned} \text{Channel utilisation} &= \frac{(\text{service time})}{(\text{service time} + \text{idle time})} \\ &= \frac{\tau}{\langle \text{IAT} \rangle} \\ &= \frac{(L/R_b)}{\langle \text{IAT} \rangle} \\ &= \frac{1}{\text{BU}} \end{aligned} \tag{4.19}$$

For instance, considering a value BU = 10 000, the channel utilisation is $10^{-4} = 0.01\%$. With such a low utilisation of the channel, the advantage gained in terms of capacity reduction from demand assignment is lost.

Bursty traffic is routed most efficiently if:

– VSATs transmit at once whenever they get traffic from the user terminals. This forbids any connection set-up delay;
– the capacity derived from the utilisation of transponder bandwidth is shared between all VSATs to allow statistical multiplexing of the bursts.

These conditions are satisfied in random time division multiple access schemes, often named ALOHA.

4.6.5 Random time division multiple access

4.6.5.1 Principle

Random time division multiple access, also named ALOHA, has been introduced in section 4.6.1. There are two modes: unslotted ALOHA, and slotted ALOHA (S-ALOHA) [HA86 p362–378][SCH77, Chapter 13]. With unslotted ALOHA, VSATs can transmit at any time, which means that they are not synchronised. With S-ALOHA, VSATs transmit in time slots, which means that they are synchronised but not coordinated, in the sense that while transmitting in a given time slot they ignore whether other VSATs are transmitting or not in the same time slot.

Figure 4.30 illustrates the principle of S-ALOHA. Every carrier is transmitted in the form of a burst with duration equal to that of a time slot. Every carrier burst conveys a packet of data. Synchronisation between VSATs is derived from the signals transmitted by the hub station and received on the outbound link.

Transmission of a packet is initiated by a message being generated by a user terminal attached to the VSAT. The length of the message may not coincide with the length of a packet. If too small, it must be completed by dummy bits. If too large, it must be conveyed over several packets. As users are not coordinated, messages as well as carrier bursts are generated at random.

The transmission efficiency of the S-ALOHA protocol is measured by the normalised throughput S, expressed in number of packets successfully transmitted per packet length. The average rate of

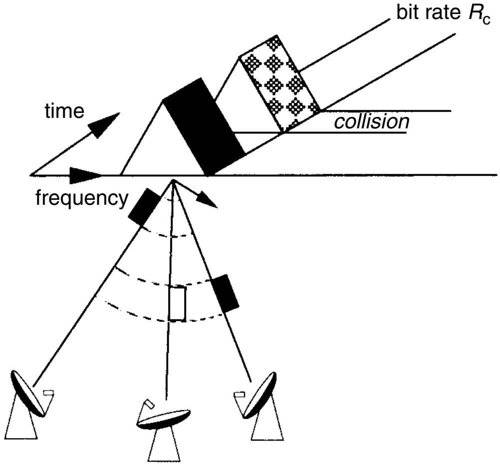


Figure 4.30 Principle of S-ALOHA

delivered bits $\langle R_c \rangle$, assuming packet length L bits, packet duration τ and transmitted bit rate $R_c = L/\tau$ is:

$$\langle R_c \rangle = S\left(\frac{L}{\tau}\right) = SR_c \tag{4.20}$$

It can be shown [TAN89 p 22], [HA86 p 362] that S depends on the average offered traffic, G (in packets per packet length), consisting of newly generated and retransmitted packets, and the number, N, of VSATs in the network:

$$S = G\left(1 - \frac{G}{N}\right)^{N-1} \tag{4.21}$$

As N becomes infinite, the above expression becomes:

$$S = G\,\mathrm{e}^{-G} \tag{4.22}$$

These equations are represented in Figure 4.31.

From Figure 4.31, one can see that the normalised throughput converges rapidly to the infinite population case and that the maximum normalised throughput for an infinite number of VSATs is equal to $1/\mathrm{e} = 0.368$, or 37%, which is poor.

With unslotted ALOHA, expression (4.22) is replaced by:

$$S = G\,\mathrm{e}^{-2G} \tag{4.23}$$

which leads to similar curves, but with an even lower maximum normalised throughput $S_{max} = 1/2\mathrm{e} = 0.184 = 18\%$.

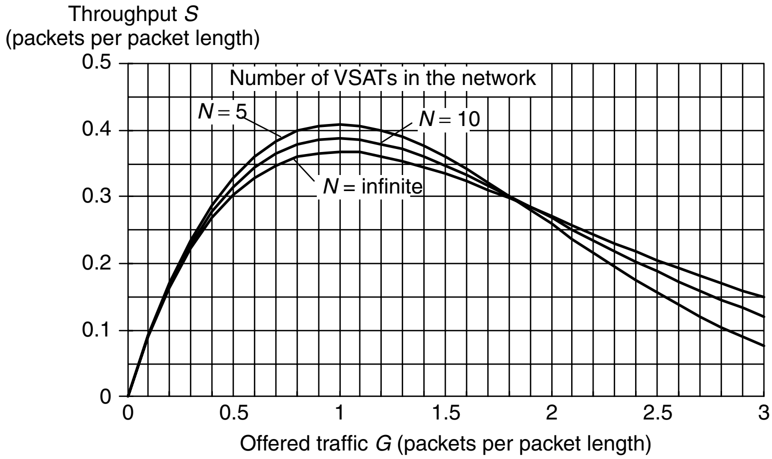


Figure 4.31 Throughput S versus offered traffic G for VSAT inbound link using S-ALOHA multiple access scheme

The low throughput of ALOHA schemes could be expected as VSATs may transmit at random in any time slot. The advantage of S-ALOHA is a higher throughput than unslotted ALOHA, and, as will be demonstrated in section 4.6.6, a lower average delay at low throughput (i.e. with highly bursty traffic), than that achieved with demand assignment FDMA or TDMA.

For a given transmission bit rate R_c (bs^{-1}), the number, N, of VSATs that can be installed in the network given the packet generation rate, λ(s^{-1}), per VSAT and the length, L, of a packet relates to the normalised throughput, S:

$$N = \frac{SR_c}{\lambda L} \tag{4.24}$$

The above expression is displayed in Figure 4.32 for various packet lengths and practical values of $S = 0.1$, and $R_c = 64$ kbs^{-1}.

4.6.5.2 Random access limitations

Instability

Packets suffering collisions must be retransmitted. VSATs awaiting retransmission are said to be 'backlogged'. The retransmitted packets generate an excess of traffic which adds to the newly generated packets. Therefore, the offered traffic G exceeds the successful traffic, or throughput S. Referring to Figure 4.31, should the value of G exceed one, which corresponds to the maximum value of S, then as the throughput S decreases for any increase of G, collisions

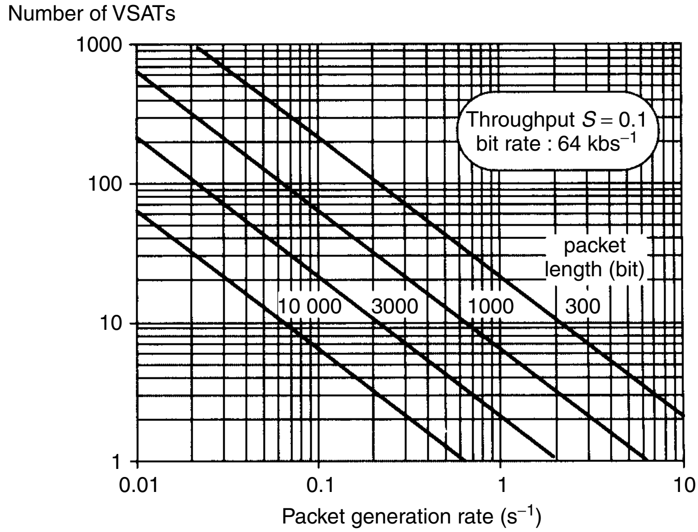


Figure 4.32 Number of VSATs in the network as a function of the packet generation rate for various packet lengths (in bits). A practical throughput $S = 0.1$ is considered, and a transmission bit rate $R_C = 64$ kbs^{-1}

become even more frequent and the network evolves towards a situation of high G and low S where VSATs constantly retransmit the same packets.

Retransmission procedures are:

1. Fixed retransmission probability: the backlogged terminal retransmits a packet with a fixed probability during each slot. This is simple but may be unstable.
2. Adaptive strategies: the feedback channel is observed and the retransmission probability is adapted according to the history of the channel.
3. Heuristic retransmission: the retransmission probability is adjusted according to the number of retransmission attempts that have already been performed for the current packet.

In order to avoid instability, the design of a VSAT network should consider a practical throughput S of 5 to 15%. Figure 4.32, which considers $S = 0.1$ can be used to determine the appropriate number of VSATs.

Long message case

Successful transmission of long messages generated by the user terminals entails successful transmission of several consecutive packets. It can be shown that the throughput of an S-ALOHA scheme when long messages made of many consecutive packets must be retransmitted is given by:

$$S = \frac{G\,\mathrm{e}^{-G}}{1 + G^2} \tag{4.25}$$

The above formula indicates that the throughput compared to the case of single packet S-ALOHA, as given by expression (4.22), is lowered by a factor $(1 + G^2)$, and has a maximum value of 0.137 at $G = 0.414$.

In order to overcome the above limitations, variants of the random time division multiple access schemes have been proposed and will now be presented.

4.6.5.3 Selective reject ALOHA

Messages are transmitted asynchronously as in unslotted ALOHA, but are partitioned into a finite number of short packets, each with its own acquisition preamble and header. The protocol exploits the fact that on an unslotted channel most collisions are partial, so that uncollided portions of messages encountering conflict can be recovered by the receiver, and only the packets actually encountering conflict are retransmitted, in a way similar to the selective repeat protocol of Figure 4.7(c). The maximum throughput has been shown to approach 0.368, irrespective of the message length distribution. In practice, the need for acquisition preamble and header in each packet limits the maximum useful throughput to the range of 0.2–0.3. The critical need for low packet overhead requires burst modems with short acquisition times.

4.6.5.4 Reservation/random TDMA

The inbound packets are transmitted in the time slots of a TDMA frame. This TDMA frame is shared by a group of VSATs in the network, as illustrated in Figure 4.22. The S-ALOHA protocol is used to convey messages short enough to be sent in a single packet. When an arriving message exceeds the slot size, a request for the necessary number of dedicated slots is made via S-ALOHA, while the remaining packets are queued at the VSAT waiting for a time slot assignment message from the hub for the remaining part of the message. All participating VSATs receiving the outbound carrier from the hub are informed of the specific slots reserved for the requesting VSAT and will refrain from transmitting packets in these slots. The requesting VSAT now can transmit the remaining part of the message without collision.

The maximum number of time slots in a TDMA frame to be used for reservation is limited to a predetermined value so that a certain number of slots are always kept free for random use. This precludes the occurrence of a lockout situation.

The requests for reservation can be either transmitted within a specific time slot divided into smaller slots, to be used on a contention basis with the ALOHA technique, within dedicated time slots, or piggy backed on the first packet of the message. Figure 4.33 illustrates the latter case.

The arrival of a long message triggers the emission of a reservation message. Should the transmission of reservation messages be achieved on a contention basis using the S-ALOHA protocol, a high degree of correlation exists between the reservation traffic fluctuations and the data traffic load. This correlation may generate instability, as data traffic may display high instantaneous fluctuations. To alleviate the problems caused by the correlation, it has been proposed that a reservation message convey reservation requests for more than one data message [ZEI91]. Not only does this reduce the load of the reservation channel, but the reservation protocol becomes unconditionally stable by simply indicating in the reservation message the number of data slots required for all the data messages waiting in the buffer of the VSAT.

4.6.5.5 Random access with notification

Satellite transponder capacity is divided into two portions: a signalling channel shared in fixed TDMA mode by multiple VSATs, each with a dedicated signalling slot, and a data channel to carry all data packets. Signalling and data channels are divided into fixed duration frames, and each frame is divided into a number of time

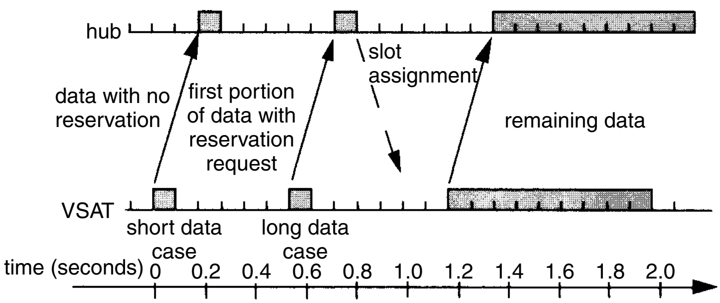


Figure 4.33 Reservation/random TDMA

slots. Frames and slots are numbered and synchronised among all VSATs via centralised control from the hub.

For the data channel, two types of slot are considered: random access and reservation. New packets use random access slots, while collided and errored packets will be allotted reserved slots. Over the separate TDMA signalling channel, every VSAT notifies the hub of the number of packets transmitted by the VSAT in a particular frame. If a packet experiences a collision, the hub recognises that the number of packets successfully received is less than the number that were reported as sent. The hub then allots the required number of slots to the VSAT in some future frame for contention-free retransmission in the reservation mode.

4.6.6 Delay analysis

Delay means the time it takes for a message to be transferred from the user terminal to the host computer. The delay is a random variable, and it is common to characterise it by its average value. However, for many applications it is not sufficient to design a VSAT network on the basis of the average delay. Quantities such as the 95th percentile of delay are of equal or greater significance.

For delay analysis, it is assumed that every user generates messages according to a Poisson process (see Appendix 1). Messages are stored at the VSAT in a buffer of infinite capacity. A message is transmitted by the VSAT to the hub in the form of packets with length L.

4.6.6.1 Delay components

The message delay components for the inbound link and the outbound link are as follows:

1. Delays at the VSAT:
   - polling and/or queuing at the VSAT-to-user terminal interface;
   - VSAT processing delay (formatting of packets);
   - protocol induced delays: with random access, this is the waiting time for retransmission if a collision has been encountered; with demand assignment, this is the waiting time for allocation;
   - time to send messages, called service time, which depends on the message size and the bit rate.

2. Satellite link transmission delay: this corresponds to the propagation time of the carrier, and is given in Figure 2.22.
3. Delays at the hub:
   - polling and/or queuing at the hub-to-host interface;
   - hub processing delay (formatting of packets);
   - time to send messages, called service time, which depends on the message size and the bit rate.

The results presented below deal with packet delay, defined as the elapsed time between the instant when a packet is being transmitted by a source VSAT to the instant when the last byte of the packet is received by the destination hub. Processing time at the VSAT and at the hub is neglected.

It should be remarked that the considered packet delay is induced by the multiple access protocol only. It does not include delays possibly induced by retransmissions initiated by the data link control protocol. However, if the bit error rate on the satellite link is low enough, and selection of an appropriate data link control protocol has been made, retransmissions due to errors are rare. It can then be considered that the packet delay induced by the multiple access protocol is representative of the actual delay.

4.6.6.2 Comparison of FDMA, TDMA and S-ALOHA

Expressions for the average packet delay are as follows:

FDMA [HA86 p357][TAN89 p636]

1. For exponentially distributed packet length:

$$T_{\mathrm{FDMA}} = T_{\mathrm{p}} + \frac{1}{\dfrac{R_{\mathrm{FDMA}}}{L} - \lambda} \quad \text{(s)} \tag{4.26}$$

2. For constant packet length:

$$T_{\mathrm{FDMA}} = T_{\mathrm{p}} + \frac{2 - \dfrac{\lambda}{\dfrac{R_{\mathrm{FDMA}}}{L}}}{2\left(\dfrac{R_{\mathrm{FDMA}}}{L} - \lambda\right)} \quad \text{(s)} \tag{4.27}$$

where T_{FDMA} is the average delay (s) for FDMA, T_{p} is the satellite link propagation delay (about 0.25 s), R_{FDMA} the carrier transmission rate

(bs^{-1}), L the length of a packet (bits) and λ the average generation rate of packets (s^{-1}) per carrier.

If the network has N VSATs and every VSAT transmits K carriers, then the total network capacity is $R = NKR_{FDMA}$. Therefore, $R_{FDMA} = R/NK$.

TDMA [HA86 p361]

1. For exponentially distributed packet length:

$$T_{TDMA} = T_p + \frac{1}{\frac{R_{TDMA}}{NL} - \lambda} - \frac{T_F}{2} + \frac{T_F}{N} \quad \text{(s)} \tag{4.28}$$

2. For constant packet length:

$$T_{TDMA} = T_p + \frac{2 - \frac{\lambda}{\frac{R_{TDMA}}{NL}}}{2\left(\frac{R_{TDMA}}{NL} - \lambda\right)} - \frac{T_F}{2} + \frac{T_F}{N} \quad \text{(s)} \tag{4.29}$$

where T_{TDMA} is the average delay (s) for TDMA, T_p is the satellite link propagation delay (about 0.25 s), T_F the frame duration (s), L the length of a packet (bits), λ the average generation rate of packets (s^{-1}) per VSAT, R_{TDMA} the carrier transmission rate (bs^{-1}) and N the number of VSATs sharing the network capacity $R = R_{TDMA}$.

If the network has N VSATs, every VSAT transmits at a rate R_{TDMA}, equal to the network capacity R, and its time share of the capacity is R_{TDMA}/N.

Demand assignment FDMA and TDMA

Assuming there is no call blocking, the requested connection is set up after an initial delay equal to the round trip propagation delay, equal to $2T_p$ (about 0.5 s). Once the connection is set up, the above formulas are applicable.

S-ALOHA [HA86 p369]

Assuming the number of VSATs to be large enough (typically larger than 10) for equation (4.22) to hold:

$$T_{S-ALOHA} = T_p + \frac{3\tau}{2} + (e^G - 1)(2T_p + \frac{(k+1)\tau}{2} + \frac{\tau}{2} \quad \text{(s)} \tag{4.30}$$

where $T_{S-ALOHA}$ is the average delay (s) for S-ALOHA, T_p the satellite link propagation delay (about 0.25 s), τ the packet duration (s): $\tau = L/R_{TDMA}$, where R_{TDMA} is the carrier transmission rate (bs^{-1}), G the offered traffic load (packet per time slot) and k the maximum retransmission interval (time slots).

Figure 4.34 displays packet delay curves for a constant length packet traffic, with packet length $L = 1000$ bits, and packet generation rate $\lambda = 0.1\ s^{-1}$ as a function of the number of VSATs in the network. Three multiple access protocols are compared: FDMA with delay given by (4.27), TDMA with delay given by (4.29) and S-ALOHA with delay given by (4.30). For FDMA, it is assumed that a VSAT transmits one carrier only ($K = 1$). For S-ALOHA, three values of the maximum retransmission interval have been selected: $k = 10$, 50 and 100 slots. The network capacity is constant and equal to $R = 100\ kbs^{-1}$. Thus, $R_{FDMA} = 100\ kbs^{-1}/KN = 100\ kbs^{-1}/N$, and $R_{TDMA} = R = 100\ kbs^{-1}$. The TDMA frame duration T_F varies with the number of VSATs: $T_F =$ (number of VSATs) $\times L/R_{TDMA} =$ number of VSATs $\times 10^{-3}$ seconds.

Clearly, S-ALOHA delivers a shorter delay than FDMA and TDMA. This demonstrates its greater ability to convey bursty traffic.

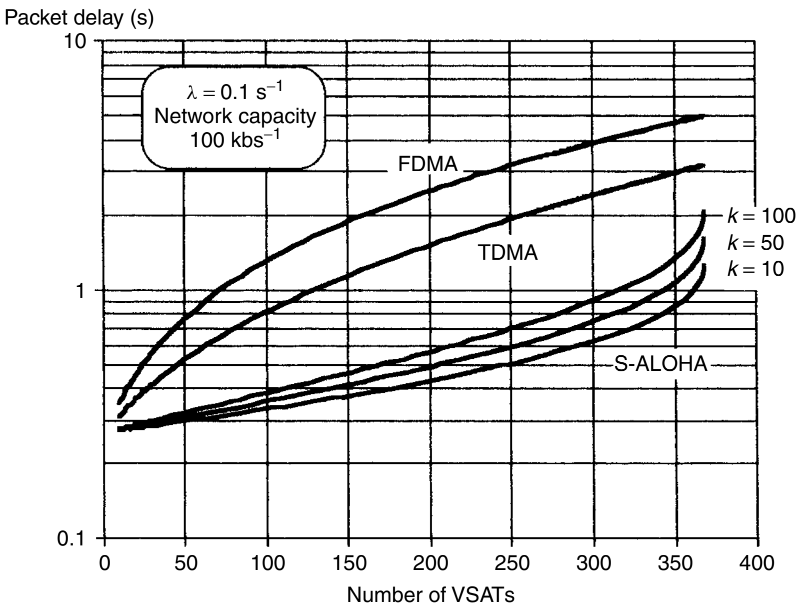


Figure 4.34 Packet delay for a constant length packet traffic, with packet length $L = 1000$ bits, and packet generation rate $\lambda = 0.1\ s^{-1}$ per VSAT as a function of the number of VSATs in the network. The network capacity is $R = 100\ kbs^{-1}$ With S-ALOHA, k represents the maximum retransmission interval in time slots

The normalised throughput S is given by:

$$S = \frac{N\lambda L}{R} \tag{4.31}$$

where N is the number of VSATs, and R the network capacity. With the selected values, $S = 10^{-3}N$. A safe design with the S-ALOHA protocol would be to limit the number of stations to below, say, 150, in order to maintain the throughput value below 15%.

It is also of interest to view the average packet delay when the size of the network is given. Equations (4.27), (4.29) and (4.30) are used again for two network sizes: 50 VSATs (Figure 4.35) and 100 VSATs (Figure 4.36). In both figures the total network capacity is the same as in Figure 4.34, i.e. $R = 100\ \text{kbs}^{-1}$.

4.6.6.3 Comparison of other protocols

Figure 4.37 presents a comparison of the delay performance of selective reject ALOHA with that of unslotted ALOHA, S-ALOHA and DAMA/TDMA [RAY88]. The latter corresponds to the variant of the reservation/random TDMA protocol where the requests for reservation are transmitted within a specific time slot divided into smaller slots, to be used on a contention basis using S-ALOHA.

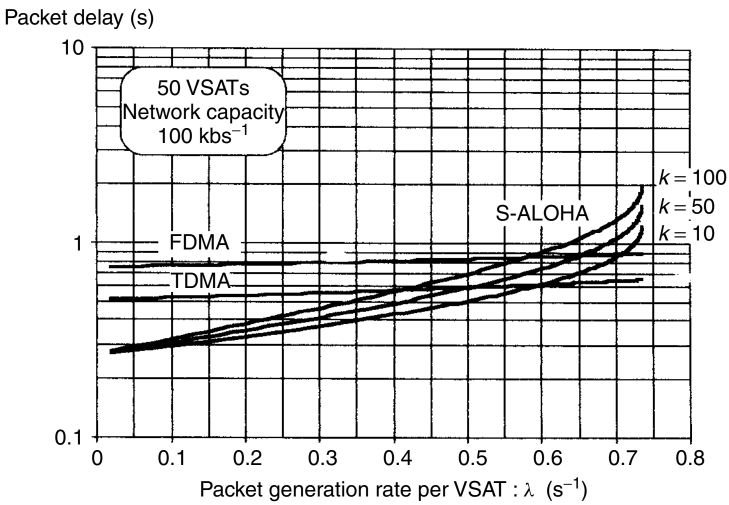


Figure 4.35 Average packet delay for a constant length packet traffic as a function of the packet generation rate λ per VSAT. Packet length is $L = 1000$ bits. The network comprises 50 VSATs. The network capacity is $R = 100\ \text{kbs}^{-1}$ With S-ALOHA, k represents the maximum retransmission interval in time slots

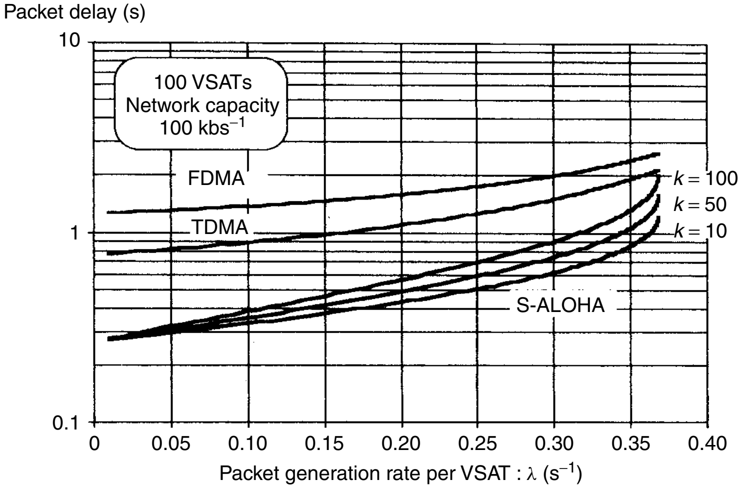


Figure 4.36 Average packet delay for a constant length packet traffic as a function of the packet generation rate λ per VSAT. Packet length is $L = 1000$ bits. The network comprises 100 VSATs. The network capacity is $R = 100\ \text{kbs}^{-1}$. With S-ALOHA, k represents the maximum retransmission interval in time slots

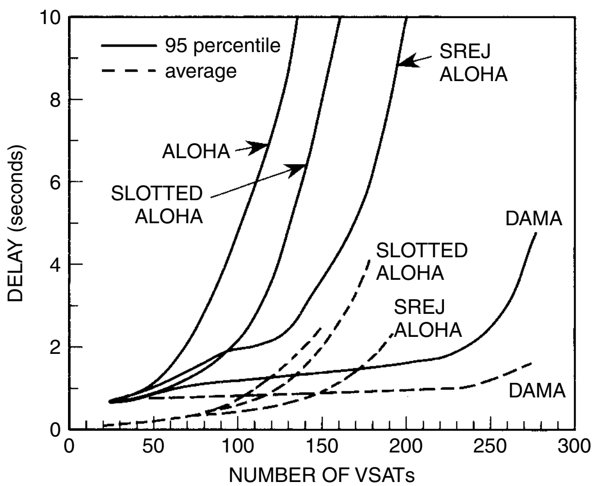


Figure 4.37 Average and peak delay versus number of VSATs for candidate multiaccess protocols. (Reproduced from (RAY88) by permission of the Institute of Electrical and Electronics Engineers, Inc, ©1988 IEEE)

The results presented in Figure 4.37 have been obtained from computer simulations, considering a channel data rate of 56 kbs^{-1}, a message generation rate per VSAT of 0.07 s^{-1}, a message length distribution adjusted to a truncated exponential law, with a maximum length of 256 bytes (2048 bits). Selective reject ALOHA outperforms

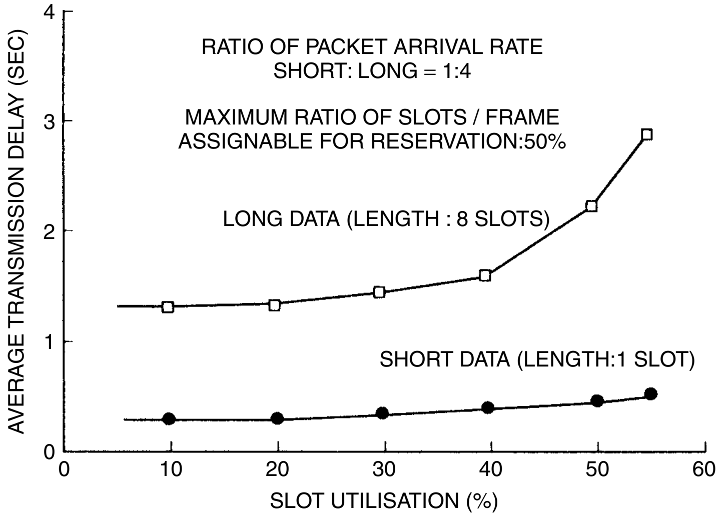


Figure 4.38 Delay versus throughput for mixed mode transmission. (Reproduced from (FUJ86) by permission of the Institute of Electrical and Electronics Engineers, Inc, © 1986 IEEE)

all its competitors at low throughput (small size networks). As throughput increases (larger size networks), reservation/random TDMA displays a lower delay.

Figure 4.38 presents results obtained for the other variant of reservation/random TDMA protocol, where a reservation message is piggy backed on the first packet of the message [FUJ86]. A mixture of random access traffic (short data) and reservation traffic (long data) are transmitted. In the case shown here, the delay for the short data is minimised in spite of the coexistence of relatively large amounts of long data even where total slot utilisation is as high as 40%.

Figure 4.39 presents results obtained for the random access with notification (RAN) protocol [CHI88]. It can be seen that the RAN protocol has a better average delay than reservation TDMA, up to approximately 64% throughput, at which point the number of random access slots reaches zero and the protocol changes to a pure demand assignment TDMA scheme.

4.6.7 Conclusion

The above discussion provides guidelines for the selection of the multiple access technique appropriate to a given traffic type. It should be kept in mind that the selection of such a technique should

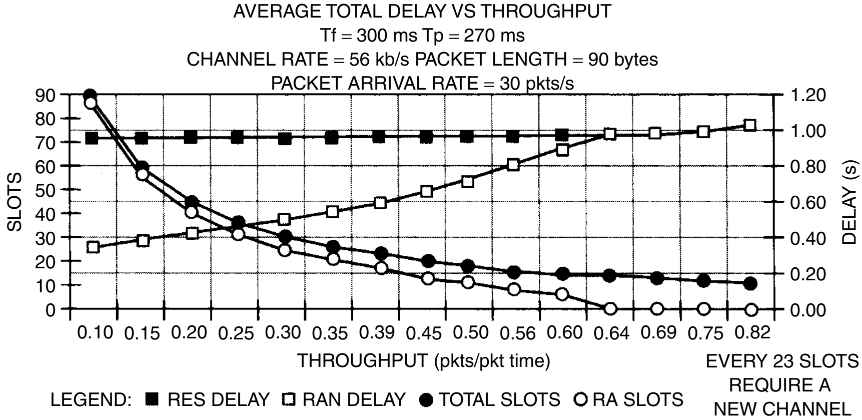


Figure 4.39 Performance of the random access with notification (RAN) protocol. (Reproduced from (CHI88) by permission of the Institute of Electrical and Electronics Engineers, Inc, © 1988 IEEE)

maintain the minimum implementation cost compatible with the required performance.

The combined FDMA–TDMA scheme of Figure 4.22 offers a convenient trade-off between the utilised transponder bandwidth and the power requirement for VSAT transmitters. It also allows a flexible implementation of techniques, from demand assignment to pure random access.

- For stream traffic, demand assigned TDMA offers a high throughput at the expense of an acceptable delay.
- For bursty traffic, a dedicated connection would be poorly used, and random time division multiple access is more efficient. S-ALOHA offers low delay at high burstiness, and can be combined with reservation techniques in order to avoid the inherent limitations of S-ALOHA at low burstiness.

These techniques are used in most present VSAT networks that support a mix of both interactive and batch applications.

CDMA is well suited for low data rate applications, to meet interference restrictions, and for applications that require spectrum spreading for frequency coordination (for instance C-band VSAT networks).

Other protocols, not discussed here because they are too complex for a cost effective implementation in VSAT networks, have been proposed. The extensive overview provided by [RAY87] makes a highly recommendable reference.

4.7 NETWORK DESIGN

4.7.1 Principles

Network design aims to translate the customer's requirements in terms of information transport performance and network availability into network configuration, space segment requirements, back-up hardware elements and channel configurations.

Most of the customer's requirements have been discussed in Chapter 3, section 3.3. It has been mentioned how important it is to start with reliable information on traffic characteristics and to get a clear perspective of the user's performance goals in terms of link availability, response time, etc.

Fortunately, most often VSAT networks replace existing leased line data networks, so that the designer can start with a comprehensive network diagram and user locations. He also needs to know the detailed characteristics of the host processing system: type, communications hardware interface, number of circuits, access protocols and applications software.

Traffic data are more difficult to obtain. Ideally, what is needed is, by type of application and by user location: the number of busy hour transactions, the average size of transaction in and out, the time of day for file transfer and for interactive traffic, etc.

The customer will be asked to specify his performance objectives: average and percentile (for example 95%) delay for interactive applications, and, for some applications, special requirements such as delay jitter. His requirements on availability will determine radio frequency link margins and terrestrial back-up connections. They will also determine back-up hardware, maintenance policy and back-up services such as shared or portable hub, portable VSATs and alternative remote network management systems.

From these inputs, the designer will define the entire hardware configuration at the hub and at remote VSATs with respect to the number of ports, bit rate, protocols and other parameters. The design should make due allowance for possible expansion during the years following the initial installation and operation.

The satellite space segment will be defined in detail. The following information should be included: type of carriers and transmitted bit rate per carrier, number of inbound carriers per outbound (TDM) carrier, total number of inbound and outbound carriers, multiple access control with respect to contention, reservation or a combination thereof.

The network design is then used to determine the cost of the network, with possible feedback to the customer's requirements.

4.7.2 Guidelines for preliminary dimensioning

The design of a network can be supported by guidelines as an outcome from simulations performed on a prevalent network architecture [ZEI91a]. An example is now presented which deals with the typical star network shown in Figure 4.40, for interactive enquiry–response applications (see Table 1.2).

Every VSAT has a number of cluster control units (CCUs) connected to it via local or intermediate range terrestrial links (access lines). The bit rate on these access lines is referred to as R_{cv}. Every CCU, in turn, has a number of computer terminals (or display terminals) attached to it. The host computer is connected to the hub station via a front end processor (FEP).

Due to the low volume and bursty nature of VSAT inquiry traffic, the inbound transponder access protocol is slotted ALOHA. The outbound link operates in a TDM (time division multiplex) mode. The bit rates of the inbound and outbound links are referred to as R_{vh} and R_{hv} respectively. The hub station is connected to the FEP through a high capacity terrestrial link. The bit rate of this link is referred to as R_{hf}.

Polling emulation, as described in section 4.5.3, is used in the two terrestrial links. The data link control over the satellite link is kept simple by assuming that every enquiry message represents an independent packet over the inbound link. Error-free links are assumed, and retransmissions take place due to collisions in the

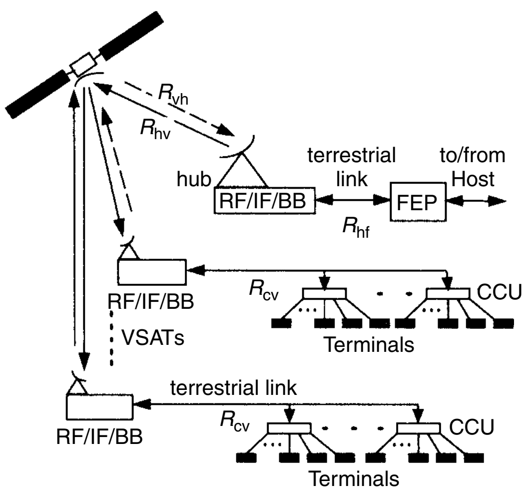


Figure 4.40 Network configuration. RF: radio frequency; FEP: front end processor; IF: intermediate frequency; CCU: cluster control unit; BB: baseband; R_{cv}: bit rate in CCU–VSAT link; R_{vh}: bit rate in VSAT–hub link; R_{hf}: bit rate in hub–FEP link; R_{hv}: bit rate in hub-VSAT link

random access. When collisions occur, the packets are scheduled to be retransmitted after a random delay chosen from a uniform distribution. Positive acknowledgement of successful enquiry messages is required since VSATs cannot reliably receive their own transmissions. This implies a minimum time-out delay of approximately 0.5 to 0.6 s before a retransmission can be initiated. A flow control process is also required in the inbound channel to prevent congestion or deadlock situations. This is achieved by a sliding window technique with a window size of seven messages.

The enquiry message length is 20 bytes and the response message length is 256 bytes. However, since the messages are transacted by a data link control layer, there is additional overhead associated with transmission of every message. Further, one considers a single response for every inputted enquiry. The generation of enquiry messages at the CCUs is approximated by Poisson processes.

The network dimensioning refers to the process of determining the minimum bit rate required for R_{cv}, R_{vh}, R_{hf} and R_{hv}, given the network configuration and the average response time required. The network configuration includes the number of VSATs (N), the number of CCUs per VSAT (M), and the average message generation rate at every CCU (λ msgh^{-1}).

The CCU–VSAT and the hub–FEP links can each be viewed as a bus conveying enquiry and response messages altogether. The length of the enquiry message is 26 bytes (20 data bytes and 6 bytes for addressing) and the length of the response message is 256 bytes. The offered traffic on these links is:

$$\text{CCU–VSAT link:} \quad T_{cv}(\text{bs}^{-1}) = M\lambda\frac{(26+256)\times 8}{3600} \tag{4.32}$$

$$\text{hub–FEP link:} \quad T_{hf}(\text{bs}^{-1}) = M\lambda N\frac{(26+256)\times 8}{3600} \tag{4.33}$$

The VSAT–hub enbound link conveys enquiry messages only, with a length equal to 32 bytes (26 as delivered by the CCU and 6 bytes for addressing). The hub–VSAT outbound link conveys response messages only, with a length equal to 262 bytes (256 data bytes and 6 bytes for addressing). The offered traffic on each of these links is:

$$\text{VSAT–hub link:} \quad T_{vh}(\text{bs}^{-1}) = M\lambda N\frac{32\times 8}{3600} \tag{4.34}$$

$$\text{hub–VSAT link:} \quad T_{hv}(\text{bs}^{-1}) = M\lambda N\frac{262\times 8}{3600} \tag{4.35}$$

The bit rate of every link will now be normalised to the offered traffic of this link. The network is first dimensioned for a given traffic configuration. The resulting dimensioning parameters are then used for other traffic configurations. This allows the determination of the limits of validity of the parameters and derivation of dimensioning guidelines.

It is considered that an acceptable response time should be of the order of 2 to 3 seconds. In the following, the dimensioning procedure will be demonstrated considering a specified mean response time of 2.5 seconds. Of course, the procedure is valid for other values of the mean response time.

Figure 4.41 displays the effect of the bit rate of any link on the mean delay. These results are obtained by simulation.

Curve 1 illustrates the influence of the CCU–VSAT link bit rate, R_{cv}, on the network response time. All other links are overdimensioned to reduce their influence. This curve shows that the CCU–VSAT link bit rate has a strong influence on the overall response time. Note that for $R_{cv} = 4800 \text{ bits}^{-1} \approx 2T_{cv}$ (normalised bit rate = 2), the overall average delay is less than 2 seconds, and it is needless to exceed that normalised bit rate.

Curve 2 illustrates the effect of the VSAT–hub link bit rate, R_{vh}, on the network response time, setting $R_{cv} = 4800 \text{ bits}^{-1} \approx 2T_{cv}$, and for

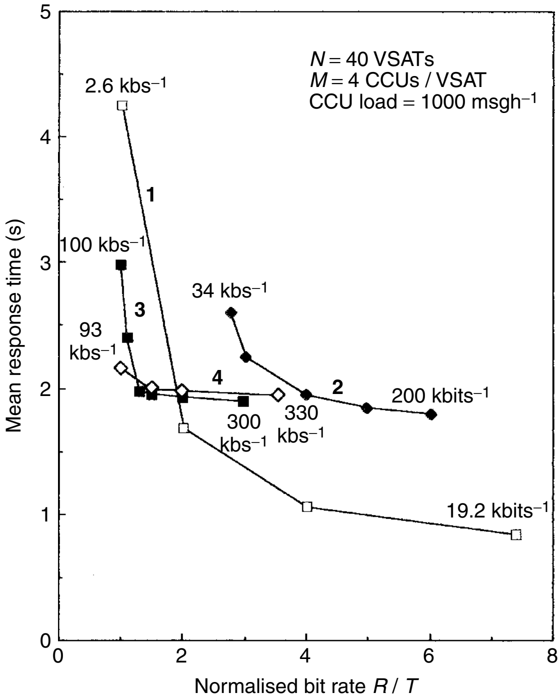


Figure 4.41 Mean response time versus bit rate of network links normalised to traffic volume in the link

overdimensioned hub–FEP and hub–VSAT links. It can be noted that when the inbound link bit rate exceeds 4 times the offered traffic in this link ($R_{\mathrm{vh}} = 46\ \mathrm{kbs}^{-1} \approx 4T_{\mathrm{vh}}$), the effect of the bit rate of this channel on the global average network response time becomes negligible. One of the reasons for this good behaviour of the inbound link operated in slotted–ALOHA is that small fixed length messages are considered for the enquiry traffic. If variable length and/or long enquiry messages were considered, the inbound link would have more effect on the global average network response time.

Curve 3 illustrates the effect of the hub–FEP link bit rate, R_{hf}, on the network response time. This curve is obtained for $R_{\mathrm{cv}} = 4800\ \mathrm{bs}^{-1} \approx 2T_{\mathrm{cv}}$ and for $R_{\mathrm{vh}} = 46\ \mathrm{kbs}^{-1} \approx 4T_{\mathrm{vh}}$. The hub–VSAT link is kept overdimensioned. The curve shows that when the capacity of the hub–FEP link exceeds 1.3 times the offered traffic in this link, the link has no more influence on the delay. From the selection of the above values, this link can be dimensioned with $R_{\mathrm{hf}} = 130\ \mathrm{kbs}^{-1} \approx 1.3T_{\mathrm{hf}}$.

Curve 4 illustrates the effect of the hub–VSAT link bit rate, R_{hv}, on the network response time. This curve is obtained for a hub–FEP link bit rate of $130\ \mathrm{kbs}^{-1} \approx 1.3T_{\mathrm{hf}}$. The bit rates of the CCU–VSAT link and the VSAT–hub link are the same as for curve 3. The curve indicates that the hub–VSAT link displays a very small effect on the network response time. Hence, the bit rate of the outbound link can be dimensioned to be equal to the offered traffic $T_{\mathrm{hv}}(R_{\mathrm{hv}} = T_{\mathrm{hv}} = 93\ \mathrm{kbs}^{-1})$.

(1) Effect of $R_{\mathrm{cv}}(R_{\mathrm{vh}} = 200, R_{\mathrm{hf}} = 300, R_{\mathrm{hv}} = 300\ \mathrm{kbs}^{-1})$;
(2) Effect of $R_{\mathrm{vh}}(R_{\mathrm{cv}} = 4.8, R_{\mathrm{hf}} = 300, R_{\mathrm{hv}} = 300\ \mathrm{kbs}^{-1})$;
(3) Effect of $R_{\mathrm{hf}}(R_{\mathrm{cv}} = 4.8, R_{\mathrm{vh}} = 46, R_{\mathrm{hv}} = 300\ \mathrm{kbs}^{-1})$;
(4) Effect of $R_{\mathrm{hv}}(R_{\mathrm{cv}} = 4.8, R_{\mathrm{vh}} = 46, R_{\mathrm{hf}} = 130\ \mathrm{kbs}^{-1})$

In conclusion, the normalised bit rates required in the four links are:

2 in the CCU–VSAT link,

4 in the VSAT–hub link,

1.3 in the hub–FEP link,

1 in the hub–VSAT link.

The resulting mean response time with this dimensioning is 2.16 s, which fulfils the 2.5 s objective. If a smaller delay is desired, only the bit rate of the CCU–VSAT link needs to be increased, since increasing the bit rates of the other links does not really improve the overall delay.

The above dimensioning procedure is only intended to provide good approximations of the required bit rates. Giving the exact values is of little interest as, in practice, one has to use standardised bit rates (4.8, 9.6, 19.2, 56, 64 kbs^{-1} etc). For example, if the above dimensioning procedure leads to a required bit rate in the CCU–VSAT link of 6 kbs^{-1}, the actual value to be implemented is the next higher standard bit rate which is 9.6 kbs^{-1}. Therefore, the actual response time should be smaller than the response time obtained by simulation.

The sensitivity of this dimensioning has been explored and is reported in the reference article [ZEI91a]. Different CCU loads and numbers of VSATs have been considered. It has been shown that, to avoid too large a delay for a practical range of parameter variations, one is led to consider a minimum network dimension, that is, $R_{cv} =$ 4800 $bits^{-1}$, $R_{vh} = 19.2$ kbs^{-1}, $R_{hf} = 56$ kbs^{-1}, and $R_{hv} = 56$ kbs^{-1}.

In conclusion, given any network configuration (M, N, λ) and a delay requirement of 2.5 seconds, the network is dimensioned as follows:

1. Compute the offered traffic in each link $T_{cv}, T_{vh}, T_{hf}, T_{hv}$ according to formulas (4.32) to (4.35).
2. Compute the preliminary bit rate required in each link: $R_{cv} = 2T_{cv}$, $R_{vh} = 4T_{vh}$, $R_{hf} = 1.3T_{hf}$, and $R_{hv} = T_{hv}$.
3. Select, for each bit rate, the next higher standard bit rate.
4. If any of the resulting bit rates is smaller than the minimum bit rate (i.e. $R_{cv} < 4800$ bs^{-1}, $R_{vh} < 19.2$ kbs^{-1}, $R_{hf} < 56$ kbs^{-1}, or $R_{hv} < 56$ kbs^{-1}), select the minimum bit rate.

4.7.3 Example

This procedure will now be applied to an example.

Consider a network characterised by $N = 50$ VSATs, $M = 6$ CCU/VSAT, and CCU load $\lambda = 500$ $msgh^{-1}CCU^{-1}$. This traffic configuration results in the following offered traffic in the four network links: $T_{cv} = 1880$ bs^{-1}, $T_{vh} = 10.7$ kbs^{-1}, $T_{hf} = 94$ kbs^{-1}, and $T_{hv} = 87.4$ kbs^{-1}. Using the dimensioning guidelines of the last section, we have:

1. The bit rate of the CCU–VSAT link R_{cv} should be twice T_{cv} ($R_{cv} =$ 3760 bs^{-1}). However, this value is smaller than the minimum bit rate considered for this link which is 4800 bs^{-1}. Hence, select for this link in the network a bit rate of 4800 bs^{-1}.

2. The bit rate of the inbound channel R_{vh} should be at least four times the offered traffic T_{vh} in this link. The required bit rate is then equal to 42.8 kbs^{-1}. Hence, select a bit rate of 56 kbs^{-1} which is the next higher standard bit rate.
3. The bit rate of the hub–FEP link should be 1.3 times T_{hf}. This gives $R_{hf} = 122.2$ kbs^{-1}. The next higher standard bit rate is 128 kbs^{-1}.
4. In the hub–VSAT link, the required bit rate is equal to the offered traffic T_{hv} (87.4 kbs^{-1}). Use the next higher standard bit rate, i.e. 128 kbs^{-1}.

Running the simulation with these values gives an average response time of 1.9 s, which fulfils the requirement of an average response time less than 2.5 s.

4.8 CONCLUSION

This chapter has described the organisation of a VSAT network, both its physical configuration and its protocol configuration.

It has been shown that protocols that perform well on terrestrial networks may cause poor channel efficiency when used on satellite links, as a consequence of higher delay. Therefore, it is important to perform protocol conversion at the interface between the hub station and the host computer or the remote terminals and the VSATs.

A critical aspect for achieving good throughput is the error recovery technique which must be compatible with the satellite channel delay and error characteristics. Error recovery may be carried out at the link level where the VSATs and the hub station assume the retransmission. An alternative is to recover errors at the transport level where retransmission of packets in error is performed by the remote terminals or the host computer.

Implementation of link level error recovery requires large transmitting and receiving buffers at VSATs and a hub that can handle the number of packets equivalent to at least one round trip time. Doing it at the transport level avoids such buffers but introduces additional packet delay, as the time-out interval is longer.

It has been shown that the inherent delay (0.25 s for one hop) and the typical bit error rate encountered on satellite links (10^{-6}–10^{-7} for 99.9% of time) are no severe impediment to providing the user terminal with an acceptable service quality.

One must be aware of how important it is to select a multiple access protocol according to the user application. For stream traffic, demand assignment FDMA or TDMA is a good choice. FDMA systems, especially SCPC systems, allow operation at a low transmission rate, hence they have low transmission power requirements. This favours low VSAT equipment cost. Unfortunately, multicarrier operation of the satellite transponder leads to unavoidable generation of intermodulation products, as will be discussed in the next chapter, unless operating the transponder at reduced power. But this may offset the advantage gained from FDMA. TDMA systems are known to exhibit the best efficiency and the highest flexibility, but at the expense of a high transmission rate. The best choice lies in between, combining FDMA and TDMA: for instance, VSATs can be organised in groups operating at different frequencies, with VSATs in a group sharing a given frequency band and accessing the satellite transponder in TDMA (Figure 4.22).

For bursty traffic, it has been shown that a combination of random TDMA and reservation techniques allows one to cope with a high range of burstiness. For instance, a combined scheme of S-ALOHA for transport of small data packets (high burstiness) and slot reservation for transport of long messages (low burstiness) gives today's VSAT networks the required flexibility for supporting a mix of both interactive and batch applications to the best user satisfaction.

Today, CDMA has not yet found widespread distribution in the non-military world, apart from some low data rate applications, to meet interference restrictions, and for applications that require spectrum spreading for frequency coordination (for instance C-band VSAT networks).

5

Radio frequency link analysis

The previous chapter addressed the networking techniques which allow connection set-up and reliable transfer of information from one user terminal to another. This chapter will address information transfer at the *physical level*.

Figure 5.1 reproduces an excerpt from Figure 4.6 in order to put into perspective the respective topics of Chapter 4 and the present one. Chapter 4 dealt with the peer layers of the hub and VSAT interface within the VSAT network at data link control and satellite channel access control levels. The present chapter focuses on the physical layer, which involves coding for forward error correction (FEC) and modulation.

Indeed, the satellite channel conveys information by means of modulated radio frequency carriers which are relayed by the satellite transponder and then received by the destination station. Noise contaminates the received carriers. Therefore, the retrieved baseband signals are also contaminated: analogue signals are noisy and data may contain erroneous bits.

Basically, it is not feasible to provide error-free transmission at the physical layer level. The only hope is to limit the bit error rate (BER) to an acceptable level constrained by cost considerations. It is the job of the upper layers, and especially the data link layer, to ensure error-free transmission by means of automatic repeat request protocols. The job is easier when the physical layer already provides 'clean' information, thanks to a low enough bit error rate. As the BER decreases, the performance of the channel improves, as illustrated in Figure 4.8.

VSAT Networks, 2nd Edition. G. Maral
 ISBN: 0-470-86684-5

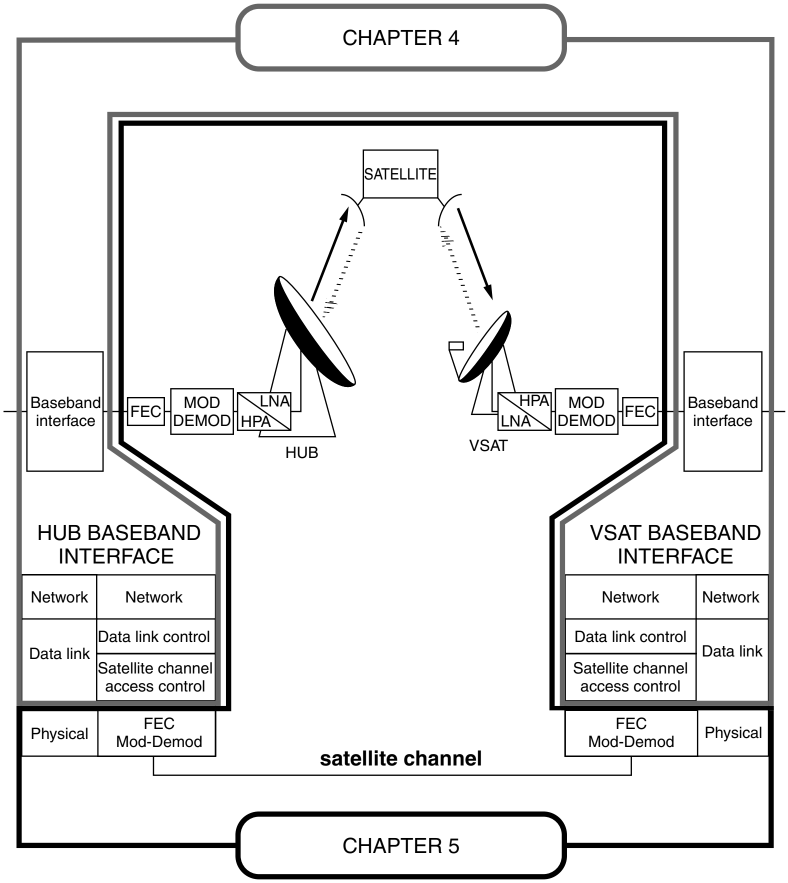


Figure 5.1 Topics covered in Chapters 4 and 5, respectively. FEC: forward error correction; MOD: modulation; DEMOD: demodulation; HPA: high power amplification, LNA: low noise amplification

This chapter aims to provide the means to calculate the quality of the information content delivered to the data link control layer. The quality of digital information is measured by the bit error rate, which is the ratio of the number of bits received in error to the total number of received bits. The bit error rate depends on the type of modulation and coding performed, and on the carrier to noise power spectral density ratio at the input of the receiver. This ratio, C/N_0, can be considered as a quality measure of the radio frequency link.

5.1 PRINCIPLES

The link analysis will be performed in the context of a star shaped network, as illustrated in Figure 5.2. The transmitting VSATs located

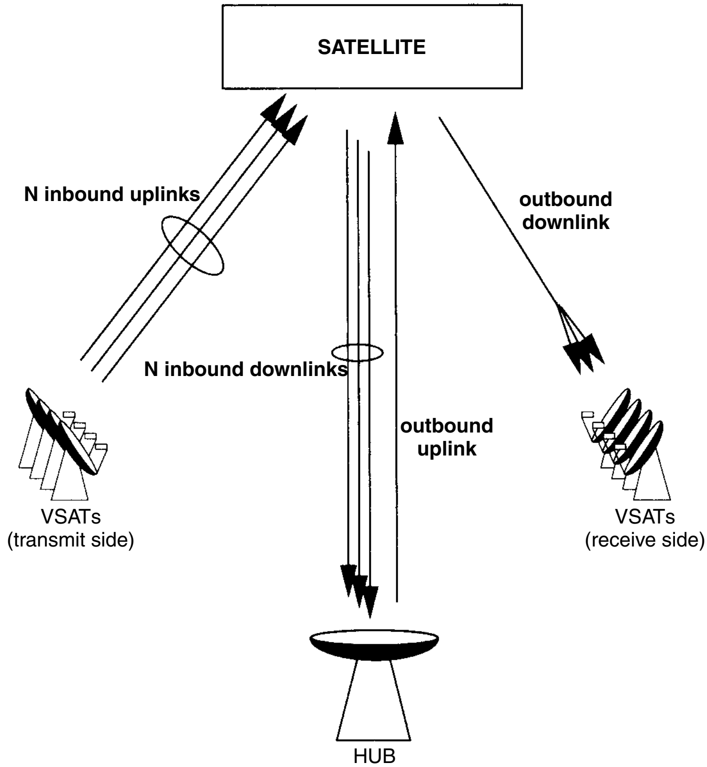


Figure 5.2 Network configuration

within the coverage of the receiving antenna of the satellite generate N inbound carriers. These carriers are relayed by the satellite transponder to the hub station. The hub station communicates with the VSATs by means of a single outbound carrier which is modulated by a time division multiplexed (TDM) stream of bits received by all VSATs within the coverage of the transmitting antenna of the satellite, thanks to the broadcasting capability of the satellite within its coverage area.

A carrier originating from a transmitting station and received by the satellite transponder at the uplink frequency, is amplified by the satellite transponder and frequency translated before being transmitted and received by the earth stations tuned to the downlink frequency. This carrier is corrupted by noise with different origins as discussed below.

5.1.1 Thermal noise

Thermal noise is present on the uplink and the downlink and is produced by natural sources. First, we have the radiation produced by radiating bodies and captured by the receiving antennas. With satellite communications, the principal sources of radiation are the earth for the satellite antenna, and the sky for the earth station antenna. Such noise is called 'antenna noise'. Another source of thermal noise is the noise generated by the receiver components.

5.1.2 Interference noise

Some interference is to be expected from systems sharing the same frequency bands, either satellite-based systems or terrestrial ones. Interference is introduced on the uplink, where the receiving satellite antenna is illuminated by carriers transmitted by earth stations belonging to an adjacent geostationary satellite system, or by terrestrial microwave relays. Interference is also introduced on the downlink where the receiving earth station antenna captures carriers transmitted by adjacent satellites or terrestrial microwave relays. This interference acts as noise if the undesired carrier spectrum overlaps with that of the wanted carrier. The problem may be of special importance on the downlink as the small size of the VSAT station and its resulting large beamwidth makes it more sensitive to reception of off-axis carriers. This is why it is preferable that VSAT networks operate within exclusive bands (see section 1.6.4).

The above interference is generated by transmitters others than those operating within the considered VSAT network. Interference is also generated within the considered VSAT network. This is sometimes called 'self-interference'. For example, some VSAT networks incorporate earth stations operating on two orthogonal polarisations at the same frequency. It may also be that the satellite is a multibeam satellite, with stations transmitting on the same polarisation and frequency but in different beams. These techniques are referred to as 'frequency reuse' techniques and are used to increase the capacity of satellite systems without consuming more bandwidth [MAR02 pp 334–336]. However, the drawback is an increased level of interference due to imperfect cross polarisation isolation of antennas in the case of frequency reuse by orthogonal polarisation, and imperfect beam to beam isolation in the case of spatial frequency reuse.

5.1.3 Intermodulation noise

From Figure 5.2, one can see that the satellite transponder supports several carriers, either N if the inbound carriers and the outbound ones are fed to separate transponders, or $N+1$ if they share the same transponder, which is usually the case. With an access scheme like TDMA, where carriers are transmitted sequentially within a frame period (see Figure 4.20), only one of these carriers is amplified at a given instant by the transponder. However, with access schemes such as FDMA (see Figures 4.16 to 4.19) or CDMA (see Figure 4.23), where carriers are continuously transmitted by the earth stations, the transponder amplifies several carriers simultaneously in a so-called 'multicarrier mode'. It is also the case when a hybrid access mode such as FDMA/TDMA is implemented (see Figure 4.22). The difference resides in the number of simultaneous carriers. This has two consequences: first, the output power of the satellite transponder is shared between the simultaneous carriers and this reduces by as many the power available to each carrier, secondly the presence of simultaneous carriers in the non-linear amplifying device of the transponder causes the generation of intermodulation products in the form of signals at frequencies f_{IM}, which are linear combinations of the P input frequencies [MAR02, p 286]. Thus:

$$f_{\mathrm{IM}} = m_1 f_1 + m_2 f_2 + \cdots + m_{\mathrm{P}} f_{\mathrm{P}} \quad \text{(Hz)} \tag{5.1}$$

where $m_1, m_2, \ldots, m_{\mathrm{P}}$ are positive or negative integers. The quantity X, called the order of an intermodulation product is defined as:

$$X = |m_1| + |m_2| + \cdots + |m_{\mathrm{P}}| \tag{5.2}$$

As the centre frequency of the pass-band transponder is large compared with its bandwidth, only those odd-order intermodulation products with $\sum m_{\mathrm{i}} = 1$ fall within the channel bandwidth. Moreover, the power of the intermodulation products decreases with the order of the product. Thus, in practice, only third-order products and, to a lesser extent, fifth order products are significant. Intermodulation products are transmitted on the downlink along with the wanted carrier, but no useful information can be extracted from them. They act as noise, as a fraction of the overall intermodulation product power falls into the bandwidth of the earth station receiver tuned to the wanted carrier. They can be modelled as white noise, with constant power spectral density $N_{0\mathrm{IM}}$ given by:

$$N_{0\mathrm{IM}} = \frac{N_{\mathrm{IM}}}{B_N} \quad (\mathrm{WHz}^{-1}) \tag{5.3}$$

where N_{IM} is the intermodulation power measured at the transponder output within the noise bandwidth B_{N} of the earth station receiver.

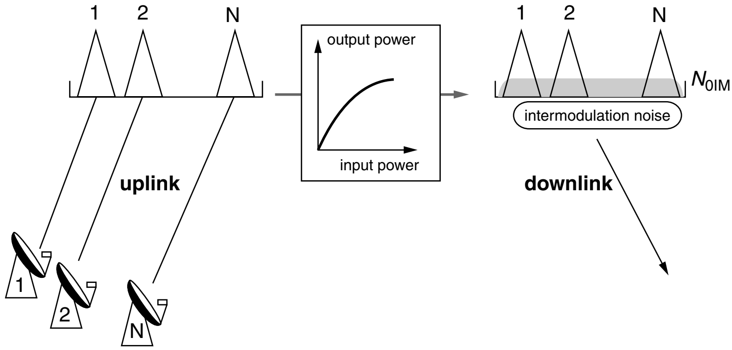


Figure 5.3 Multicarrier operation with FDMA access by *N* carriers

Figure 5.3 illustrates the above discussion and shows how intermodulation products can be accounted for in the form of an equivalent white noise with power spectral density equal to $(N_0)_{IM}$.

5.1.4 Carrier power to noise power spectral density ratio

The noise contributions are to be considered in relationship to the wanted carrier they corrupt. Therefore, one should specify carrier power to noise power spectral density ratios at the point in the link where noise corrupts the carrier.

Figure 5.4 indicates at which point in the link a given quantity is relevant, and Table 5.1 specifies notations and definitions for the corresponding carrier power to noise power spectral density ratio. In both Figure 5.4 and Table 5.1 G_{Xpond} is the power gain of the transponder for carrier power C_U at the transponder input.

5.1.5 Total noise

At the receiver input of the receiving station in Figure 5.4, the demodulated carrier power is C_D, and the power spectral density

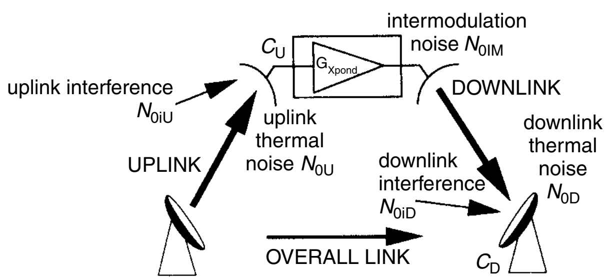


Figure 5.4 Point in the link where quantities used in Table 5.1 are relevant

Table 5.1 Carrier power to noise power spectral density according to the considered noise contribution(see Figure 5.4 for point in the link where notation and definition applies)

Origin of noise	Notation for noise power spectral density (WHz^{-1})	Notation and definition for carrier power to noise power spectral density ratio
Uplink thermal noise	N_{0U}	$(C/N_0)_U = C_U/N_{0U}$
Downlink thermal noise	N_{0D}	$(C/N_0)_D = C_D/N_{0D}$
Uplink interference	N_{0iU}	$(C/N_{0i})_U = C_U/N_{0iU}$
Downlink interference	N_{0iD}	$(C/N_{0i})_D = C_D/N_{0iD}$
Intermodulation noise	N_{0IM}	$(C/N_0)_{IM} = G_{Xpond}C_U/N_{0IM}$

N_{0T} of the corrupting noise is that of the total noise contribution. This total contribution builds up from:

– the uplink thermal noise and uplink interference noise retransmitted on the downlink by the satellite transponder. On their way from the satellite transponder input to the earth station receiver input, they are subject to power gains and losses which amount to a total gain G_{TE}. This power gain is the product of the transponder power gain G_{Xpond} and the gain G_D from transponder output to earth station receiver input (which in practice is much less than one in absolute value, so it should actually be considered as a loss):

$$G_{TE} = G_{Xpond} \times G_D \tag{5.4}$$

Therefore, the respective contributions of uplink thermal noise and uplink interference noise at the earth station receiver input are $G_{TE}N_{0U}$ and $G_{TE}N_{0iU}$. Note that G_{Xpond} has been defined as the transponder power gain for carrier power C_U at transponder input. The transponder being non-linear, the actual transponder gain depends on the power of the considered input signal. Hence, G_{Xpond} has different values for the noise and for the carrier. This is referred to as the 'capture effect'. However, for simplicity, this will not be considered here;

– the intermodulation noise generated at the transponder output and transmitted on the downlink. Hence, its contribution at the earth station receiver input is G_DN_{0IM};

– the downlink thermal noise and downlink interference with respective contributions at the earth station receiver input N_{0D} and N_{0iD}.

The total noise power spectral density at the earth station receiver input is given by:

$$N_{0T} = \sum N_{0j} \quad (WHz^{-1}) \tag{5.5}$$

where N_{0j} are the above individual contributions at the earth station receiver input.

The carrier power to noise power spectral density ratio at the earth station receiver input C_D/N_{0T} conditions the quality of the baseband signal delivered to the user terminal in terms of BER. This ratio relates to the overall link from station to station and will be noted $(C/N_0)_T$ (T for total). $(C/N_0)_T$ can be calculated as follows:

$$\left(\frac{C}{N_0}\right)_T^{-1} = \left(\frac{N_{0T}}{C_D}\right) = \sum \frac{N_{0j}}{C_D} = \frac{(G_{TE}N_{0U} + G_{TE}N_{0iU})}{C_D} + \frac{G_D N_{0IM}}{C_D} + \frac{N_{0D}}{C_D} + \frac{N_{0iD}}{C_D} \quad (\mathrm{Hz}^{-1}) \tag{5.6}$$

Consider that $C_D = G_{TE}C_U = G_{Xpond} \times G_D \times C_U$, equation (5.6) now becomes:

$$\left(\frac{C}{N_0}\right)_T^{-1} = \frac{N_{0U}}{C_U} + \frac{N_{0iU}}{C_U} + \frac{N_{0IM}}{G_{Xpond}C_U} + \frac{N_{0D}}{C_D} + \frac{N_{0iD}}{C_D} = \left(\frac{C}{N_0}\right)_U^{-1} + \left(\frac{C}{N_0}\right)_D^{-1} + \left(\frac{C}{N_0}\right)_{IM}^{-1} + \left(\frac{C}{N_{0i}}\right)_U^{-1} + \left(\frac{C}{N_{0i}}\right)_D^{-1} \quad (\mathrm{Hz}^{-1}) \tag{5.7}$$

Note that introducing the actual transponder gain depending on signal power at transponder input instead of the same value G_{Xpond} for the noise and for the carrier would introduce a corrective term to the values of $(C/N_0)_U$ and $(C/N_{0i})_U$ in the above equation.

The following sections provide means for the determination of the terms implied in the calculation of $(C/N_0)_T$ according to equation (5.7). Sections 5.2 and 5.3 discuss the parameters involved in the calculation of uplink $(C/N_0)_U$ and downlink $(C/N_0)_D$. Section 5.4 discusses the intermodulation and the parameters involved in the calculation of $(C/N_0)_{IM}$. Section 5.5 is dedicated to interference analysis and means to calculate $(C/N_{0i})_U$ and $(C/N_{0i})_D$. Section 5.6 recapitulates the previous terms in expression (5.7) for the overall link $(C/N_0)_T$. Section 5.7 deals with bit error rate determination. Section 5.8 demonstrates how power and bandwidth can be exchanged through the use of forward error correction. Section 5.9 gives an example of a calculation for VSAT networks.

5.2 UPLINK ANALYSIS

Figure 5.5 illustrates the geometry of the uplink. In order to calculate the value of $(C/N_0)_U$ in the worst case, the transmitting earth station is assumed to be located at the edge of the uplink coverage, defined as the contour where the satellite receiving antenna has a constant gain defined relative to its maximum value at boresight, for instance −3 dB, corresponding to a reduction by a factor two of the gain compared to its maximum. From Table 5.1, the ratio $(C/N_0)_U$ is defined as:

$$\left(\frac{C}{N_0}\right)_U = \frac{C_U}{N_{0U}} \quad \text{(Hz)} \tag{5.8}$$

where C_U is the power of the received carrier at the input to the satellite transponder. N_{0U} is the noise power spectral density and relates to the uplink system noise temperature T_U given by (5.32):

$$N_{0U} = kT_U \quad \text{(W/Hz)} \tag{5.9}$$

where k is the Boltzman constant: $k = 1.38 \times 10^{-23}\ \text{JK}^{-1}$; $k(\text{dBJK})^{-1} = 10 \log k = -228.6\ \text{dBJK}^{-1}$.

$(C/N_0)_U$ can be expressed [MAR02, p251] as:

$$\left(\frac{C}{N_0}\right)_U = \text{IBO}_1 \left(\frac{C}{N_0}\right)_{Usat} \quad \text{(Hz)}$$
$$\left(\frac{C}{N_0}\right)_U \text{(dBHz)} = \text{IBO}_1\text{(dB)} + \left(\frac{C}{N_0}\right)_{Usat} \text{(dBHz)} \tag{5.10}$$

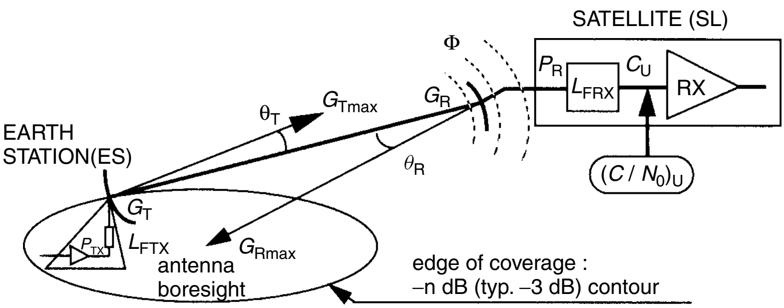


Figure 5.5 Geometry of the uplink P_{TX}: transmitter output power; L_{FTX}: feeder loss from transmitter to antenna; G_T: earth station antenna transmit gain in direction of satellite; θ_T: earth station antenna depointing angle; G_{Tmax}: earth station antenna transmit gain at boresight; Φ: power flux density at satellite antenna; G_R: satellite antenna receive gain at edge of coverage; θ_R: satellite antenna half beamwidth angle; P_R received power at antenna output; L_{FRX}: feeder loss from satellite antenna to receiver input; C_U: carrier power at receiver input; RX: receiver

where IBO_1 is the input back-off per carrier for the considered carrier, as defined in Appendix 6 by expression (A6.5), and $(C/N_0)_{Usat}$ is the value required to saturate the satellite transponder, and is given by:

$$\left(\frac{C}{N_0}\right)_{Usat} = \Phi_{sat}\left(\frac{1}{G_1}\right)\left(\frac{G}{T}\right)_{SL}\left(\frac{1}{k}\right) \quad (Hz) \qquad (5.11)$$

$$\left(\frac{C}{N_0}\right)_{Usat}(dBHz) = \Phi_{sat}(dBWm^{-2}) - G_1(dBi) + \left(\frac{G}{T}\right)_{SL}(dBK^{-1}) - 10\log k\ (dBJK)$$

where Φ_{sat} is the power flux density required to saturate the satellite transponder (see section 5.2.1), $(G/T)_{SL}$ is the figure of merit of the satellite receiving equipment (see section 5.2.4), G_1 is the gain of an ideal antenna with area equal to 1 m^2:

$$G_1 = \frac{4\pi}{\lambda^2} = 4\pi\left(\frac{f}{c}\right)^2 \qquad (5.12)$$

$$G_1(dBi) = 10\log 4\pi - 20\log\lambda = 10\log 4\pi + 20\log\left(\frac{f}{c}\right)$$

λ is the wavelength (m), f is the frequency (Hz), c is the speed of light: $c = 3\times 10^8\ ms^{-1}$, k is the Boltzmann constant: $k = 1.38\times 10^{-23}\ JK^{-1}$; $k(dBJK^{-1}) = 10\log k = -228.6\ dBJK^{-1}$.

5.2.1 Power flux density at satellite distance

The power flux density Φ is defined in Appendix 5. Assume the satellite to be at distance R from a transmitting earth station, with effective isotropic radiated power $EIRP_{ES}$. Then the power flux density at satellite level is:

$$\Phi = \frac{EIRP_{ES}}{4\pi R^2} \quad (Wm^{-2}) \qquad (5.13)$$

$$\Phi\ (dBWm^2) = EIRP_{ES}\ (dBW) - \log 4\pi R^2$$

The power flux density can also be calculated from G_1 given by (5.12) and the uplink path loss L_U, discussed in section 5.2.3:

$$\Phi = \frac{EIRP_{ES}G_1}{L_U} \quad (Wm^{-2}) \qquad (5.14)$$

$$\Phi\ (dBWm^{-2}) = EIRP_{ES}\ (dBW) + G_1(dBi) - L_U\ (dB)$$

Satellite transponder input characteristics are given in terms of saturated power flux density Φ_{sat}, which designates the power flux density required to saturate the transponder. From expression (5.13), it can be seen that Φ is controlled by the transmitting earth station $EIRP_{ES}$. Should the transmitting earth station $EIRP_{ES}$ be allocated to a single carrier, the power of that carrier at the transponder input determines IBO or IBO_1, as defined in Appendix 6 by equations (A6.2) and (A6.5) respectively. For example:

$$IBO_1 = \frac{\Phi_1}{\Phi_{sat}} \tag{5.15}$$

$$IBO_1(dB) = \Phi_1(dBWm^{-2}) - \Phi_{sat}(dBWm^{-2})$$

Should N stations be transmitting simultaneously, the powers of their individual carriers combine at the transponder input, and the total flux density is given by the sum of their individual contributions, each calculated from (5.13) or (5.14):

$$\Phi_t = \sum \Phi_i \quad i = 1, 2, \ldots , N$$

The transponder total input back-off, defined in Appendix 6 by equation (A6.7), is then given by:

$$IBO_t = \frac{\Phi_t}{\Phi_{sat}} \tag{5.16}$$

$$IBO_t(dB) = \Phi_t(dBWm^{-2}) - \Phi_{sat}\ (dBWm^{-2})$$

5.2.2 Effective isotropic radiated power of the earth station

From the definition given in Appendix 5, the effective isotropic radiated power of the earth station $EIRP_{ES}$ is expressed as:

$$EIRP_{ES} = P_T G_T \quad (W) \tag{5.17}$$

$$EIRP_{ES}\ (dBW) = P_T\ (dBW) + G_T\ (dBi)$$

where P_T is the power fed to the transmitting antenna, and G_T is the earth station antenna transmit gain in the pertinent direction.

Figure 5.6 is an enlargement of the transmitting earth station represented in Figure 5.5. The earth station transmitter TX with output power P_{TX} feeds power P_T to the antenna through a feeder with feeder loss L_{FTX}. The antenna displays a transmit gain G_{Tmax} at boresight, and a reduced transmit gain G_T in the direction of the satellite as a result of the transmit depointing off-axis angle θ_T.

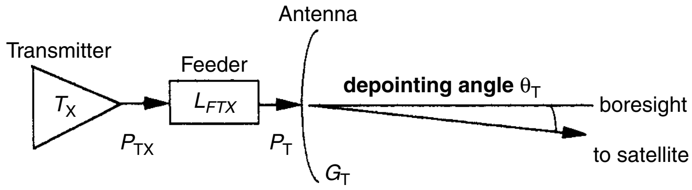


Figure 5.6 Transmitting earth station components

The transmitter output power P_{TX} is smaller than or equal to the transmitter output rated power P_{TXmax}, depending on the transmitter output back-off.

In order to calculate the actual gain G_T, one needs to know more about the antenna gain pattern. Appendix 4 defines the antenna gain pattern and its most important parameters, the maximum gain G_{max} and the half power beamwidth θ_{3dB}, which are expressed as:

$$G_{max} = \eta_a \left(\frac{\pi Df}{c}\right)^2 \tag{5.18}$$

$$G_{max}(\text{dBi}) = 10 \log \eta_a \left(\frac{\pi Df}{c}\right)^2$$

and:

$$\theta_{3dB} = \frac{70c}{fD} \quad \text{(degrees)} \tag{5.19}$$

where:

η_a = antenna efficiency (typically 0.6)

D = antenna diameter (m)

f = frequency (Hz)

c = speed of light = 3×10^8 ms^{-1}

Figures 5.7 and 5.8 display values of these parameters for typical hub and VSAT antenna diameters. For the transmit gain and half power beamwidth values, one should consider 14 GHz and 6 GHz for the frequency values.

Should there be no feeder loss, the power fed to the antenna would be P_{TX}, and should the antenna be perfectly pointed, its transmit gain in the direction of the satellite would be equal to G_{Tmax}. Therefore, $EIRP_{ES}$ would be maximum and equal to:

$$EIRP_{ESmax} = P_{TX}G_{Tmax} \quad \text{(W)} \tag{5.20}$$

$$EIRP_{ESmax}(\text{dBW}) = P_{TX}\ (\text{dBW}) + G_{Tmax}\ (\text{dBi})$$

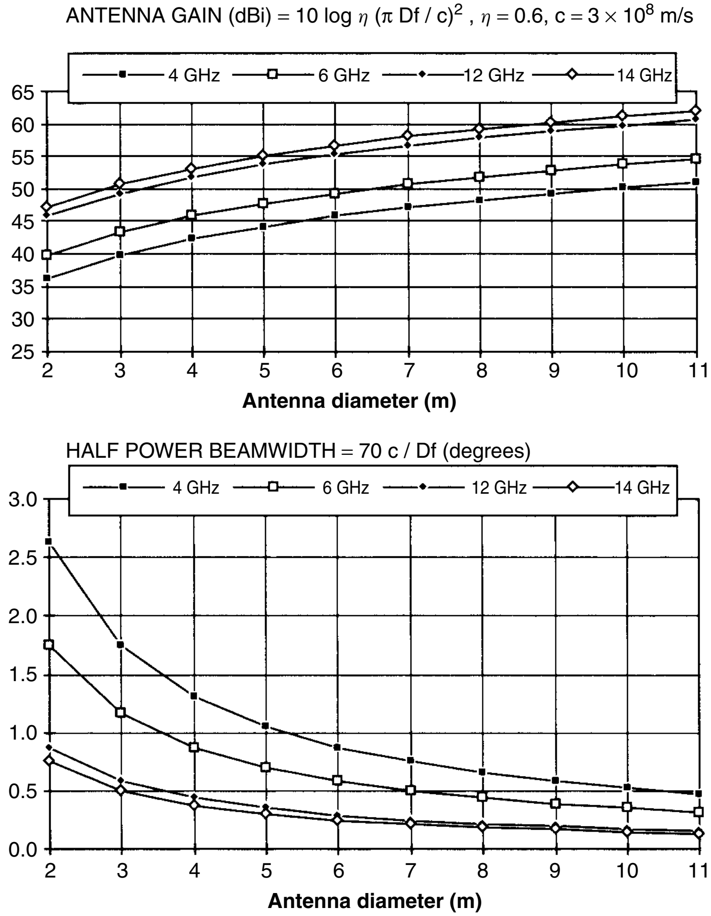


Figure 5.7 Antenna gain and half power beamwidth for typical hub station antenna diameters

Figure 5.9 displays the maximum EIRP values that can be achieved from a given combination of transmitter power and antenna diameter. This is to be considered as an upper bound, as provided by ideal transmitting equipment that is perfectly pointed. The actual EIRP value depends on the magnitude of the losses, which will now be discussed.

The power P_T fed to the antenna can be calculated from the known values of the transmitter output power P_{TX} and the feeder loss L_{FTX}:

$$P_T = \frac{P_{TX}}{L_{FTX}} \quad \text{(W)} \tag{5.21}$$

$$P_T\,(\text{dBW}) = P_{TX}\,(\text{dBW}) - L_{FTX}\,(\text{dB})$$

Typically, L_{FTX} is of the order of 0.1 to 0.3 dB.

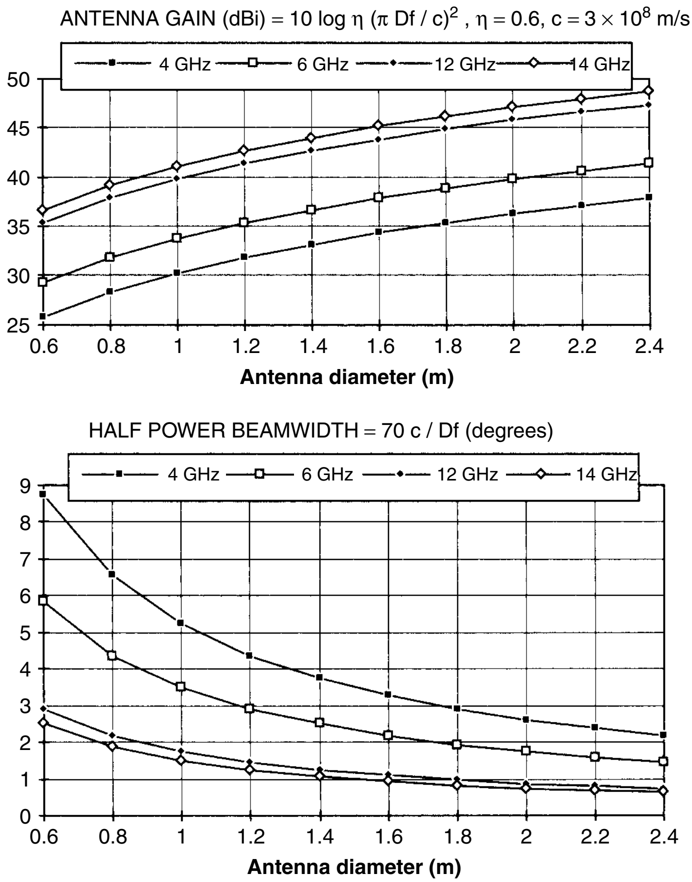


Figure 5.8 Antenna gain and half power beamwidth for typical VSAT antenna diameters

Similarly, G_T can be calculated from the maximum value of the earth station antenna gain G_{Tmax} and the transmission depointing loss L_T:

$$G_T = \frac{G_{Tmax}}{L_T} \tag{5.22}$$

$$G_T\,(\text{dBi}) = G_{Tmax}\,(\text{dBi}) - L_T\,(\text{dB})$$

where:

$$G_{Tmax} = \eta_a \left(\frac{\pi D f}{c}\right)^2 \tag{5.23}$$

$$G_{Tmax}\,(\text{dBi}) = 10 \log \eta_a \left(\frac{\pi D f}{c}\right)^2$$

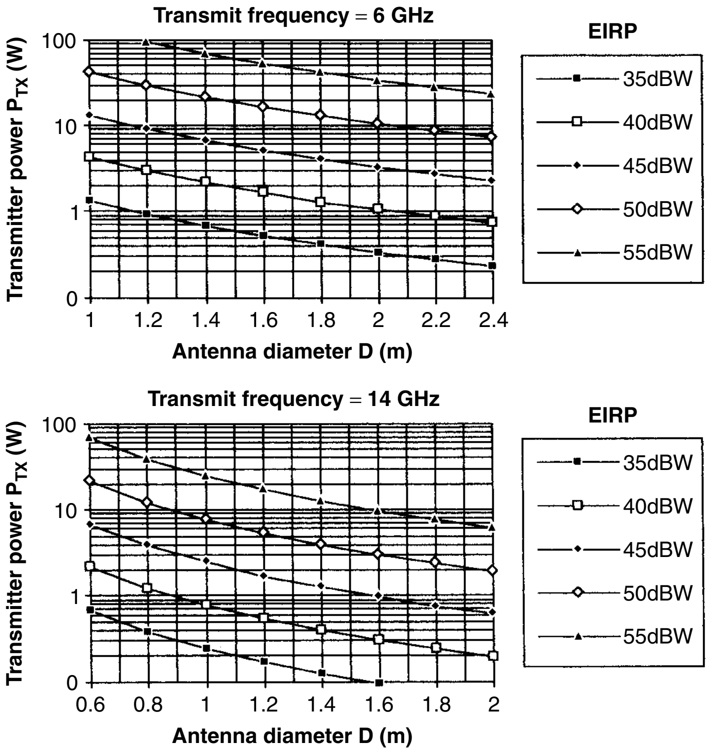


Figure 5.9 Maximum achievable $\text{EIRP}_{\text{ESmax}}$ at 6 GHz and 14 GHz for given antenna size and transmitter power

and

$$\eta_a = \text{antenna efficiency (typically 0.6)}$$
$$D = \text{antenna diameter (m)}$$
$$f = \text{frequency (Hz)}$$
$$c = \text{speed of light} = 3 \times 10^8\ \text{ms}^{-1}$$

The depointing loss L_T can be expressed in dB as a function of the depointing angle θ_T and the transmit half power beamwidth θ_{3dB} of the antenna:

$$L_T\,(\text{dB}) = 12\left(\frac{\theta_T}{\theta_{3dB}}\right)^2 \tag{5.24}$$

where θ_{3dB} is given by (5.19).

The depointing angle θ_T is not easy to determine. Should the earth station be equipped with a tracking antenna, as would a large hub station, say with antenna diameter larger than 5 m at Ku-band and 9 m at C-band, the depointing angle is given by the tracking accuracy of the tracking equipment, and is typically of the order of $0.2\theta_{3dB}$. Therefore, the depointing loss remains smaller than 0.5 dB. But small hub stations and VSATs are equipped with fixed mount antennas. The value of the maximum depointing angle θ_{Tmax} then depends on the pointing accuracy at installation, and the subsequent motion of the geostationary satellite. An upper limit to the depointing angle value can be estimated from an analysis of the pointing procedure and the satellite motion.

Figure 5.10 represents the underlying geometry. The figure displays off-axis angles as viewed from the antenna boresight, represented by the centre of the circle. As indicated in Chapter 3, section 3.2.3, at time of installation the antenna is pointed towards the satellite by a search for a maximum received power from a satellite beacon or a downlink carrier. Once this is done the antenna is left to itself, and one assumes that the mount is robust enough to ensure a constant pointing direction with time. The radius of the circle in Figure 5.10 represents the maximum initial pointing error (IPE). Now the satellite is maintained within a so-called station keeping window (see Chapter 2, section 2.3.8). It is assumed here that the station keeping window dimensions are equal in the north–south (NS) and in the east–west (EW) directions, so that the station keeping window can be represented by the square in Figure 5.10 with half-width SKW. The worst situation occurs when the initial pointing is performed while the satellite is at an extreme position within its station keeping window, say the ES corner, and the actual direction of the antenna boresight coincides with the centre of the initial pointing error circle as positioned on the figure. Subsequently, the satellite may move to the opposite extreme position, say the NW corner. Therefore, the maximum depointing angle

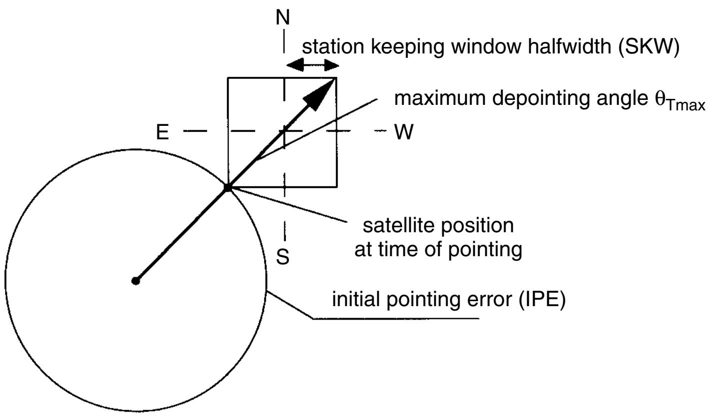


Figure 5.10 Geometry of pointing errors

θ_{Tmax} is the sum of the initial depointing error IPE and the angle corresponding to the extreme positions of the satellite, that is the diagonal of the station keeping window.

$$\theta_{Tmax} = \text{IPE} + 2(2)^{1/2}\ \text{SKW} \quad \text{(degrees)} \tag{5.25}$$

The initial pointing error IPE is typically equal to $0.2\theta_{3dB,}$ where θ_{3dB} represents the half power beamwidth of the antenna at the frequency of the received beacon or carrier used for pointing, i.e. at *downlink* frequency. Denoting this frequency by f_D, and combining (5.19), (5.24) and (5.25) results in:

$$L_{Tmax} = 12\left[0.2\left(\frac{f_U}{f_D}\right) + 2(2)^{1/2}\ \text{SKW}\ \frac{Df_U}{210 \times 10^8}\right]^2 \quad \text{(dB)} \tag{5.26}$$

Figure 5.11 displays the maximum transmit depointing loss L_{Tmax} for a C-band and a Ku-band system, assuming SKW = 0.025°, and Figure 5.12 displays the same curves, assuming SKW = 0.05°.

Comparing the curves in Figures 5.11 and 5.12 provides some insight on the impact on depointing loss of the satellite station keeping window size.

Table 5.2 summarises the above results by providing typical values.

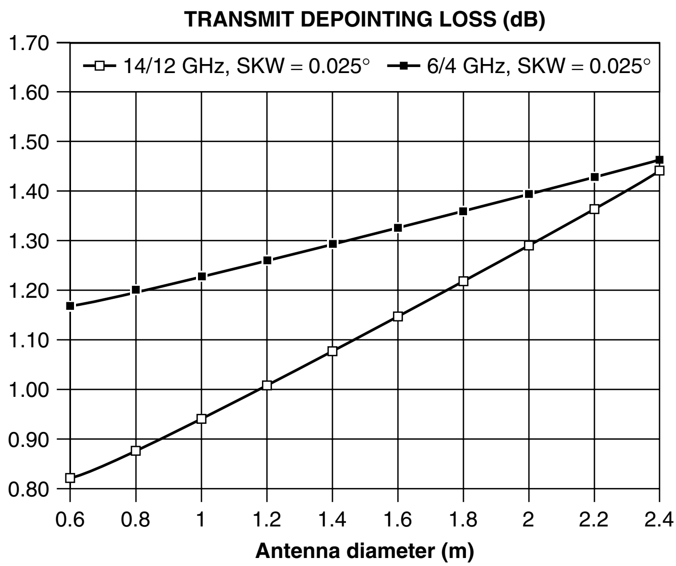


Figure 5.11 Transmit antenna gain depointing loss L_T for a C-band and a Ku-band system, assuming the half width of the station keeping window SKW = 0.025°

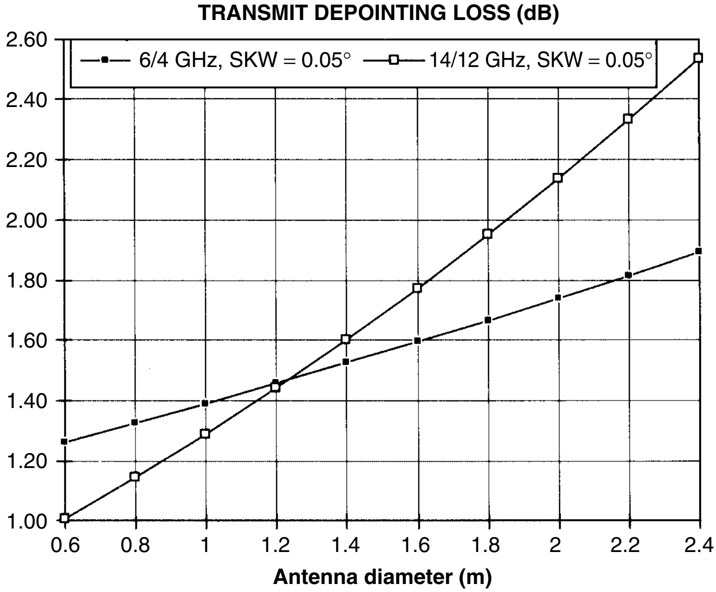


Figure 5.12 Transmit antenna gain depointing loss L_T for a C-band and a Ku-band system, assuming the half width of the station keeping window SKW = 0.05°

Table 5.2 Maximum achievable EIRP, typical magnitude of losses, and actual EIRP for hub and VSATs at Ku-band (14 GHz)

	Large hub	Small hub	VSAT	
Antenna diameter, D	10 m	3 m	1.8 m	1.2 m
Transmitter power, P_{TX}	100 W	10 W	1 W	1 W
Maximum EIRP	81.1 dBW	61.6 dBW	46.2 dBW	42.7 dBW
Feeder loss, L_{FTX}	0.2 ± 0.1 dB	0.2 ± 0.1 dB	0.2 ± 0.1 dB	0.2 ± 0.1 dB
Depointing loss, L_T	0.5 ± 0.1 dB	2.4 ± 0.7 dB	1.6 ± 0.3 dB	1.2 ± 0.2 dB
Actual EIRP	80.4 ± 0.2 dBW	59 ± 0.8 dBW	44.4 ± 0.4 dBW	41.3 ± 0.3 dBW

5.2.3 Uplink path loss

The uplink path loss, L_U, is the overall attenuation of the carrier power on its way from the earth station transmitting antenna to the satellite receiving antenna. It can be shown that this attenuation has two components, the free space loss, L_{FS}, defined in Appendix 5, and the atmospheric loss, L_A, so that the path loss can be expressed as [MAR02, Chapter 5]:

$$L_U = L_{FS}L_A \tag{5.27}$$

$$L_U(\text{dB}) = L_{FS}(\text{dB}) + L_A\ (\text{dB})$$

The free space loss depends on the frequency f and on the distance R between the earth station and the satellite:

$$L_{FS} = \left(\frac{4\pi Rf}{c}\right)^2 = \left(\frac{4\pi R_0 f}{c}\right)^2 \left(\frac{R}{R_0}\right)^2 \tag{5.28}$$

$$L_{FS}(\text{dB}) = 10\log\left(\frac{4\pi R_0 f}{c}\right)^2 + 10\log\left(\frac{R}{R_0}\right)^2$$

where c is the speed of light ($c = 3 \times 10^8\ \text{ms}^{-1}$) and R_0 is the satellite height ($R_0 = 35\,786$ km for a geostationary satellite).

The ratio $(R/R_0)^2$ is a geometric factor which takes into account the position of the earth station relative to the subsatellite point on the earth surface, and is expressed as (see section 2.3.3):

$$\left(\frac{R}{R_0}\right)^2 = 1 + 0.42(1 - \cos l \cos L) \tag{5.29}$$

where l and L are respectively the difference in latitude and in longitude between the earth station and the subsatellite point. Notice that the subsatellite point of a geostationary satellite is on the Equator, and hence its latitude is zero, so l identifies with the earth station latitude, while L should be taken as the actual difference in longitude between that of the earth station and that of the satellite meridian.

For a geostationary satellite, Figure 5.13 gives the variation in dB of the first term of equation (5.28) as a function of frequency, and Figure 5.14 gives the variation in dB of the second term of equation (5.28) as a function of the earth station location. The free space loss L_{FS} (in dB) is calculated by adding the values obtained from these two figures. The second term appears to be a small corrective term to the first one.

The attenuation of radio frequency carriers in the atmosphere, denoted by L_A, is due to the presence of gaseous components in the troposphere, water (rain, clouds, snow and ice) and the ionosphere. Water plays an important role, especially at Ka-band, as it has an absorption line at 22.3 GHz. Gaseous components and water in the form of vapour are constantly present in the atmosphere. Water is occasionally present in the form of rain and as such produces attenuation and cross polarisation of the radio wave, i.e. transfer of part of the energy transmitted in one polarisation to the orthogonal polarisation state.

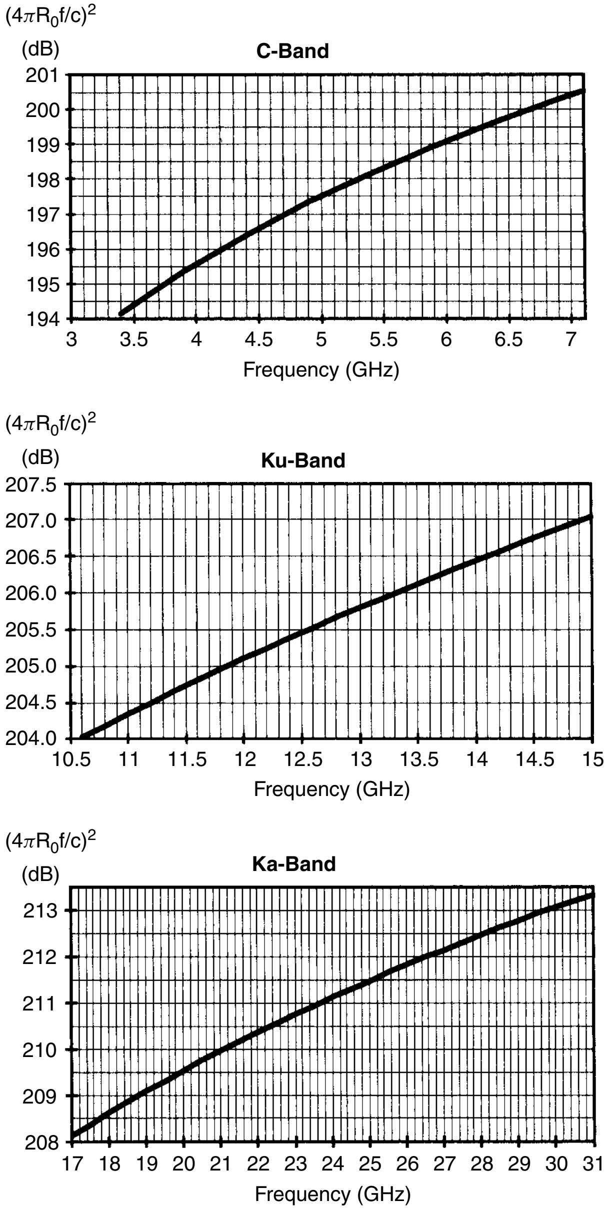


Figure 5.13 Variation in dB of $(4\pi R_0 f/c)^2$ as a function of frequency for a geostationary satellite ($R_0 = 35\,786$ km)

It is convenient to consider power loss L_A as the result of two attenuation terms:

$$L_A = A_{AG} A_{RAIN} \tag{5.30}$$

$$L_A(\text{dB}) = A_{AG}(\text{dB}) + A_{RAIN}(\text{dB})$$

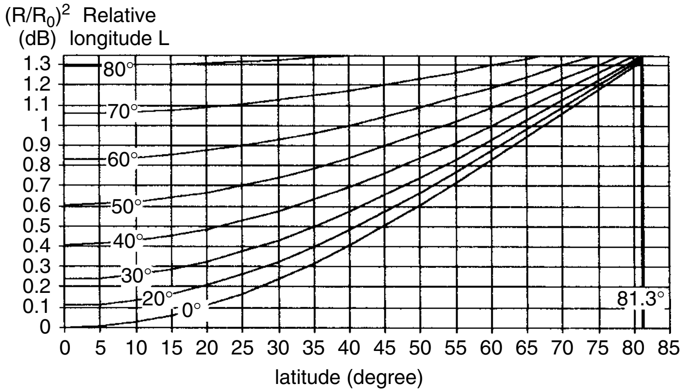


Figure 5.14 Variation in dB of $(R/R_0)^2$ as a function of the position of the earth station with respect to the subsatellite point for a geostationary satellite

where A_{AG} is the always present attenuation due to the atmosphere during 'clear sky' conditions (no rain) and A_{RAIN} is the additional and occasional attenuation due to rain.

The attenuation A_{AG} depends on frequency and elevation angle, and is higher at low elevation angles as a result of the increased path length of the radio wave in the atmosphere. One can consider that the attenuation A_{AG} is, for elevation angles greater than 10°, negligible at C-band, less than 0.5 dB at Ku-band, and less than 1 dB at Ka-band.

The attenuation A_{RAIN} is to be considered in relationship to *rainfall rate,* also called *rain intensity,* expressed in mmhour^{-1}, elevation angle and frequency. Relevant techniques for the determination of rain attenuation A_{RAIN} for various time percentages are presented in [MAR02 pp 224–236]. A_{RAIN} increases with rainfall rate and can reach high values when small percentages of time are considered. Rainfall rate depends on the climate and hence on the considered location. The percentage of time a given value of rain attenuation A_{RAIN} is exceeded can be calculated from the rainfall rate, $R_{0.01}$, expressed in mmhr^{-1}, exceeded at the considered location for 0.01% of an average year. Figure 5.15 shows a world map with contours of constant value for $R_{0.01}$. These are simplified contours, just to show typical trends and underline the fact that the equatorial zones are faced with rainfall rate values higher than temperate climate zones (near the equator $R_{0.01}$ is of the order of 100 mmhr^{-1}, and less than 50 mmhr^{-1} at latitudes higher than 45°). More accurate data and maps in digital formats can be obtained from the International Telecommunication Union (ITU), and can be viewed in [MAR02 pp 225–230]. A_{RAIN} also depends on the length of the path of the radio frequency wave through the atmosphere, and therefore on

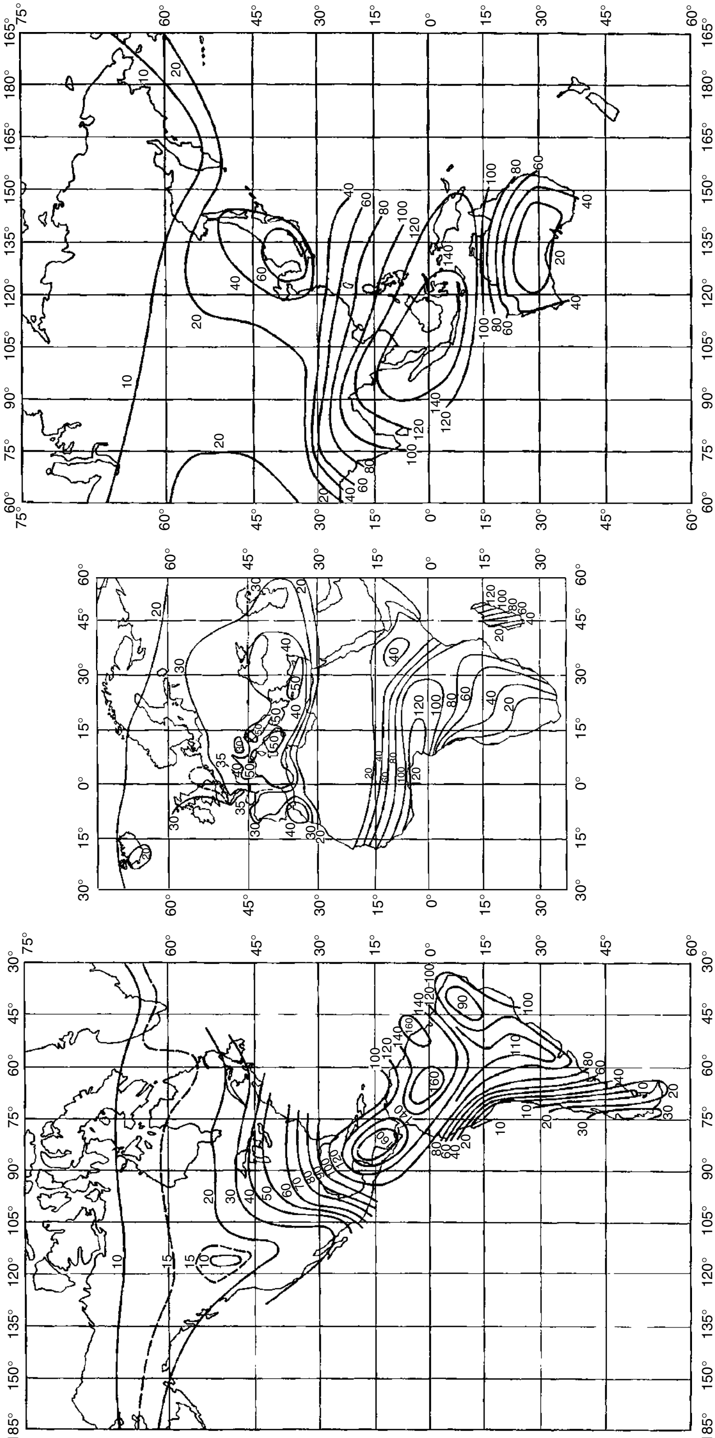


Figure 5.15 World map with contours of constant value of R_{001}, rainfall rate exceeded for 0.01% of time

the elevation angle of the antenna, which depends on the relative positioning of the earth station with respect to the satellite. For earth stations far away from the satellite, and therefore pointing at low elevation angle, the path through rain is longer, and the attenuation is higher. Finally, A_{RAIN} increases with frequency from C-band to Ka-band. Vertical polarisation is more attenuated by rain than horizontal polarisation as a result of the non-spherical shape of raindrops.

Figures 5.16 and 5.17 display the percentage of time the value of rain attenuation indicated on the horizontal axis is exceeded for different frequencies of a vertically polarised wave, considering elevation angle $E = 35°$. Figure 5.16 relates to $R_{0.01} = 30$ mmhr^{-1} and therefore is representative of temperate climate zones. Figure 5.17 relates to $R_{0.01} = 100$ mmhr^{-1}, and therefore is representative of tropical or equatorial climate zones.

Attenuation due to rain clouds and fog is usually small compared with that due to rain precipitation, except for clouds and fog with a high water concentration. For an elevation angle $E = 20°$, it is negligible at C-band, typically 0.5 to 1.5 dB at Ku-band, and 2 to 4 dB at Ka-band. This attenuation, however, is observed for a greater time percentage. Attenuation due to ice clouds is smaller still. Dry snow has little effect. Although wet snowfalls can cause greater attenuation than the equivalent rainfall rate, this situation is very rare and has little effect on attenuation statistics. The degradation of antenna characteristics due to accumulation of snow and ice on the dish may be more significant than the effect of snow along the path.

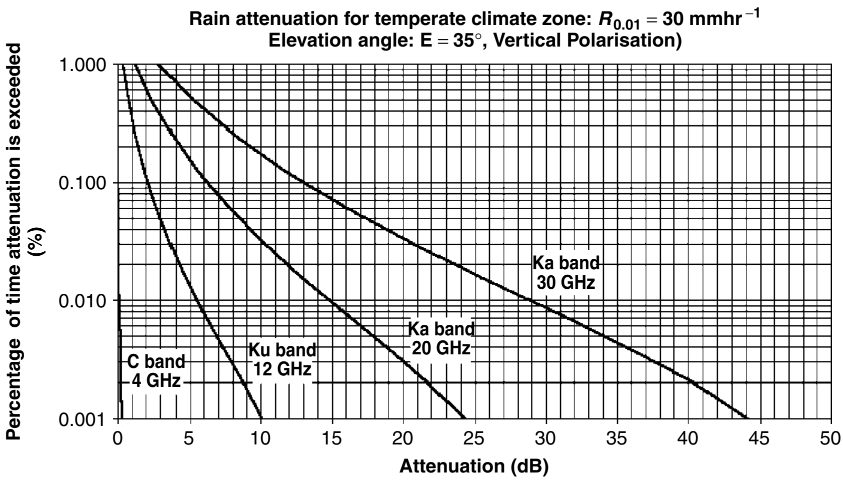


Figure 5.16 Percentage of time attenuation is exceeded for temperate climate zones where the rainfall rate exceeded for 0.01% of an average year is $R_{0.01} = 30$ mmhr^{-1}. Elevation angle $E = 35°$. Earth station latitude = 47.4°N, longitude = 5°E, altitude = 300 m. Satellite longitude = 13°E

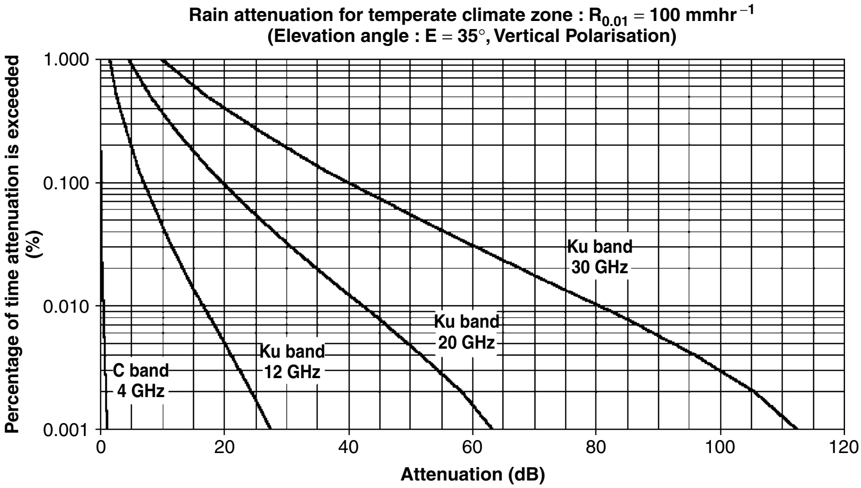


Figure 5.17 Percentage of time attenuation is exceeded for equatorial and tropical climate zones where the rainfall rate exceeded for 0.01% of an average year is $R_{0.01} = 100\ \text{mmhr}^{-1}$. Elevation angle $E = 35°$. Earth station in Florida, USA: latitude = 28°N, longitude = 82° W, altitude = 300 m. Satellite longitude = 41.4° W

Another effect of rain is depolarisation of the wave. As a consequence, where frequency reuse by orthogonal polarisation is used, a part of the carrier power transmitted in one polarisation is transferred to the orthogonal polarisation, causing cross polarisation interference. This is discussed in section 5.5 in relation to interference.

5.2.4 Figure of merit of satellite receiving equipment

The figure of merit $(G/T)_{SL}$ of the satellite receiving equipment incorporates the composite gain from the antenna to the satellite receiver input, and the uplink system noise temperature. As a factor in the expression (5.10) or (5.11) for $(C/N_0)_U$, it indicates the capability of the satellite receiving equipment to build up a high value of $(C/N_0)_U$. Its expression is given by:

$$\left(\frac{G}{T}\right)_{SL} = \left(\frac{G_{Rmax}}{L_R L_{pol}}\right)_{SL} \left(\frac{1}{L_{FRX}}\right)_{SL} \left(\frac{1}{T_U}\right) \quad (\text{K}^{-1}) \quad (5.31)$$

$$\left(\frac{G}{T}\right)_{SL} (\text{dBK}^{-1}) = G_{Rmax}(\text{dBi}) - L_R(\text{dB}) - L_{pol}(\text{dB}) - L_{FRX}(\text{dB}) - 10\log T_U$$

where G_{Rmax} is the satellite antenna receive gain at boresight and L_{R} is the off-axis gain loss corresponding to reception from a station at the edge of coverage. If the considered coverage contour corresponds to that of the half-power beamwidth, then $L_{\mathrm{R}} = 2$ and L_{R} (dB) = 3 dB. L_{pol} is the gain loss as a result of possible polarisation mismatch between the antenna and the received wave. Methods for evaluating this loss are given in [MAR02, p 205 and p 391]. A practical value is $L_{\mathrm{pol}} = 0.1$ dB. L_{FRX} is the feeder loss from the antenna to the receiver input, typically 1 dB. T_{U} is the uplink system noise temperature, and is given by:

$$T_{\mathrm{U}} = \frac{T_{\mathrm{A}}}{L_{\mathrm{FRX}}} + T_{\mathrm{F}}\left(1 - \frac{1}{L_{\mathrm{FRX}}}\right) + T_{\mathrm{R}} \quad \text{(K)} \tag{5.32}$$

where T_{A} is the satellite antenna noise temperature, T_{F} is the temperature of the feeder and T_{R} is the satellite receiver effective input noise temperature. Practical values are: $T_{\mathrm{A}} = 290$ K, $T_{\mathrm{F}} = 290$ K, $L_{\mathrm{FRX}} = 1$ dB, $T_{\mathrm{R}} = 500$ K, hence $T_{\mathrm{U}} = 790$ K.

The maximum satellite antenna gain G_{Rmax} depends on the coverage and is typically 20 dBi for global coverage and 38 dBi for a narrow spot beam coverage. Therefore, $(G/T)_{\mathrm{SL}}$ ranges from -13 dBK^{-1} for a global coverage to $+5$ dBK^{-1} for a spot beam coverage as indicated in Table 2.2.

5.3 DOWNLINK ANALYSIS

Figure 5.18 illustrates the geometry of the downlink. In order to calculate the worst case value of $(C/N_0)_{\mathrm{D}}$, the transmitting earth station is assumed to be located at the edge of the downlink coverage, defined as the contour where the satellite receiving antenna has a constant gain defined relative to its maximum value at boresight, for instance -3 dB, corresponding to a reduction by a factor two of the gain compared to its maximum. From Table 5.1, the ratio $(C/N_0)_{\mathrm{D}}$ is defined as:

$$\left(\frac{C}{N_0}\right)_{\mathrm{D}} = \frac{C_{\mathrm{D}}}{N_{0\mathrm{D}}} \quad \text{(Hz)} \tag{5.33}$$

where C_{D} is the power of the received carrier at the input to the earth station receiver. $N_{0\mathrm{D}}$ is the power spectral density of noise and relates to the downlink system noise temperature T_{D}:

$$N_{0\mathrm{D}} = \mathrm{k}\, T_{\mathrm{D}} \quad (\mathrm{WHz}^{-1}) \tag{5.34}$$

where k is the Boltzman constant: $\mathrm{k} = 1.38 \times 10^{-23}\ \mathrm{JK}^{-1}$; k (dBJK^{-1}) $= 10 \log k = -228.6$ dBJK^{-1}.

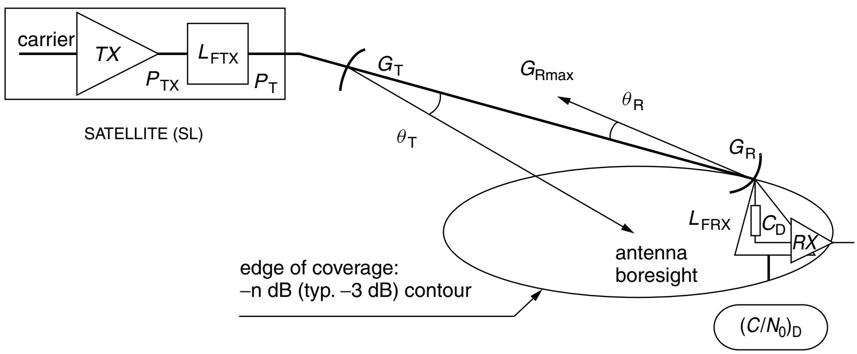


Figure 5.18 Geometry of the downlink. P_{TX}: carrier power at satellite transmitter output; L_{FTX}: feeder loss from satellite transmitter to antenna; P_T: carrier power fed to the satellite antenna G_T: satellite antenna transmit gain in direction of earth station; θ_T: satellite antenna half beamwidth angle; G_{Rmax}: earth station antenna receive gain at boresight; θ_R: earth station antenna depointing angle; L_{FRX}: feeder loss from earth station antenna to receiver input; C_D: carrier power at receiver input; RX: receiver

$(C/N_0)_D$ can be expressed [MAR93, p 65] as:

$$\left(\frac{C}{N_0}\right)_D = \mathrm{OBO}_1\left(\frac{C}{N_0}\right)_{\mathrm{Dsat}} \quad (\mathrm{Hz}) \tag{5.35}$$

$$\left(\frac{C}{N_0}\right)_D (\mathrm{dBHz}) = \mathrm{OBO}_1(\mathrm{dB}) + \left(\frac{C}{N_0}\right)_{\mathrm{Dsat}} (\mathrm{dBHz})$$

where OBO_1 is the output back-off per carrier for the considered carrier, as defined in Appendix 6 by expression (A6.4). OBO_1 depends on IBO_1. A simplified model is given by (A6.10). $(C/N_0)_{\mathrm{Dsat}}$ is the value of $(C/N_0)_D$ obtained at transponder saturation, and is given by:

$$\left(\frac{C}{N_0}\right)_{\mathrm{Dsat}} = \mathrm{EIRP}_{\mathrm{SLsat}}\left(\frac{1}{L_D}\right)\left(\frac{G}{T}\right)_{\mathrm{ES}}\left(\frac{1}{\mathrm{k}}\right) \quad (\mathrm{Hz}) \tag{5.36}$$

$$\left(\frac{C}{N_0}\right)_{\mathrm{Dsat}} (\mathrm{dBHz}) = \mathrm{EIRP}_{\mathrm{SLsat}}(\mathrm{dBW}) - L_D(\mathrm{dB}) + \left(\frac{G}{T}\right)_{\mathrm{ES}} (\mathrm{dBK}^{-1}) - 10\log k(\mathrm{dBJK}^{-1})$$

$\mathrm{EIRP}_{\mathrm{SLsat}}$ is the effective isotropic radiated power of the satellite (W) when operated at saturation (see section 5.3.1). L_D is the downlink path loss (see section 5.3.3). $(G/T)_{\mathrm{ES}}$ is the figure of merit of the earth station (see section 5.3.4). k is the Boltzman constant: $\mathrm{k} = 1.3 \times 10^{-23}\ \mathrm{JK}^{-1}$; k $(\mathrm{dBJK}^{-1}) = 10\log k = -228.6\ \mathrm{dBJK}^{-1}$.

The above expressions assume that the satellite transmits a noise-free carrier so that the noise which corrupts the carrier at the earth station receiver is downlink noise only, with neither contribution from the uplink noise relayed by the transponder, nor from interference.

5.3.1 Effective isotropic radiated power of the satellite

From the definition given in Appendix 5, the effective isotropic radiated power of the satellite is expressed as:

$$\mathrm{EIRP_{SL}} = P_T G_T \quad \mathrm{(W)} \tag{5.37}$$

$$\mathrm{EIRP_{SL}\,(dBW)} = P_T\,\mathrm{(dBW)} + G_T\,\mathrm{(dBi)}$$

where P_T is the power fed to the transmitting antenna and G_T is the satellite antenna transmit gain along the constant gain coverage contour.

The power P_T fed to the antenna depends on the power P_{TX} at the output of the transponder amplifier and the feeder loss L_{FTX}:

$$P_T = \frac{P_{TX}}{L_{FTX}} \quad \mathrm{(W)} \tag{5.38}$$

$$P_T\,\mathrm{(dBW)} = P_{TX}\,\mathrm{(dBW)} - L_{FTX}\,\mathrm{(dB)}$$

$\mathrm{EIRP_{SLsat}}$ is obtained when the transponder amplifier operates at saturation. The actual satellite EIRP depends on the total output back-off:

$$\mathrm{EIRP_{SL}} = \mathrm{OBO_t EIRP_{SLsat}} \quad \mathrm{(W)} \tag{5.39}$$

$$\mathrm{EIRP_{SL}\,(dBW)} = \mathrm{OBO_t\,(dB)} + \mathrm{EIRP_{SLsat}\,(dBW)}$$

$\mathrm{OBO_t}$ is a function of $\mathrm{IBO_t}$, as discussed in Appendix 6.

Table 2.2 gives typical values of satellite transponder $\mathrm{EIRP_{SLsat}}$ according to the type of coverage, assuming that the transmitted carrier benefits from the saturated transponder output power.

5.3.2 Power Flux density at earth surface

According to the definition given in Appendix 5, the power flux density generated by satellite transmission of a carrier at the earth surface is equal to:

$$\Phi = \frac{\mathrm{EIRP_{SL}}}{4\pi \mathrm{R}^2} \quad (\mathrm{Wm^{-2}}) \tag{5.40}$$

$$\Phi\,(\mathrm{dBWm^{-2}}) = \mathrm{EIRP_{SL}\,(dBW)} - \log 4\pi R^2$$

Radio regulations impose limits on the power flux density at the earth surface produced by satellites; according to expression (5.40), such limits translate into maximum values of $EIRP_{SL}$. The objective here is to limit the level of interference from an interfering satellite onto an earth station in the wanted satellite network. This aspect is discussed in more detail in section 5.5.

5.3.3 Downlink path loss

The downlink path loss L_D is the overall attenuation of the carrier power on its way from the satellite transmitting antenna to the earth station receiving antenna. As for the uplink, it builds up from two components, the free space loss, L_{FS}, and the atmospheric loss, L_A, so that the path loss can be expressed as:

$$L_D = L_{FS}L_A \tag{5.41}$$

$$L_D\,(\mathrm{dB}) = L_{FS}\,(\mathrm{dB}) + L_A\,(\mathrm{dB})$$

The downlink free space loss L_{FS} depends on the downlink frequency f and on the distance R between the earth station and the satellite, with an expression identical to (5.28) used for the uplink free space loss. The distance R is taken into account by the same geometric factor $(R/R_0)^2$ as for the uplink, so expression (5.29) can be used as well.

Values can be obtained from Figures 5.13 and 5.14, using the appropriate downlink frequency, and relative coordinates of the receiving earth station.

The atmospheric loss L_A is again the result of the combined effect of attenuation due to atmospheric gases A_{AG} and attenuation due to rain A_{RAIN}. Values for A_{AG} have been discussed in the context of the uplink. Calculation of A_{RAIN} can be performed according to [MAR02, pp 224–236]. The curves in Figures 5.16 and 5.17 can be considered for a preliminary evaluation of the value of A_{RAIN} exceeded for a given percentage of time depending on the climatic region

5.3.4 Figure of merit of earth station receiving equipment

Figure 5.19 displays the components of the receiving equipment: the antenna, the feeder from antenna to receiver and the receiver. The receiver comprises a low noise amplifier (LNA), a down converter

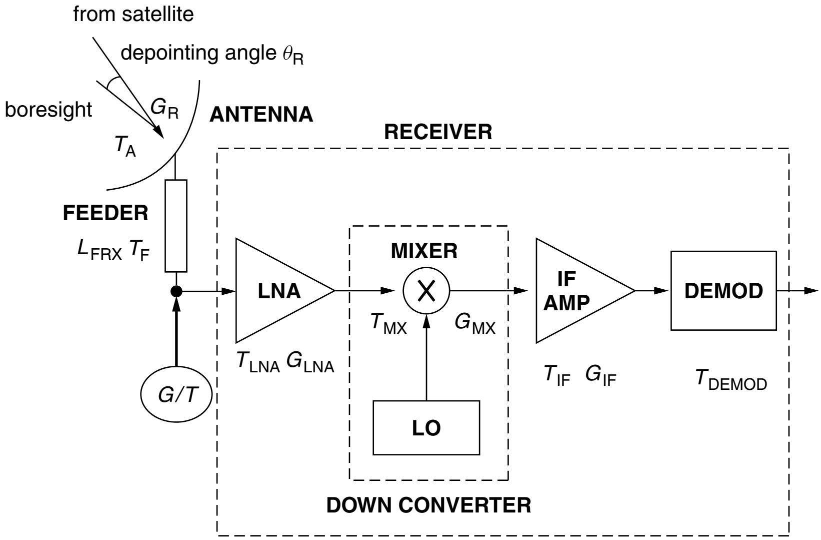


Figure 5.19 Earth station receiving equipment components. G_R: receive antenna gain in the direction of the satellite; T_A: antenna noise temperature; L_{FRX}: feeder loss; T_F: feeder temperature; LNA: low noise amplifier with effective input noise temperature T_{LNA} and power gain G_{LNA}; MIXER with effective input noise temperature T_{MX} and power gain G_{MX}; IF AMP: intermediate frequency amplifier with effective input noise temperature T_{IF} and power gain G_{IF}; DEMOD: demodulator with input noise temperature T_{DEMOD}

constituted of a mixer and a local oscillator (LO), an intermediate frequency amplifier (IF AMP) and a demodulator (DEMOD).

As can be seen from expressions (5.35) and (5.36), the figure of merit $(G/T)_{ES}$ of this equipment measures its ability to build up a high value of $(C/N_0)_D$. Its expression is given by:

$$\left(\frac{G}{T}\right)_{ES} = \left(\frac{G_{Rmax}}{L_R L_{pol}}\right)_{ES} \left(\frac{1}{L_{FRX}}\right)_{ES} \left(\frac{1}{T_D}\right) \quad (K^{-1}) \quad (5.42)$$

$$\left(\frac{G}{T}\right)_{ES} (dBK^{-1}) = G_{Rmax}\,(dBi) - L_R\,(dB) - L_{pol}\,(dB) - L_{FRX}\,(dB) - 10\log T_D$$

G_{Rmax} is the earth station antenna receive gain at boresight. L_R is the off-axis gain loss corresponding to the depointing angle θ_R. L_{pol} is the gain loss corresponding to antenna polarisation mismatch, typically 0.1 dB. L_{FRX} is the feeder loss from the antenna to the receiver input, typically 0.5 dB. T_D is the downlink system noise temperature, given by:

$$T_D = \frac{T_A}{L_{FRX}} + T_F\left(1 - \frac{1}{L_{FRX}}\right) + T_R \quad (K) \qquad (5.43)$$

T_A is the earth station antenna noise temperature. T_F is the temperature of the feeder, typically 290 K. T_R is the earth station receiver effective input noise temperature.

The earth station antenna receive gain at boresight is given by:

$$G_{Rmax} = \eta_a \left(\frac{\pi Df}{c}\right)^2 \tag{5.44}$$

$$G_{Rmax}\,(\text{dBi}) = 10 \log \eta_a \left(\frac{\pi Df}{c}\right)^2$$

Values of G_{Rmax} are given in Figures 5.7 and 5.8, where the value of frequency to be considered is that of the downlink.

The actual receive gain in the direction of the satellite is:

$$G_R\,(\text{dBi}) = G_{Rmax}\,(\text{dBi}) - L_R\,(\text{dB}) - L_{pol}\,(\text{dB}) \tag{5.45}$$

where the off-axis receive gain loss L_R corresponding to the depointing angle θ_R is given by:

$$L_R\,(\text{dB}) = 12\left(\frac{\theta_R}{\theta_{3dB}}\right)^2 \tag{5.46}$$

where θ_{3dB} is the half power beamwidth of the receive radiation pattern as given by expression (5.19) and the curves of Figures 5.7 and 5.8, considering the actual value of the downlink frequency.

The determination of the maximum value θ_{Rmax} of θ_R has been discussed in section 5.2.2 and is illustrated in Figure 5.10. Actually, θ_{Rmax} is equal to θ_{Tmax} as given by expression (5.25). Values of L_{Rmax} (dB) are readily calculated from (5.24) by replacing θ_T by θ_R and considering $\theta_R = \theta_{Rmax}$.

Expression (5.43) shows that the downlink system noise temperature depends on the earth station antenna noise temperature T_A. The antenna noise temperature T_A represents the overall contribution of noise components captured by the antenna. Two situations are to be considered:

- Antenna noise temperature for *clear sky conditions* (Figure 5.20). The antenna captures the noise radiated by the sky with temperature T_{SKY} and a contribution T_{GROUND} from the ground in the vicinity of the earth station. The overall contribution is given by:

$$T_A = T_{SKY} + T_{GROUND} \quad (\text{K}) \tag{5.47}$$

- antenna noise temperature for *rain conditions* (Figure 5.21). Rain acts as an attenuator with attenuation A_{RAIN} and average medium

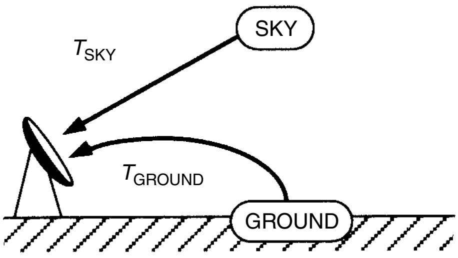


Figure 5.20 Components of earth station antenna noise temperature for clear sky conditions

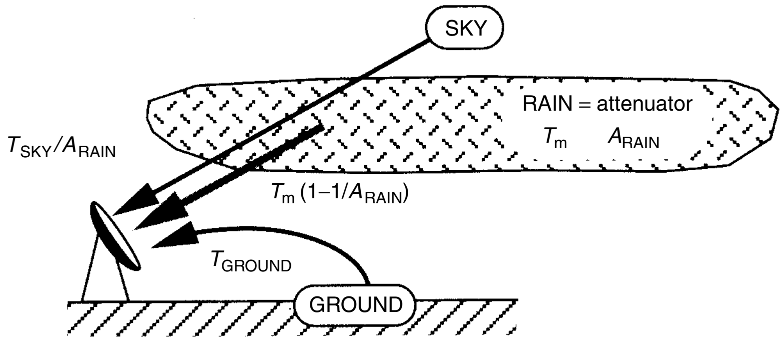


Figure 5.21 Components of earth station antenna noise temperature for rain conditions

temperature T_m. It attenuates the contribution from the clear sky, and generates its own noise with noise temperature T_m $(1 - 1/A_{RAIN})$ at the output of the attenuation process. The noise contribution from the ground in the vicinity of the earth station is considered not to be modified by rain. The overall contribution is given by:

$$T_A = \frac{T_{SKY}}{A_{RAIN}} + T_m\left(1 - \frac{1}{A_{RAIN}}\right) + T_{GROUND} \quad \text{(K)} \qquad (5.48)$$

Table 5.3 displays typical values of the clear sky noise T_{SKY} at two elevation angles ($E = 10°$ and $E = 35°$) for a standard atmosphere. The contribution of the ground in the vicinity of the earth station depends on the characteristics of the antenna (mounting, diameter), the sidelobe radiation pattern and the frequency. A typical value is $T_{GROUND} = 30$ K.

Figure 5.22 displays the variation of the antenna noise temperature T_A with rain attenuation A_{RAIN}. The antenna noise temperature increases with rain attenuation.

Table 5.3 Clear sky noise contribution T_{SKY} (standard atmosphere)

Frequency	$E = 10°$	$E = 35°$
4 GHz	10 K	4 K
12 GHz	20 K	7 K

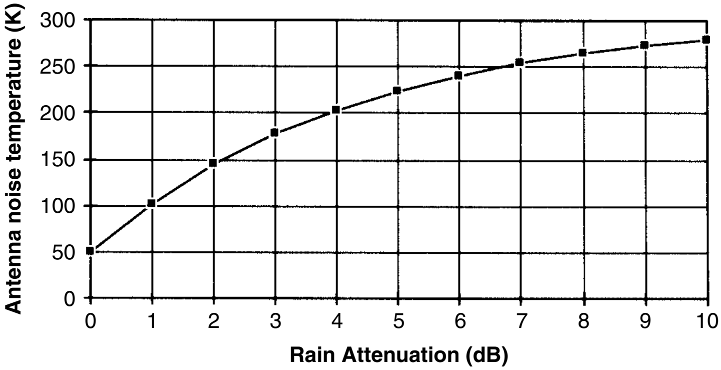


Figure 5.22 Variation of antenna noise temperature with rain attenuation (frequency 12 GHz, $T_{SKY} = 20$ K, $T_{GROUND} = 30$ K)

The earth station receiver effective input noise temperature T_R can be calculated from Friis's formula:

$$T_R = T_{LNA} + \frac{T_{MX}}{G_{LNA}} + \frac{T_{IF}}{G_{LNA}G_{MX}} + \frac{T_{DEMOD}}{G_{LNA}G_{MX}G_{IF}} \quad (\text{K}) \qquad (5.49)$$

In the above formula, all values of noise temperature and gain are to be expressed in absolute value (not in dB). Usually, the LNA gain is high enough (typically 50 dB $= 10^5$) for all terms but the first one to be negligible. Therefore: $T_R \approx T_{LNA}$, with a typical value equal to 30 K at C-band and 80 K at Ku-band.

From the above set of expressions, one can see that the figure of merit $(G/T)_{ES}$ of an earth station is maximal when there is no depointing loss, no feeder loss, no polarisation mismatch and no rain attenuation. This maximum value is given by:

$$\left(\frac{G}{T}\right)_{ESmax} = (G_{Rmax})_{ES}\left(\frac{1}{T_{Dmin}}\right) \quad (\text{K}^{-1}) \qquad (5.50)$$

$$\left(\frac{G}{T}\right)_{ESmax} (\text{dBK}^{-1}) = G_{Rmax}(\text{dBi}) - 10\log T_{Dmin}$$

where:

$$T_{Dmin} = T_{SKY} + T_{GROUND} + T_R \quad (\text{K}) \qquad (5.51)$$

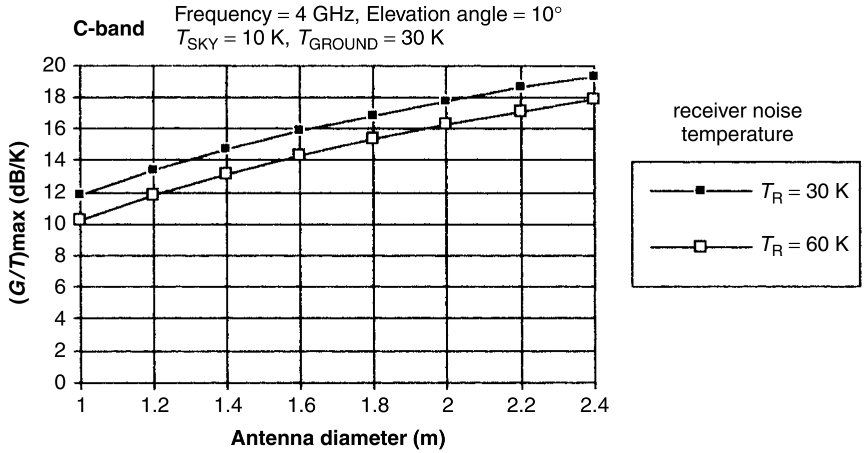


Figure 5.23 Variations of $(G/T)_{ESmax}$ with antenna diameter at 4 GHz

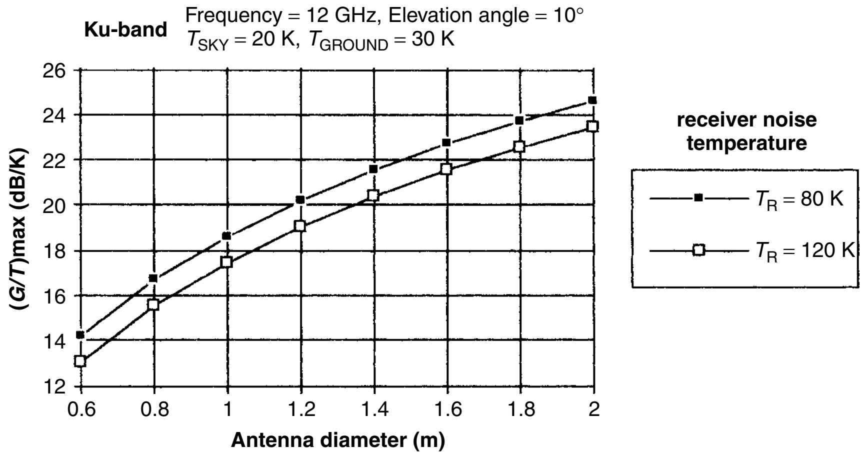


Figure 5.24 Variations of $(G/T)_{ESmax}$ with antenna diameter at 12 GHz

Figures 5.23 and 5.24 display variations of $(G/T)_{ESmax}$, with antenna diameter respectively at 4 GHz (C-band) and 12 GHz (Ku-band), considering the receiver noise temperature T_R as a parameter.

When a more representative value is sought, the impact of depointing and feeder losses should be introduced along with the impact of rain attenuation. The resulting degradation of $(G/T)_{ES}$ with respect to $(G/T)_{ESmax}$ can be evaluated from:

$$\left(\frac{G}{T}\right)_{ES} (\text{dBK}^{-1}) = \left(\frac{G}{T}\right)_{ESmax} (\text{dBK}^{-1})$$

$$- L_{Rmax}(\text{dB}) \qquad \text{depointing loss}$$

$$- L_{pol}(\text{dB}) \qquad \text{polarisation mismatch loss}$$

$$- \text{DELTA(dB)} \qquad \text{combined effect of feeder loss and rain attenuation} \tag{5.52}$$

where:

$$L_{\text{Rmax}}\,(\text{dB}) = 12\left(\frac{\theta_{\text{Rmax}}}{\theta_{3\text{dB}}}\right)^2 \quad (\text{dB}) \tag{5.53}$$

$$L_{\text{pol}} = \text{typically } 0.1 \text{ dB}$$

$$\text{DELTA(dB)} = L_{\text{FRX}} + 10\log T_{\text{D}} - 10\log T_{\text{Dmin}} \tag{5.54}$$

In expression (5.54), T_{D} represents the downlink system noise temperature as given by (5.43), with T_{A} given by (5.48).

For a VSAT, typical values of L_{Rmax} are 0.6 dB at C-band and 1 dB at Ku-band. Tables 5.4 and 5.5 display typical values of DELTA.

Table 5.4 Typical values of DELTA at 4 GHz (C-band): elevation angle = 35°; $T_{\text{SKY}} = 4$ K; $T_{\text{GROUND}} = 30$ K; feeder loss $L_{\text{FRX}} = 0.5$ dB; station keeping window half-width SKW = 0.025°

	C-band	A_{RAIN}	
		0 dB	0.6 dB*
T_{R}	30 K	2.1 dB	3.3 dB
	60 K	1.6 dB	2.6 dB

*0.6 dB corresponds to the value of attenuation due to rain which is exceeded for 0.01% of the time in a tropical or equatorial climate zone where $R_{0.01} = 100$ mmhr^{-1}. (see Figure 5.17)

Table 5.5 Typical values of DELTA at 12 GHz (Ku-band): elevation angle = 35°; $T_{\text{SKY}} = 7$ K; $T_{\text{GROUND}} = 30$ K; feeder loss $L_{\text{FRX}} = 0.5$ dB; station keeping window half-width SKW = 0.025°

	Ku-band	A_{RAIN}	
		0 dB	0.6 dB*
T_{R}	80 K	1.4 dB	4.9 dB
	120 K	1.2 dB	4.1 dB

*6 dB corresponds to the value of attenuation due to rain which is exceeded for 0.01% of the time in a temperate climate zone where $R_{0.01} = 30$ mmhr^{-1} (see Figure 5.16)

Table 5.6 Achievable G/T with current VSAT technology at C-band (elevation angle $E = 35°$)

	C-band	
Antenna diameter	1.8 m	2.4 m
G/T clear sky	15 dBK^{-1}	17 dBK^{-1}
G/T rain (0.01%)*	13 dBK^{-1}	16 dBK^{-1}

*Tropical or equatorial climate location

Table 5.7 Achievable G/T with current VSAT technology at Ku-band (elevation angle $E = 35°$)

	Ku-band	
Antenna diameter	1.2 m	1.8 m
G/T clear sky	19 dBK^{-1}	22 dBK^{-1}
G/T rain (0.01%)*	15 dBK^{-1}	19 dBK^{-1}

*Temperate climate location

Table 5.8 Typical G/T for hub stations (elevation angle $E = 35°$)

	C-band		
Antenna diameter	3 ma^{+}	6 ma^{+}	10 m
G/T clear sky	19 dBK^{-1}	25 dBK^{-1}	29 dBK^{-1}
G/T rain (0.01%)*)	18 dBK^{-1}	23 dBK^{-1}	28 dBK^{-1}
	Ku-band		
Antenna diameter	3 ma†	6 m	10 m
G/T clear sky	26 dBK^{-1}	33 dBK^{-1}	37 dBK^{-1}
G/T rain (0.01%)**	22 dBK^{-1}	29 dBK^{-1}	34 dBK^{-1}

*Tropical or equatorial climate location.
**Temperate climate location.
†No tracking.

From the above results, it is possible to work out the data in Tables 5.6 and 5.7. These indicate achievable values of G/T for VSATs depending on antenna diameter, using an elevation angle $E = 35°$. Table 5.8 indicates typical values for hub stations.

5.4 INTERMODULATION ANALYSIS

When operated in a non-linear mode, the transponder amplifier generates intermodulation products, and the transponder output power is shared not only between the amplified carriers but also with the intermodulation products. This is most pronounced at and near transponder saturation, when the total input back-off is equal to, or approaches, zero (see Appendix 6 for definition of back-off). Intermodulation products of modulated carriers are transmitted along with the wanted carrier and act as noise for the wanted carrier as received by the destination earth station.

Figure 5.25 gives typical curves for the carrier power to intermodulation noise power density ratio $(C/N_0)_{\mathrm{IM}}$ at the destination earth station receiver input, as a function of the transponder amplifier total input back-off, assuming n equally powered carriers.

An approximate formula for the curves in Figure 5.25 is [ITU88 p90]:

$$\left(\frac{C}{N_0}\right)_{\mathrm{IM}} = 79 - 10\log n - 1.65(\mathrm{IBO_t\,(dB)} + 5)\ \ \mathrm{IBO_t} < -5\ \mathrm{dB(dBHz)} \tag{5.55}$$

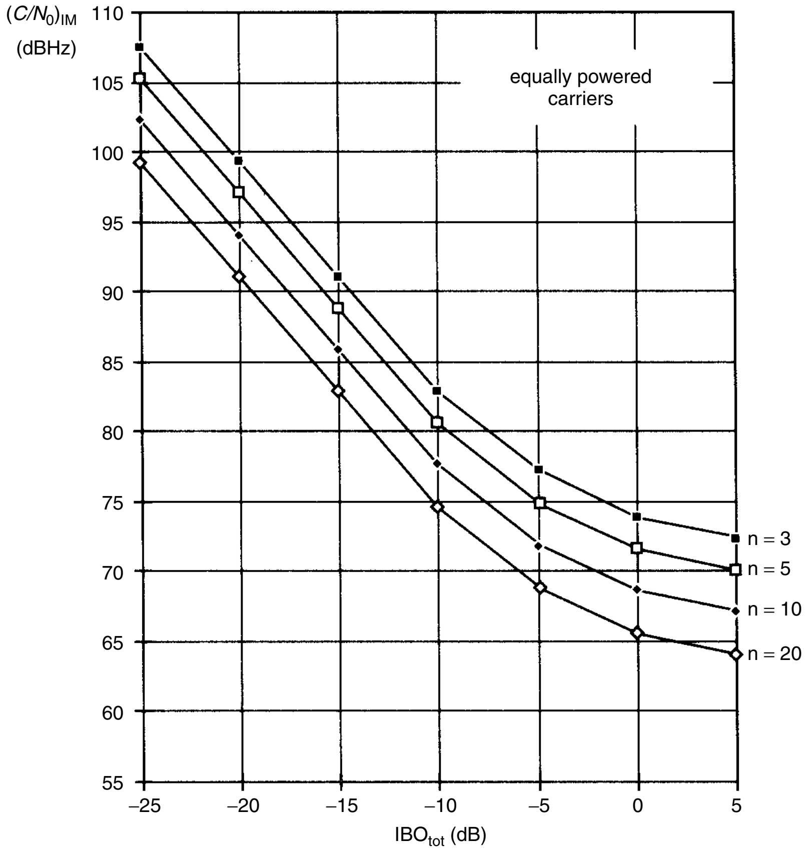


Figure 5.25 $(C/N_0)_{\mathrm{IM}}$ as a function of the total input back-off

5.5 INTERFERENCE ANALYSIS

5.5.1 Expressions for carrier-to-interference ratio

Interference is unwanted radio frequency energy introduced at the end receiver of the wanted link. Figure 5.26 illustrates the spectrum of a wanted carrier contaminated by an interfering carrier. The receiver with equivalent noise bandwidth B_{N} captures carrier power C, and interference power N_i. The carrier-to-interference power ratio is C/N_i.

The amount of interference power into the receiver depends on the overlap of the carrier spectra, as illustrated in Figure 5.26.

The receiver noise bandwidth B_N is matched to the bandwidth occupied by the wanted carrier with power C. Power C is given by:

$$C = \mathrm{PSD}_{\mathrm{w}} B_{\mathrm{N}} \quad \mathrm{(W)} \tag{5.57}$$

where PSD_w is the wanted carrier power spectral density (WHz^{-1}).

The interfering carrier has a power spectral density $\mathrm{PSD}_{\mathrm{i}}$ (WHz^{-1}) and occupies a band B_{i}. The amount of interfering power N_{i} captured by the receiver is equal to the spectrum overlap area:

$$N_i = \mathrm{PSD}_{\mathrm{i}} \min(B_{\mathrm{i}}, B_{\mathrm{N}}) \quad \mathrm{(W)} \tag{5.58}$$

where $\min(B_{\mathrm{i}}, B_{\mathrm{N}})$ represents the smallest of the two bandwidths, B_{i} or B_{N}.

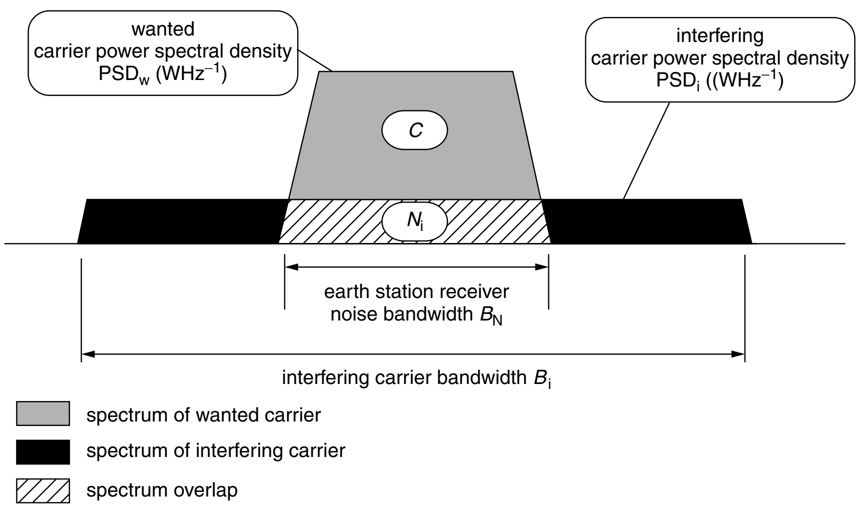


Figure 5.26 Spectrum of received carrier contaminated by an interfering carrier

The expression for the carrier-to-interference power ratio C/N_i results from (5.57) and (5.58):

$$\frac{C}{N_i} = \frac{PSD_w B_N}{PSD_i \min(B_i, B_N)} \tag{5.59}$$

Carrier power spectral density, PSD, is proportional to the EIRP density defined as EIRP/B, where B is the carrier bandwidth (Hz). Indeed:

$$PSD = \left(\frac{EIRP}{B}\right)\left(\frac{1}{L}\right) G_{RX} \quad (WHz^{-1}) \tag{5.60}$$

where:

L is the satellite link path loss,

G_{RX} is the composite receive gain expressed as:

$$G_{RX} = G_{Rmax}/L_R L_{pol} L_{FRX} \tag{5.61}$$

G_{Rmax} is the antenna receive gain at boresight,

L_R is the off-axis gain loss,

L_{pol} is the gain loss corresponding to antenna polarisation mismatch,

L_{FRX} is the feeder loss from antenna to receiver input.

The above derivations are approximate as no account is made of the actual shape of the modulated carrier spectrum. A more realistic approach should also consider that the centre frequency of the spectrum of the wanted carrier and the interfering carrier may not be coincident.

5.5.2 Types of interference

Interference can be:

- *self-interference*, i.e. transmissions by the wanted satellite system which hosts the considered VSAT network; or
- *external interference*, i.e. transmissions from systems sharing the same frequency bands as the wanted satellite system. Candidate interfering systems are other satellite systems and terrestrial microwave systems. C-band and most of Ku-band are shared by satellite and terrestrial microwave systems (see Figure 1.17). It is therefore convenient to operate VSAT networks in bands that are allocated to satellite links on a primary and exclusive basis,

as this eliminates the interference from terrestrial microwave systems. However, this does not protect a VSAT network from being interfered with by other satellite-based networks.

5.5.3 Self-interference

Self-interference is caused by frequency reuse and imperfect filtering. The first type is often called *co-channel interference* (CCI), the second type *adjacent channel interference* (ACI).

5.5.3.1 Co-channel interference

Co-channel interference (CCI) is a drawback of frequency reuse techniques which aim to increase the capacity of a satellite system without demanding more frequency band. These techniques have been presented in Chapter 2, section 2.1.5.

Co-channel interference originates in imperfect isolation between geographically separated beams or between orthogonal polarisations reusing the same frequency band. Therefore, it strongly depends on the antenna characteristics of the earth station and the satellite.

Beam to beam interference

Figure 5.27 illustrates the generation of co-channel interference as a result of imperfect isolation between beams in a multibeam satellite system. The VSAT network is assumed to be located within beam 1. An earth station belonging to some other network is located in beam 2. As a result of the non zero gain of the antenna gain pattern defining beam 1 in the direction of this earth station, some of its emitted power (upward arrow in black) is received in beam 1 (upward arrow in grey). This power enters the transponder conveying the inbound and outbound links of the VSAT network. As frequency reuse implies that both beams use the same frequency band, there is a chance of carrier spectrum overlap between the carrier emitted by the station of beam 2 and one or several of the VSAT network uplinked carriers (bidirectional arrows in black representing inbound and outbound links). This overlap constitutes uplink interference into the VSAT network.

The transponder dedicated to beam 2 transmits carriers to stations of beam 2. The downward arrow in black in beam 2 represents one of these carriers. A part of its power is emitted in beam 1 (downward arrows in grey in beam 1) and the part of its spectrum that overlaps

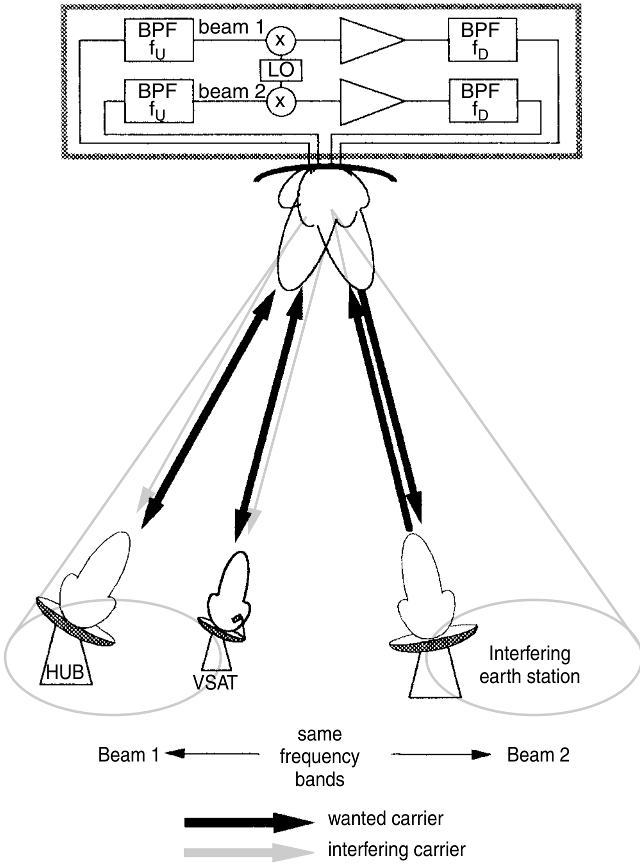


Figure 5.27 Generation of co-channel interference in a multibeam satellite system

with the spectrum of either the outbound or any of the inbound carriers of the VSAT network constitutes downlink interference to the VSAT network.

Expressions for carrier-to-co-channel interference power ratio C/N_i can be derived from (5.59) and (5.60):

Uplink co-channel interference:

$$\left(\frac{C}{N_\mathrm{i}}\right)_\mathrm{U} = \frac{\mathrm{EIRP_{ESw}}}{\mathrm{EIRP_{ESi}}}\frac{B_\mathrm{i}}{\min[B_\mathrm{i}, B_\mathrm{N}]}\frac{L_\mathrm{Ui}}{L_\mathrm{Uw}}\frac{G_\mathrm{RXw}}{G_\mathrm{RXi}} \tag{5.62a}$$

$$\begin{aligned}\left(\frac{C}{N_\mathrm{i}}\right)_\mathrm{U}(\mathrm{dB}) = {} & \mathrm{EIRP_{ESw}}\,(\mathrm{dBW}) - \mathrm{EIRP_{ESi}}\,(\mathrm{dBW}) + 10\log B_\mathrm{i} \\ & - 10\log\min[B_\mathrm{i}, B_\mathrm{N}] + L_\mathrm{Ui}\,(\mathrm{dB}) - L_\mathrm{Uw}\,(\mathrm{dB}) \\ & + G_\mathrm{RXw}\,(\mathrm{dBi}) - G_\mathrm{RXi}\,(\mathrm{dBi})\end{aligned} \tag{5.62b}$$

where $\mathrm{EIRP}_{\mathrm{ESw}}$ is the EIRP of the wanted station of beam 1 in the direction of the satellite. $\mathrm{EIRP}_{\mathrm{ESi}}$ is the EIRP of the interfering station of beam 2 in the direction of the satellite. L_{Uw} and L_{Ui} are the uplink path loss for the wanted carrier and the interfering one, respectively. G_{RXw} and G_{RXi} are the satellite composite receive gain of beam 1, given by (5.61), for the wanted and interfering carriers, respectively. The worst case should consider that the wanted station is at the edge of beam 1 and the interfering station at the edge of beam 2, as near as possible to the wanted station, so that the difference G_{RXw} (dBi) $-$ G_{RXi} (dBi) is minimal. The difference value can be obtained from the satellite coverage charts, or if not available, from the reference satellite antenna gain patterns given in the ITU-R Recommendation 672 and Report 558.

Downlink co-channel interference:

$$\left(\frac{C}{N_{\mathrm{i}}}\right)_{\mathrm{D}} = \frac{\mathrm{EIRP}_{\mathrm{SL1ww}}}{\mathrm{EIRP}_{\mathrm{SL2iw}}} \frac{B_{\mathrm{i}}}{\min[B_{\mathrm{i}}, B_N]} \tag{5.63a}$$

$$\left(\frac{C}{N_{\mathrm{i}}}\right)_{\mathrm{D}} (\mathrm{dB}) = \mathrm{EIRP}_{\mathrm{SL1ww}}\,(\mathrm{dBW}) - \mathrm{EIRP}_{\mathrm{SL2iw}}\,(\mathrm{dBW}) + 10\log B_{\mathrm{i}} - 10\log\min[B_{\mathrm{i}}, B_{\mathrm{N}}] \tag{5.63b}$$

where $\mathrm{EIRP}_{\mathrm{SL1ww}}$ is the satellite EIRP in beam 1 for the wanted carrier in the direction of the wanted station. Its value depends on the output back-off for that carrier, $\mathrm{OBO}_{\mathrm{w}}$, and the satellite EIRP in beam 1 at saturation in the direction of the wanted station, $\mathrm{EIRP}_{\mathrm{SL1wsat}}$:

$$\mathrm{EIRP}_{\mathrm{SL1ww}}\,(\mathrm{dBW}) = \mathrm{OBO}_{\mathrm{w}}\,(\mathrm{dB}) + \mathrm{EIRP}_{\mathrm{SL1wsat}}\,(\mathrm{dBW}) \tag{5.64}$$

$\mathrm{EIRP}_{\mathrm{SL2iw}}$ is the satellite EIRP in beam 2 for the interfering carrier in the direction of the wanted station. Its value depends on the output back-off for that carrier, $\mathrm{OBO}_{\mathrm{i}}$, and the satellite EIRP in beam 2 at saturation, $\mathrm{EIRP}_{\mathrm{SL2wsat}}$ in the direction of the wanted station:

$$\mathrm{EIRP}_{\mathrm{SL2iw}}\,(\mathrm{dBW}) = \mathrm{OBO}_{\mathrm{i}}\,(\mathrm{dB}) + \mathrm{EIRP}_{\mathrm{SL2wsat}}\,(\mathrm{dBW}) \tag{5.65}$$

The worst case should consider that the wanted station is at the edge of beam 1 and the interfering station at the edge of beam 2, as near as possible to the wanted station so that the difference $\mathrm{EIRP}_{\mathrm{SL1ww}}(dBW) - \mathrm{EIRP}_{SL2iw}$ (dBW) is minimal.

If one assumes that beams 1 and 2 have equal saturated maximum EIRP and identical gain pattern, then:

$$\mathrm{EIRP}_{\mathrm{SL1wsat}}\,(\mathrm{dBW}) = \mathrm{EIRP}_{\mathrm{SL2wsat}}\,(\mathrm{dBW}) + G_{\mathrm{T1w}}\,(\mathrm{dBi}) - G_{\mathrm{T2w}}\,(\mathrm{dBi}) \tag{5.66}$$

where G_{T1w} (dBi) and G_{T2w} (dBi) represent respectively the transmit antenna gain in the direction of the wanted station for the common antenna gain pattern of beams 1 and 2.

Combining expressions (5.63) to (5.66):

$$\left(\frac{C}{N_i}\right)_D = \frac{OBO_w}{OBO_i}\frac{B_i}{\min[B_i, B_N]} \tag{5.67a}$$

$$\left(\frac{C}{N_i}\right)_D \text{(dB)} = \text{OBO}_w\text{ (dB)} - \text{OBO}_i\text{ (dB)} + 10\log B_i - 10\log\min[B_i, B_N] + G_{T1w}\text{ (dBi)} - G_{T2w}\text{ (dBi)} \tag{5.67b}$$

Example

Consider a multibeam satellite with the same EIRP and antenna pattern in all beams. Table 5.9 displays the parameters for a wanted VSAT, the interfering station, and satellite relevant characteristics, and gives results of $(C/N_i)_U$ and $(C/N_i)_D$ calculations.

Overall link co-channel interference

Co-channel interference may contaminate both the uplink and the downlink portions of the VSAT network inbound or outbound link. The overall link carrier-to-co-channel interference power ratio is given by:

$$\left(\frac{C}{N_i}\right)_T^{-1} = \left(\frac{C}{N_i}\right)_U^{-1} + \left(\frac{C}{N_i}\right)_D^{-1} \tag{5.68}$$

where the absolute values (not in dB) of $(C/N_i)_U$ and $(C/N_i)_D$, calculated according to expressions (5.62a) and (5.67a) respectively, should be used.

With the values calculated in Table 5.9, this gives $(C/N_i)_T = \{10^{-1.8} + 10^{-1.8}\}^{-1} = 31.55 = 15$ dB.

For any link, the carrier power-to-co-channel interference power spectral density ratio (C/N_{0i}) is given by:

$$\frac{C}{N_{0i}} = \left(\frac{C}{N_i}\right)B_N \quad \text{(Hz)} \tag{5.69}$$

$$\frac{C}{N_{0i}}\text{ (dBHz)} = \left(\frac{C}{N_i}\right)\text{ (dB)} + 10\log B_N$$

Cross polarisation interference

Co-channel interference is also caused by orthogonally polarised carriers reusing the same frequency band within a given beam.

Table 5.9 Carrier-to-co-channel interference power ratio calculation

		EIRP	1	$EIRP_{ESw}$	40 dBW
UPLINK	VSAT	Receiver noise equivalent bandwidth	2	B_N	100 kHz = 50 dBHz
		Uplink path loss	3	L_{Uw}	207 dB
		EIRP	4	$EIRP_{ESi}$	65 dBW
	Interfering station	Interfering carrier bandwidth	5	B_i	2 MHz = 63 dBHz
		Uplink path loss	6	L_{Ui}	207 dB
	Satellite	Gain pattern	7	$G_{RXw} - G_{RXi}$	30 dB
		$(C/N_i)_U$	8	Expression (5.62) = 1 − 2 − 3 − 4 + 5 + 6 + 7	18 dB
DOWNLINK	Satellite	Wanted carrier output back-off	9	OBO_w	−30 dB
		Interfering carrier back-off	10	OBO_i	−5 dB
		Antenna gain difference	11	$G_{T1w} - G_{T2w}$	30 dB
		$(C/N_i)_D$	12	Expression (5.67) = 9 − 10 + 5 − 2 + 11	18 dB

Should the interfering stations be part of the considered VSAT network, then the VSAT network operation requires the use of two transponders, one for each group of carriers with the same polarisation. This is generally not justified for VSAT networks, hence all VSATs and the hub within the geographical coverage of the beam operate on the same polarisation. Cross polarisation interference then originates in carriers transmitted by stations of other networks using the same satellite system and operating within the same frequency bands as the considered VSAT network but on the orthogonal polarisation. These stations may be located within the same beam, and their carriers are then conveyed by the cross-polar satellite transponder, as illustrated in Figure 5.28.

The level of interference is conditioned by the cross polarisation isolation (XPI) of earth station and satellite antennas. The cross polarisation of an antenna is defined in Appendix 4. Figure 5.29 illustrates how energy from one polarisation contaminates the

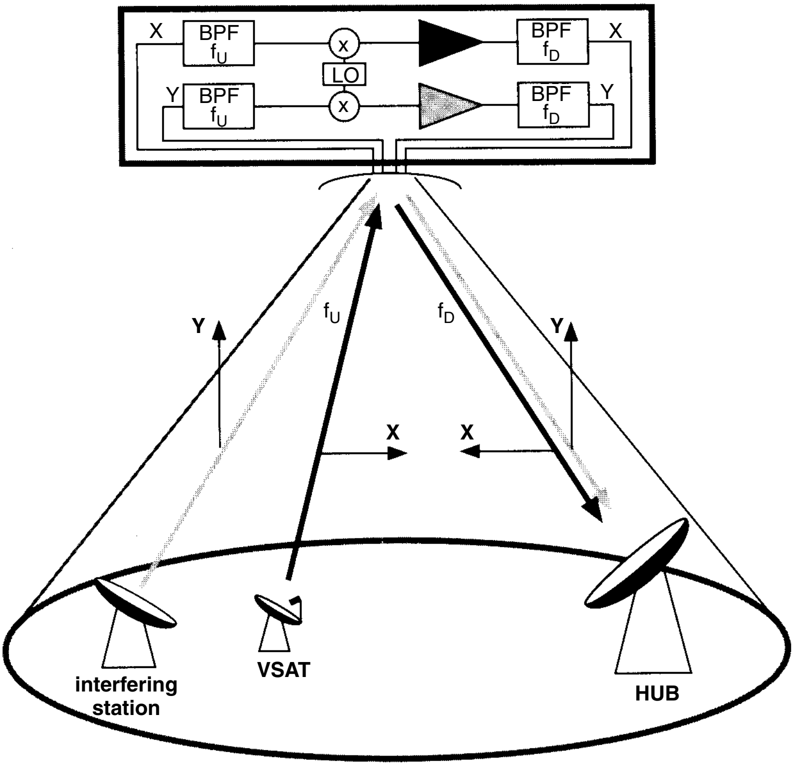


Figure 5.28 Frequency reuse based on transmission of two carriers at the same frequency with orthogonal polarisations X and Y. BPF: band pass filter

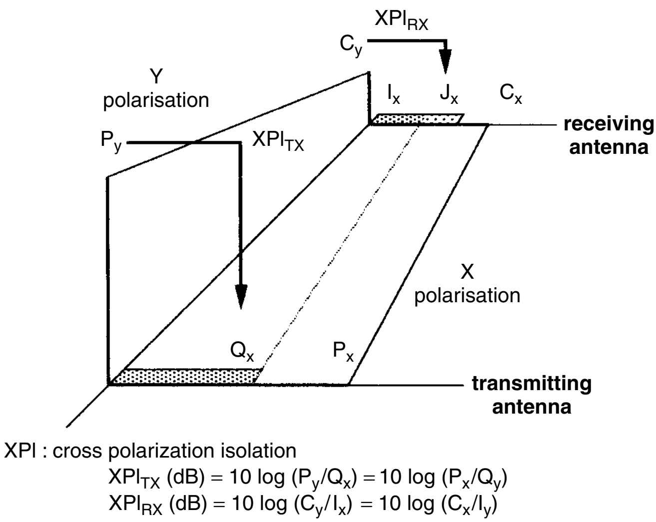


Figure 5.29 Contribution of energy to the horizontal component from the vertical one at the transmitting and receiving ends of a satellite link. TX: transmitting antenna; RX: receiving antenna

orthogonal one altogether at the transmitting and the receiving ends of a satellite link.

Carriers are transmitted at the same frequency, one on each X and Y polarisation. P_x and P_y are the respective transmitted powers. Due to imperfect cross polarisation isolation, a contribution Q_x is transmitted along with P_x. The receiving antenna receives the wanted carrier C_x to which is added I_x, the interfering contribution from Q_x. On the orthogonal polarisation, it receives C_y, and due to imperfect cross polarisation isolation, a contribution J_x is generated, which is added to I_x, so the total interference power is $I_x + J_x$. A similar process occurs on the Y-component. Table 5.10 displays the different power components and the resulting carrier to interference ratio.

From Table 5.10 an expression for the carrier-to-interference power ratio can be derived:

$$\frac{C_x}{(I_x + J_x)} = \left[\frac{P_y}{P_x}10^{-\frac{\mathrm{XPI_{TX}}}{10}} + \frac{C_y}{C_x}10^{-\frac{\mathrm{XPI_{RX}}}{10}}\right]^{-1} \tag{5.70}$$

$$\frac{C_y}{I_y + J_y} = \left[\frac{P_x}{P_y}10^{-\frac{\mathrm{XPI_{TX}}}{10}} + \frac{C_x}{C_y}10^{-\frac{XPI_{\mathrm{RX}}}{10}}\right]^{-1} \tag{5.71}$$

where values for XPI are in dB.

Example

Assuming the same power transmitted on both polarisations, $P_x = P_y$, the same path loss for both polarisations, $C_x = C_y$, and the same cross polarisation isolation for the transmitting and the

Table 5.10 Components of wanted and interference carrier power and resulting carrier-to-cross polarisation interference power ratio (enter values for XPI in dB)

	X polarisation	Y polarisation
Transmitted power (W)	P_x	P_y
Cross polar transmitted power (W)	$Q_x = P_y 10^{-(\mathrm{XPI_{TX}}/10)}$	$Q_y = P_x 10^{-(\mathrm{XPI_{TX}}/10)}$
Received carrier power (W)	C_x	C_y
Generated cross polar interference (W)	$I_x = C_y 10^{-(\mathrm{XPI_{RX}}/10)}$	$I_y = C_x 10^{-(\mathrm{XPI_{RX}}/10)}$
Received co-polar interference (W)	$J_x = Q_x C_x / P_x$	$J_y = Q_y C_y / P_y$
Carrier-to-cross polarisation interference power ratio	$C_x/(I_x + J_x)$	$C_y/(I_y + J_y)$

receiving antenna, $XPI_{TX} = XPI_{RX} = 27$ dB, then the carrier-to-cross-polarisation interference power ratio for the considered link is:

$$\frac{C}{N_i} = \frac{C_x}{I_x + J_x} = \frac{C_y}{I_y + J_y} = \{10^{-2.7} + 10^{-2.7}\}^{-1} = 250.6 = 24 \text{ dB}$$

However, it may be that the power transmitted by the interfering station is much higher than the power of a transmitting VSAT. Assume, for instance, that the interfering station transmits on the Y polarisation with an EIRP 10 dB higher than that of a VSAT on its inbound X-polarised link, then $P_y/P_x = C_y/C_x = 10$ and C/N_i reduces to 14 dB. Similar considerations apply to the downlink.

The above calculations do not include depolarisation of the wave on its way from the transmitter to the receiver, which occurs in the atmosphere in the presence of rain, or ice clouds. Rain induced depolarisation results from differential attenuation and differential phase shift between two characteristic orthogonal polarisations. These effects originate in the non-spherical shape of raindrops.

The relationship between cross polarised discrimination XPD (see Appendix 4 for definition) and the co-polarised path attenuation A_{RAIN} is important for predictions based on attenuation statistics (as in Figures 5.16 and 5.17).

The cross polarisation discrimination XPD (p) not exceeded for p% of the time is given by:

$$XPD(p) = XPD_{rain} - C_{ice} \quad (\text{dB}) \tag{5.72}$$

where XPD_{rain} is the cross polarisation discrimination due to rain and C_{ice} is the contribution of ice clouds, respectively given by:

$$XPD_{rain} = C_f - C_A + C_\tau + C_\theta + C_\sigma \quad (\text{dB}) \tag{5.73}$$

$$C_{ice} = \frac{XPD_{rain}(0.3 + 0.1 \log p)}{2} \quad (\text{dB}) \tag{5.74}$$

where:

$$C_f = 30 \log f$$

$$C_A = V(f) \log A_{RAIN}(p)$$

$$V(f) = 12.8 f^{0.19} \text{ for } 8 \leq f \leq 20 \quad (\text{GHz})$$

$$V(f) = 22.6 \text{ for } 20 \leq f \leq 35 \quad (\text{GHz})$$

$$C_\tau = -10 \log[1 - 0.484(1 + \cos 4\tau)]$$

where f is the frequency (GHz) and τ is the tilt angle of the linearly polarised electric field vector with respect to the horizontal (for circular polarisation use $\tau = 45^\circ$).

$$C_\theta = -40\log(\cos E) \text{ for } E \leq 60^\circ$$

where E is the elevation angle.

$$C_\sigma = 0.0052\sigma^2$$

where σ is the standard deviation of the raindrop inclination angle distribution, expressed in degrees; σ takes the values 0°, 5°, 10° and 15° for $p = 1\%$, 0.1%, 0.01% and 0.001% of the time, respectively.

Expression (5.73) is in agreement with long term measurements for $8\,\text{GHz} \leq f \leq 35\,\text{GHz}$ and elevation angle $E \leq 60^\circ$. For lower frequencies down to 4 GHz, one can calculate $\text{XPD}_1(p)$ at frequency $f_1(8\,\text{GHz} \leq f_1 \leq 30\,\text{GHz})$ according to expression (5.73) and derive $\text{XPD}_2(p)$ at frequency $f_2(4\,\text{GHz} \leq f_2 \leq 8\,\text{GHz})$ from the following semi-empirical formula:

$$\text{XPD}_2(p) = \text{XPD}_1(p) - 20\log\left[\frac{f_2[1 - 0.484(1 + \cos 4\tau_2)]^{0.5}}{f_1[1 - 0.484(1 + \cos 4\tau_1)]^{0.5}}\right]$$

where τ_1 and τ_2 are the respective polarisation tilt angles at frequencies f_1 and f_2.

Snow (dry or wet) causes similar phenomena. Ice clouds, where high altitude ice crystals are in a region close to the 0°C isotherm, also contribute to cross polarisation. However, in contrast to rain and other hydrometeors, this effect is not accompanied by attenuation. It causes a reduction in the value of XPD given by (5.74)

When all sources of depolarisation are included and assuming equally powered carriers on both polarisations, the carrier-to-cross polarisation interference power ratio for a given link (uplink or downlink) can be calculated as:

$$\frac{C}{N_\text{i}} = \Delta\{10^{-(\text{XPI}_\text{TX}/10)} + 10^{-(\text{XPD}/10)} + 10^{-\text{XPI}_\text{RX}/10}\}^{-1} \qquad (5.75)$$

where $\Delta = C_\text{x}/C_\text{y}$ is the ratio of the co-polar wanted carrier power to the cross polar interfering carrier power (assuming the same path loss for both carriers).

The overall link (from station to station) carrier-to-cross polarisation interference ratio $(C/N_\text{i})_\text{T}$ is obtained by adding uplink and downlink interference power at the earth station receiver input:

$$\left(\frac{C}{N_\text{i}}\right)_\text{T}^{-1} = \left(\frac{C}{N_\text{i}}\right)_\text{U}^{-1} + \left(\frac{C}{N_\text{i}}\right)_\text{D}^{-1} \qquad (5.76)$$

where the absolute values of $(C/N_\text{i})_\text{U}$ and $(C/N_\text{i})_\text{D}$, each calculated according to expression (5.75), should be used.

For any link, the carrier power-to-cross polarisation interference power spectral density ratio (C/N_{0i}) is given by:

$$\frac{C}{N_{0i}} = \frac{C}{N_i} B_N \quad \text{(Hz)} \tag{5.77}$$

$$\frac{C}{N_{0i}} \text{ (dBHz)} = \frac{C}{N_i} \text{ (dB)} + 10 \log B_N$$

5.5.3.2 Adjacent channel interference

Adjacent channel interference (ACI) originates in a part of the power of a carrier adjacent to a given carrier being captured by a satellite transponder or an earth station receiver tuned to the frequency of the carrier considered. Figure 5.30 illustrates the situation where ACI

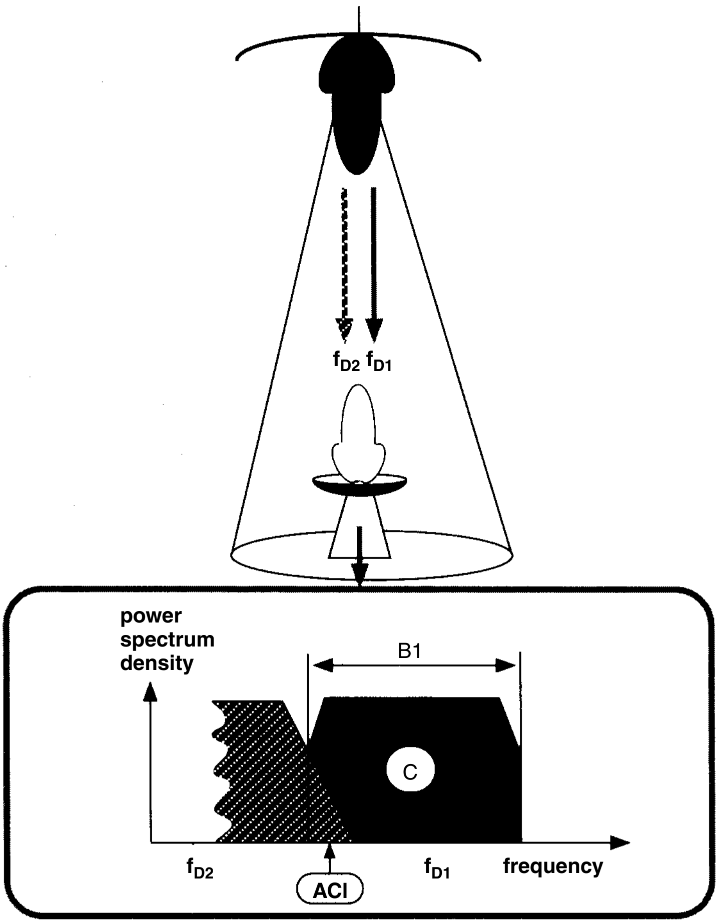


Figure 5.30 Downlink adjacent channel interference (ACI)

affects a downlink carrier. Part of the spectrum of the adjacent carrier at frequency f_{D2} falls into the earth station receiver tuned to the carrier at frequency f_{D1}. The corresponding carrier-to-interference power ratio is determined by the area labelled C with respect to the area labelled ACI. Typical values of C/ACI for VSAT networks are in the range 25 to 30 dB. Adjacent channel interference can be reduced by adopting larger guard bands between carriers, at the expense of a larger utilised bandwidth on the satellite transponder.

5.5.4 External interference

External interference originates outside the considered satellite system: transmissions by other satellites and stations, be it earth stations or terrestrial ones (microwave relays on the earth). Figure 1.17 illustrates the geometry of external interference. Actually, VSAT networks operating at Ku-band can be operated in the exclusive bands of Figure 1.15, and this protects those networks from microwave relay interference.

5.5.4.1 Interference from adjacent satellite systems

The interference generated by an earth station into an adjacent satellite, or by an adjacent satellite into an earth station, depends on their respective antenna radiation patterns. Actual antenna patterns should be considered but they may not be known at the early planning stages of a VSAT network. Therefore, reference patterns may be used instead. For VSATs and hub station antennas, one can refer to the ITU-R Recommendation S672. Figure 5.31 displays a typical VSAT antenna radiation pattern and a reference sidelobe pattern. All reference patterns specify a sidelobe envelope level of the form $A - B \log \theta$ (dBi) where θ is the off-axis angle. In Figure 5.31, $A = 29$ dBi for $2.5^\circ \leq \theta < 7^\circ$ and $A = 32$ dBi for $9.2^\circ \leq \theta < 48^\circ$, and $B = 25$.

In order to calculate the level of interference encountered at the receiver end of a satellite link, it is necessary to know the angular separation between two geostationary satellites as seen by an earth station. Figure 5.32 displays the geometry where the satellites are separated in longitude by an angle α. The angle θ represents the angular separation between the two geostationary satellites as seen by the earth station.

The distance d between the two satellites is given by:

$$d^2 = R_w^2 + R_a^2 - 2R_w R_a \cos\theta \qquad (\mathrm{m}^2) \qquad (5.78)$$

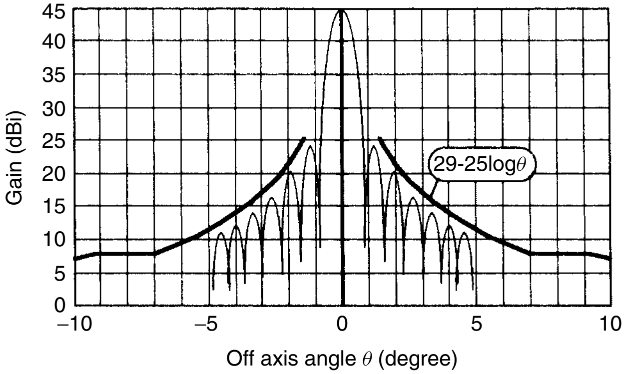


Figure 5.31 Typical VSAT antenna radiation pattern and sidelobe reference pattern

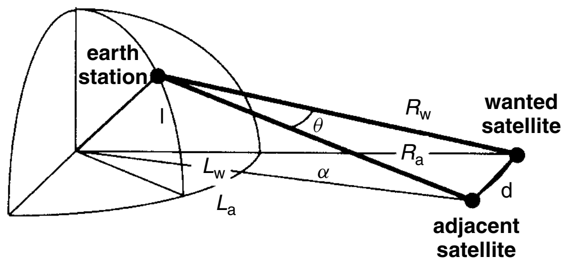


Figure 5.32 Geometry of earth station with respect to the wanted and the adjacent satellite

where:

R_w is the slant range from the earth station to the wanted satellite,

R_a is the slant range from the earth station to the adjacent satellite.

On the other hand, $d/2 = (R_e + R_0)\sin(\alpha/2)$, resulting in:

$$d^2 = 4(R_e + R_0)^2 \sin^2(\alpha/2) = 2(R_e + R_0)^2(1 - \cos\alpha) \qquad (\mathrm{m}^2) \tag{5.79}$$

where:

R_e is the earth radius: $R_e = 6378$ km

R_0 is the satellite altitude: $R_0 = 35\,786$ km

Combining (5.78) and (5.79) leads to:

$$\theta = \mathrm{arc}\cos\left[\frac{R_w^2 + R_a^2 - 2(R_e + R_0)^2(1 - \cos\alpha)}{2R_w R_a}\right] \quad \text{(degrees)} \tag{5.80}$$

Table 5.11 θ/α ratio for $\alpha = 4^\circ$ orbital separation and various values of the earth station latitude l and relative longitude L_w with respect to the wanted satellite

Latitude l	Relative longitude L_w		
	0°	10°	20°
0°	1.18	1.17	1.16
10°	1.17	1.17	1.15
20°	1.16	1.16	1.15
30°	1.15	1.14	1.13
40°	1.12	1.12	1.11
50°	1.10	1.10	1.09

R_w and R_a are given by expression (2.8), and depend on the latitude l of the earth station and the relative longitudes L_w and L_a with respect to the wanted and the adjacent satellite respectively. Note that $L_a - L_w = \alpha$. Table 5.11 provides some values of the θ/α ratio for an orbital separation of 4° and different locations of the earth station.

It can be seen that for a practical situation where the earth station is not too far away from the wanted satellite, the θ/α ratio remains in the range 1.1 to 1.18. Therefore, a practical rule of thumb is:

$$\theta = 1.15\alpha \tag{5.81}$$

Uplink interference analysis

The uplink interference deals with the case where a satellite transponder receives a wanted carrier from an earth station located within the coverage of its receiving antenna and some carrier power at the same frequency from an interfering station normally transmitting to an adjacent satellite, as illustrated in Figure 5.33. The worst case is assumed, i.e. the wanted earth station is located at the edge of coverage of the satellite it transmits to, and the interfering earth station is located at the centre of coverage.

From (5.59) and (5.60), the expression for carrier-to-interference power ratio $(C/N_i)_U$ is:

$$\left(\frac{C}{N_i}\right)_U = \frac{\text{EIRP}_{\text{ESw}}}{\text{EIRP}_{\text{ESi}}} \frac{B_i}{\min[B_i, B_N]} \frac{L_{\text{Ui}}}{L_{\text{Uw}}} \frac{G_{\text{RXw}}}{G_{\text{RXi}}} \tag{5.82}$$

$$\begin{aligned}\left(\frac{C}{N_i}\right)_U (\text{dB}) = {} & \text{EIRP}_{\text{ESw}}\ (\text{dBW}) - \text{EIRP}_{\text{ESi}}\ (\text{dBW}) + 10\log B_i \\ & - 10\log\min[B_i, B_N] + L_{\text{Ui}}\ (\text{dB}) - L_{\text{Uw}}\ (\text{dB}) \\ & + G_{\text{RXw}}\ (\text{dBi}) - G_{\text{RXi}}\ (\text{dBi})\end{aligned}$$

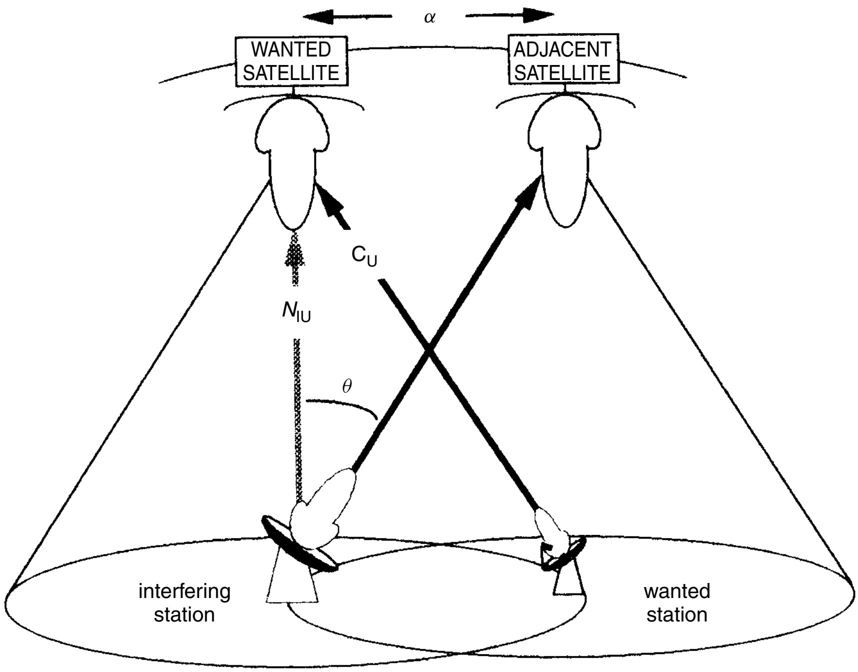


Figure 5.33 Uplink interference from adjacent satellite systems

where $\mathrm{EIRP}_{\mathrm{ESw}}$ is the EIRP for the wanted carrier in the direction of the wanted satellite. $\mathrm{EIRP}_{\mathrm{ESi}}$ is the EIRP of the interfering carrier in the direction of the wanted satellite. L_{Uw} and L_{Ui} are the uplink path loss for the wanted carrier and the interfering one, respectively. G_{RXw} and G_{RXi} are the satellite composite receive gain values given by (5.61) in the direction of the wanted and interfering earth stations, respectively.

The antenna radiation pattern of the interfering earth station is such that:

$$\mathrm{EIRP}_{\mathrm{ESi}}\,(\mathrm{dBW}) = \mathrm{EIRP}_{\mathrm{ESi,max}}\,(\mathrm{dBW}) - G_{\mathrm{Ti,max}}\,(\mathrm{dBi}) + 29 - 25\log\theta \tag{5.83}$$

where $\mathrm{EIRP}_{\mathrm{ESi,max}}$ is the maximum value of the EIRP allocated to the interfering carrier of bandwidth B_{i} by the interfering earth station with maximum transmitting antenna gain $G_{\mathrm{Ti,max}}$.

Assuming limited geographical area: L_{Ui} (dB) = L_{Uw} (dB). Considering the worst case, where the wanted earth station is at the edge of coverage and the interfering one at the centre of coverage:

$$G_{\mathrm{SLw}}\,(\mathrm{dBi}) = G_{\mathrm{SLi}}\,(\mathrm{dBi}) - 3\ \mathrm{dB}$$

Putting $\theta = 1.15\alpha$ according to (5.81), the expression for $(C/N_{\mathrm{i}})_{\mathrm{U}}$ is:

$$\left(\frac{C}{N_{\mathrm{i}}}\right)_{\mathrm{U}} (\mathrm{dB}) = \mathrm{EIRP}_{\mathrm{ESw}}\,(\mathrm{dBW}) - \mathrm{EIRP}_{\mathrm{ESi,max}}\,(\mathrm{dBW}) + G_{\mathrm{Ti,max}}\,(\mathrm{dBi}) - 32 + 25\log(1.15\alpha) + 10\log B_{\mathrm{i}} - 10\log\min[B_{\mathrm{i}}, B_{\mathrm{N}}] \quad (5.84)$$

The corresponding carrier to interference spectral power density ratio $(C/N_{0\mathrm{i}})_{\mathrm{U}}$ is given by:

$$\left(\frac{C}{N_{0\mathrm{i}}}\right)_{\mathrm{U}} (\mathrm{dBHz}) = \left(\frac{C}{N_{\mathrm{i}}}\right)_{\mathrm{U}} + 10\log B_{\mathrm{N}} \quad (5.85)$$

Example

Consider a satellite angular separation $\alpha = 4^{\circ}$. The wanted station is a VSAT (1.2 m antenna, operating at 14 GHz) with an $\mathrm{EIRP}_{\mathrm{ESw}} =$ 40 dBW. The receiver noise bandwidth is $B_{\mathrm{N}} = 100$ kHz. The interfering earth station is a 5 m antenna station with $G_{\mathrm{Ti,max}} = 55$ dBi and $\mathrm{EIRP}_{\mathrm{ESi,max}} = 70$ dBW. The interfering carrier bandwidth is $B_{\mathrm{i}} = 2$ MHz. From (5.82), $(C/N_{\mathrm{i}})_{\mathrm{U}} = 23$ dB.

Downlink interference analysis

Figure 5.34 illustrates the worst case for downlink interference.

From (5.59) and (5.60), the expression for carrier-to-interference power ratio $(C/N_{\mathrm{i}})_{\mathrm{D}}$ is:

$$\left(\frac{C}{N_{\mathrm{i}}}\right)_{\mathrm{D}} = \frac{\mathrm{EIRP}_{\mathrm{SLww}}}{\mathrm{EIRP}_{\mathrm{SLiw}}}\,\frac{B_{\mathrm{i}}}{\min[B_{\mathrm{i}}, B_{\mathrm{N}}]}\,\frac{L_{\mathrm{Di}}}{L_{\mathrm{Dw}}}\,\frac{G_{\mathrm{RXw}}}{G_{\mathrm{RXi}}}$$

$$\left(\frac{C}{N_{\mathrm{i}}}\right)_{\mathrm{D}} (\mathrm{dB}) = \mathrm{EIRP}_{\mathrm{SLww}}\,(\mathrm{dBW}) - \mathrm{EIRP}_{\mathrm{SLiw}}\,(\mathrm{dBW}) + 10\log B_{\mathrm{i}} - 10\log\min[B_{\mathrm{i}}, B_{\mathrm{N}}] + L_{\mathrm{Di}}\,(\mathrm{dB}) - L_{\mathrm{Dw}}\,(\mathrm{dB}) + G_{\mathrm{RXw}}\,(\mathrm{dBi}) - G_{\mathrm{RXi}}\,(\mathrm{dBi}) \quad (5.86)$$

where $\mathrm{EIRP}_{\mathrm{SLww}}$ is the wanted satellite EIRP for the wanted carrier in the direction of the wanted station. $\mathrm{EIRP}_{\mathrm{SLiw}}$ is the interfering satellite EIRP for the interfering carrier in the direction of the wanted station. L_{Dw} and L_{Di} are the uplink path loss for the wanted carrier and the interfering one, respectively. G_{RXw} and G_{RXi} are the wanted station composite receive gain values given by (5.61) in the direction of the wanted and interfering satellites, respectively.

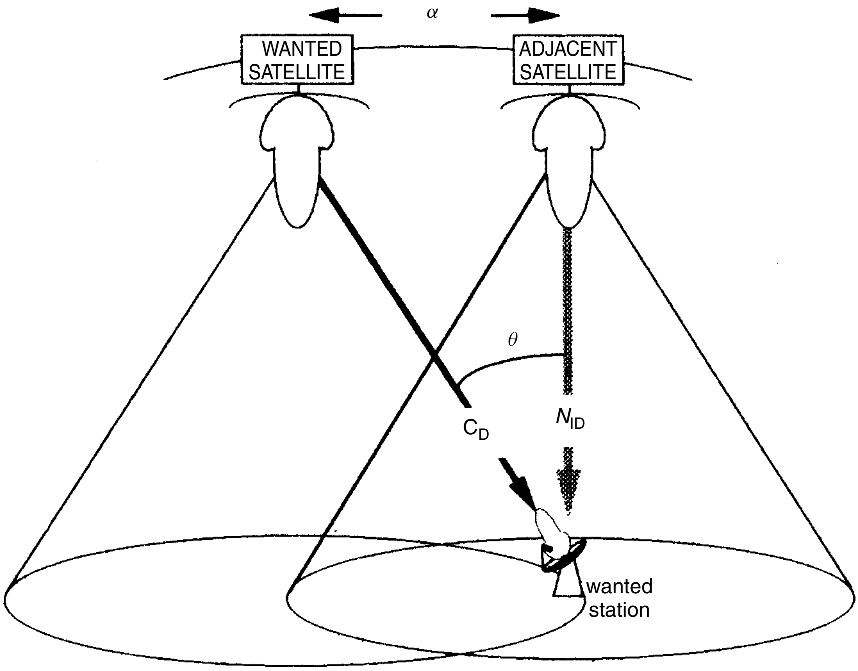


Figure 5.34 Downlink interference from adjacent satellite systems

For a wanted station at the edge of coverage of the wanted satellite and at the centre of coverage of the interfering one (worst case):

$$\mathrm{EIRP}_{\mathrm{SLw}}\ (\mathrm{dBW}) = \mathrm{EIRP}_{\mathrm{SLw,max}}\ (\mathrm{dBW}) - 3\ \mathrm{dB}$$

$$\mathrm{EIRP}_{\mathrm{SLi}}\ (\mathrm{dBW}) = \mathrm{EIRP}_{\mathrm{SLi,max}}\ (\mathrm{dBW})$$

Assuming limited geographical area: $L_{\mathrm{Di}}\ (\mathrm{dB}) = L_{\mathrm{Dw}}(\mathrm{dB})$. The antenna radiation pattern of the wanted station is such that:

$$G_{\mathrm{RXi}}\ (\mathrm{dBi}) = G_{\mathrm{RXw}}\ (\mathrm{dBi}) - G_{\mathrm{RXmax}}\ (\mathrm{dBi}) + 29 - 25\log\theta\ (\mathrm{dB}) \quad (5.87)$$

where G_{RXmax} (dBi) is the maximum receive composite gain of the wanted station.

Putting $\theta = 1.15\alpha$ according to (5.81), the expression for $(C/N_{\mathrm{i}})_{\mathrm{D}}$ is:

$$\left(\frac{C}{N_{\mathrm{i}}}\right)_{\mathrm{D}} (\mathrm{dB}) = \mathrm{EIRP}_{\mathrm{SLw,max}}\ (\mathrm{dBW}) - \mathrm{EIRP}_{\mathrm{SLi,max}}\ (\mathrm{dBW}) + 10\log B_{\mathrm{i}} - 10\log\min[B_{\mathrm{i}}, B_{\mathrm{N}}] + G_{\mathrm{RXmax}}\ (\mathrm{dBi}) - 32 + 25\log(1.15\alpha) \quad (5.88)$$

The corresponding carrier-to-interference spectral power density ratio $(C/N_{0i})_D$ is given by:

$$\left(\frac{C}{N_{0i}}\right)_D \text{(dBHz)} = \left(\frac{C}{N_i}\right)_D + 10\log B_N \qquad (5.89)$$

Example

Consider a satellite angular separation $\alpha = 4°$. The wanted station is a VSAT (1.2 m antenna, operating at 12 GHz) with $G_{RXmax} =$ 41 dBi, receiving the outbound link from the wanted satellite with $EIRP_{SLw,max}$ (dBW) = 25 dBW. The receiver noise bandwidth is $B_N = 500$ kHz. The interfering satellite transmits a TV carrier with $EIRP_{SLi,max}$ (dBW) = 52 dBW and bandwidth = 26 MHz. From (5.88), $(C/N_i)_D = 15.7$ dB.

5.5.4.2 Terrestrial interference

Bands allocated to satellite communications are most often also allocated to terrestrial microwave links. The only restricted frequency bands are available at Ku-band and Ka-band for exclusive use for satellite communications (see Figure 1.17). In those bands, VSAT networks are protected against terrestrial interference. At C-band, terrestrial interference is a major concern to VSAT networks. Indeed, the smaller the earth station antenna and the lower the frequency, the larger the beamwidth. Therefore small VSATs are sensitive to interference, especially at C-band. For VSATs being used for business applications and installed at the user's premises, possibly in city centres, terrestrial interference at C-band may be important. The most convenient way to protect the network from interference is to use spread spectrum communications. The calculation of the interference in a city environment is quite complex, and proceeding to site measurements before installing VSATs and hub would be recommended.

5.5.5 Conclusion

As a result of its small size, the VSAT antenna provides limited interference rejection. Terrestrial interference can be avoided by using exclusive bands at 14/12 GHz and 30/20 GHz. Adjacent satellite interference still remains. Should the adjacent satellites transmit carriers with high EIRP on the frequency of the outbound link, then the VSATs may receive an unacceptable level of interference.

The inbound links, as a result of the low EIRP of the transmitting VSATs, are received at the input of the satellite transponder with low power. Should other stations transmit on the orthogonal polarisation, the uplink part of the inbound link may suffer from cross-polarisation interference, especially if those stations are not properly aligned with the satellite antenna polarisation. The downlink part of the inbound link will also be contaminated by the cross polarisation interference generated by the carriers conveyed by the cross-polar transponder. It is therefore desirable, when planning a VSAT network, to check the occupancy status of the cross-polar transponder.

Rain or ice induced cross-polar interference may contribute in a sensitive way to the overall level of interference a small percentage of the time. This is to be considered when addressing link availability.

5.6 OVERALL LINK PERFORMANCE

The overall link is that from station to station. It comprises two portions: the uplink from the transmitting station to the satellite, and the downlink from the satellite to the receiving station. It is considered here that uplinked carriers are amplified and frequency converted by the satellite transponder before being transmitted on the downlink (no on-board demodulation and remodulation).

The overall carrier power-to-noise power spectral density ratio at the earth station receiver input $(C/N_0)_T$ measures the overall link performance. Table 5.12 provides a summary of the main equations to be used for evaluating $(C/N_0)_T$. The meaning of each of the notations is given in Table 5.13.

Table 5.12 Summary of main equations for evaluating the overall link carrier power-to-noise power spectral density ratio $(C/N_0)_T$

Equation	Unit	Source equation
Overall link quality		
$(C/N_0)_T^{-1} = (C/N_0)_U^{-1} + (C/N_0)_D^{-1} + (C/N_0)_{IM}^{-1} + \Sigma(C/N_{0i})_U^{-1} + \Sigma(C/N_{0i})_D^{-1}$ (all terms in absolute units, not dB)	Hz	(5.7)
Uplink calculations		
$(C/N_0)_U$ (dBHz) = IBO_1 (dB) + $(C/N_0)_{Usat}$ (dBHz)	dBHz	(5.10)
IBO_1 (dB) = Φ (dBWm^{-2}) − Φ_{sat} (dBWm^{-2})		(5.15)
Φ (dBWm^{-2}) = $EIRP_{ES}$ (dBW) + G_1 (dB) − L_U (dB)	dBHz	(5.14)
$(C/N_0)_{Usat}$ (dBHz) = Φ_{sat} (dBWm^{-2}) − G_1(dBi) + $(G/T)_{SL}$ (dBK^{-1}) + 228.6	dBHz	(5.11)
G_1 (dBi) = $10\log 4\pi + 20 log(f_U/c)$	dBi	(5.12)

Table 5.12 *(continued)*

Equation	Unit	Source equation
Transponder transfer calculations		
Φ_t (Wm^{-2}) = $\Sigma\Phi_i$ (Wm^{-2}) $i = 1, 2, \ldots, N$ (all terms in absolute units, not dB)	Wm^{-2}	(5.16)
IBO_t (dB) = Φ_t ($dBWm^{-2}$) − Φ_{sat} ($dBWm^2$)	dB	(5.16)
OBO_t (dB) = 0 if $-5\,dB < IBO_t < 0\,dB$; 0.9 (IBO_t(dB) + 5) if $IBO_t < -5\,dB$		(A6.10)
Downlink calculations		
$(C/N_0)_D$ (dBHz) = OBO_1 (dB) + $(C/N_0)_{Dsat}$ (dBHz)	dBHz	(5.35)
OBO_1 (dB) = 0.9 (IBO_1 (dB) + 5) if $IBO_t < -5\,dB$ (assumes linear transfer)		(A6.10)
$(C/N_0)_{Dsat}$ (dBHz) = $EIRP_{SLsat}$ (dBW) − L_D (dB) + $(G/T)_{ES}$ (dBK^{-1}) + 228.6	dBHz	(5.36)
Intermodulation calculations		
$(C/N_0)_{IM}$ (dBHz) = 79 − 10 log n − 1.65 (IBO_t (dB) + 5) if $IBO_t < -5\,dB$	dBHz	(5.55)
IBO_t = IBO_1 (dB) + 10 log n where n is the number of equally powered carriers		(A6.9)
Co-channel interference calculations		
$(C/N_{0i})_U$ (dB) = $EIRP_{ESw}$ (dBW) − $EIRP_{ESi}$ (dBW) + 10 log B_i − 10 log min[B_i, B_N] + L_{Ui} (dB) − L_{Uw} (dB) + G_{RXw} (dBi) − G_{RXi} (dBi)	dB	(5.62)
$(C/N_{0i})_U$ (dBHz) = $(C/N_i)_U$ (dB) + 10 log B_N	dBHz	(5.69)
$(C/N_i)_D$ (dB) = $EIRP_{SL1ww}$ (dBW) − $EIRP_{SL2iw}$ (dBW) + 10 log B_i − 10 log min[B_i, B_N]	dB	(5.63)
$(C/N_{0i})_D$ (dBHz) = $(C/N_i)_D$ (dB) + 10 log B_N	dBHz	(5.69)
Cross polarisation interference		
$(C/N_i)_U = \Delta_U\{10^{-(XPI_{TX}/10)} + 10^{-(XPD/10)} + 10^{-(XPI_{RX}/10)}\}^{-1}$		(5.75)
$(C/N_{0i})_U$ (dBHz) = $(C/N_i)_U$ (dB) + 10 log B_N	dBHz	(5.77)
$(C/N_i)_D = \Delta_D\{10^{-(XPI_{TX}/10)} + 10^{-(XPD/10)} + 10^{-(XPI_{RX}/10)}\}^{-1}$		(5.75)
$(C/N_{0i})_D$ (dBHz) = $(C/N_i)_D$ (dB) + 10 log B_N	dBHz	(5.77)
Adjacent satellite interference		
$(C/N_i)_U$ (dB) = $EIRP_{ESw}$ (dBW) − $EIRP_{ESi,max}$ (dBW) + $G_{Ti,max}$ (dBi) − 32 + 25 log(1.15α) + 10 log B_i − 10 log min[B_i, B_N]	dB	(5.84)
$(C/N_{0i})_U$ (dBHz) = $(C/N_i)_U$ (dB) + 10 log B_N	dBHz	(5.85)
$(C/N_i)_D$ (dB) = $EIRP_{SLw,max}$ (dBW) − $EIRP_{SLi,max}$ (dBW) + 10 log B_i − 10 log min[B_i, B_N] + G_{RXmax} (dBi) − 32 + 25 log(1.15α)	dB	(5.88)
$(C/N_{0i})_D$ (dBHz) = $(C/N_i)_D$ + 10 log B_N	dBHz	(5.89)

Table 5.13 Meaning of symbols displayed in Table 5.12

Symbol	Meaning	Unit	Source equation
B_N	Receiver noise equivalent bandwidth	Hz	(5.57)
B_i	Interfering carrier bandwidth	Hz	(5.58)
$(C/N_0)_D$	Downlink carrier power-to-noise power spectral density ratio	Hz	(5.35)
$(C/N_0)_{Dsat}$	Same as above, at saturation	Hz	(5.36)
$(C/N_{0i})_D$	Downlink carrier power-to-interference noise power spectral density ratio	Hz	(5.89)
$(C/N_0)_{IM}$	Carrier power-to-intermodulation power spectral density ratio	Hz	(5.55)
$(C/N_{0i})_U$	Uplink carrier power-to-interference noise power spectral density ratio	Hz	(5.85)
$(C/N_0)_T$	Overall link (from station to station) carrier power-to-total noise power spectral density ratio	Hz	(5.7)
$(C/N_0)_U$	Uplink carrier power-to-noise power spectral density ratio	Hz	(5.10)
$(C/N_0)_{Usat}$	Same as above, at saturation	Hz	(5.11)
$EIRP_{ES}$	EIRP of earth station	W	(5.17)
$EIRP_{ESi}$	EIRP of interfering earth station	W	(5.17)
$EIRP_{ESw}$	EIRP of wanted earth station	W	(5.17)
$EIRP_{SLsat}$	EIRP of satellite transponder at saturation	W	(5.37)
$EIRP_{SL1ww}$	EIRP of satellite in beam 1 for wanted carrier in direction of wanted station	W	(5.37)
$EIRP_{SL2iw}$	EIRP of satellite in beam 2 for the interfering carrier in the direction of the wanted station	W	(5.37)
G_{RXi}	Satellite composite receive gain of beam 1 for interfering carrier		(5.67)
G_{RXw}	Satellite composite receive gain of beam 1 for wanted carrier		(5.67)
G_1	Gain of an ideal antenna with area equal to 1 m^2		(5.12)
$(G/T)_{ES}$	Figure of merit of earth station receiving equipment	K^{-1}	(5.42)
$(G/T)_{SL}$	Figure of merit of satellite receiving equipment	K^{-1}	(5.31)
IBO_t	Total input back-off		(5.16)
IBO_1	Input back-off per carrier		(5.15)
L_U	Uplink path loss		(5.27)
L_{Ui}	Uplink path loss for interfering carrier		(5.27)
L_{Uw}	Uplink path loss for wanted carrier		(5.27)
L_D	Downlink path loss		(5.41)
OBO_t	Total output back-off with multicarrier operation mode		(A6.6)
OBO_1	Output back-off per carrier with multicarrier operation mode		(A6.10)
XPD	Cross polar discrimination		(A4.6)
XPI_{RX}	Receive antenna cross polarisation isolation		(A4.5)
XPI_{TX}	Transmit antenna cross polarisation isolation		(A4.5)
Δ	Ratio of co-polar wanted carrier power to cross-polar interfering carrier power		(5.75)
Φ_{sat}	Power flux density at saturation (Wm^{-2}) for satellite transponder	Wm^{-2}	(5.11)(5.14)

5.7 BIT ERROR RATE DETERMINATION

The overall radio frequency link performance is given by the value of $(C/N_0)_T$, which can be calculated from equation (5.7) introducing the parameters discussed in the previous sections.

The discussion will now focus on ways to derive the quality of the signal delivered to the user terminal, that is the bit error rate (BER).

The BER is a function of the dimensionless ratio E_b/N_0 where E_b is the energy per information bit and N_0 the overall link noise power spectral density, i.e. $N_0 = N_{0T}$ as expressed by equation (5.5). The energy per information bit is defined as the energy accumulated at the receiver from reception of carrier power C_D during a time interval equal to the time it takes to receive an information bit. At information bit rate R_b, the time it takes to receive an information bit is $T_b = 1/R_b$. Therefore:

$$E_b = \frac{C_D}{R_b} \quad \text{(J)} \tag{5.90}$$

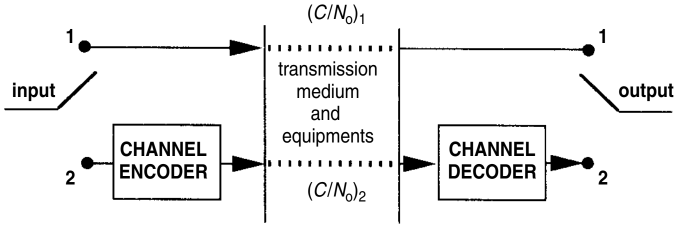
and

$$\begin{aligned} \frac{E_b}{N_0} &= \frac{C_D/R_b}{N_{0T}} = \frac{(C/N_0)_T}{R_b} \\ \frac{E_b}{N_0}\,\text{(dB)} &= \left(\frac{C}{N_0}\right)_T \text{(dBHz)} - 10\log R_b \end{aligned} \tag{5.91}$$

The relationship between E_b/N_0 and BER depends on the type of modulation and forward error correction (FEC) scheme used. FEC techniques are based on transmitting redundant bits along with the information bits, in such a way as to allow the receiver to decode and correct some of the erroneous bits. These techniques are presented in [MAR02, p 169–182]. It must be made clear that, contrary to the ARQ protocols presented in Chapter 4, section 4.5, FEC does not guarantee error-free transmission. FEC only reduces the BER, given $(C/N_0)_T$, at the expense of a larger bandwidth.

Figure 5.35 displays a comparison between an uncoded link and a coded one using BPSK or QPSK modulation. Curve 1 represents the uncoded link performance and curves 2 the coded ones. It is assumed here that both links deliver data to the user at the same rate R_b and the same BER. The coding gain G_{cod} is defined as the E_b/N_0 saving introduced by the coding scheme at the considered BER, as illustrated in Figure 5.35 for a BER of 10^{-6} and a code rate $\rho = 7/8$:

$$G_{cod}\,\text{(dB)} = \left(\frac{E_b}{N_0}\right)_1 \text{(dB)} - \left(\frac{E_b}{N_0}\right)_2 \text{(dB)} \tag{5.92}$$



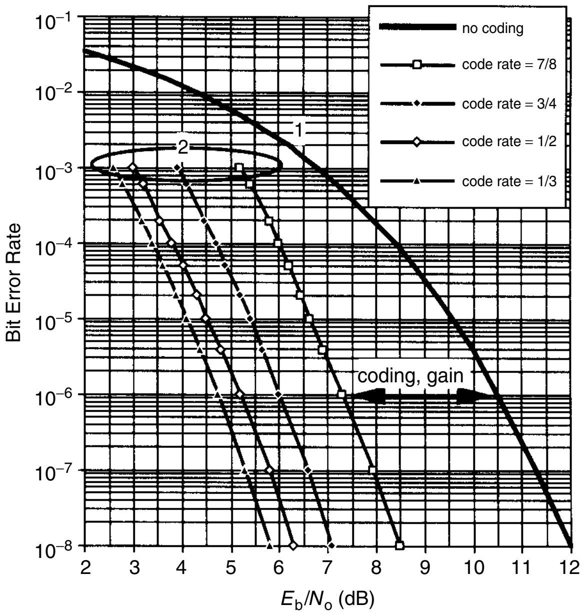


Figure 5.35 Comparison between uncoded and coded links for convolutional encoding and typical Viterbi decoder performance

This saving translates into a reduction in the required $(C/N_0)_T$:

$$\text{No coding:}\quad \left(\frac{C}{N_0}\right)_{T1} \text{(dBHz)} = \left(\frac{E_b}{N_0}\right)_1 \text{(dB)} + 10\log R_b \quad (5.93)$$

$$\text{With coding:}\quad \left(\frac{C}{N_0}\right)_{T2} \text{(dBHz)} = \left(\frac{E_b}{N_0}\right)_2 \text{(dB)} + 10\log R_b \quad (5.94)$$

The reduction in $(C/N_0)_T$ is equal to:

$$\Delta\left(\frac{C}{N_0}\right)_T \text{(dB)} = \left(\frac{C}{N_0}\right)_{T1} \text{(dBHz)} - \left(\frac{C}{N_0}\right)_{T2} \text{(dBHz)}$$
$$= \left(\frac{E_b}{N_0}\right)_1 \text{(dB)} - \left(\frac{E_b}{N_0}\right)_2 \text{(dB)} = G_{cod} \quad (5.95)$$

The coding gain is higher with low values of code rate.

5.8 POWER VERSUS BANDWIDTH EXCHANGE

For a carrier conveying an information bit rate R_b, using forward error correction (FEC) with code rate ρ, and modulation with spectral efficiency Γ (ratio of bit rate to bandwidth), the required bandwidth is:

$$B = \frac{R_b}{\Gamma \rho} \qquad (5.96)$$

Hence, a larger bandwidth is required for low values of code rate, given the information bit rate R_b and the modulation scheme. However, with low code rate one benefits from a larger coding gain and therefore a lower requirement on E_b/N_0 which translates, according to equation (5.95), into a reduced requirement for the received carrier power. This suggests a potential power versus bandwidth exchange, which will now be demonstrated through a specific example.

Consider a VSAT transmitting at an information bit rate $R_b = 64\,\text{kbs}^{-1}$. BPSK modulation is used with spectral efficiency $\Gamma = 0.7\,\text{bs}^{-1}\text{Hz}^{-1}$. The required bit error rate BER $= 10^{-7}$. Table 5.14 gives the required values of E_b/N_0 depending on the code rate, as given by Figure 5.35, the corresponding values of $(C/N_0)_T$, according to formulas (5.93) and (5.94), and bandwidth as given by formula (5.96).

Figure 5.36 illustrates the power versus bandwidth exchange, which is experienced by changing the code rate. This flexible exchange is paramount to any design of a VSAT network.

5.9 EXAMPLE

The purpose of this section is to perform a typical dimensioning of radio frequency links according to the principles exposed in this chapter.

Figure 5.37 shows the architecture of the considered VSAT network.

Table 5.14 Impact of code rate on required values of radio frequency link parameters (BPSK, $R_b = 64\,\text{kbs}^{-1}$, BER $= 10^{-7}$)

Code rate	Required E_b/N_0	Required $(C/N_0)_T$	Required bandwidth
1	11.3 dB	59.4 dBHz	92 kHz
7/8	8 dB	56.1 dBHz	104 kHz
3/4	6.6 dB	54.7 dBHz	122 kHz
2/3	6.2 dB	54.3 dBHz	137 kHz
1/2	5.8 dB	54 dBHz	183 kHz

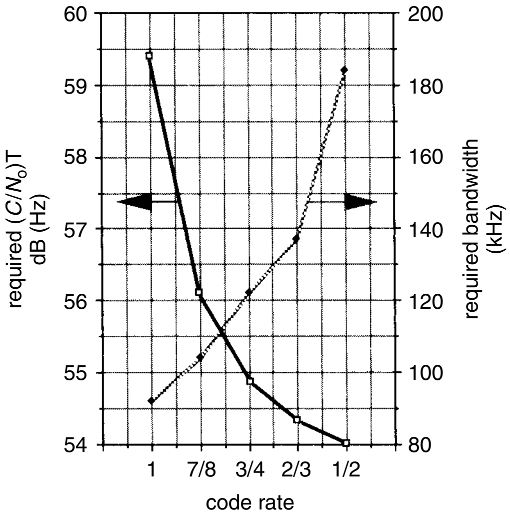


Figure 5.36 Illustration of the power versus bandwidth exchange for $R_b = 64\ \text{kbs}^{-1}$ and BER $= 10^{-7}$

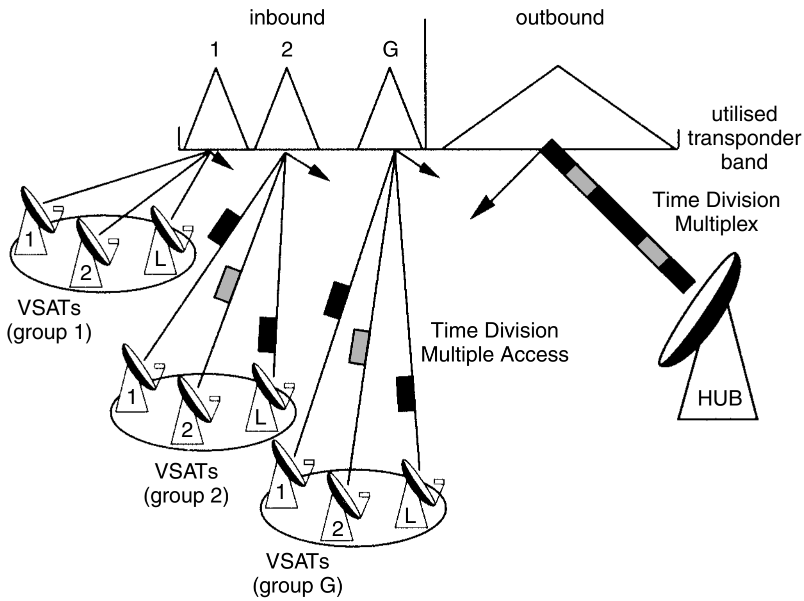


Figure 5.37 Architecture of the considered VSAT network

Uplink and downlink frequencies are all approximated respective to $f_U = 14.25$ GHz and $f_D = 12.7$ GHz. The service requirement is a BER $= 10^{-7}$, which entails a required $E_b/N_0 = 11.3$ dB without coding. The earth stations are assumed to be located so that the minimum elevation angle for all stations is $E_{min} = 35°$.

For instance, the earth station can be at medium latitude (about 45°) and within 20° of relative longitude with respect to the satellite.

Table 5.15 displays the network parameters; Table 5.16 gives the satellite, VSAT and hub characteristics; Table 5.17 indicates link parameter calculations for clear sky and rain conditions (the selected value of rain attenuation, $A_{RAIN} = 6$ dB, is exceeded for 0.01% of the time in Europe climatic region H, at an elevation angle $E = 35°$).

The network comprises $G = 3$ groups of $L = 33$ VSATs each, so that the total number of VSATs in the network is 99. Each group is being allocated a frequency band for its inbound links. The VSATs of a group access this frequency band in a reservation/random TDMA mode. Therefore, the inbound links are made of bursts of inbound carriers. These bursts convey information bits at a rate $R_{binb} = 64\ \text{kbs}^{-1}$. Coding with code rate $\rho_{inb} = 1/2$ is used, with coding gain 5.5 dB at the required BER $= 10^{-7}$. Therefore, the inbound transmitted bit rate within each burst is $R_{cinb} = R_{binb}/\rho_{inb} = 128\ \text{kbs}^{-1}$. The modulation is BPSK, with spectral efficiency $\Gamma = 0.7\ \text{bs}^{-1}.\text{Hz}^{-1}$. Therefore, according to (5.96) the utilised bandwidth per inbound link is $B_{inb} = 183$ kHz.

The outbound link is a time division multiplex (TDM) at information bit rate $R_{boutb} = 256\ \text{kbs}^{-1}$ with code rate $\rho_{outb} = 1/2$. The transmitted bit rate is $R_{coutb} = 512\ \text{kbs}^{-1}$. The modulation is BPSK.

Table 5.15 Network parameters

1	Type of access	FDMA/TDMA	
2	Number of groups of VSATs	G	3
3	Number of VSATs per group	L	33
4	Number of VSATs in network	N	99
5	Inbound information bit rate	R_{binb}	$64\ \text{kbs}^{-1}$
6	Inbound code rate	ρ_{inb}	0.5
7	Inbound transmission rate	R_{cinb}	$128\ \text{kbs}^{-1}$
8	Inbound modulation	BPSK	
9	Inbound modulation spectral efficiency	Γ	$0.7\ \text{bs.}^{-1}\text{Hz}^{-1}$
10	Outbound information bit rate	R_{boutb}	$256\ \text{kbs}^{-1}$
11	Outbound code rate	ρ_{outb}	0.5
12	Outbound transmission bit rate	R_{coutb}	$512\ \text{kbs}^{-1}$
13	Outbound modulation	BPSK	
14	Outbound modulation spectral efficiency	Γ	$0.7\ \text{bs.}^{-1}\text{Hz}^{-1}$
15	Percentage for guard bands		20%
16	Utilised transponder band		1536 kHz

Notes: Line 4 = Line 2 × Line 3
Line 7 = Line 5/Line 6
Line 12 = Line 10/Line 11
Line 16 = (1 + Line 15/100) × (Line 2 × Line 7/Line 9 + Line 12/Line 14) according to (5.97)

Table 5.16 VSAT network satellite and earth station parameters

Satellite (SL) parameters	Reference	Symbol	Value	Unit
Saturation power flux density	(5.11)	Φ_{sat}	−85	$dBWm^{-2}$
Edge of coverage saturation EIRP	(5.37)	$EIRP_{SLsat}$	43	dBW
Edge of coverage satellite G/T	(5.31)	$(G/T)_{SL}$	2.5	dBK^{-1}
Transponder bandwidth		B_{Xpond}	36	MHz
VSAT parameters				
Antenna diameter	App. 4	D_{VSAT}	1.8	m
HPA rated power		$P_{TXVSATsat}$	1	W
Transmit gain (max)	(5.23)	$G_{TmaxVSAT}$	46.4	dBi
TX implementation losses: feeder loss	§5.2.2	$L_{FTX,VSAT}$	0.2	dB
depointing loss	(5.24)	$L_{TmaxVSAT}$	1.2	dB
HPA output back-off	(A6.1)	OBO_{VSAT}	−6	dB
HPA output power	(A6.1)	P_{TXVSAT}	0.25	W
EIRP	(5.17)	$EIRP_{VSAT}$	39	dBW
Receive gain (max)	(5.44)	$G_{RmaxVSAT}$	45.4	dBi
RX implementation losses: feeder loss	§5.3.4	$L_{FRX,VSAT}$	0.5	dB
depointing loss	(5.46)	$L_{RmaxVSAT}$	0.9	dB
Figure of merit (clear sky)	§5.3.4	$(G/T)_{VSAT}$	22.3	dBK^{-1}
Figure of merit (rain)	§5.3.4		18.8	dBK^{-1}
Hub parameters				
Antenna diameter	App. 4	D_{HUB}	5.5	m
HPA rated power		P_{TXHUB}	5	W
Transmit gain (max)	(5.23)	$G_{TmaxHUB}$	57.1	dBi
TX implementation losses: feeder loss	§5.2.2	$L_{FTX,HUB}$	0.2	dB
depointing loss (tracking)	§5.2.2	$L_{TmaxHUB}$	0.5	dB
HPA output back-off	(A6.1)	OBO_{HUB}	−8	dB
HPA output power	(A6.1)	P_{TXHUB}	0.8	W
EIRP	(5.17)	$EIRP_{HUB}$	54.6	dBW
Receive gain (max)	(5.44)	$G_{RmaxHUB}$	55.1	dBi
RX implementation losses: feeder loss	§5.3.4	$L_{FRX,VSAT}$	0.5	dB
depointing loss (tracking)	§5.2.2	$L_{RmaxVSAT}$	0.5	dB
Figure of merit (clear sky)	§5.3.4	$(G/T)_{HUB}$	32.5	dBK^{-1}
Figure of merit (rain)	§5.3.4	$(G/T)_{HUB}$	29	dBK^{-1}

According to (5.96), the utilised bandwidth for the outbound link is $B_{outb} = 732$ kHz.

Assuming a 20% guard band between carriers, the total utilised transponder bandwidth is:

$$B = 1.2(GB_{inb} + B_{outb}) = 1.2 \times (3 \times 183 + 732)\ \text{kHz} = 1.54\ \text{MHz}. \quad (5.97)$$

Table 5.17 Link parameter calculations

		Inbound		Outbound	
		Clear sky	Rain on VSAT uplink	Clear sky	Rain on VSAT downlink
Uplink					
1	Earth station EIRP (per carrier)	39 dBW	39 dBW	54.6 dBW	54.6 dBW
2	Path loss at 35° elevation angle	208 dB	214 dB	208 dB	208 dB
3	Gain of ideal 1m^2 antenna	44.5 dBi	44.5 dBi	44.5 dBi	44.5 dBi
4	Operating flux density (per carrier)	−124.5 dBWm^{-2}	−130.5 dBWm^{-2}	−108.9 dBWm^{-2}	−108.9 dBWm^{-2}
5	Saturation flux density	−85 dBWm^{-2}	−85 dBWm^{-2}	−85 dBWm^{-2}	−85 dBWm^{-2}
6	Input back-off (per carrier)	−39.5 dB	−45.5 dB	−23.9 dB	−23.9 dB
7	Total input back-off	−23.5 dB	−23.6 dB	−23.5 dB	−23.6 dB
8	Satellite G/T	2.5 dBK^{-1}	2.5 dBK^{-1}	2.5 dBK^{-1}	2.5 dBK^{-1}
9	Uplink C/N_0 at saturation	101.6 dBHz	101.6 dBHz	101.6 dBHz	101.6 dBHz
10	Uplink C/N_0 (per carrier)	62.1 dBHz	56.1 dBHz	77.7 dBHz	77.7 dBHz
Downlink					
11	Saturated satellite EIRP	43 dBW	43 dBW	43 dBW	43 dBW
12	Path loss at 35° elevation angle	207 dB	207 dB	207 dB	213 dB
13	Earth station G/T	32.5 dBK^{-1}	32.5 dBK^{-1}	22.3 dBK^{-1}	18.8 dBK^{-1}
14	Downlink C/N_0 at saturation	97.1 dBHz	97.1 dBHz	86.9 dBHz	77.4 dBHz
15	Total output back-off	−16.6 dB	−16.7 dB	−16.6 dB	−16.7 dB
16	Output back-off (per carrier)	−31.0 dB	−36.4 dB	−17.0 dB	−17.0 dB
17	Small signal suppression	0.5 dB	0.5 dB	0 dB	0 dB
18	Operating EIRP (per carrier)	11.5 dBW	6.1 dBW	26.0 dBW	26.0 dBW
19	Downlink C/N_0 (per carrier)	65.6 dBHz	60.2 dBHz	69.9 dBHz	60.7 dBHz
20	*Intermodulation*	104.7 dBHz	104.9 dBHz	119.3 dBHz	124.9 dBHz
21	*Interference*	72.8 dBHz	67.4 dBHz	77.6 dBHz	77.6 dBHz
22	*Total link C/No (per carrier)*	60.2 dBHz	54.4 dBHz	68.7 dBHz	60.3 dBHz
23	Required E_b/N_o (no coding)	11.3 dB	11.3 dB	11.3 dB	11.3 dB
24	Coding gain	5.5 dB	5.5 dB	5.5 dB	5.5 dB
25	Required E_b/N_o (with coding)	5.8 dB	5.8 dB	5.8 dB	5.8 dB
26	*Required C/No*	53.9 dBHz	53.9 dBHz	59.9 dBHz	59.9 dBHz

(*continued overleaf*)

Table 5.17 (*continued*)

		Inbound		Outbound	
		Clear sky	Rain on VSAT uplink	Clear sky	Rain on VSAT downlink
27	*Margin*	6.3 dB	0.5 dB	8.8 dB	0.4 dB
28	Transponder power usage per carrier	0.08%	0.02%	2.02%	2.02%
29	Transponder power usage	2.16%	2.13%	2.16%	2.13%
30	Transponder bandwidth usage per carrier	0.51%	0.51%	2.03%	2.03%
31	Transponder bandwidth usage	4.27%	4.27%	4.27%	4.27%

Notes

Line 2: equation (5.27) $L_{FS} = 206.6$ dB (Figure 5.13) + 0.9 dB (Figure 5.14)
$L_A = 0.5\,\text{dB} + A_{RAIN}$; clear sky: $A_{RAIN} = 0\,dB$, rain: $A_{RAIN} = 6$ dB
Line 4: equation (5.14) = Line 1 − Line 2 + Line 3
Line 6: equation (5.15) = Line 4 − Line 5
Line 7: equation (5.16) $= 10\log(3 \times 10^{IBO_{inb}(dB)}/10 + 10^{IBO_{outb}(dB)/10})$ if clear sky $= 10\log(2 \times 10^{IBO_{inb}(dB)}/10 + 10^{IBO_{inbrain}(dB)/10} + 10^{IBO_{outbrain}/10})$ if rain
Line 9: equation (5.11) = Line 5 − Line 3 + Line 8 + 228.6
Line 10: equation (5.10) = Line 9 + Line 6
Line 12: equation (5.41) $L_{FS} = 205.6$ dB (Figure 5.13) + 0.9 dB (Figure 5.14)
$L_A = 0.5\,\text{dB} + A_{RAIN}$; clear sky: $A_{RAIN} = 0$ dB, rain: $A_{RAIN} = 6$ dB
Line 14: equation (5.36) = Line 11 − Line 12 + Line 13 + 228.6
Line 15: equation (A6.10) = 0.9 (Line 7 + 5)
Line 16: equation (A6.10) = 0.9 (Line 6 + 5)
Line 17: small signal suppression takes into account the capture effect (see section 5.1.4)
Line 18: Line 11 + Line 16 − Line 17
Line 19: equation (5.35) = Line 14 + Line 16 − Line 17
Line 20: equation (5.55) = 79 − 10 log 3 − 1.65(Line7 + 5) for inbound links = previous result − $EIRP_{inb}$ + $EIRP_{outb}$ for outbound links
Line 21: considers downlink adjacent satellite interference only, see equation (5.88), with $\alpha = 4°$, $EIRP_{SLi,max} = 52$ dBW-, and $B_i = 26$ MHz
= Line 18 − 52 + 10 log Bi + G_{RXmax} − 32 + 25 log 4.6°, where $G_{RXmax} = 54.6$ dBi inbound, 44.9 dBi outbound.
Line 22: equation (5.7) $= -10\log(10^{-\text{Line10}/10} + 10^{-\text{Line19}/10} + 10^{-\text{Line20}/10} + 10^{-\text{Line21}/10})$
Line 25 = Line 23 − Line 24
LIne 26 = Line 25 + 10 log R_b where $R_b = 64\,\text{kbs}^{-1}$ inbound, $256\,\text{kbs}^{-1}$ outbound
Line 27 = Line 22 − Line 26
Line 28 = $100\ 10^{OBO_{inb}/10}$
Line 29 = $100\ 10^{OBO_t/10}$
Line 30 = $100\ B_N/B_{Xpond}$
Line 31 = 100 ((Line 16 of Table 5.15)$/B_{Xpond}$) according to equation (5.98)

The percentage of transponder bandwidth utilised by the VSAT network is:

$$\frac{100B}{B_{\mathrm{Xpond}}} = 100\left(\frac{1.54}{36}\right) = 4.3\% \tag{5.98}$$

Appendices

APPENDIX 1: TRAFFIC SOURCE MODELS

The purpose of traffic source modelling is to provide mathematical tools that can be used in simulations, and also give some insight into the physical behaviour of the device acting as a source. For any network design and performance evaluation, it is important to have models. Such models should be matched to the actual traffic, measured at these interfaces.

The model should include the following features:

- the rate at which messages are generated;
- the time between messages, also named interarrival time;
- the duration (in seconds) or length (in bits) of the message.

The Poisson model is very popular and useful. The defining assumptions are as follows:

- the probability of one message being generated in a small time interval $\Delta t (\Delta t \to 0)$ is proportional to that interval, and therefore equal to $\lambda \Delta t$, where λ is a constant;
- Δt is considered small enough so that there cannot be more than one arrival in Δt, and the probability of no message being generated in Δt is equal to $(1 - \lambda \Delta t)$.

VSAT Networks, 2nd Edition. G. Maral
 ISBN: 0-470-86684-5

Message generation

Based on these assumptions, it can be shown that the probability of K messages being generated in an interval T, much larger than Δt, is given by:

$$P(K) = (\lambda T)^K \, e^{-\lambda T}/K! \tag{A1.1}$$

The average number of messages being generated in T seconds is then:

$$\langle K \rangle = \lambda T \tag{A1.2}$$

So, the Poisson parameter introduced as a proportionality factor for the small interval Δt turns out to be the average number of messages generated per unit time as well. Figure A1.1 represents the probability $P(K)$ of K messages being generated during time T as a function of λT. It reaches a maximum equal to $(K/e)^K/K!$ for $\lambda T = K$.

Interarrival time (IAT)

It can also be shown that the time τ between messages is a continuously distributed exponential random variable:

$$f(\tau) = \lambda e^{-\lambda \tau} \quad (s^{-1}) \tag{A1.3}$$

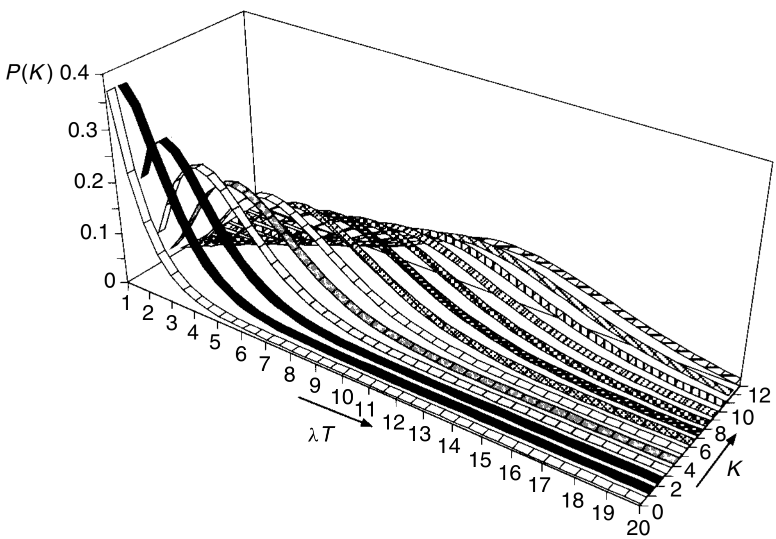


Figure A1.1 Probability $P(K)$ of K messages being generated during time T as a function of λT

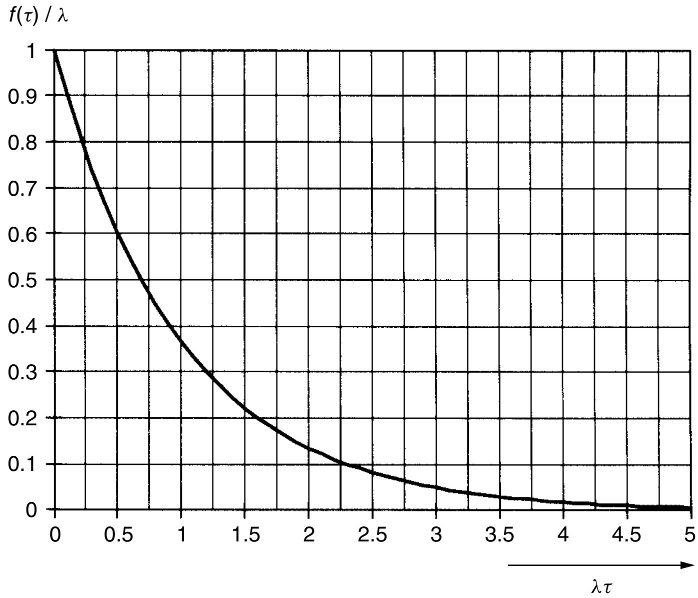


Figure A1.2 Normalised probability density function of message interarrival time

where $f(\tau)$ is the probability density function of τ, and is represented in Figure A1.2.

The average interrarival time is:

$$\langle \mathrm{IAT} \rangle = \int_0^{\infty} \tau f(\tau)\, \mathrm{d}\tau = \frac{1}{\lambda} \quad \text{(s)} \tag{A1.4}$$

Message length

Exponential distribution

The probability for a message to be L bits in length is given by:

$$P(L) = \mu \mathrm{e}^{-\mu} L \tag{A1.5}$$

The average message length is:

$$\langle L \rangle = \frac{1}{\mu} \tag{A1.6}$$

Geometric distribution

The probability for a message to be L bits in length is given by:

$$P(L) = pq^{L-1} \tag{A1.7}$$

where p is the probability of having a one-bit message, and $q = 1 - p$.
The average message length is:

$$\langle L \rangle = \frac{1}{p} \quad \text{(A1.8)}$$

APPENDIX 2: AUTOMATIC REPEAT REQUEST (ARQ) PROTOCOLS

Notation

D = number of data bits per frame to be conveyed from source to destination
L = length of frame (bits) $= D + H$
H = total number of bits in the frame header (and trailer if any)
R_b = information bit rate over the connection (bs^{-1}),
BER = bit error rate (s^{-1})
P_f = probability for a frame to be in error $= 1 - (1 - \text{BER})^L$
T_{RT} = round trip time $= L/R_b + T_p + A/R_b + T_p = (A + L)/R_b + 2T_p \approx L/R_b + 2T_p$
T_p = propagation time from link source node to link destination node
A = length (in bits) of acknowledgement frame (considered negligible compared to L).

Channel efficiency

The channel efficiency is defined as the ratio of the throughput THRU to the information bit rate R_b of the source:

$$\eta_c = \frac{\text{THRU}}{R_b} \quad \text{(A2.1)}$$

Stop-and-wait (SW) protocol

The delivery time of a frame is equal to the average number $\langle M_{SW} \rangle$ of transmissions multiplied by the round trip time T_{RT} between each transmission (one error-free frame is transmitted each T_{RT} interval). Therefore, the throughput is equal to:

$$\text{THRU} = \frac{D}{(T_{RT}\langle M_{SW}\rangle)} \quad (bs^{-1}) \quad \text{(A2.2)}$$

and the channel efficiency is given by:

$$\eta_{\text{cSW}} = \frac{D}{(R_{\text{b}} T_{\text{RT}} \langle M_{\text{SW}} \rangle)} \tag{A2.3}$$

The probability $P(k)$ for a frame to be transmitted k times is having a frame correctly received after $k-1$ unsuccessful tries:

$$P(k) = (1 - P_{\text{f}}) P_{\text{f}}^{k-1} \tag{A2.4}$$

The average number of transmissions $\langle M_{\text{SW}} \rangle$ is equal to:

$$\begin{aligned}
\langle M_{\text{SW}} \rangle &= \sum_{k=0}^{\infty} kP(k) \\
&= (1 - P_{\text{f}}) \sum_{k=0}^{\infty} kP_{\text{f}}^{k-1} \\
&= (1 - P_{\text{f}}) \frac{\mathrm{d}}{\mathrm{d}P_{\text{f}}} \left(\sum_{k=0}^{\infty} P_{\text{f}}^{k} \right) \\
&= (1 - P_{\text{f}}) \frac{\mathrm{d}}{\mathrm{d}P_{\text{f}}} \left(\frac{1}{1 - P_{\text{f}}} \right) \\
&= \frac{1}{1 - P_{\text{f}}} \qquad \text{(A2.5)}
\end{aligned}$$

Therefore:

$$\eta_{\text{cSW}} = \frac{D(1 - P_{\text{f}})}{R_{\text{b}} T_{\text{RT}}} \tag{A2.6}$$

Go-back-*N* (GBN) protocol

Each NACK received by the sender initiates the retransmission of all subsequent frames transmitted after the non-acknowledged one. Let us assume that the sender, upon receipt of an ACK, is allowed to transmit W frames and stop until reception of the next ACK. If a NACK is received, the number of frames that are retransmitted is W.

The delivery time of a frame is equal to the average number $\langle M_{\text{GBN}} \rangle$ of transmissions of that frame multiplied by its duration, L/R_{b}. Hence, the channel efficiency is given by:

$$\eta_{\text{cGBN}} = \frac{D}{L \langle M_{\text{GBN}} \rangle} \tag{A2.7}$$

Table A2.1

Number of transmitted frames	Number of unsuccessful transmissions E	Probability
1	0	$P(E=0)=(1-P_f)$
$W+1$	1	$P(E=1)=(1-P_f)P_f$
$2W+1$	2	$P(E=2)=(1-P_f)P_f^2$
$\vdots$	$\vdots$	$\vdots$
$kW+1$	k	$P(E=k)=(1-P_f)P_f^k$

The probabilities associated with the number of transmissions of a given frame are given in Table A2.1.

The average number of transmissions of a given frame is equal to:

$$\begin{aligned}\langle M_{\text{GBN}}\rangle &= \sum_{k=0}^{\infty}(kW+1)(1-P_f)P_f^k \\ &= W(1-P_f)\sum_{k=0}^{\infty}kP_f^k + \sum_{k=0}^{\infty}(1-P_f)P_f^k\end{aligned}$$

Considering

$$\begin{aligned}\sum_{k=0}^{\infty}kP_f^k &= P_f\sum_{k=0}^{\infty}kP_f^{k-1} \\ &= P_f\frac{d}{dP_f}\left(\sum_{k=0}^{\infty}P_f^k\right) \\ &= P_f\frac{d}{dP_f}\left(\frac{1}{1-P_f}\right) \\ &= \frac{P_f}{(1-P_f)^2}\end{aligned}$$

and

$$\sum_{k=0}^{\infty}(1-P_f)P_f^k = \sum_{k=0}^{\infty}(P_f^k - P_f^{k+1}) = 1$$

then

$$\langle M_{\text{GBN}}\rangle = WP_f/(1-P_f) + 1 = (1+(W-1)P_f)/(1-P_f) \qquad \text{(A2.8)}$$

One can verify that $W=1$ corresponds to the stop-and-wait case, as given by (A2.5).

According to (A2.7):

$$\eta_{\text{cGBN}} = \left(\frac{D}{L}\right)\frac{(1-P_{\text{f}})}{(1+(W-1)P_{\text{f}})} \tag{A2.9}$$

The maximum value of η_{cGBN} is obtained for continuous transmission. This occurs if the window W exceeds the maximum number of frames on the link. A frame lasts L/R_{b} and the link can hold $T_{\text{RT}}/(L/R_{\text{b}})$ frames. Therefore, the condition for continuous transmission is:

$$W > \frac{R_{\text{b}}T_{\text{RT}}}{L} \tag{A2.10}$$

The channel efficiency is bounded by:

$$\begin{aligned}\eta_{\text{cGBN}} &< (D/L)\frac{(1-P_{\text{f}})}{(1+((T_{\text{RT}}R_{\text{b}}/L)-1)P_{\text{f}})} \\ &= \frac{D(1-P_{\text{f}})}{(L(1-P_{\text{f}})+T_{\text{RT}}R_{\text{b}}P_{\text{f}})}\end{aligned} \tag{A2.11}$$

Selective-repeat (SR) protocol

Upon reception of a NACK, the sender retransmits the erroneous frame only. At the receiver, frames must be stored and resequenced.

The delivery time of a frame is equal to the average number $\langle M_{\text{SR}}\rangle$ of transmissions of that frame multiplied by its duration L/R_{b}. $\langle M_{\text{SR}}\rangle$ is equal to $\langle M_{\text{SW}}\rangle$. Hence:

$$\langle M_{\text{SR}}\rangle = \frac{1}{(1-P_{\text{f}})} \tag{A2.12}$$

$$\eta_{\text{cSR}} = \frac{D(1-P_{\text{f}})}{L} \tag{A2.13}$$

APPENDIX 3: INTERFACE PROTOCOLS

This appendix intends to give brief information on most of the common protocols. For more details, the reader is invited to refer to the specialised literature, such as, for instance, [TUG-82][TAN-89] and the relevant texts of the EIA (Electronic Industries Association) and the ITU-T (formerly CCITT).

ASYNC (Asynchronous Communications)

Each information character or block of data is individually synchronised, usually by the use of start and stop elements. The gap between each character or block is not necessarily of a fixed length. Asynchronous data are usually produced by low speed terminals with bit rates up to a few kbs^{-1}.

BISYNC (Binary Synchronous Communications), also termed BSC

A set of control characters and control character sequences for synchronised transmission of binary-coded data between devices in a data communications system (see HDLC).

HDLC (High Data Level Protocol)

A Layer 2 (data link) protocol which rules orderly transfer of information between interfaced computers or terminals. The basic functions of HDLC are:

- to establish and terminate a connection between two terminals;
- to assure the message integrity through error detection, request for retransmission, and positive or negative acknowledgements;
- to identify the sender and the receiver through polling or selection;
- to handle special control functions such as requests for status, station reset, reset acknowledgement, start, start acknowledgement, and disconnection.

PAD (Packet Assembler–Disassembler)

Applies to exchange of serial data streams with character-mode terminal and the packetising–depacketising of the corresponding data exchanged with the ITU-T X25 terminal. Among the basic functions of the PAD are:

- assembly of characters into packets destined for the X25 Data Terminal Equipment (DTE);
- disassembly of the user data field of packets destined for the start–stop mode DTE (asynchronous transmission in which a group of code elements corresponding to a character signal is preceded by a start element and followed by a stop element);
- handling of virtual call set-up and clearing, resetting and interrupt procedures;

- generation of service signals;
- a mechanism for forwarding packets when the proper conditions exist, such as when a packet is full or an idle time expires;
- a mechanism for transmitting data characters, including start, stop, and parity elements as appropriate to the start–stop DTE;
- a mechanism for handling a 'break' signal from the start–stop DTE.

RS232

A layer 1 (physical layer) protocol standard as well as an electrical standard specifying handshaking functions between the Data Terminal Equipment (DTE) and the Data Circuit-terminating Equipment (DCE) over short distances (up to 15 m) at low speed data rates (upper limit of 20 kbs^{-1}).

RS232 makes use of a 25 pin connector. A positive voltage between +5 and +25 V represents a logic 0, and a negative voltage between −5 and −25 V represents logic 1.

The ITU-T counterparts of RS232 are V24 and V28.

RS422

A layer 1 (physical layer) protocol standard. It is a differential balanced voltage interface standard capable of higher data rates over longer distances than those specified in RS232.

The ITU-T counterparts of RS422 are Recommendations V11 and X27.

RS449

Expands specifications of RS232 to higher data rates and longer distances (for instance, 2 Mbs^{-1} over 60 m cables). The mechanical, functional and procedural interfaces are given in RS449, but the electrical interfaces are given by RS423 for unbalanced transmission (all circuits share a common ground) and RS422 for balanced transmission (each one of the circuits requires two wires with no common ground). RS449 makes use of a 37 pin connector.

SDLC (Synchronous Data Link Control)

An IBM variant of HDLC.

SNA/SDLC

See SDLC

TCP/IP (Transmission Control Protocol/Internet Protocol)

A set of protocols from layer 3 to 5 (network/transport/session layers) developed to allow cooperating computers to share resources across a network. Among the basic functions of TCP/IP are:

- file transfer;
- remote login;
- computer mail;
- access to distributed databases, etc.

V11 (also X27)

Deals with the electrical characteristics of balanced double-current interchange circuits operating with data signalling rates up to 10 Mbs^{-1}. It is similar to RS422.

V24

A list of definitions for interchange circuits between Data Terminal Equipment (DTE) and Data Circuit-terminating Equipment (DCE) for transfer of binary data, control and timing signals. The definitions are applicable to synchronous and asynchronous data communications. It is similar to RS232.

V28

Defines the electrical characteristics for unbalanced double-current interchange circuits operating below 20 kbs^{-1}. Binary 1 corresponds to voltages lower than -3 V. Binary 0 corresponds to voltages higher than $+3$ V. It is similar to RS232.

V35

Defines interchange circuits for data transmission at 48 kbs^{-1}. Practice has established V35 as a standard for interface circuits operating at 48, 56 and 64 kbs^{-1} using a 34 pin connector. It is similar to RS232, with slight differences (no external transmit clock).

X3

Describes the basic functions and the user-selectable functions of the packet assembler–disassembler (PAD). It applies to exchange of serial data streams to/from a character-mode terminal from/to an X25 terminal (transport layer).

X21

Applies to the DTE/DCE interface for synchronous operation on public data networks. It defines functions at all three lower layers of the network.

X25

Defines a set of protocols for block transfer between a host computer and a packet switching network. For layer 1 (physical layer), X25 specifies layer 1 of X21. For layer 2 (data link layer), a link access protocol (LAP) is defined using the principles of high level data link control (HDLC). This layer provides the function of error and flow control for the access link between the DTE and the network. Each frame has a check sequence to detect errors, and error frames are retransmitted when requested by the receiving end or by time out. Flow control is accomplished through the sending of receiver ready and receiver not ready commands. Layer 3 of X25 (network layer) defines the packet formats and control procedures for exchange of information between a DTE and the network. X25 provides the capability of multiplexing up to 4096 logical channels, or virtual circuits, on a single access link. Each channel can be used for virtual calls or a permanent virtual circuit. Each packet exchanged across the interface has its associated logical channel number identified, and each logical channel operates independently of the others. The data packets are also identified by sequence numbers which are used for flow control within individual logical channels. The sequence numbering may be based upon either modulo 8 for normal operation or modulo 128 for extended transmission delay conditions. Data packets are limited to a maximum data field length (nominally 128 octets, with possible extension up to 1024 octets).

X28

Defines the interface for the start–stop mode terminals accessing the packet assembler–disassembler (PAD) on a public data network. It specifies procedures for establishing an access information path between a start–stop DTE and a PAD, and for character interchange and service initialisation between them, as well as for the exchange of control information. It also summarises PAD commands and service signals.

X29

Specifies the procedures for the exchange of control information and user data between an X25 DTE and a packet assembler–disassembler (PAD).

APPENDIX 4: ANTENNA PARAMETERS

Gain

Definition

The gain of an antenna is the ratio of the power radiated (or received) per unit solid angle by the antenna in a given direction to the power radiated (or received) per unit solid angle by an isotropic antenna fed with the same power.

Maximum gain

The gain is maximal in the direction of maximum radiation (the electromagnetic axis of the antenna, also called the *boresight*) and has a value given by:

$$G_{\mathrm{max}} = \left(\frac{4\pi}{\lambda^2}\right) A_{\mathrm{eff}} \qquad \text{(A4.1)}$$

where $\lambda = c/f$. c is the velocity of light ($c = 3 \times 10^8\ \mathrm{ms}^{-1}$) and f is the frequency of the electromagnetic wave. A_{eff} is the *effective aperture area* of the antenna. For an antenna with a circular aperture or reflector of diameter D and geometric surface $A = \pi D^2/4$, $A_{\mathrm{eff}} = \eta_a A$, where η_a is the efficiency of the antenna (a typical value for antenna technology and frequencies used in VSAT networks is $\eta_a = 0.6$). Hence:

$$G_{max} = \eta_a \left(\frac{\pi D}{\lambda}\right)^2 = \eta_a \left(\frac{\pi Df}{c}\right)^2 \quad \text{(A4.2)}$$

$$G_{max}\,(\text{dBi}) = 10 \log \left[\eta_a \left(\frac{\pi D}{\lambda}\right)^2\right] = 10 \log \left[\eta_a \left(\frac{\pi Df}{c}\right)^2\right] (\text{dBi})$$

Antenna radiation pattern

The *antenna radiation pattern* indicates the variations of gain with direction. For an antenna with a circular aperture or reflector, this pattern has rotational symmetry about its boresight and can be represented by its variation within any plane containing the boresight. Figure A4.1 displays a typical pattern which can be represented either in polar coordinates (Figure A4.1(a)) or in Cartesian coordinates (Figure A4.1(b)).

Figure A4.1 reveals the major lobe which contains the direction of maximum gain G_{max} at boresight ($\theta = 0°$), and the sidelobes with smaller secondary maxima.

Half power beamwidth

It is convenient to characterise the width of the antenna radiation pattern by the angle between the directions in which the gain falls to half its maximum value. This angle is called the *3 dB beamwidth* θ_{3dB}. A practical formula for θ_{3dB} is:

$$\theta_{3dB} = 70 \left(\frac{\lambda}{D}\right) = 70 \left(\frac{c}{fD}\right) \quad \text{(degrees)} \quad \text{(A4.3)}$$

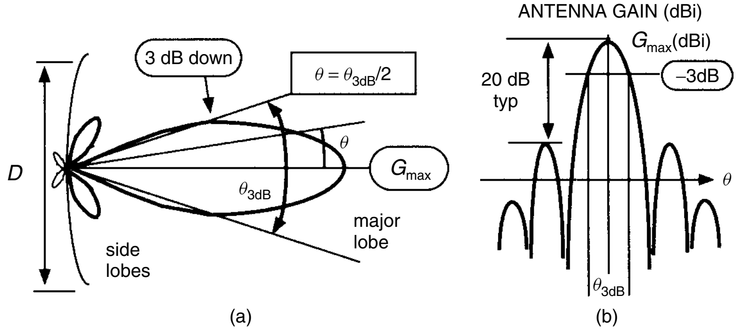


Figure A4.1 Antenna radiation pattern, (a) polar coordinates; (b) Cartesian coordinates (dB on vertical scale)

It can be noted that θ_{3dB} increases with decreasing D, which indicates that a small aperture antenna displays a large beamwidth.

Depointing loss

In a direction θ near to the boresight, say between 0 and $\theta_{3dB}/2$, the value of the gain is given by:

$$G(\theta)(\text{dBi}) = G_{\max}(\text{dBi}) - 12\left(\frac{\theta}{\theta_{3dB}}\right)2 \quad (\text{dBi}) \tag{A4.4}$$

Polarisation

The wave radiated by an antenna consists of an electric field component and a magnetic field component. These two components are orthogonal and perpendicular to the direction of propagation of the wave. They vary in time at the frequency of the wave. By convention the *polarisation* of the wave is defined by the direction of the electric field. In general, the direction of the electric field is not fixed and its amplitude is not constant. During one period, the projection of the extremity of the vector representing the electric field onto a plane perpendicular to the direction of propagation of the wave describes an ellipse, as illustrated in Figure A4.2. The polarisation is said to be elliptical.

Two waves are in *orthogonal polarisation* if their electrical fields describe identical ellipses in opposite directions. In particular, the following can be obtained:

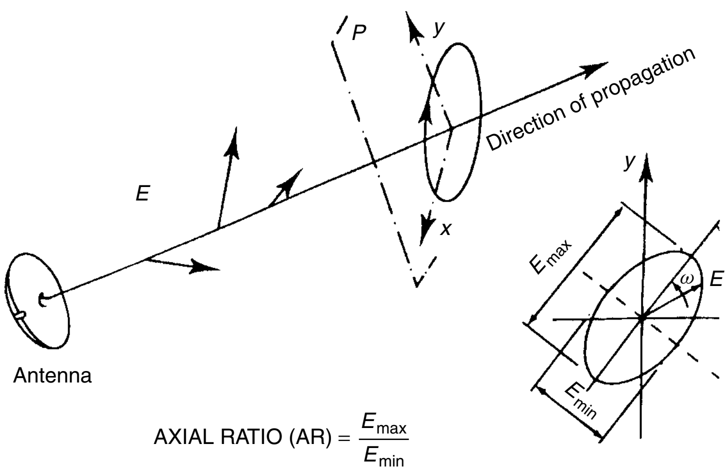


Figure A4.2 Polarisation of an electromagnetic wave

– two orthogonal circular polarisations are described as right-hand circular and left-hand circular (the direction of rotation is for an observer looking in the direction of propagation);
– two orthogonal linear polarisations are described as horizontal and vertical (relative to a local reference).

An antenna designed to transmit or receive a wave of given polarisation cannot transmit or receive in the orthogonal polarisation. This allows two simultaneous links to be set up at the same frequency between the same two locations, so called 'frequency reuse' by orthogonal polarisation. To achieve this, either two polarised antennas are installed at each end or, preferably, one antenna designed for operation with the two specified polarisations may be used. This practice must, however, take account of imperfections in the antennas and the possible depolarisation of the waves by the transmission medium (the atmosphere, especially with rain, in the case of satellite links). These effects introduce mutual interference of the two links. This situation is illustrated in Figure A4.3 which relates to the case of two orthogonal linear polarisations (the illustration is also valid for any two orthogonal polarisations).

Let a and b be the amplitudes, assumed to be equal, of the electric fields of the two waves transmitted simultaneously with linear polarisation, a_c and b_c the amplitudes received with the same polarisation, and a_x and b_x the amplitudes received with orthogonal polarisations. The following concepts are defined:

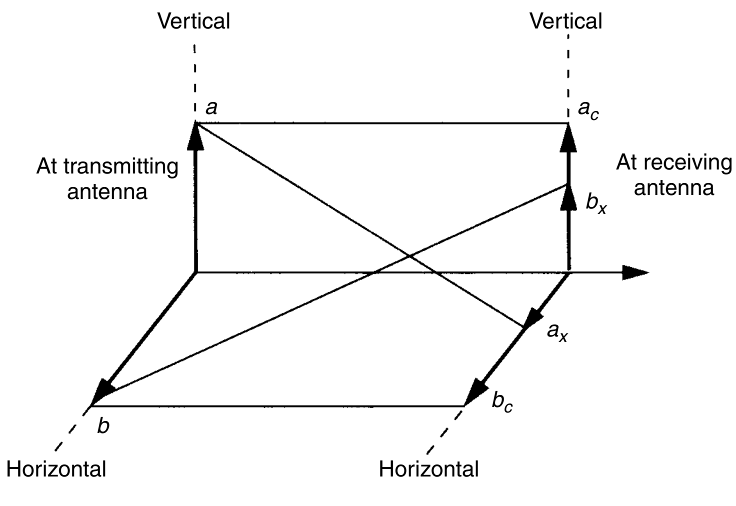


Figure A4.3 Amplitude of the transmitted and received electric field for the case of two orthogonal polarisations

– the *cross polarisation isolation*:

$$\text{XPI} = \frac{a_c}{b_x} \text{ or } \frac{b_c}{a_x}$$

$$\text{XPI (dB)} = 10\log\left(\frac{a_c}{b_x}\right) \text{ or } 10\log\left(\frac{b_c}{a_x}\right) \quad \text{(A4.5)}$$

– the *cross polarisation discrimination* (when a single polarisation is transmitted):

$$\text{XPD} = \frac{a_c}{a_x} \quad \text{(A4.6)}$$

In practice, XPI and XPD are comparable and are often confused within the term *isolation*.

The values and relative values of the components vary as a function of the direction relative to the antenna boresight. The antenna is thus characterised, for a given polarisation, by a radiation pattern for the nominal polarisation (co-polar) and a radiation pattern for the orthogonal polarisation (cross-polar). Cross polarisation discrimination is usually maximal on the antenna axis and degrades for directions other than boresight.

APPENDIX 5: EMITTED AND RECEIVED POWER

Effective isotropic radiated power (EIRP) of a transmitter

The power radiated per unit solid angle by an isotropic antenna fed from a radio frequency source of power P_T is given by $P_T/4\pi$ (Wsteradian^{-1}).

In a direction where the value of transmission gain is G_T, any antenna radiates a power per unit solid angle equal to $G_T(P_T/4\pi) = (P_T G_T)/4\pi$ (Wsteradian^{-1}).

The product $P_T G_T$ is called the *effective isotropic radiated power* (EIRP). It is expressed in W.

$$\text{EIRP} = P_T G_T \quad \text{(W)} \quad \text{(A5.1)}$$

$$\text{EIRP(dBW)} = P_T\,\text{(dBW)} + G_T\,\text{(dBi)}$$

Power flux density at receiver

A surface of effective area A situated at a distance R from the transmitting antenna subtends a solid angle A/R^2 at the transmitting antenna. It receives a power equal to the product of the power radiated per unit solid angle $(P_T G_T)/4\pi$ and the considered solid angle A/R^2:

$$P_R = \left(\frac{P_T G_T}{4\pi}\right)\left(\frac{A}{R^2}\right) = \Phi A \quad \text{(W)} \tag{A5.2}$$

$$P_R\,\text{(dBW)} = \Phi\,(\text{dBWm}^{-2}) - 10\log A$$

where Φ is called the *power flux density*:

$$\Phi = \frac{P_T G_T}{4\pi R^2} \quad (\text{Wm}^{-2}) \tag{A5.3}$$

$$\Phi\,(\text{dBWm}^{-2}) = \text{EIRP (dBW)} - 10\log(4\pi R^2)$$

Figure A5.1 summarises the above derivations.

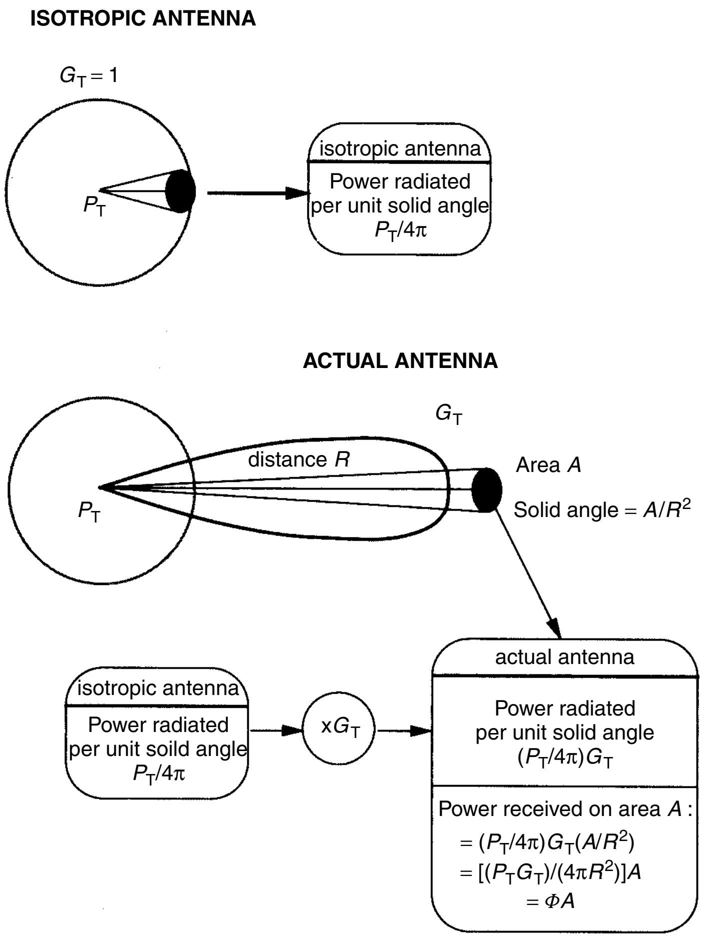


Figure A5.1 Power flux density at receiver

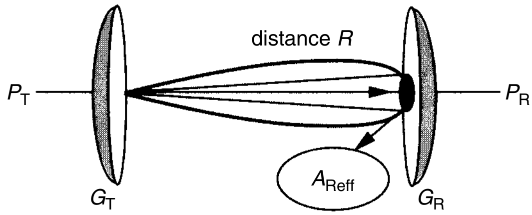


Figure A5.2 Power captured by a receiving antenna

Power available at the output of the receiving antenna

Figure A5.2 represents a receiving antenna of effective aperture area A_{Reff} located at a distance R from a transmitting antenna.

From (A5.2) the receiving antenna captures a power equal to:

$$P_R = \Phi A_{Reff} = \left(\frac{P_T G_T}{4\pi}\right)\left(\frac{A_{Reff}}{R^2}\right) \quad \text{(W)} \qquad \text{(A5.4)}$$

The effective aperture area of an antenna is expressed as a function of its receiving gain G_R by the expresssion:

$$A_{Reff} = \frac{G_R}{(4\pi/\lambda^2)} \quad (\text{m}^2) \qquad \text{(A5.5)}$$

The ratio $4\pi/\lambda^2$ can be looked upon as the gain of an ideal antenna with an effective aperture area equal to 1 m^2.

Denoting $G_1 = 4\pi/\lambda^2$:

$$A_{Reff} = \frac{G_R}{G_1} \quad (\text{m}^2) \qquad \text{(A5.6)}$$

From (A5.4) and (A5.6) one can derive an expression for the received power:

$$P_R = \Phi A_{Reff} = \Phi\left(\frac{G_R}{G_1}\right) \quad \text{(W)} \qquad \text{(A5.7)}$$

$$P_R\ (\text{dBW}) = \Phi(\text{dBWm}^{-2}) + G_R\ (\text{dBi}) - G_1\ (\text{dBi}) \quad (\text{dBW})$$

From (A5.4) and (A5.6) one can derive another expression for the received power:

$$P_R = \frac{P_T G_T}{1/L_{FS}} G_R \quad \text{(W)} \qquad \text{(A5.8)}$$

$$P_R\ (\text{dBW}) = \text{EIRP}\ (\text{dBW}) - L_{FS}\ (\text{dB}) + G_R\ (\text{dBi}) \quad (\text{dBW})$$

where $L_{FS} = (4\pi R/\lambda)^2$ is called the *free space loss*, and represents the ratio of the transmitted to the received power in a link between two isotropic antennas.

APPENDIX 6: CARRIER AMPLIFICATION

Carrier amplification takes place at the transmitting earth station (a VSAT or the hub) and on-board the satellite, within each transponder. Power amplifiers used are either solid state power amplifiers (SSPAs) or traveling wave tubes (TWTs). Both types act as non-linear devices when operated near saturation, where the output power is maximal. The non-linearity has two aspects: a decreasing power gain, as the output power comes to saturation, and a variation in the phase of the amplified carrier relative to the input phase.

Figure A6.1 displays typical transfer characteristics for a power amplifier. All quantities are normalised to their respective values at saturation, when the amplifier is operated in a single carrier mode.

Denoting by P_o^1 the output power, and by P_i^1 the input power (1 stands for single carrier drive, o for output, i for input), and $(P_o^1)_{sat}$ and $(P_i^1)_{sat}$ those quantities at saturation, one defines the output back-off (OBO) and the input back-off (IBO) as:

$$\text{OBO} = \frac{P_o^1}{(P_o^1)_{sat}} \tag{A6.1}$$

$$\text{or OBO (dB)} = 10\log\left\{\frac{P_o^1}{(P_o^1)_{sat}}\right\}$$

$$\text{IBO} = \frac{P_i^1}{(P_i^1)_{sat}} \tag{A6.2}$$

$$\text{or IBO (dB)} = 10\log\left\{\frac{P_i^1}{(P_i^1)_{sat}}\right\}$$

The values available from the TWT manufacturer are the output power at saturation $(P_o^1)_{sat}$ and the power gain at saturation G_{sat}. From these two quantities, one can derive $(P_i^1)_{sat}$ as:

$$(P_i^1)_{sat} = \frac{(P_o^1)_{sat}}{G_{sat}} \tag{A6.3}$$

For example, a 50 W TWT, with a 55 dB gain, displays an input power at saturation $(P_i^1)_{sat} = 50\ W/10^{5.5} = 158\ \mu W$.

With the above definitions, the values for OBO (dB) and IBO (dB) are negative in the normal range of operation, i.e. below saturation.

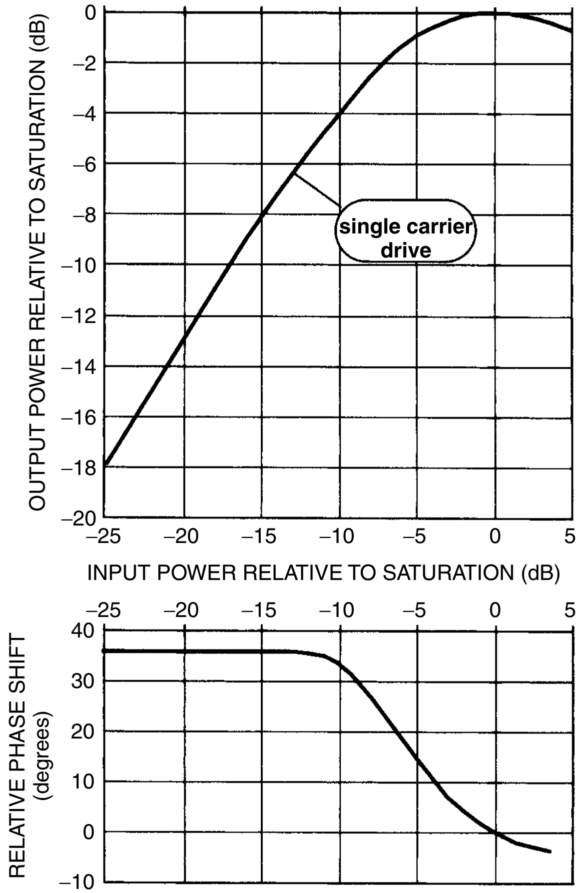


Figure A6.1 Power amplifier characteristics: single carrier operation

Note that some people define OBO and IBO as the inverses of expressions (5.53) and (5.54). OBO (dB) and IBO (dB) values are then positive.

The aspect of the curves in Figure A6.1 is non-linear. When operated in a multicarrier mode, the non-linearity generates intermodulation and the TWT output power is shared, not only between the amplified carriers, but also with the intermodulation products (see section 5.1.3). Denoting by P_o^n and P_i^n, respectively, the output and input power of one carrier among the n amplified ones, one can define:

– the output back-off *per carrier*:

$$\mathrm{OBO_1} = \text{single carrier output power/single carrier output power at saturation}$$

$$= \frac{P_o^n}{(P_o^1)_{sat}} \tag{A6.4}$$

or:

$$\text{OBO}_1(\text{dB}) = 10\log\left\{\frac{P_o^n}{(P_o^1)_{sat}}\right\}$$

– the input back-off *per carrier*:

$$\text{IBO}_1 = \text{single carrier input power/single carrier input power at saturation}$$

$$= \frac{P_i^n}{(P_i^1)_{sat}} \tag{A6.5}$$

or:

$$\text{IBO}_1(\text{dB}) = 10\log\left\{\frac{P_i^n}{(P_i^1)_{sat}}\right\}$$

– the *total* output back-off:

$$\text{OBO}_t = \text{sum of all carrier output power/single carrier input power at saturation}$$

$$= \frac{\Sigma P_o^n}{(P_o^1)_{sat}} \tag{A6.6}$$

or:

$$\text{OBO}_t\ (\text{dB}) = 10\log\left\{\frac{\Sigma P_o^n}{(P_o^1)_{sat}}\right\}$$

– the *total* input back-off:

$$\text{IBO}_t = \text{sum of all carrier input power/single carrier input power at saturation}$$

$$= \frac{\Sigma P_i^n}{(P_i^1)_{sat}} \tag{A6.7}$$

or:

$$\text{IBO}_t\ (\text{dB}) = 10\log\left\{\frac{\Sigma P_i^n}{(P_i^1)_{sat}}\right\}$$

With n equally powered carriers:

$$\text{OBO}_1 = \text{OBO}_t/n \tag{A6.8}$$

$$\text{or OBO}_1(\text{dB}) = \text{OBO}_t(\text{dB}) - 10\log\ n$$

$$\text{IBO}_1 = \text{IBO}_t/n \tag{A6.9}$$

$$\text{or IBO}_1(\text{dB}) = \text{IBO}_t(\text{dB}) - 10\log\ n$$

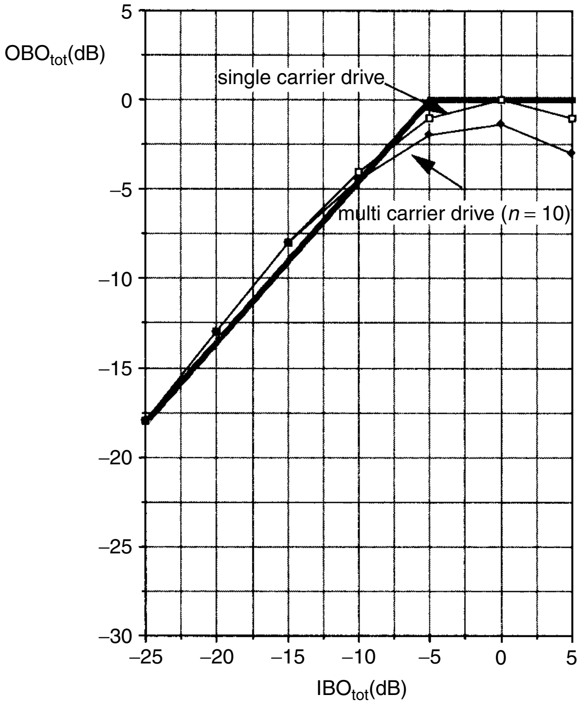


Figure A6.2 OBO_t as a function of IBO_t

Figure A6.2 gives typical variations of OBO_t as a function of IBO_t. A simple but useful model involves approximating the curves by the two segments:

$$OBO_t(dB) = 0.9(IBO_t(dB) + 5) \qquad IBO_t < -5\text{ dB}$$

$$OBO_t(dB) = 0\text{ dB} \qquad -5\text{ dB} < IBO_t < 0\text{ dB} \qquad \text{(A6.10)}$$

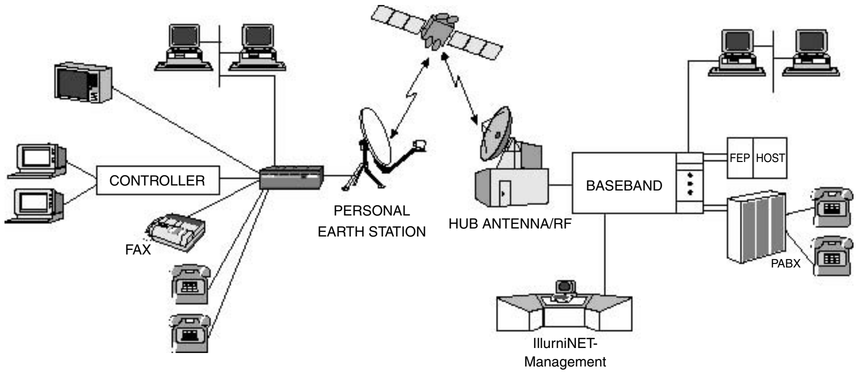
APPENDIX 7: VSAT PRODUCTS

This appendix aims to introduce some popular VSAT products. The list of presented products is by no means exhaustive, and information is subject to change. It is therefore recommended that the reader refers to the most recent information released by the manufacturer.

Hughes Network Systems (HNS)

Address: 11717 Exploration Lane, Germantown, MD 20878, USA
www.hns.com.

Personal earth station (PES)



Technical specifications

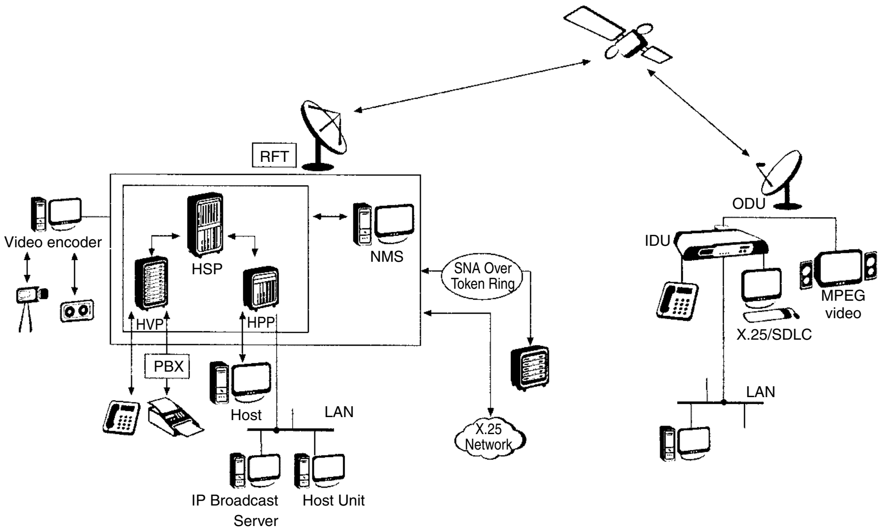
Frequency	
Ku-band, C-band	
Data rates	
Asynchronous:	Up to 19.2 kbs^{-1}
Synchronous:	1.2–64 kbs^{-1} (Standard rates)
Ports	
Standard:	Up to 4 serial ports with LAN Video IF port, 950–1700 MHz
Optional:	2 voice ports
Interfaces	
Data:	RS-232, RS-422, or V.35
LAN:	Ethernet: 10BaseT Token-ring: Type 1, Type 3
Voice:	RJ 11 two-wire loop start
Antenna	
Ku-band:	0.98, 1.2, 1.8, and 2.4 metres
C-band:	1.8 and 2.4 metres
RF Power:	0.5, 1.0 and 2 watt (Ku-band) 2 watt (C-band)
Outroute	
512, 128 Kbs^{-1}	
Inroute	
256, 128, 64 Kbs^{-1}	

Protocol support	
Ethernet (10 Mbs^{-1})	
Token-ring:	4/16 Mbs^{-1} (optional)
Transparent Bridging: SDLC (PU4-PU2, PU4-PU4)	
SDLC to Token-ring	
X.25	
BSC 3270	
Bit and byte transparent	
HASP	
Frame Transparent	
X.3/X.28.X.29 PAD	
Broadcast	
Telnet	
SLIP/PPP	
TCP/IP	
Specialised protocols	
Bit error rate	
1×10^{-7}-at threshold	
1×10^{-9} typical	
Operating temperature	
Outdoor equipment	−30°C to +55°C
Indoor equipment	+10°C to +40°C
Power	
90–264 VAC, 47–63 Hz	
−24 VDC	

Gilat Satellite Networks

www.gilat.com

Skystar advantage



Technical specifications

Network	Architecture		Interactive, star configuration
	Capacity		Up to 34 000 VSATs
	Protocol support		TCP/IP, X 25, Async (X.3/X.28/X.29), SDLC and more
	IP functionalities		TCP, UDP, RIP V1, RIP V2, IRDP, ARP, ICMP, Classes (A, B, C, D), Subnetting and classless addressing, UDP, IGMP
	RF frequency		Ku-band, Ext. Ku-band, C-band, Ext. C-band
Hub station	User ports	Interface:	RS-232, RS-422, V.35, Token-ring or Ethernet
		Information bit rate:	110 bs^{-1} to 512 kbs^{-1} (on serial ports), DCE or DTE
		Data format:	Synchronous or asynchronous
	Outbound carner		Statistical multiplexing
		Number, of carriers:	Configurable
		Bit rate:	64 kbs^{-1} to 8 Mbs^{-1}, up to 24 Mbs^{-1} aggregate
		Error connection:	Concatenated Viterbi and Reed Sobmon
	BER performance		Better than 10^{-12}
	modulator	Modulation scheme:	BPSK or QPSK
Remote terminal	Inbound carner		Proprietary combined TDMA and FDMA (FTDMA)
		Bit rate:	9.6, 19.2, 38.4, 76.8 and 153.6 kbs^{-1}, software configurable. Optional dual bit rate
		Modulation scheme	MSK
	Outdoor unit	Antenna size (Typical):	Ku-band: 0.55 to 1.20 m, C-band: 1.80 m
		SSPA power:	Ku-band 0.5, 1 and 2 watt
			Extended Ku-band 1 watt
			C, Extended C-band 2 watt
		Up-converter:	Compact
		Operating temperature:	-40° to $+60^{\circ}$C
		Humidity:	Up to 100%
	Indoor unit:	Basic unit ports:	2 serial ports plus Ethernet
		Expansion cards:	4 serial ports, Token-ring, voice, video, USE and customised cards, field upgradable

Port information bit rate:	50 bs^{-1} to 128 kbs^{-1} (on serial ports), DCE or DTE	
Interface:	Serial:	RS-232, X.21
	Token-ring:	UTP RJ45, STP DB9
	Ethernet:	10BaseT
	Voice:	2-wire FXS RJ11
	Video:	BNC for composite video, S-video (Mini DIN)
	Audio:	3.5 mm mini jack
Operating voltage:	AC:	Autorange 100–240 V
	DC (optional):	24 V–48 V, 12 V
Power consumption:	Less than 25 W, including ODU	
Dimensions:	6 cm (h) × 40 cm (w) × 34 cm (d)	
Weight:	3.9 kg	
Operating temperature:	−10° to 60°C (weatherised version optional)	
Humidity:	Up to 9.5%, non-condensing	

References

[ABR92] Abramson, N. (1992) Fundamentals of packet multiple access for satellite networks, *IEEE Journal on Selected Areas in Communication*, **10**(2), 309–316.

[CHI88] Chitre, D.M. and McCoskey, J.S. (1988) VSAT networks: architecture, protocols and management, *IEEE Communications Magazine*, **26**(7), 28–38.

[EVA99] Evans, B.G. (Ed.) (1999) *Satellite Communications Systems*, 3rd edition, IEE, London.

[EVE92] Everett, J. (Ed.) (1992) *VSATs: Very Small Aperture Terminals*, IEE Telecommunications Series 28, Peter Pelegrinus.

[FER90] Ferrari, D. (1990) Client requirements for real time communications services, *IEEE Communications Magazine*, **11**, 65–72.

[FUJ86] Fujii, A., Teshigawara, Y., Tejima, S. and Matsumoto, Y. (1986) *AA/TDMA–Adaptive satellite access method for mini earth station networks*, GLOBECOM 86, paper 42.4, Houston.

[HA86] Ha, Tri T. (1986) *Digital Satellite Communications*, MacMillan.

[ITU88] (1988) *Handbook on Satellite Communications*, International Telecommunication Union, Geneva.

VSAT Networks, 2nd Edition. G. Maral

ISBN: 0-470-86684-5

[ITU00] (2000) *Radiocommunications Regulations*, International Telecommunication Union, Geneva.

[JON88] Jones, L. (1988) VSAT technology for today and for the future, Part V: Planning and implementing the network, *Communications News*, **25**(2), 44–47.

[MAR93] Maral, G. and Bousquet, M. (1993) *Satellite Communications Systems*, 2nd edition, John Wiley & Sons, Chichester.

[MAR02] Maral, G. and Bousquet, M. (2002) *Satellite Communications Systems*, 4th edition, Wiley, Chichester.

[RAY87] Raychauduri, D. and Joseph, K. (1987) Ku-band satellite data networks using very small aperture terminals. Part 1: Multi access protocols, *International Journal of Satellite Communications*, **5**, 195–212.

[RAY88] Raychauduri, D. and Joseph, K. (1988) Channel access protocols for Ku-band VSAT networks: a comparative study, *IEEE Communications Magazine*, **26**(5), 34–44.

[SAL88] Salamoff, S. (1988) Real world applications prove benefits, *Communications News*, **25**(1), 38–42.

[SCH77] Schwarz, M. (1977) *Computer Communications Network Design and Analysis*, Prentice Hall.

[SMI72] Smith, F.L. (1972) A nomogram for look angles to geostationary satellites, *IEEE Transactions on Aerospace and Electronic Systems*, **AES 5**, 394.

[TAN89] Tanenbaum, A.S. (1989) *Computer Networks*, 2nd edition, Prentice Hall.

[TUG82] Tugal, D. and Tugal, O. (1982) *Data Transmission*, McGraw Hill.

[ZEI91] Zein, T. and Maral, G. (1991) Stabilized Aloha reservation protocol in a DAMA satellite network, *Proceedings of the Second European Conference on Satellite Communications*, Liège, Belgium, 121–126, ESA SP-332.

[ZEI91a] Zein, T., Maral, G. and Jabbari, B. (1991) Guidelines for a preliminary dimensioning of a transaction oriented VSAT network, *International Journal of Satellite Communications*, **9**, 391–397.

Index

VSAT Networks, 2nd Edition. G. Maral

ISBN: 0-470-86684-5